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

Michael Thaler*

September 12, 2019

Abstract

When people receive information about the world around them, the information often evokes both what they currently believe and what they are motivated to believe. This paper theoretically and experimentally explores motivated reasoning, the impact of motivation on inference. I analyze a model of motivated reasoning in which people distort their updating process towards particular beliefs that they are more motivated to hold. To test predictions of the model, I create a new experimental design in which people make inferences about the veracity of customized news sources. These sources are tailored to subjects' current beliefs such that there is nothing to infer from the signal sent, but motivated reasoning would lead to directional distortions. Using a large online experiment, I find evidence for politically-driven motivated reasoning on eight topics: immigration, upward mobility, racial discrimination, crime, gender-based math ability, climate change, gun laws, and beliefs about other subjects. There is evidence for ego-driven motivated reasoning on beliefs about own performance, but only for men. I also find that motivated reasoning leads people to become more polarized, less accurate, and overconfident in their beliefs.

JEL classification: C91; D83; D84; D91; L82

Keywords: Motivated reasoning; biased beliefs; polarization; overconfidence; fake news

*Harvard University. Email: michaelthaler@g.harvard.edu.
I would like to especially thank my advisors, Alberto Alesina, Christine Exley, David Laibson, and Matthew Rabin, for invaluable support and hyphen-use expertise. Thank you also to Kyle Chauvin, Liz Engle, Ben Enke, Ed Glaeser, Michael Hiscox, Alejandro Lagomarsino, Bart Lipman, Gautam Rao, Marta Serra-Garcia, Jesse Shapiro, Richard Thaler, Mattie Toma, and seminar participants for helpful comments. I am grateful for funding support from the Harvard Business School Research Fellowship, the Eric M. Mindich Research Fund for the Foundations of Human Behavior, and the Bradley Graduate Fellowship.

1 Introduction

"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).

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 the economy, 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. Flynn, Nyhan, and Reifler 2017; Ortoleva and Snowberg 2015). As shown by Gerber and Huber 2009 and Meeuwis et al. 2019, these beliefs 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 often form their posterior beliefs by incorporating their prior beliefs and the relative likelihood of seeing that information. If we only observe beliefs at a snapshot in time, two people's disagreement can be consistent with several explanations: They may have had different priors, they may differently assess the informativeness of the news, or they may have different updating processes. The first two channels are often relevant in politicized settings. First, Democrats and Republicans often have different preconceived notions in such settings, 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 emphasizes the third channel: People differently update, 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 what beliefs they find attractive to hold (*motivated beliefs*). In the model of *motivated reasoning* developed in this paper, people mistakenly interpret information through the lens of these motivated beliefs.[2] Agents update using a modified Bayes' rule, weighing priors and likelihoods as a Bayesian would, but act as if motivated beliefs are an extra signal.

[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).

[2] The model will consider a motivated belief function that maps distributions of states to the real numbers, drawing comparisons to expected utility.

The model shows how motivated reasoning leads agents to over-trust news that reinforces their biases, can cause belief polarization, and on politicized issues leads partisans to become overconfident about their beliefs.

Designing an experiment that identifies motivated reasoning 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 objective is to test the model by constructing an environment in which people have the *same* prior and the *same* subjective likelihood, but motivated beliefs are exogenously randomized. In this setting, differences in updating cannot be due to Bayes' rule, and cannot be due to general over- or under-weighting of priors or likelihoods.[3]

The context of this experiment involves how people infer the veracity of news sources that send messages about economic-, social-, and ego-related topics. Not only is this setting useful for identifying motivated reasoning, it is also a relevant one in the United States today. 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" (Gallup 2018; Knight Foundation 2018).[4] I run the experiment on Amazon Mechanical Turk with approximately 1000 subjects, eliciting over 10000 assessments. Subjects are given factual questions about nine *politicized* topics (on economic and social issues), three neutral topics, and one *ego-relevant* topic (on own performance in the experiment). The list of topics and hypothesized motivated beliefs (motives) is in Table 1.

Specifically, the experimental design has two main steps. Firstly, subjects are given a variety of factual, politically-relevant questions with numerical answers; on each question, the median of subjects' belief distribution is elicited, so that subjects think the true answer is equally likely to be above or below their median. Secondly, they 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 told which source the message came from; instead, they infer the source's veracity from the message. Since messages relate the answer to the subjective median, rational subjects would believe that it is equally likely for either source to report either message. The message is therefore *completely uninformative*, given their own stated beliefs, about the veracity of the news source, and a Bayesian would not update. However, a

[3]As discussed later, it also shows that motivated reasoning cannot be fully explained by utility-maximizing beliefs (Brunnermeier and Parker 2005; Benabou and Tirole 2011; Mobius et al. 2014).

[4]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 name.

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

Table 1: The list of topics and pre-hypothesized motives in the experiment. The first nine topics are called politicized topics. On the computer, each topic is a hyperlink that links to the exact question wording in Appendix C.

subject who engages in motivated reasoning will trust 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 and ego, so that agents will assess Pro-Party or Pro-Own Performance news to be more truthful than Anti-Party or Anti-Own Performance news. It also predicts that for people with erroneous current beliefs, the direction of this error predicts the direction of their motivated beliefs, since their beliefs are incorrect in part because of motivated beliefs.

The main result of the experiment is that Bayesian updating is rejected in favor of politically- and ego-driven 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 news is 9 percentage points (s.e. 0.7 percentage points) more likely to be True News than Anti-Party news. Similarly, Pro-Own Performance veracity assessments are 6 percentage points (s.e. 2 percentage points) higher than Anti-Own Performance assessments. As expected with politically-motivated reasoning, this gap increases in the partisanship of the subject, and assessments on neutral topics lie in between Pro-Party news and Anti-Party news assessments. This design allows for enough statistical power to test motivated reasoning for each topic individually; for nine of the ten topics, the main effect is significant at the $p = 0.001$ level. I also find that the main results are robust to a host of alternative explanations.[5] On each of eight politicized 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.[6]

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

[5]In particular, 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 misunderstanding what a median is. Importantly, there is also evidence for asymmetric updating regarding their beliefs about the initial question. Subjects are 9 percentage points (s.e. 1 percentage point) more likely to change their beliefs in the direction of the message if the news is Pro-Motive than if the news is Anti-Motive, 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.

[6]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 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]. Results from these papers can be explained by motivated reasoning. This literature is discussed in greater detail in Section 6.

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 assessments to news that (falsely) accentuates their error compared to news that (truthfully) brings them closer to the correct answer. Indeed, in the experiment, people trust the error-accentuating 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 even controlling for whether the news is Pro-Party or Anti-Party.[7]

Thirdly, the theory explains how motivated reasoning can lead to more confident but less accurate political beliefs from partisans. It predicts that people will be worse at assessing political news than assessing neutral news and this gap will be larger among partisans. However, partisans will be excessively confident in their assessments. Towards the end of the experiment, subjects are asked to rate their performance relative to 100 other subjects, and both predictions are supported: compared to moderates, partisans *expect* that they have performed better, but *actually* perform worse on news assessments. This suggests that motivated reasoning may provide a link between overconfidence and partisanship, a relationship documented in Ortoleva and Snowberg 2015.

As predicted by the model, motivated reasoning also significantly impacts how people change their beliefs about the politicized topics themselves, leading to belief polarization. Partisans become more politically polarized despite receiving uninformative signals, and 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 misupdating from information because of their gap in trust of Pro-Party and Anti-Party news; it also shows that informational content is *not* a necessary condition for polarization.[8] Politically-motivated reasoning helps reconcile the stylized facts that the ideological polarization of beliefs is high, while the ideological polarization of information acquisition is not especially high (Gentzkow and Shapiro 2011).[9]

[7]The results of these last two paragraphs correspond to the pre-hypothesized "Specific Aims" in my Foundations of Human Behavior grant application. The exact wording of these aims are as follows:

- **Aim 1:** To test whether Republicans and Democrats are more likely to believe news that aligns with their party's stances than news that doesn't when their prior beliefs are the same: Motivated reasoning.
- **Aim 2:** To determine which topics lead to stronger or weaker motivated reasoning.
- **Aim 3:** To test whether people with more extreme ideologies and opinions motivatedly reason more, and to test whether people believe Fake News is more likely to be true than True News.

[8]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).

[9]Gentzkow and Shapiro 2006 and Gentzkow, Wong, and Zhang 2018 provide alternative theoretical explanations with Bayesian agents who have different priors, but these results are not predicted by their models.

There are no other sizable demographic heterogeneities in motivated reasoning on the politicized topics, neither in direction or magnitude, controlling for party preference.[10] After party controls, demographic treatment effects are statistically indistinguishable from zero, and estimates are precise enough to rule out even modest effect sizes. This result suggests that motivated reasoning is a ubiquitous bias, and even on many issues that are explicitly about demographic groups, such as gender and math ability, racial discrimination, and income mobility, many motivated beliefs about the outside world are principally driven by politics.

However, when subjects are asked about their own performance relative to other subjects, there is a large gender heterogeneity. Here, men motivatedly reason to believe they outperformed others, and women do not motivatedly reason in either direction on average.[11] Motivated reasoning can help explain the gender gap in overconfidence, and more broadly indicates that politically-driven motivated reasoning is a more universal phenomenon than ego-driven motivated reasoning in the United States today.

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.[12] The design for these papers typically involves giving subjects informative signals and testing for asymmetric updating from "Good" and "Bad" news, and thus requires disentangling motivated biases from nonmotivated biases such as underinference from information and prior-confirming bias. 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 used to test motivated reasoning on a wide variety of topics.

The rest of the paper proceeds as follows: Section 2 develops the model of motivated reasoning, generating testable predictions. Section 3 introduces the experimental design and

[10] Demographics include race, gender, income, age, education, religion, and location.

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

[12] It is worth noting that there is more consistant 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 Appendix B.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.

hypotheses corresponding to these predictions. Section 4 discusses further details of the experiment and data.

Section 5 analyzes the main results from the experiment. The primary specifications and results are about over-trusting Pro-Party news and error-accentuating news. I also show a number of robustness checks, evidence for belief polarization, heterogeneities, and how motivated reasoning leads to underperformance and overconfidence.

Section 6 delves into each of the individual topics discussed in Table 1 and discusses how these results relate to prior topic-specific evidence on updating from information. Section 7 concludes and proposes future directions of work.

Appendix A provides additional results that are omitted from the main text. Appendix B discusses results from a willingness-to-pay treatment and structurally estimates a version of the motivated reasoning model. Appendix C lists the exact questions, answers, and sources that subjects see. The online appendices include additional robustness checks, as well as the entire experiment flow with screenshots of each page.

2 Model and Predictions

This section introduces and develops a model of motivated reasoning in which agents treat motivated beliefs as an additional signal when receiving information. The model predicts that people will irrationally over-trust news that supports their motivated beliefs, 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 updating. It also generates secondary predictions under additional functional form assumptions, showing how motivated reasoning affects belief polarization, underperformance, and overconfidence.

2.1 A Model of Motivated Reasoning

I describe the updating process of an agent who has *motivated beliefs*. As an example of *politically*-motivated beliefs, a Republican may be motivated to believe that murder and manslaughter rates increased during the presidency of Barack Obama, and a Democrat may be motivated to believe that rates decreased. As an example of *ego*-motivated beliefs, both Republicans and Democrats may be motivated to believe that they are more knowledgeable than others about such issues.

Motivated reasoning posits that agents misupdate from information about events that they are motivated to believe. In particular, I will define motivated reasoning by extending

the framework of Kahan 2016a in which agents update using a modified Bayes' rule. They act as if they receive an additional signal that tells them what they are motivated to believe; the likelihood of this signal is proportional to the agents' motivated belief function.

To formalize, compare a Bayesian agent (she) to a motivated-reasoning agent (he) when they receive the same signal x about the probability that a state is equal to θ.[13] The Bayesian sets her posterior to be proportional to her prior times the likelihood of the signal:

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

Taking logs and introducing $\neg\theta$ as the complementary state to θ gives the Bayesian logit updating process:

$$\text{logit } \mathbb{P}(\theta|x) = \text{logit } \mathbb{P}(\theta) + \log\left(\frac{\mathbb{P}(x|\theta)}{\mathbb{P}(x|\neg\theta)}\right), \tag{1}$$

The motivated reasoner updates similarly, but he incorporates her prior, likelihood, and motivated beliefs:

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

where $M(\theta) : \Theta \to \mathbb{R}_+$.[14] Define $m(\theta) \equiv \log M(\theta)$ and take log odds ratios to get the motivated-reasoning logit updating process, which will be central to the rest of this section:

$$\text{logit } \mathbb{P}(\theta|x) = \text{logit } \mathbb{P}(\theta) + \log\left(\frac{\mathbb{P}(x|\theta)}{\mathbb{P}(x|\neg\theta)}\right) + \varphi(m(\theta) - m(\neg\theta)). \tag{2}$$

$m(\theta) : \Theta \to \mathbb{R}$ is denoted the **motive** function, i.e. how much the agent is motivated to believe the state is θ.

It will be useful to treat the motive function cardinally in order to study uncertainty. That is, m can be thought of as an *expected* motive function to mirror the standard expected utility function u. (Note that I use cardinality here in the manner of cardinal utility, and not in the sense of mathematical cardinality. As with utility, motive orderings are preserved up to positive affine transformations.)

As discussed before, the motivated reasoner acts as if he receives both the actual signal

[13] Similar definitions follow with continuous states: $\mathbb{P}(\theta)$ is replaced by $f(\theta)$.

[14] Note that there is also a change in the proportionality constant between Bayes and motivated reasoning, but this is not a function of θ.

(x) and the reminder of how much he wants to believe the state is θ (m). He weights the motive signal by parameter $\varphi \geq 0$, called **inference flexibility**. When $\varphi = 0$, the agent updates rationally; when $\varphi > 0$, the agent motivatedly reasons.[15]

Importantly, we will assume that the motive function does not depend on the signal structure. Examples of motives include anticipation, ego, optimism, and believing that your political party is better than its opposition; this paper focuses on motives related to politics and ego.

2.2 Identifying Motives

We now use the above framework to make inferences about people's motives by looking at their updating process. By fixing priors and likelihoods, the difference in updating between motivated and unmotivated questions provides information about motives, and this subsection gives a procedure for identifying a part of the motive function.

Consider an agent with prior on a state θ, $F(\theta)$. 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 μ, and the true probability is indeed zero.

To preview the experimental design that will be developed in Section 3, consider an agent who now receives a message from one of two news sources, True News (TN) or Fake News (FN), and is asked to predict the probability that the message comes from TN. 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	G	L
Fake News	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:

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

[15] Here, φ is treated as an exogenous parameter. It may be a function of signal x and signal set X, but does not depends on motives. For further discussion of φ, see Appendix B.

such that he maximizes utility by stating his subjective belief a.

We can now look at how a Bayesian and a motivated reasoner update about the news source. Given message G, the Bayesian uses Equation (1):

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

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

However, the motivated reasoner uses Equation (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 assessments)
The procedure above identifies motivated reasoning from Bayesian updating:

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

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

In this paper, states will be real numbers and motives will typically be assumed to be monotonic; 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 $m > 0$.[16]

[16]Strictly monotonic motives posit that people are more motivated to have 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 / outside their 50 percent confidence interval.

In the experiment, many predictions will 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 (politics and ego) predict $m(\theta|\theta > \mu) - m(\theta|\theta < \mu)$, and therefore predict $a|G - a|L$.

2.3 Inferring Motives from Beliefs

When motives are unobservable, this structure also allows us to infer about agents' motives by looking at the initial belief μ. The idea is that the error in beliefs can be partly explained by motivated reasoning, and therefore the direction of the error predicts the direction of the motive function. Loosely, a motivated reasoner with $\mu > \theta$ is more likely to have an increasing motive function than a motivated reasoner with $\mu < \theta$, and so they will have previously updated differently from a signal drawn from the same distribution. When they then make news assessments using the structure above, agents will trust news that *accentuates* the error in their beliefs compared to news that *mitigates* the error.

For instance, consider an agent with an unknown motive function whose beliefs are erroneously too high and has median $\mu > \theta$. If this error is due to misinference from past signals, then the direction of the error predicts the direction of motivated reasoning. This implies that he will give a higher veracity assessment to a G message than an L message, in spite of the fact that G is Fake and L is True.

More formally, there is a state $\theta \in (\theta_L, \theta_H)$ with $-\infty < \theta_L < \theta_H < \infty$. Consider a Bayesian and motivated reasoner with diffuse prior $\theta \sim U(\theta_L, \theta_H)$, and suppose that the motivated reasoner has motive $m(\theta)$ that is strictly monotonic in θ. They receive a public signal about θ, $z = \theta + \epsilon_z$ with $\epsilon_z \sim F_z$, and suppose that this signal has a positive probability of taking values close to the truth: For all $\delta > 0$, there exists some $\delta' > 0$ such that $F_z(\delta) - F_z(0) > \delta'$ and $F_z(0) - F_z(-\delta) > \delta'$.

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

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

In the context of the design above, 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-accentuating messages more than error-mitigating messages when motivated reasoning plays a role.

Therefore, since error-mitigating messages are True News and error-accentuating messages are Fake News, this implies that agents give higher assessments to Fake News than True News, even when 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 prior-confirming bias). Practically, it also suggests that excessive trust in Fake News will be more prominent when people hold stronger motivated beliefs.

2.4 Motivated Reasoning, Underperformance, and Overconfidence

Psychologically, this theory posits that motivated reasoning is an *error* in updating. That is, 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.

For simplicity, we assume a linear motive, $m(\theta) = m \cdot \theta$, so that steepness is captured by $|m|$. In the political context, $|m|$ can be thought of as increasing in political partisanship. Using the quadratic utility from above, agents' assessments lead to utility that is decreasing in $|m|$. This implies that motivated reasoners underperform Bayesians, who update the same way as motivated reasoners who have $m = 0$.

Fact 3 (Steeper motives lead to underperformance)
For all $\varphi > 0$ and prior $p \in (0, 1)$, $\mathbb{E}[u(a; m)]$ 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 4 (Steeper motives lead to overconfidence)
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 Appendix A.1.

To intuitively understand why this is generally true, consider a partisan (with a steeper motive) and a moderate (with a less steep motive). The partisan will move her assessments substantially 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 fairly low, because when partisans update more from Pro-Party (Anti-Party) news, they may end up below (above) 1/2, and this leads to underconfidence.

Combining these results, this implies that political partisans will be more overconfident in their beliefs as a causal result of motivated reasoning. With further symmetry assumptions, this also implies that the most confident groups of agents (i.e. partisans) will be the ones who differ the *most* about beliefs.

[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. However, as mentioned later, the experiment in this paper does not allow much room for self-deception because the true state is revealed soon after beliefs are elicited.

3 Experimental Design

3.1 Summary, Timeline, and Topics

The primary goal of the experimental design is to identify motivated reasoning as a bias in updating. On the issues in Table 1, people have preconceived beliefs that may differ and reflect something about what they are motivated to believe. As such, the experiment is designed to take people's current beliefs and construct an environment in which they have the same priors over a state and receive information with the same subjective likelihood, but different hypothesized motivated beliefs.

To fix ideas, consider the following question 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?*

For this question, most Republicans guess larger numbers than most Democrats. Indeed, Appendix A.4 shows that there is a party belief gap for *every* politicized topic in the direction predicted in Table 1. While a belief gap is suggestive, this by itself is not evidence for any mechanism. To test the hypothesis that subjects bias their updating in the direction of their political preference, we see whether subjects find news more trustworthy if it says they should change their beliefs even more in the "Republican" or "Democratic" direction, and to argue that this trust discrepancy is due to motivated reasoning.

The main test of this in the experiment involves three steps:

1. **Beliefs:** Subjects are asked to guess the answers to questions like the refugee 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 instructions pages are in the Online Appendix.
2. **News:** Subjects receive a binary message from one of two 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.

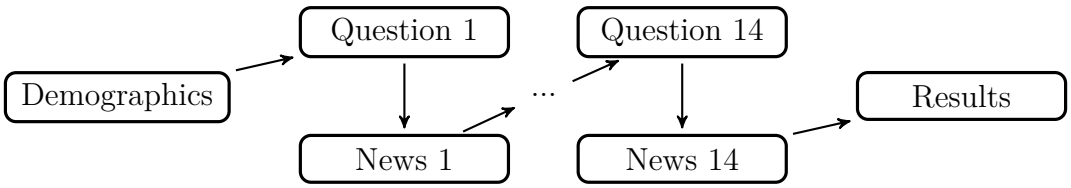
The message says either "The answer is **greater than** your previous guess of [previous guess]." or "The answer is **less than** your previous guess of [previous guess]." Note that the message space is *different* for each subject since subjects have different priors. These customized messages are designed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

For the Crime Under Obama question, "greater than" corresponds to Pro-Republican News and "less than" to Pro-Democratic News. For subjects who give a higher rating to Republicans than Democrats, "greater than" is Pro-Party news and "less than" is Anti-Party news, and vice versa for subjects who give a higher rating to Democrats than Republicans.

3. **Assessment:** After receiving the message, subjects assess the probability that the source was True News on a scale from 0/10 to 10/10 and are incentivized to state their true belief. This is the main outcome measure. The page is identical to the beliefs page but the guess boxes are replaced with assessment choices. The effect of variation in news on veracity assessments is the primary dependent variable for much of this paper. An example of News / Assessment for a subject who gives a guess of 57 is in the Online Appendix.

The general point of this setup is that subjects receive messages that compare the answer to their median, so they should not rationally update their assessment based on the message. This is described in detail in Section 2.2.

Subjects see 14 questions in the experiment, the 13 in Table 1, and one 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.[18]

[18]The quiz questions were: (1) The 2016 Democratic vice-presidential nominee; (2) The winner of the 2017 Alabama Senate special election; (3) The 2017 Supreme Court nominee; (4) The president of France

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 knew that they would see this page and were forced to go through it before exiting the study and receiving payment.[19]

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 subjects' performance compared to other Republican subjects' performance on Questions 1-12.[20]

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 WTP treatment (discussed in the appendix), which occurs for Question 12, does not overlap with the comprehension check.

All of the specific question wordings are in Appendix C. Screenshots for every page are in the Online Appendix.

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 gives a $x/10$ percent chance of winning the bonus.[21]

Questions Page

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.

as of 2018. 43 percent got every question correct.

[19]Subjects spend an average of 71 seconds on this 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.

[20]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'.

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

- *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 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.[22]

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 0.75 quantile and lb to be the 0.25 quantile 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.

See the Online Appendix for a screenshot of this Question page, for the question instructions page, and for the question points system shown to subjects.

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**]."

This is the main treatment variation.

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

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 given each assessment and news source.

All subjects are told that True News *always* tells the truth and Fake News *never* tells the truth, and that sources are iid. We will see that subjects' assessments are consistent with iid sources; the content of previous messages does not affect the assessment of current messages.[23]

See the Online Appendix for a screenshot of this news assessment page, the news assessment instructions, and for the points system shown to subjects.

Occasionally, a subject will correctly guess the answer. If this happens, she skips the news assessment page and moves on to the next question.[24] Exact correct guesses are rare however, because the answers are usually quite precise.

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 purpose of assessing veracity and that they do not differentially value messages on

[23]One-third of subjects are told that True News and Fake News are equally likely. This does not noticeably impact assessments or treatment effects on any question except for the random number question; on that question, subjects who are given this prior are significantly more likely to report $a = 0.5$.

[24]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**]."

politicized and neutral topics (indicating naivete about their motivated reasoning). For more detailed instructions and results, see Appendix B.

3.3 Hypotheses

This subsection summarizes the predictions from Section 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 Pro-Motive news than Anti-Motive news, and (2) the predicted direction of motives is as in Fact 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 ∈ (a|Anti-Party news, a|Pro-Party news).*
- *(a|Pro-Party news − a|Anti-Party news) increases in partisanship.*

As an implication of this, we should expect 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-accentuating news compared to error-mitigating news, as in Fact 2. Since error-accentuating 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 non-neutral 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.*

A related prediction is that subjects will give higher assessments to news that moves them away from the mean population belief compared to news that moves them towards the mean. For the Second Guess subsample, this leads to a form of belief polarization; subjects are more likely to update in the direction of more extreme beliefs than in the direction of more moderate beliefs.

Lastly, motivated reasoning predicts a form of underperformance and overconfidence, as in Fact 3 and Fact 4. As a direct implication of the previous hypothesis, motivated reasoners will earn fewer points with stronger motivated beliefs. In the political realm, the severity of this 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 3 (Politically-motivated reasoning and partisan overconfidence)

- *Average points scored on news assessments will be decreasing in partisanship on politicized topics, but not on neutral topics.*
- *Overall performance relative to other subjects will be decreasing in partisanship.*
- *Predicted performance relative to other subjects will be increasing in partisanship.*

In 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 given a 50-50 prior to begin with.

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 participants tend to have more diverse demographics than students in university laboratories with respect to politics (Levay, Freese, and Druckman 2016). The experiment was coded using oTree, an open-source software based on the Django web application framework (Chen, Schonger, and Wickens 2016).

The study was offered to MTurk workers currently living in the United States. 1387 subjects were recruited and answered at least one question, and 1300 subjects completed the study.[25] Of these subjects, 987 (76 percent) passed simple attention and comprehension checks, and the rest are dropped from the analyses.[26]

[25]This number is updated from an earlier version of this paper, and does not include subjects who reached the Results page but did not click the button to complete the study. Such subjects may not have seen the content on the Results page and are therefore dropped.

[26]In order to pass these checks, subjects needed to perfectly answer the comprehension check question in Appendix C (by giving a correct answer, correct bounds, and answering the news assessment with certainty). In addition, many questions stated bounds (by asking for percentages or fractions); subjects needed to always answer within these bounds. Main results are robust to inclusion of these subjects.

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.[27] These subjects will be labeled as "Pro-Dem," "Pro-Rep," and "Neutral," respectively, and for most analyses the Neutral subjects will be dropped. Results are similar if ideology or party affiliation is used instead, though many more subjects will then be 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.

Treatments were cross-randomized so that 2/3 of subjects would not receive a prior about the veracity of the news source, and 1/3 of subjects would be told this prior was 1/2 True News, 1/2 Fake News; independently, 1/2 of subjects would receive the willingness-to-pay treatment and 1/2 would be in the second-guess treatment. Of the non-Neutral subjects, 66 percent did not receive a prior and 34 percent did; 49 percent were in the WTP treatment and 51 percent were in the second-guess treatment.

There are 13 questions for each subject, so there are a total of 11661 guesses to questions for the 897 non-neutral subjects. There are 11443 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.[28] I drop these 218 observations for news assessment analyses. There are 7902 news assessments on politicized topics, 891 on the own performance topic, and 2650 on neutral topics.

The balance table for the Pro-Party / Anti-Party treatment is in Appendix A.5. Since this randomization is within subject, treatments are expected to be balanced across demographics. Pro-Party news / Anti-Party news are indeed balanced across all topics, round numbers, demographics, and treatment groups. Importantly, the shares of Anti-Party and Pro-Party news are also balanced. This suggests that there was no differential attrition in the experiment by treatment.[29]

[27] This is typical on MTurk (Levay, Freese, and Druckman 2016).

[28] 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.

[29] Relatedly, there were 76 subjects with a party preference who had received at least one message but did not complete the study. Of these, 24 had received more Anti-Party news, 23 had received more Pro-Party news, and 29 had received the same number of each.

5 Results

5.1 Raw Data

This subsection shows that the raw data supports the main predictions of the model, and later subsections show the related regressions. 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 political Fake News will be trusted more than political True News.

The mean assessment of Pro-Party news is 62.0 percent (standard error (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).[30] The difference between every pair of these is highly significant ($p < 0.001$ each).[31]

Figure 1 shows the subject-demeaned assessments by news type (Pro-Party, Anti-Party, Neutral) and subject type (Partisan, Moderate). Subjects give higher average assessments to Pro-Party than to Neutral news, and higher to Neutral than to Anti-Party news, and this difference is larger for partisans. Appendix Figure 4 shows the empirical distribution of assessments for Pro-Party and Anti-Party news on politicized topics. Subjects give higher assessments to Pro-Party news than to Anti-Party news: the former distribution stochastically dominates the latter. Next, Figure 2 shows the subject-demeaned assessments by news type (Pro-Party, Anti-Party, Neutral) and news veracity (True, Fake). The raw data strongly suggest that subjects give higher assessments to Fake News than to True News on politicized topics but not on neutral topics. 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 5 shows the empirical distribution of assessments for True News and Fake News on politicized questions. Subjects give higher assessments to Fake News than to True News: the former distribution stochastically dominates the latter.

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

[31] 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, perhaps subjects motivatedly reason to trust what they are told in general. Further work can explore this latter channel.

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.[32]

In particular, the main specification for politically-motivated reasoning in Table 2 looks at assessments a for subject i, question topic q, and round r with FE for i, q, and r when all news is Pro-Party or Anti-Party:

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

Results are qualitatively identical if we use $\text{logit}(a_{iqr})$ instead of a_{iqr}, as suggested by the model. This form is used for ease of interpretation.

Hypothesis 1 claims that motivated reasoning leads to higher assessments for Pro-Party and lower for Anti-Party; so, Table 2 includes a specification with indicators for both Pro-Party (vs. Neutral) news and Anti-Party (vs. Neutral) news. It also predicts that the Pro-Party / Anti-Party gap is increasing in partisanship, so another specification includes an term interacting partisanship (the absolute difference in party ratings) with Pro-Party news.

Hypothesis 2 claims that for politicized news, subjects will trust Fake News more than True News, so the final two specifications in Table 2 include a dummy for True News.

Both Hypothesis 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.

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.

There is significant evidence for politically-motivated reasoning on all but one of the hypothesized topics asked about.[33] I also include 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.[34] I leave further discussion of individual

[32]Three subjects randomly received no Pro-Party news; two subjects randomly received no Anti-Party news; and all subjects received Neutral news.

[33]This is the main instance where the oversampling of Pro-Dem has an effect, as Pro-Rep subjects motivatedly reason about Media Bias in the Pro-Party direction, while Pro-Dem subjects do not. Also, on the performance topic, only male subjects seem to motivatedly reason towards believing they outperformed others.

[34]There is one amusing caveat to this: Californians of all parties appear to substantially motivatedly reason to think that the center of the US is farther west. They assess "Farther West" news to be *15 percentage points* more likely True News than "Farther East" news (s.e. 5 percentage points; $p = 0.003$). While the magnitude is large, it is unclear how seriously to take this result; it was not hypothesized, and

topics for Section 6.

These results show that people are updating about the veracity of the news source based on the messages received at a substantial rate, and show that we can infer about motives from erroneous priors. The psychology of motivated reasoning and prior-confirming bias both predict that people will trust news that accentuates errors rather than mitigates them, but only motivated reasoning makes the differential prediction in motive-relevant topics compared to neutral topics.[35]

There is evidence that both Republican and Democratic partisans motivatedly reason.[36] But there is a substantial heterogeneity between parties and topics; on many topics, only subjects of one party seems to directionally motivatedly reason, and on two topics, both parties seem to motivatedly reason in the *same* direction but by different magnitudes (Climate Change: both towards Pro-Dem; Media Bias: both towards Pro-Rep).

Consistent with hypotheses, partisans engage in more politically-motivated reasoning than moderates on nearly every topic. In the appendix, Figure 6 splits the topic-by-topic estimates Figure 3 into partisans and moderates. On each politicized topic for which there is evidence that moderates motivatedly reason, the coefficient is larger for partisans than moderates. There is no notable difference on neutral topics.

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, should be more likely to update from Pro-Party news than from Anti-Party news. This test is useful as a robustness check, but also helps us better understand how these messages affect subjects' beliefs about these issues.

As hypothesized, on politicized questions, subjects change in the direction of Pro-Party messages more frequently than from Anti-Party messages.[37]

We can also use these changes to find evidence for belief polarization by showing that subjects are more likely to change their guess away from the mean belief in the population than towards it. Specifically, on politicized topics, subjects are more likely to change their

there is no evidence for similar effects for other west coast states nor opposite effects for east coast states.

[35]Additionally, note that all messages tell subjects that their answer is incorrect. Models of prior-confirming bias would not predict asymmetric trust of such news sources.

[36]However, Pro-Dem subjects tend to have much wider assessment gaps than Pro-Rep subjects. Much of this is due to very moderate Pro-Rep subjects who actually motivatedly reason in the *Pro-Dem* direction on politicized issues. This may be a feature of Mechanical Turk; subjects are more liberal, controlling for party, compared to ANES data. It may also be a feature of the specific topics chosen, which may yield stronger motivated beliefs among Pro-Dem subjects than Pro-Rep subjects.

[37]The intensive margin is similar, but is more complicated to define across questions.

guesses in the direction of a *polarizing* message (one that tells them that their guess is further away from the mean) than from the opposite message. An explanation for this is that polarizing messages are more likely to be in the direction of motives.

It is worth noting that both motivated reasoning and belief polarization can be almost entirely explained by differences in news assessments; when controlling for assessments, these effects essentially vanish. Table 2 summarizes these results.

Since belief changes and news assessments are consistent with each other, this validates the news assessment measure of motivated reasoning. That is, how someone assesses the veracity of a news source is the main determining factor of how she changes her beliefs in this setting.

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.

5.4 Alternative Explanations and Robustness Checks

5.4.1 Misunderstanding "median" and skewed priors

It is reasonable to expect that subjects don't fully understand the concept of a median and answer with their mean belief instead. This would not directionally impact the news assessment results in general, unless the prior distribution were notably skewed. We can use where 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.

34.3 percent of guesses are exactly halfway between the upper and lower bound. In the appendix, Table 8 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 "unskewed" priors. All the results are qualitatively the same, and none of the coefficients are significantly different compared to the full sample.

5.4.2 The independence of news sources

The interpretation of P(True News) in the model and analysis assumes that subjects treat news sources as independent from each other. 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.[38]

[38]See the online appendix for the exact wording.

In Table 9, 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 are treating news sources as independent.

5.4.3 Misunderstanding "Fake News"

First, suppose that subjects believe that messages from Fake News are actually from "Random News" and randomly correct and incorrect, instead of *always* being incorrect. In this experiment, that would not affect any predictions about assessments.[39]

A Bayesian still has an ex ante prior that Pro-Party and Anti-Party messages are equally likely, and does not update about P(True News) given either message. A motivated reasoner who is motivated to believe that the answer is large still updates to think that P(True | Pro-Party) > P(True | Anti-Party).

The more complicated situation is if subjects believe 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 | Pro-Party truth than Pro-Party news | Anti-Party truth.

To test this, we can again look at how subjects change their guesses in Section 5.3. 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 Section 5.3, we see that subjects are similarly likely to update from Pro-Party and Anti-Party news *after controlling for their assessments.*[40] While the data are too imprecise to rule out that there exist subjects who expect Fake News to be asymmetric, this effect seems to be small and not explain much of the asymmetry in updating.

5.4.4 Incorrect initial guesses

Firstly, while it can sometimes be in subjects' best interests to intentionally misreport μ_q in order to earn more points on news assessment questions, I find no evidence of this.

[39] It would affect the likelihood that a subject changes her guess in the direction of the message, but not the relative treatment effects of receiving Pro-Party and Anti-Party news.

[40] This null result is also true for the intensive margin.

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;[41] the difference is 0.8 points (s.e. 1.0) and insignificant ($p = 0.383$). Within-subject tests, controlling for topic, and controlling for linear round trends do not change the null result.

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 harder to rule out. If this is unbiased noise such that the probability that a subject is equally likely to state her Q quantile and her $1 - Q$ quantile, then the main results do not change. However, a bias in the opposite direction from ones party can lead to similar results. For instance, if subjects on average bias their initial guesses towards the population mean, then they may rationally trust Pro-Party news and Fake News more than Anti-Party news and True News.[42]

Psychologically, 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 in which purely time spent affects the extremity of beliefs seems unlikely to explain these results, as the main treatment effect is uncorrelated with time spent on the question page. 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 explanation 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.

5.4.5 Expressive preferences

There is recent evidence showing that people in an experiment may forgo payment in order to make political statements (Bursztyn et al. 2019). In this experiment, if subjects have a preference for stating Pro-Party signals, both their initial guesses and their news assessments

[41] 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.

[42] I would like to thank David Laibson for bringing this alternative explanation to my attention.

will be too much 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.

As seen in Section 5.3, subjects are more likely to change their guesses in the Pro-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.

5.4.6 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 2 to subjects in each treatment. Estimates naturally become noisier, but the direction of every estimate is identical. There is no evidence that any treatment affected motivated reasoning.

It is also possible that subjects learn over the course of the experiment that they motivatedly reason. There is also 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-Motive news than Anti-Motive news. I also restrict Table 2 to Rounds 1-6 and Rounds 7-12, and effects are in the same direction.

5.5 Heterogeneity by Non-Party Demographics

There are two types of heterogeneities to consider: Heterogeneity in the direction of motivated reasoning, and heterogeneity in its magnitude. The main finding is that neither noticeably depend on any non-political demographics, and that we can rule out even moderately large effects.

First, we consider what the *direction* depends on by interacting the direction of the news with observable demographics of race, gender, income, age, education, partisan lean of state, and religiosity. Table 4 starkly shows that, controlling for party preference, *none* of the other demographics have any significant effect on the direction of motivated reasoning.

Not only are other demographics not statistically significant from zero, they are all statistically significant 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 no statistically significant demographic effects, and the party effect seems to dominate.

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 5 takes as given the motives in Table 1 and interacts the predicted direction of motivated reasoning with demographics.

Politically, we see clear 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 either direction on average. This party difference may also be better explained by direction instead of magnitude of bias, as the online sample is liberally skewed conditional on party preference; for instance, only 60 percent of Pro-Rep moderates approved of President Trump's performance.

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.

However, there is a notable exception not captured here: While men do not motivatedly reason more than women on *politicized* topics, Figure 3 showed an enormous gender heterogeneity on the relative performance topic. This suggests that there are not large gender differences in the magnitude of the bias of motivated reasoning, nor differences in motivated political beliefs, but instead are large differences in ego-related motivated beliefs.

5.6 Underformance and Overconfidence

On news assessment questions, subjects typically score worse than if they ignored the message entirely. This is primarily explained by two main 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 Hypothesis 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 6 shows that scores are especially lower for politicized (and performance) topics compared

to neutral topics.[43]

The lower-than-75 scores on neutral topics can be explained by subjects updating with noise.[44] 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 6, 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.

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.[45] On the relative performance question, subjects are asked to predict how many of these 100 they outscored. Hypothesis 3 posits that while more partisan subjects have lower Performance scores, they *expect* to have *higher* Performance scores.

Table 7 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 overall overconfidence; subjects expect to perform approximately at the median. Yet partisans on average score worse than the median and expect to score better than the median.

In addition, Table 7 shows that male subjects are significantly overconfident, evidence that coincides with their ego-driven motivated reasoning. In Section 6.1, I discuss more about gender heterogeneities.

5.7 Discussion

These results strongly support the hypothesis of motivated reasoning with politically-motivated beliefs compared to alternative explanations. Subjects substantially over-trust Pro-Party news and Fake News in an environment with uninformative signals, real monetary stakes, and little room for self-deception. Furthermore, political partisans of all demographics engage

[43] All results are robust to including subjects who did not see a message.

[44] 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.

[45] These overall scores are an average of scores on assessments, guesses, bounds, and either second guesses or the willingness-to-pay round. They are calculated after round 12.

in motivated reasoning, suggesting that this is a widespread bias. This bias leads to other errors and biases such as over-trust in Fake News, underperformance, and overconfidence.

Finally, the results in Section 5.3 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 states θ_q, despite receiving very uninformative news. Importantly, the polarization here does not rely on subjects receiving different news sources in the experiment; this relates to Gentzkow and Shapiro 2011, who find only modest differences in media that liberals and conservatives consume. Motivated reasoning can help explain why people polarize from similar media outlets.

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, heterogeneities across issues and subjects are testable, and the Bayesian null is more easily falsifiable.

6 Individual Topics

As previously discussed, one of the goals of this project is to show how widespread politically-motivated reasoning is; as such, the experiment is designed to be portable to many important economic, political, and social issues. Topics were chosen intentionally to relate to the literature on each issue, and this section goes into more detail on the contribution of this experiment. Recall that specific question wordings from this paper's experiment are in Appendix C.

6.1 Motivated reasoning about performance

There is a growing interest in behavioral and experimental economics about motivated reasoning, and the emphasis typically focuses on motivated beliefs about oneself. Experimentally, the main areas studied have studied beliefs about ones altruism (Exley and Kessler 2018; Di Tella et al. (2015)), beliefs about ones earnings (Mayraz 2013), and beliefs about ones performance (Mobius et al. 2014; Eil and Rao 2011).[46] This experiment contributes to the latter literature by looking at own-relative-performance motives and party-relative-performance motives.

[46]Eil and Rao 2011 also look at beliefs of ones physical attractiveness.

While on average, subjects are overconfident and motivatedly reason in the direction of thinking they outperformed others, this masks a sizeable gender disparity. This directional effect is only present among male subjects; on average, female subjects neither overplace nor motivatedly reason in either direction. Since there is an actual performance gap, the initial overconfidence gap is conceptually consistent with the notion of gender stereotyping from Bordalo et al. 2019 and updating from Coffman, Collis, and Kulkarni 2019. However, the differential motivated reasoning results are in slight contrast to Mobius et al. 2014, who find a gender difference in overconfidence but not in the asymmetry of the updating process.

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 7. Except for the highest-performing women, women of all performance levels expect to score below the median, and men of all performance levels expect to score above the median. While subjects who actually perform better also expect to perform better, the gender confidence gap is at least as large as the confidence gap between the 1st and 100th percentile performers.[47]

Motivated reasoning about relative performance is also gendered; male subjects substantially motivatedly reason towards high performance and there is no evidence of directional motivated reasoning by female subjects. Men assess Pro-Performance news as being 11 percentage points (s.e. 2 percentage points; $p < 0.001$) more likely to be True News than Anti-Performance news. Women assess Pro-Performance news as being 0.05 percentage points (s.e. 3 percentage points; $p = 0.834$) more likely to be True News. Since female subjects engage in *politically*-motivated reasoning as much as male subjects do, this suggests that the performance motive for men is substantial and positive while the performance motive for women is close to zero.

Additionally, I include one question about predicting others' performance by party. Half of subjects are told how the average Democrat scored (using pilot data) and are asked to predict how the average Republican scored, and the other half are reversed. Performance estimates of others are too low, and there is substantial motivated reasoning in the direction of believing subjects of ones party outperformed subjects of the opposing party.

6.2 Gender and ability

Sarsons 2017 studies the effect of surgery outcomes on PCP surgeon referrals and finds a substantial disparity by the gender of the surgeon. Female surgeons are punished more

[47]Linearly regressing guesses on actual performance and gender estimates that the 100th percentile subjects give guesses that are 7 percentiles (s.e. 3 percentiles) larger than 1st percentile subjects, and that male subjects give guesses that are 11 percentiles (s.e. 1 percentile) larger than female subjects.

than male surgeons for unexpectedly negative outcomes, while male surgeons are rewarded more than female surgeons for unexpectedly positive outcomes. These results are consistent with this paper's motivated reasoning model, under the assumption that PCP's (or the surgeons themselves) have motivated beliefs than men are better surgeons, since signals can be interpreted with much noise. Importantly, there is no such disparity on *expected* outcomes, since these outcomes are precisely uninformative. Other work (such as Kunda and Sinclair 1999 and Coffman, Collis, and Kulkarni 2019) show how information can exacerbate discrimination and stereotyping along racial and gender lines.

My experiment contributes to the discrimination and beliefs literature by showing that the updating process may not only be biased, but be *politically*-motivated. Consistent with prior work, I don't find *gender* heterogeneity in motivated reasoning about gender and math ability; however, I do find significant *political* heterogeneity, with Democrats being more motivated to believe that high school girls' math GPAs are high.

6.3 Upward mobility

Alesina, Stantcheva, and Teso 2018 gives subjects precise information to make them more pessimistic about upward mobility and study the effect of these beliefs on support for government intervention. They find a strong first stage for all subjects; but while the second stage is large for left-wing subjects, it is negligible or even negative on right-wing subjects.

The general results are summarized in the following table:

Ideology	First Stage	Second Stage 1	Second Stage 2
Left-wing	Large	Large	Positive
Right-wing	Large	Negligible	Negative

Second Stage 1 asks subjects whether unequal opportunity is a problem and Second Stage 2 asks about support for government tools/policies to combat inequality.

The precise signal (for the first stage) mitigates biased beliefs, but the noisy signal (for the second stage) amplifies political polarization. This is consistent with politically-motivated reasoning in which flexibility is low for the first stage and high for the second stage; it is hard to misinterpret precise news, but easy to misinfer what the government should do about it.

In my experiment, I use a similar question (upward mobility for the cohort that grew up during the Reagan tax cuts) and find evidence for motivated reasoning on the equivalent of the First Stage when subjects have flexible inference.

6.4 Racial discrimination

Haaland and Roth 2019 look at how giving subjects information about labor market discrimination affects their beliefs on discrimination and government policies. As with Alesina, Stantcheva, and Teso 2018, they find that Democrats and Republicans both tend to have first-stage updating about discrimination, but the two parties update in opposite directions about government policies like affirmative action.[48]

It appears that motivated reasoning may be a driving factor in the second-stage discrepancy. In that paper, flexibility may be low in the first stage, leading to little room for motivated reasoning, but the mapping from the existence of discrimination to the importance of government policies can lead to more flexible updating. In this similar setting, motivated reasoning affects initial beliefs in an environment that allows for high flexibility.

6.5 Gun laws

Taber and Lodge 2006 is a classic reference that has formed the basis for many recent experiments on motivated reasoning and prior-confirming biases in political science. Participants receive "confirmatory" or "disconfirmatory" evidence and are asked to rate arguments that support/oppose gun control and affirmative action. They find that subjects with strong priors rate arguments that support their priors higher. They also find evidence that strong-prior subjects spend longer reading confirmatory evidence than disconfirmatory evidence.

These results are explained with a prior-confirming bias in which people overly believe in messages that support their prior.[49] My experiment adds on to this insight by showing that this is not just due to differences in prior or opinion, but also due to politically-motivated reasoning. It also extends this framework to study trust in news about factual events (here, the change in murders before and after a substantial gun reform law in Australia), as opposed to support of opinionated stances. By looking at prior beliefs, this experiment supports the hypothesis that prior-confirming bias plays a role, but demonstrates that the effect of political preference on belief updating is at least as significant.

6.6 Immigration and refugees

My question about the impact of refugees on crime relates to several recent papers which have considered the impact of people receiving news about immigration more broadly. Recent

[48]Interestingly, they also find in an obfuscated follow-up study differential first-stage responses, indicating either experimenter demand or selective memory.

[49]Note that the terminology I use differs from that of Taber and Lodge 2006; they define confirmation bias as being about selective information exposure.

examples include Alesina, Miano, and Stantcheva 2018; Druckman, Peterson, and Slothuus 2013; Grigorieff, Roth, and Ubfal 2018; and Haaland and Roth 2018.

The evidence has been mixed. For instance, in the cross-country surveys in Alesina, Miano, and Stantcheva 2018, respondents have biased estimates about the number, origin, and economic standing of immigrants, thinking that immigrants make up a larger share of the population and are more reliant on government transfers than the truth. Correcting these misperceptions by providing information about the true share or origin of immigrants does not increase support for redistribution. While giving respondents an anecdote about an immigrant who works hard does increase support for redistribution, this treatment effect is lowest among right-wing respondents.

This is consistent with motivated reasoning in which respondents have an anti-immigration motive and right-wing respondents' motive is most extreme. Such a story would explain biased and differing initial beliefs, the minimal updating to information, and the differential updating to the hard work anecdote. It is also consistent with my data on refugees on crime, in which subjects' motive depends on their party preference and subjects' initial beliefs predict their motive.

6.7 News about politicians

Much of the political science motivated reasoning literature looks at perceptions of United States leaders. For instance, Nyhan and Reifler 2013 and Nyhan, Reifler, and Ubel 2013 look at perceptions of Mitt Romney and Barack Obama, the 2012 presidential candidates, and find differential responses to information design to correct for false beliefs. In several cases, corrective messages did not affect subjects' false beliefs at all when the corrections pushed subjects in the opposite direction of their political party's stance.

Politically-motivated reasoning makes these exact types of predictions when the corrective messages are sufficiently flexible or when subjects infer about the veracity of the messages' source. In my experiment, I focus particularly on comparing the murder and manslaughter crime rate in the United States at the beginning and end of Obama's presidency, a topic highlighted frequently in the 2016 presidential election. I am able to isolate that Republicans motivatedly reason to think the crime rate was higher at the end of Obama's presidency and that Democrats motivatedly reason to think it was lower.

6.8 Climate change

My experiment has a similar flavor to Sunstein et al. 2017, who find asymmetric updating from "good news" / "bad news" about climate change based on a measure of subjects' belief

in climate change. However, since news has a different subjective likelihood (in their case, messages are hypothetical scenarios of future scientific reports), it is hard to test motivated reasoning against Bayesian updating, and the authors give examples of how both would occur.

My experiment isolates the motivated reasoning mechanism from Bayesian updating. It also contributes to the discussion in Sunstein et al. 2017 about what a motive actually is; the authors suggest that motives may be perverse since Democrats seem motivated to believe that climate change is severe. While I do find a party difference (which is a test of politically-motivated reasoning), both Democrats and Republicans act motivated to believe that there is a scientific consensus around climate change.

7 Conclusion

Motivated reasoning plays a substantial role in people's assessment of the veracity of news and helps explain why people form inaccurate 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 belief polarization, overconfidence, and an excess trust in Fake News.

One interpretation of this paper is unambiguously bleak: People of all demographics similarly motivatedly reason, do so on essentially every topic they are asked about in the experiment, and make particularly biased inferences on issues they find important. However, there is an alternative interpretation: This experiment takes a step towards better understanding motivated reasoning, and makes it easier for future work to attenuate the bias. Using this experimental design, we can identify and estimate the magnitude of the bias; future projects that use interventions to attempt to mitigate motivated reasoning can use this estimated magnitude as an dependent variable. Since the bias does seem to decrease utility in some settings, people may have demand for such interventions.

A potential path to attenuate the bias involves understanding the determinants of inference flexibility. There is no evidence that flexibility 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 flexibility, while an agent who receives a hard-to-interpret signal will likely have higher flexibility 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 inference flexibility, we can limit the bias in people's updating process.[50]

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 could provide insight on how motives and choices interact.

Finally, while one definition of motivated beliefs is that they are beliefs that increase utility, this paper provides no evidence that people are motivated to believe that the world is good for others, and such an interpretation gives perverse implications about peoples' preferences. Do Republicans really get *utility* from believing that more Americans were murdered during Obama's presidency? Do Democrats really get *utility* from believing that there is rampant racial discrimination in labor markets? These are controversial questions, but ones that are crucial for understanding how people behave in an ultra-politicized world.

[50] Appendix B.2 discusses a simple structural estimation of flexibility under additional assumptions in which we look at the same information structure for unmotivated states. It implies that if an intervention reduces flexibility on unmotivated states, it should lead to a less-biased updating process on motivated states.

References

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*.

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

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*.

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*.

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

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*.

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

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*.

Di Tella, Rafael et al. (2015). "Conveniently Upset: Avoiding Altruism by Distorting Beliefs about Others' Altruism". In: *American Economic Review*.

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*.

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*.

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*.

Gagnon-Bartsch, Tristan, Matthew Rabin, and Joshua Schwartzstein (2018). "Channeled Attention and Stable Errors". In: *Working Paper*.

Gallup (2018). *Military, Small Business, Police Still Stir Most Confidence*. URL: https://news.gallup.com/poll/236243/military-small-business-police-stir-confidence.aspx.

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*.

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*.

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

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*.

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*.

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*.

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*.

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*.

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*.

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*.

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*.

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*.

Stone, Daniel, Matthew Gentzkow, and Jesse Shapiro (2015). "Media Bias in the Marketplace: Theory". In: *Handbook of Media Economics*.

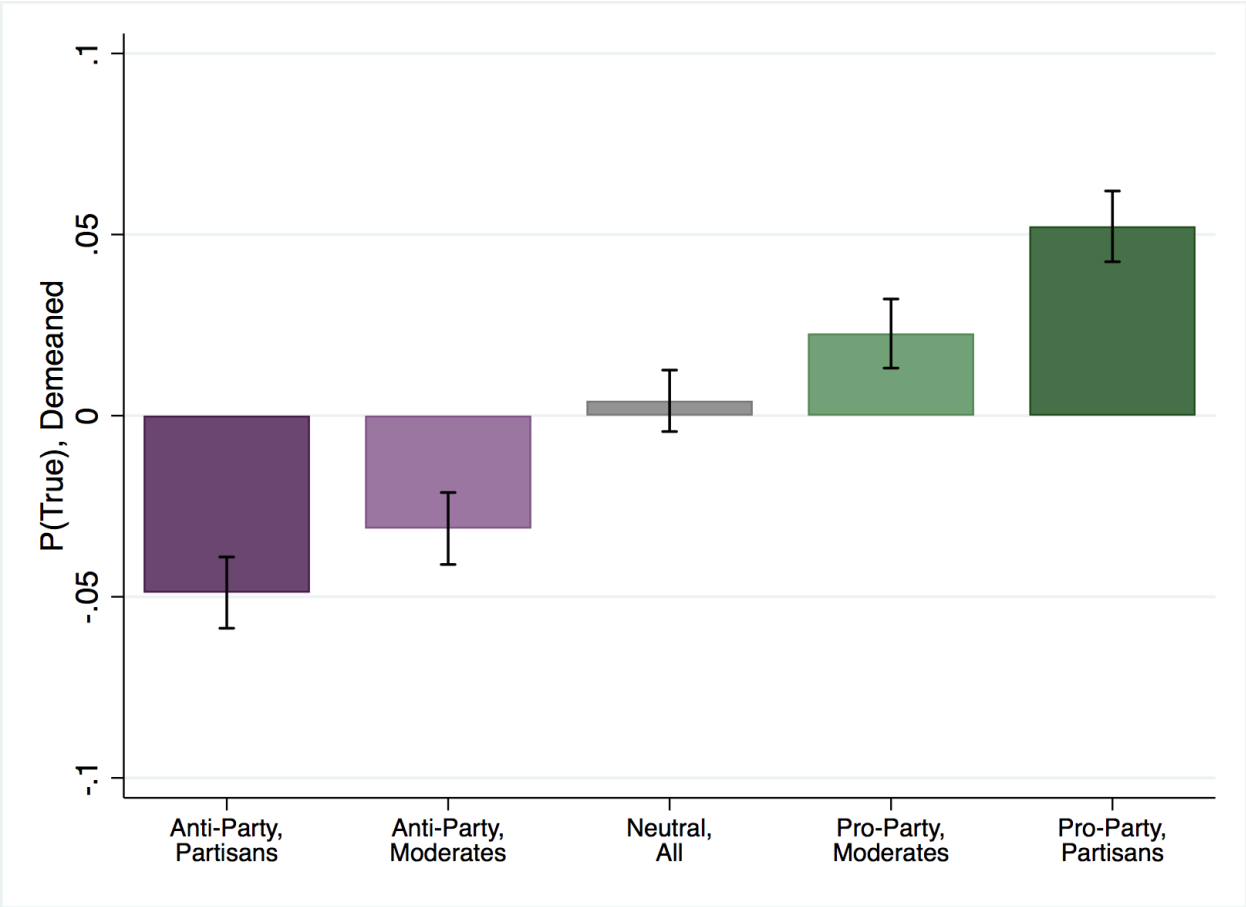
Sunstein, Cass et al. (2017). "How People Update Beliefs About Climate Change: Good News and Bad News". In: *Cornell Law Review*.

Taber, Charles and Milton Lodge (2006). "Motivated Skepticism in the Evaluation of Political Beliefs". In: *American Journal of Political Science*.

Tetlock, Philip (1983). "Accountability and the Perserverance of First Impressions". In: *Social Psychology Quarterly*.

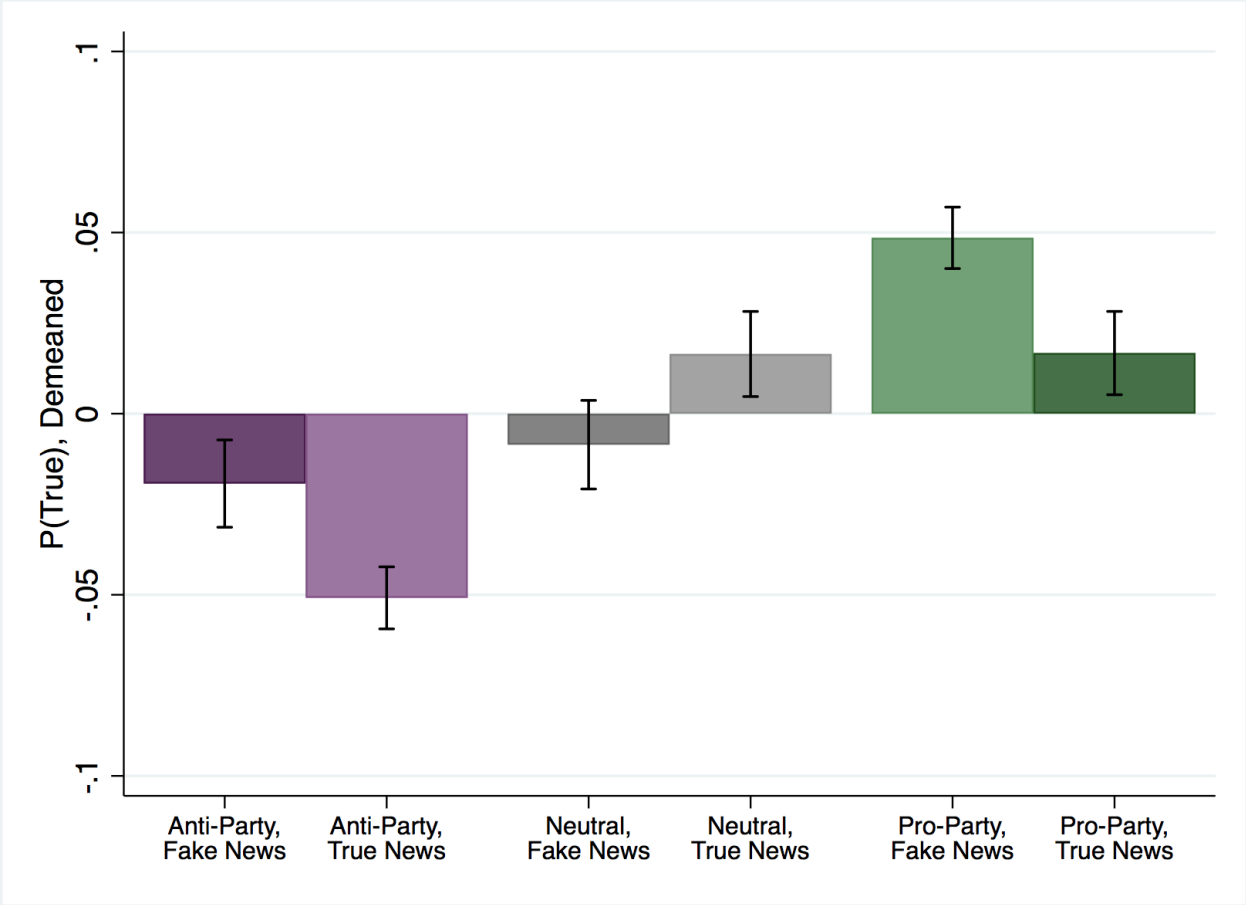
Zimmermann, Florian (2019). "The Dynamics of Motivated Beliefs". In: *American Economic Review, Forthcoming*.

Figure 1: Politically-Motivated Reasoning: Perceived Veracity by News Type and Subject Partisanship



Notes: The y-axis is stated P(True), demeaned by subject-level FE. 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. 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.

Figure 2: Motivated Reasoning and Trust in Fake News: Perceived Veracity by News Type and Actual Veracity



Notes: The y-axis is stated P(True), demeaned by subject-level FE. 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. Fake News sends messages that reinforce the direction of subjects' error; True News sends messages that go against the direction of subjects' error.

Table 2: The Effect of News Type and Actual Veracity on Perceived Veracity

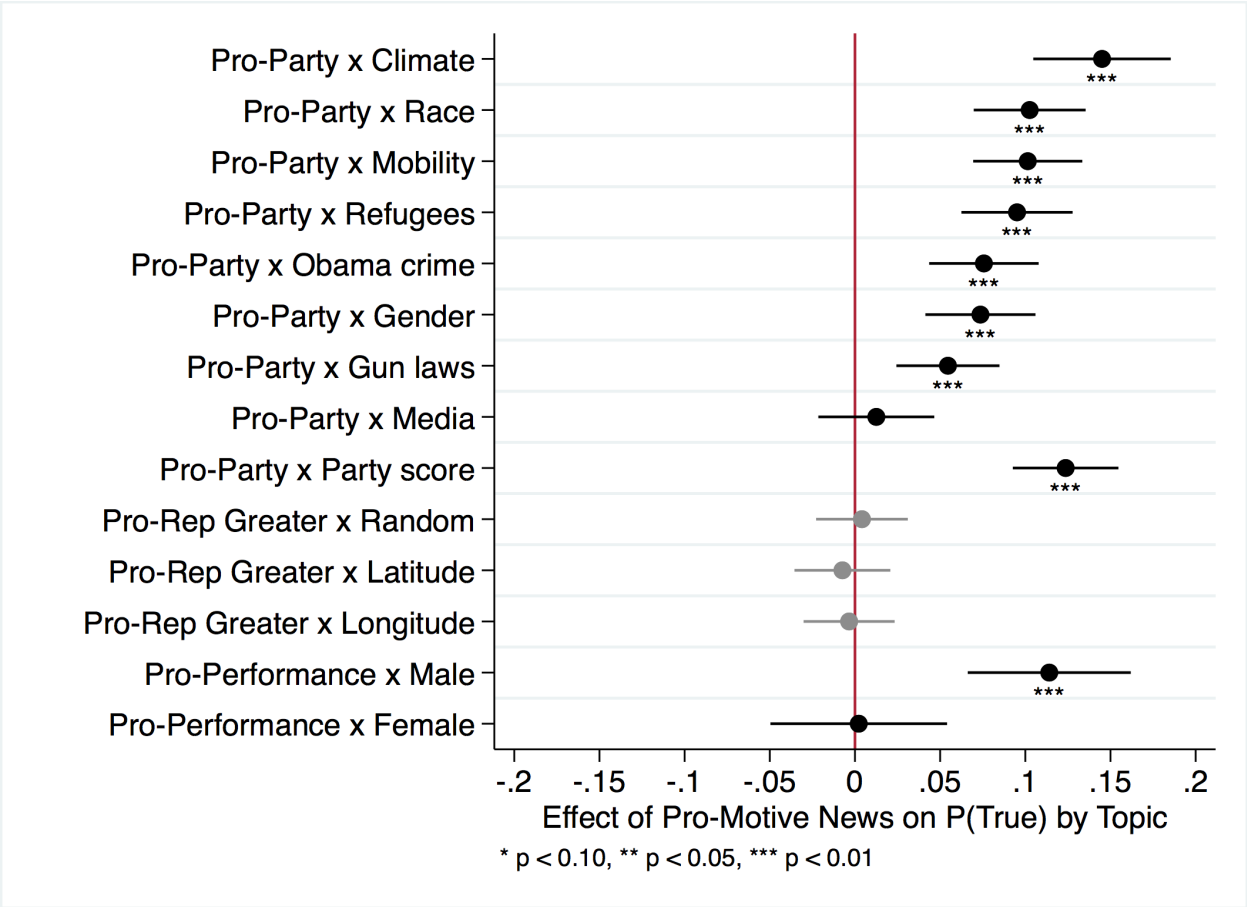
	(1)	(2)	(3)	(4)	(5)	(6)
Pro-Party News	0.091***	0.088***	0.040***	0.037***		0.077***
	(0.006)	(0.007)	(0.013)	(0.007)		(0.007)
Partisanship x			0.050***			
Pro-Party			(0.012)			
Anti-Party News				-0.048***		
				(0.007)		
True News					-0.059***	-0.034***
					(0.006)	(0.007)
Neutral News	No	No	No	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. Controls: race, gender, log(income), years of education, religion, and state. Partisanship is the absolute difference between ratings of the Republican and Democratic parties.

Figure 3: Motivated Reasoning Across Topics: Effect of Pro-Motive News on Perceived Veracity by Topic



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. Pro-Rep Greater is a placebo check to test whether Pro-Rep and Pro-Dem subjects give different assessments on neutral topics.

Table 3: Changing Guess to Follow Message Given News Type

	(1)	(2)	(3)	(4)	(5)	(6)
Pro-Party News	0.092***		0.087***	0.015		0.020
	(0.016)		(0.016)	(0.014)		(0.014)
Polarizing News		0.042***	0.020		-0.015	-0.020*
		(0.014)	(0.014)		(0.012)	(0.012)
P(True)				0.827***	0.839***	0.832***
				(0.040)	(0.039)	(0.040)
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	0.34	0.33	0.34	0.49	0.49	0.49
Mean	0.717	0.717	0.717	0.717	0.717	0.717

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. 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".

Table 4: 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. Pro-Rep: Higher rating for Republican than Democratic Party. Red State: Voted for Trump in 2016. Religious: Subject affiliates with any religion.

Table 5: 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 included for round number and topic. Only Pro-Motive / Anti-Motive news observations, defined in Table 1. Pro-R: Higher rating for Rep than Dem Party. Partisan: Above median for abs(Rep Party rating - Dem Party rating); Moderate: Below median. Red State: State voted for Trump in 2016. Religious: Subject affiliates with any religion.

Table 6: The Effect of Topic and Partisanship on News Assessment Scores

	(1)	(2)	(3)
Politicized Topic	-4.141***	-1.920*	-1.920*
	(0.622)	(1.075)	(1.028)
Partisanship x Politicized Topic		-5.085**	-3.583***
		(1.975)	(1.136)
Partisanship x Neutral Topic			1.534
			(1.617)
Round FE	Yes	Yes	Yes
Subject FE	Yes	Yes	No
Observations	11612	11612	11612
R^2	0.12	0.12	0.01
Mean	69.466	69.466	69.466

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. Subject controls are party preference, race, gender, log(income), education, religion, and home state.

Table 7: Performance and Expected Performance by Partisanship

	Performance			Expectation		
	(1)	(2)	(3)	(4)	(5)	(6)
Partisanship	-4.74*	-3.74	-4.79*	6.13***	9.30***	8.12***
	(2.60)	(2.67)	(2.68)	(2.18)	(2.16)	(2.15)
Male		3.70**	3.66**		11.76***	12.05***
		(1.68)	(1.69)		(1.31)	(1.30)
Subject controls	No	No	Yes	No	No	Yes
Observations	987	987	987	987	987	987
R^2	0.00	0.01	0.04	0.01	0.08	0.13
Mean	47.64	47.64	47.64	50.36	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 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. Subject controls are party preference, race, gender, log(income), education, religion, and home state.

A Supplementary Appendix: Additional Results

A.1 Proof of Hypothesis 3

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:

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

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

Define $\Delta_m \equiv \varphi m\left(\mathbb{E}[\theta|\theta>\mu] - \mathbb{E}[\theta|\theta<\mu]\right)$. We 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 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.2 Raw Data: Pro-Party and Anti-Party News Assessments

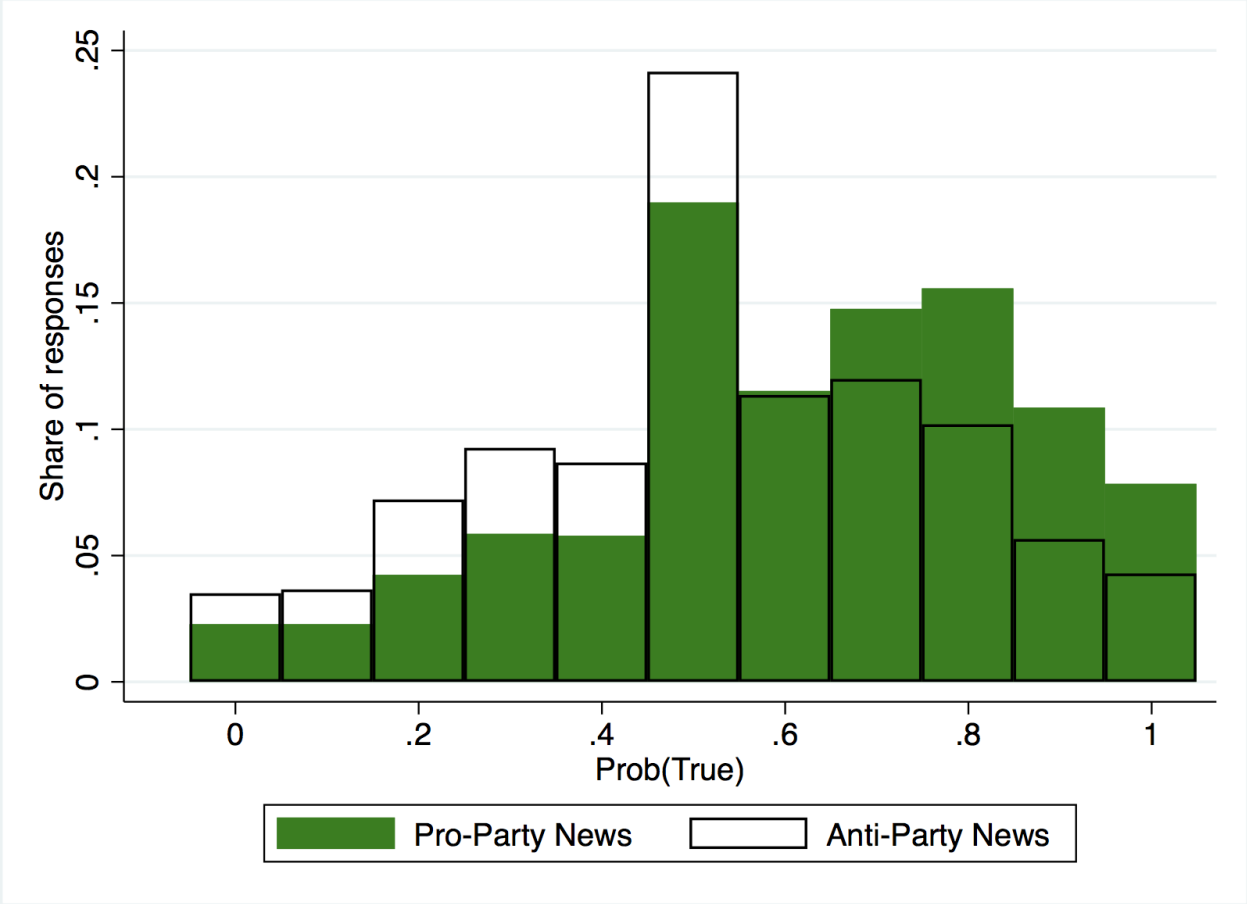


Figure 4: Histogram of Perceived Veracity of Pro-Party and Anti-Party News.

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

A.3 Raw Data: True News and Fake News Assessments

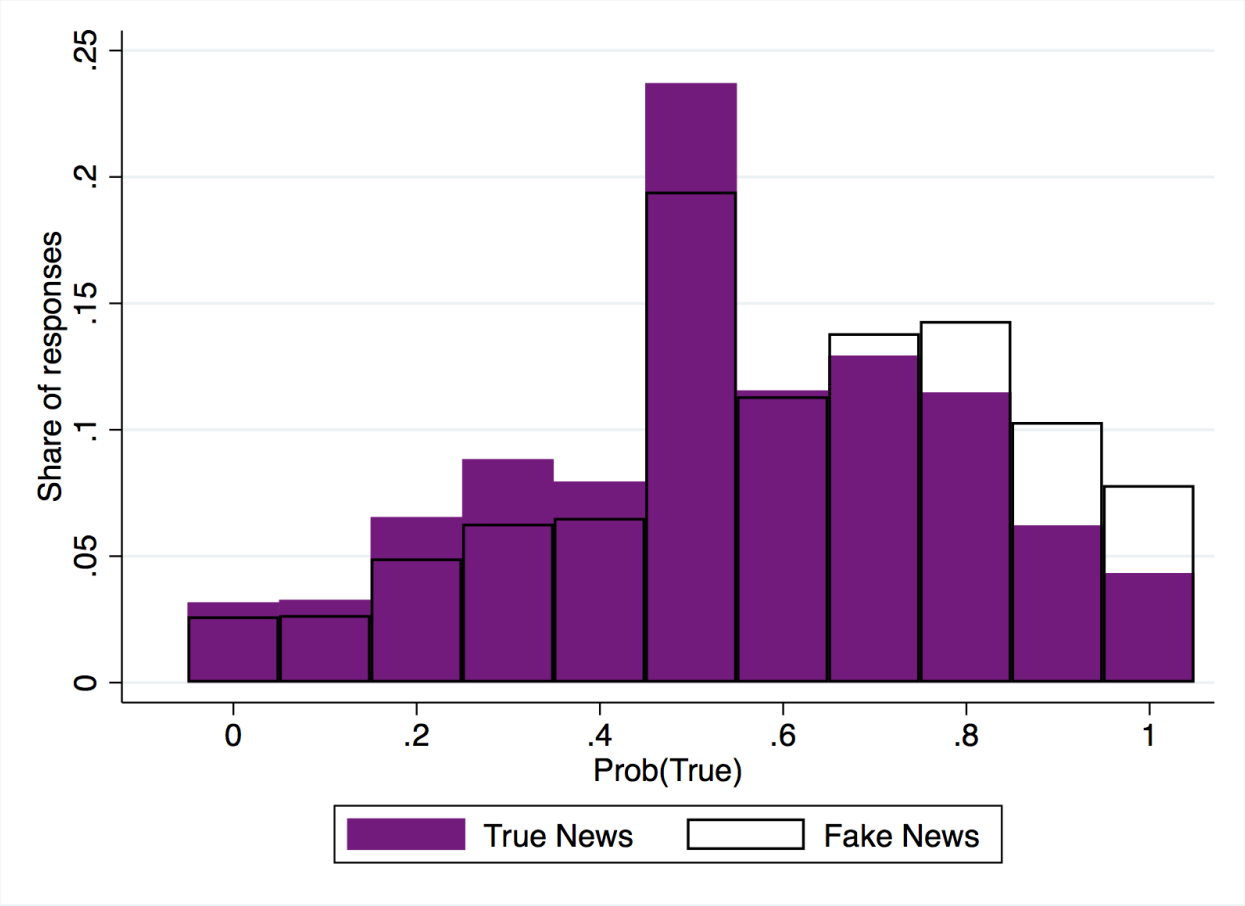


Figure 5: Histogram of Perceived Veracity of True News and Fake News on Politicized Topics.

Notes: Only Pro-Party / Anti-Party news observations, as defined in Table 1. Bayesians would give the same assessment for True News and Fake News.

A.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

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

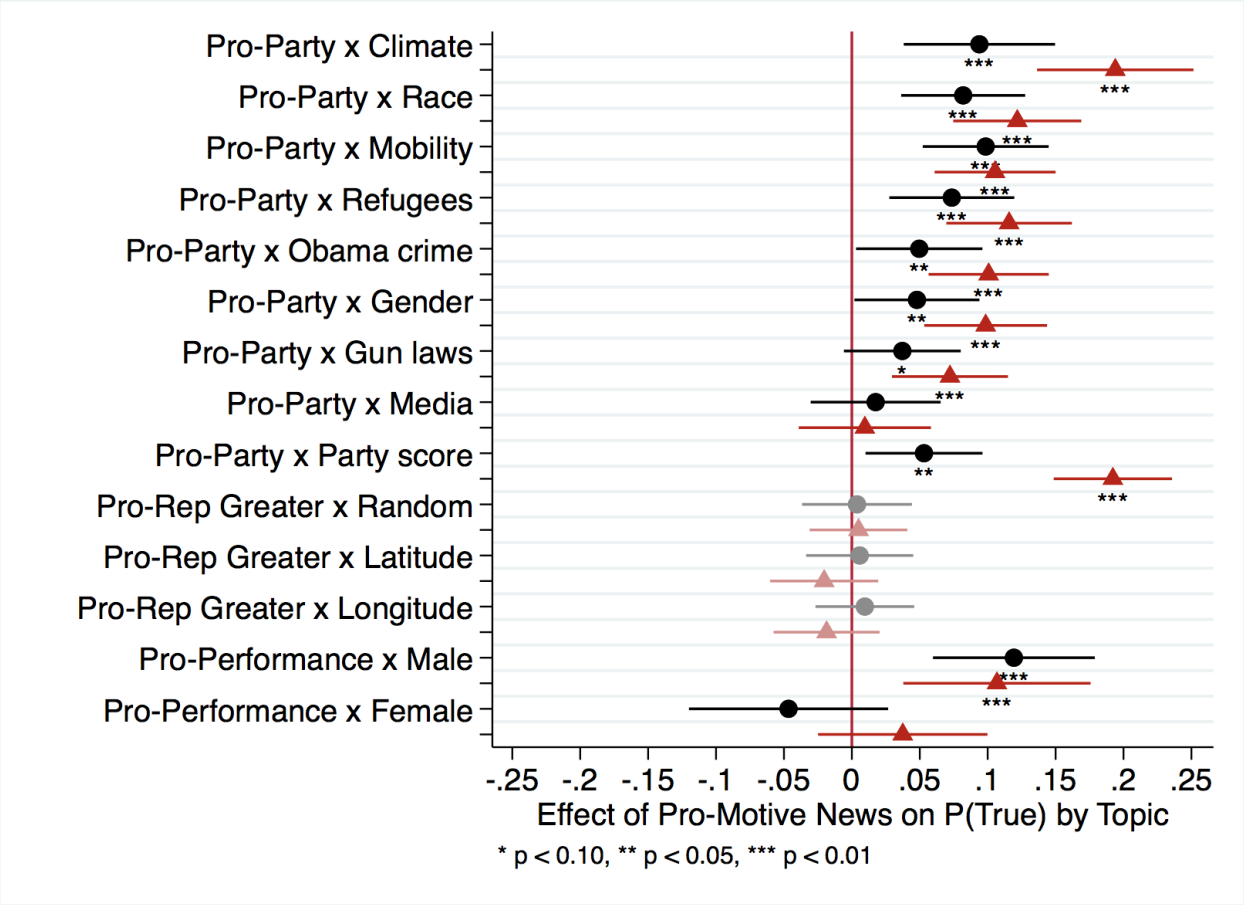
Notes: OLS, errors clustered at subject level. Party-indifferent subjects included. Guesses are winsorized at the 5% level. Third column represents mean Pro-Rep guess minus mean Pro-Dem guess. The sign of every coefficient points in the predicted motive direction from Table 1.

A.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)	
Cook PVI	-0.000	-0.003	0.003	0.155
	(0.001)	(0.001)	(0.002)	
WTP elicited	0.490	0.476	0.014	0.213
	(0.008)	(0.008)	(0.011)	
Given prior	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 Republican Party minus the rating of the Democratic Party and is between -1 and 1. Partisanship is the absolute difference in these ratings. Education is in years. Religious is 1 if subject in any religious group. Cook PVI is Cook's Partisan Voting Index of the subject's home state. WTP elicited is 1 if subject in the willingness-to-pay treatment and 0 if in the second-guess treatment. Given prior is 1 if subject is told that P(True News) is 1/2 and 0 if not.

Figure 6: Motivated Reasoning Across Topics: Effect of Pro-Motive 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-Motive (vs. Anti-Motive) 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.

Table 8: The Effect of News Type, Actual Veracity, and Skewed Prior Distributions on Perceived Veracity

	(1)	(2)	(3)	(4)
Unskewed	-0.010	-0.003	-0.011	0.004
	(0.009)	(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	-0.001	0.016	0.007
	(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			0.001	
			(0.013)	
True News				-0.027***
				(0.008)
Unskewed x True News				-0.022
				(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	7882	7882	10499	7882
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: party, race, gender, log(income), education (in years), religion, state. 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. 0.5% of obs removed for having a bound range of 0.

Table 9: The Effect of News Type, Actual Veracity, and Previous News Types on Perceived Veracity

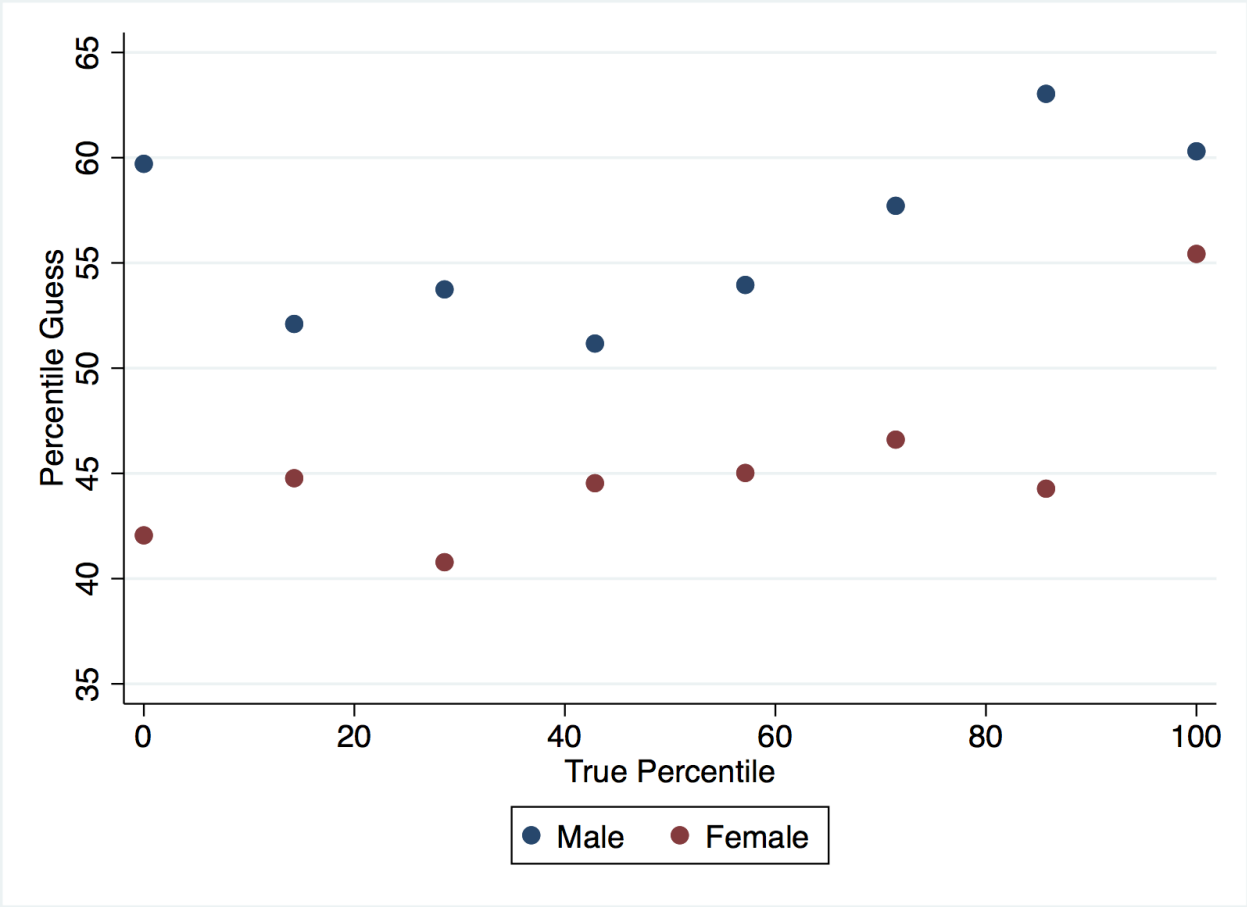
	(1)	(2)	(3)	(4)
Previous Pro-Party	-0.001	-0.001	-0.002	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)
Pro-Party News	0.087***	0.039***	0.036***	0.076***
	(0.007)	(0.013)	(0.007)	(0.007)
Partisanship x Pro-Party		0.050***		
		(0.012)		
Anti-Party News			-0.048***	
			(0.007)	
True News				-0.034***
				(0.007)
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: party, race, gender, log(income), education (in years), religion, state. 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 7: Bin-Scatter Plot of Expected Performance by Gender and 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.

Table 10: The Effect of Topic and Partisanship on Overprecision

	(1)	(2)	(3)	(4)
Politicized topics	0.139***	0.141***	0.168***	0.141***
	(0.009)	(0.017)	(0.010)	(0.018)
Partisanship x Politicized		0.061*		0.061*
		(0.031)		(0.032)
Partisanship		0.020		
		(0.031)		
Round FE	Yes	Yes	Yes	Yes
Subject FE	No	No	Yes	Yes
Observations	11844	11844	11844	11844
R^2	0.03	0.04	0.17	0.17
Mean	-0.001	-0.001	-0.001	-0.001

Standard errors in parentheses

OLS, errors clustered at subject level.

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

Notes: Dependent variable is overprecision, which takes 0.5 if the correct answer lies outside the subjects' 50% confidence interval and -0.5 otherwise. OLS, errors clustered at subject level. Party-indifferent subjects included. Politicized topics are defined in Table 1. Partisanship is the absolute difference between Republican and Democratic ratings.

Table 11: The Effect of Actual Veracity on Perceived Veracity by Overprecision

	(1)	(2)	(3)	(4)
True News	-0.061***	-0.058***	-0.058***	-0.056***
	(0.006)	(0.006)	(0.006)	(0.006)
Overprecision		0.030***		0.025***
		(0.008)		(0.008)
Overprecision x True News		-0.057***		-0.057***
		(0.011)		(0.011)
Question FE	Yes	Yes	Yes	Yes
Round FE	Yes	Yes	Yes	Yes
Subject FE	No	No	Yes	Yes
Observations	8696	8696	8696	8696
R^2	0.02	0.03	0.23	0.23
Mean	0.573	0.573	0.573	0.573

Standard errors in parentheses

OLS, errors clustered at subject level.

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

Notes: Dependent variable is news veracity assessment. OLS, errors clustered at subject level. Party-indifferent subjects included. Overprecision is a dummy variable that takes 0.5 if the correct answer lies outside the subjects' 50% confidence interval and -0.5 otherwise.

B Supplementary Appendix: Demand for News, Flexibility, and Structurally Estimating Motives

This appendix section discusses awareness of motivated reasoning and inference flexibility. 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 flexibility is related to the noisiness of the updating process. In particular, we modify Equation (2) as follows:

$$\text{logit } \mathbb{P}(\theta|x) = \text{logit } \mathbb{P}(\theta) + \log\left(\frac{\mathbb{P}(x|\theta)}{\mathbb{P}(x|\neg\theta)}\right) + \varphi(m(\theta) - m(\neg\theta)) + \epsilon, \tag{3}$$

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 flexibility.[51]

B.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 $10 bonus in the experiment.[52] 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.

[51] If flexibility 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.

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

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.

B.2 Inference Flexibility and Demand for Messages

This subsection aims to use variance in assessments and demand for messages (WTP) to show that flexibility, φ, is positive, and to argue that subjects are unaware of their directionally-motivated reasoning. This uses the parametrization from Equation (3); in this case, flexibility 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 flexbility is positive and *expected* flexibility 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-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 flexibility.[53]

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.

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

B.3 Structural Estimation

The more precise measure of flexibility allows for an analytical structural estimation of Equation (3). In particular, we restrict to *linear* motive functions $m(\theta) = m \cdot \theta$ and define flexibility φ as the standard deviation of noise in subjects' updating process as above.

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

1. $m(\theta) = 0$ for neutral topics. This allows for identification of φ through variance in assessments on neutral topics.
2. φ is fixed across questions and individuals. The former is necessary to separately identify $m(\theta)$ and φ. If φ is allowed to vary across individuals, the model is exactly identified and estimates are very unstable.[54]

Assuming that subjects have normally-distributed priors, Equation (3) 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.[55]

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

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

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.[56]

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 Appendix B.5.

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

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

[56] 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.

We end up with the following estimates:

$$\begin{aligned}
\hat{m}_{iy} &= \frac{\text{logit } a_{iy} - \text{logit } \hat{p}_i}{\hat{\varphi} R_{iy}}. \\
\text{logit } \hat{p}_i &= \frac{1}{N} \sum_n \text{logit } a_{in} \\
\hat{\varphi}^2 &= \frac{1}{IQ} \sum_{i,n} \left(\text{logit } a_{in} - \text{logit } \hat{p}_i\right)^2 .
\end{aligned} \tag{5}$$

Estimated motives are proportional to the change from logit assessment and logit prior, and decrease in flexibility. Estimated priors are equal to the average logit assessments on neutral questions. Estimated flexibility 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.[57]

Now, we can solve the set of equations in Equation (5) 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 3.

B.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 13 using three variants of the main predictions. First, the sign of the estimated motives are in the hypothesized direction from Table 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 ones 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, $p = 0.592$).[58]

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

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

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

$|m_{i,\text{Obama and crime}}|$ is 0.126, and the average $|m_{i,\text{Guns and crime}}|$ is 0.026.[59] 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 magnitude 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.

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

B.5 Structural Estimation Calculation Details

Recall the log likelihood:

$$\sum_{i,q} \log f_{iq} = \frac{IQ\log(2\pi)}{2}\log\hat{k} + \frac{1}{2\hat{k}^2}\sum_i\left[\sum_n(\text{logit } a_{in} - \text{logit } \hat{p}_i)^2 + \sum_y(\text{logit } a_{iy} - \text{logit } \hat{p}_i - \hat{k}\hat{m}_{iy}R_{iy})^2\right], \tag{6}$$

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 $\hat{m}_{iq}$:

$$\frac{\partial\left(\sum \log f_{iq}\right)}{\partial \hat{m}_{iq}} = 0 = \frac{1}{2\hat{\varphi}^2}(-2\hat{\varphi}R_{iy})(\text{logit } a_{iy} - \text{logit } \hat{p}_i - \hat{k}\hat{m}_{iy}R_{iy}) = 0$$
$$\implies \hat{m}_{iy} = \frac{\text{logit } a_{iy} - \text{logit } \hat{p}_i}{\hat{\varphi}R_{iy}}. \tag{7}$$

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:

$$\text{logit } \hat{p}_i = \frac{1}{N}\sum_n \text{logit } a_{in}. \tag{8}$$

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:

$$\hat{\varphi}^2 = \frac{1}{IQ}\sum_{i,n}\left(\text{logit } a_{in} - \frac{1}{N}\sum_{i,n'} \text{logit } a_{in'}\right)^2 = \frac{1}{IQ}\sum_{i,n}\left(\text{logit } a_{in} - \text{logit } \hat{p}_i\right)^2. \tag{9}$$

Figures for Appendix B

Table 12: Determinants of Willingness-To-Pay

	(1)	(2)	(3)	(4)
Assessment SD		22.655***	22.605***	20.343**
		(8.421)	(8.470)	(8.600)
Politicized topics	1.093		1.052	
	(1.697)		(1.701)	
Constant	8.448***	4.234**	3.466	6.890*
	(1.498)	(2.078)	(2.436)	(3.616)
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 13: 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.

C 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`