Is Journalistic Truth Dead? Measuring How Informed Voters Are about Political News

Charles Angelucci, Andrea Prat

SciencesPo
LABORATOIRE INTERDISCIPLINAIRE
D'ÉVALUATION DES POLITIQUES PUBLIQUES

LIEPP Working Paper
January 2022, nº128

Is Journalistic Truth Dead? Measuring How Informed Voters Are about Political News

Charles ANGELUCCI
MIT
cangeluc@mit.edu

Andrea PRAT
Columbia University
ap3116@gsb.columbia.edu

How to cite this publication:
ANGELUCCI, Charles, Andrea PRAT, **Is Journalistic Truth Dead? Measuring How Informed Voters Are about Political News,** *Sciences Po LIEPP Working Paper* n°128, 2022-01-20.

Is Journalistic Truth Dead? Measuring How Informed Voters Are about Political News*

Charles Angelucci† Andrea Prat‡

First Version: July 2020

This Version: December 2021

Abstract

How many voters are informed about political news mainstream journalists consider important? We develop a methodology that combines a protocol for identifying major news stories, online surveys, and the estimation of a model that disentangles individual information precision from news story salience and partisanship. We focus on news about U.S. politics in a monthly sample of 1,000 voters repeated 8 times. On average, 85% of individuals are able to distinguish the major real news story of the month from fake news. 59% of individuals confidently believe this news story to be true, 39% are uncertain, and 3% confidently believe it to be false. Our results indicate that the starkest pattern about the ability of voters to identify major news stories is not the generalized death of truth or its ideological polarization but rather its unequal distribution along socioeconomic lines.

JEL: L82, D72, D83, D90.

Keywords: media, inequalities, polarization, information

*We thank Roland Bénabou, Alessandro Bonatti, Laurent Bouton, Julia Cagé, Teresa Esteban Casanelles, Simone Galperti, Leo Goldman, Andy Guess, Emeric Henry, Amit Khandelwal, José L. Montiel Olea, Suresh Naidu, Jacopo Perego, Jett Pettus, Vincent Pons, Miklos Sarvary, Andrey Simonov, Anya Schiffrin, Carlo Schwarz, Noam Yuchtman, Sevgi Yuksel, and seminar audiences at Bocconi, Brown, UC Berkeley, UBC, Columbia University, CSEF, the University of Exeter, Harvard University, IIOC 2021, Indiana University, Johns Hopkins, Northwestern University (MEDS), the NYC Media Seminar, the University of Oslo, Sciences-Po Paris, the Stockholm School of Economics, Toronto Rotman, the 4th Summer Workshop in Political Economy at EIEF, and the Rational Inattention and Political Economy Conference at UC San Diego for useful suggestions and comments. We are grateful to Szymon Sacher for outstanding research assistance. We gratefully acknowledge funding from The Media and Technology Program at Columbia Business School. This paper formerly circulated under the title "Measuring Voters' Knowledge of Political News."

†MIT. 100 Main St, Cambridge, MA 02142. Email: cangeluc@mit.edu

‡Columbia University. 3022 Broadway, New York, NY 10027. Email: ap3116@gsb.columbia.edu

1 Introduction

The news media plays a key role in modern democracy by providing citizens with the information they need to keep government accountable. This effect has been documented in numerous contexts by a growing body of literature [see e.g., Eisensee and Strömberg, 2007, Ferraz and Finan, 2008, Gerber et al., 2009, Snyder and Strömberg, 2010, Enikolopov et al., 2011, Banerjee et al., 2012, Kendall et al., 2015, Arias et al., 2018, 2019, Labonne et al., 2019, Chen and Yang, 2019, Knight and Tribin, 2019].[1]

In recent years, the increased awareness of the importance of the news media has been accompanied by widespread concerns about voter information. A key issue has been the spread of political fake news [Lazer et al., 2018]. Misinformation spread through social media is identified as an important factor in elections around the world [e.g., Levitin, 2016, Stengel, 2019]. The potential effect of fake news is not limited to the spread of falsehoods. Vosoughi et al. [2018] document how on Twitter "false news reached more people than the truth; the top 1% of false news cascades diffused to between 1000 and 100,000 people, whereas the truth rarely diffused to more than 1000 people." The spread of fake news could also have the indirect effect of blocking the distribution of accurate information [Meyer, 2018]. As information becomes cheaper, individual attention is the binding constraint, and we may be unable to acquire accurate information because we are drowning in an ocean of irrelevant or false information [Gleick, 2011]. These "flooding" tactics can even be used on purpose to suppress inconvenient truths [Tufekci, 2017].

References to the "death of truth" and the "post-truth world" have become commonplace and are often accompanied by calls for immediate action to counter this risk [e.g., d'Ancona, 2017, Ball, 2018, Kakutani, 2018, Baggini, 2017]. As Barack Obama put it, "One of the biggest challenges that we have to our democracy is the degree to which we don't share a common baseline of facts."[2] Indeed, a number of initiatives have been launched or proposed, including media literacy efforts, engagement programs, fact-

[1] See Strömberg [2015] for a survey. Media bias also affects political outcomes [on this and related topics see, for instance, Mullainathan and Shleifer, 2005, DellaVigna and Kaplan, 2007, Gentzkow et al., 2015, Martin and Yurukoglu, 2017].

[2] Interview with David Letterman in January 2018. Obama added: "We are operating in completely different information universes. If you watch Fox News, you are living on a different planet than you are if you, you know, listen to NPR."

checking platforms, software solutions to block false statements, as well legal reform and public funding of journalism [Schiffrin, 2020, Lazer et al., 2018, Guess et al., 2020a]. Constitutional scholars have suggested that we re-think the role of the First Amendment to restrict the dissemination of falsehoods [Bollinger and Stone, 2018, Sunstein, 2019]. Wu [2018] argues that technology should make us re-assess freedom of speech: the First Amendment was designed at a time where information was scarce, but this is no longer the case ("Gone are the days when the CBS evening news might reach the nation automatically, or whatever made the front cover of the New York Times was known to all."); in a world where information flood is an important risk, the First Amendment may be "obsolete."

Given the stakes involved, it is particularly important to gather as much evidence as possible about the "post-truth" phenomenon. Recently, some scholars have begun collecting more systematic evidence of the effect of fake news on voter information [see e.g., Allcott and Gentzkow, 2017, Barrera Rodriguez et al., 2018, Nyhan, 2020a,b]. A number of papers has found that fake news may not be as widely believed or circulated as initially feared [see e.g., Grinberg et al., 2019, Allen et al., 2020a, Guess et al., 2019, 2020b,c] and that people are able to recognize their falseness [see e.g., Pennycook et al., 2019, Pennycook and Rand, 2019, Pennycook et al., 2020, 2021, Allen et al., 2020b].

However, perhaps surprisingly, there is no systematic evidence on real news. How accurate is Obama's statement that we live in different political information universes and do not share a common baseline of facts? Given a set of recent important political events, how many voters are informed about them and are able to distinguish them from fake news? The existing evidence, surveyed in the Literature Section, cannot answer this question because, while there are surveys on news information, none of them defines its "baseline of facts" but rather selects information items – and the text of such items – in unspecified ways. If we do not have a criterion to determine the set of stories people are quizzed on, we will be unable to interpret the results both in absolute terms (were the stories easy or abstruse? new or old?) and in relative terms (were items selected and written in a way that makes them more familiar or credible to certain segments of voters?).

We report the results of a large-scale repeated survey aimed at measuring to what extent U.S. citizens are informed about recent political events that mainstream journalists consider important. Our methodology can be articulated around three steps.

As mentioned above, the first step consists of identifying the "journalistic truth" over which individuals are tested. In our protocol, stories are selected by a panel of U.S. mainstream journalists, who are asked to choose the three major domestic politics news stories of the past month among the set of Reuters wires related to Federal politics.

We make no claim that news stories are selected or written in an objective way. The only goal is transparency. The subjectivity in our protocol can be ascribed to a well-defined set of actors: a large for-profit news organization like Reuters and a panel of professional journalists. We do not claim that those actors are more impartial or knowledgeable than other members of our society. We choose them because they are representative of commercial news production. Our exercise should be seen as an attempt to measure the informational distance between different segments of US voters and mainstream journalists.

We run surveys on the news selected according to the protocol above. We repeat the survey for 8 months on 8 different panels of approximately 1,000 U.S. voters. Respondents are selected by YouGov, a polling company, to produce a nationally representative sample of US adult citizens. We extend the protocol to news about the Democratic Party presidential primaries and news about Sports and Entertainment, all selected according to the same protocol.

Second, we administer financially incentivized quizzes similar to those used by, for instance, Allcott et al. [2020], Guess [2015], Prior et al. [2015], Bullock et al. [2015], and Chen and Yang [2019].[3] As part of the survey, respondents take multiple quizzes. In each quiz, respondents are incentivized to select the 3 most important news stories of the month according to our panel of journalists from a list that also contains 3 fake domestic politics news stories that could have happened in the same time span. We use two methods to generate fake news, and we run multiple surveys based on both methods. The first method consists in identifying three widely circulated fake news about the Federal Government by using snopes.com, a leading fact-checking website. For the second approach, we rely on our panel of journalists to write false statements about the Federal Government.

[3] On the role of partisan congruence and incentives to recall information accurately see Prior et al. [2015] and Bullock et al. [2015]. Both papers show that monetary incentives lead to less party cheerleader behavior in answering survey questions. On the effects of monetary incentives in surveys that measure political information see also Prior and Lupia [2008].

In the third step, we use the data to estimate the parameters of an information model. This is a necessary step because the raw response rates cannot be interpreted directly. Individuals may randomize among statements they do not know and they may use information about a statement being true or false to rule in or rule out other statements. This challenge relates to the problem of inferring underlying parameters from multiple-answer tests, and we approach it with techniques inspired by Item Response Theory [see e.g., Bock, 1972]. The main adjustment is that while that theory deals with a "vertical" parameter only (ability), our response rates are also affected by a "horizontal" parameter (partisan congruence between a respondent and a statement).

In our model, when an individual is confronted with a news story, she forms a *belief* about the story, defined – in a standard Bayesian way – as the probability she assigns to the story being true. The belief is a continuous variable with a value between zero and one that depends on: (i) features of the story like salience and partisanship; and (ii) features of the individual like her information level and partisanship. The individual uses these probabilities to select the 3 stories he or she thinks are most likely to be true.

The model yields a discrete choice specification that can be estimated with standard Bayesian techniques. The main object of interest is the posterior distribution of the individual-level information parameter, but we also obtain estimates for the salience of each story, as well as other parameters of the model such as the effect of time on information and the strength of partisan congruence. We use the model's parameter estimates to analyze individuals' ability to identify true and fake news as well as the probability they assign to news stories being true or false.

Our findings can be split into aggregate and disaggregate. Let us begin with the aggregate results, which paint a sobering picture. Journalistic truth is not dead. A majority of U.S. voters confidently identify the news stories that journalists consider most important and confidently reject widely circulated or well-crafted fake stories. However, there is also a sizeable minority of people who are unable to identify the main real news stories of the month.

We first look at individuals' ability to identify true and false statements. Our main specification predicts that, if faced with 1 typical true news story and 1 typical fake news story, on average between 81 and 84% of individuals will correctly identify the true news story. If faced with 1 true news story and 3 fake news stories, between 64 and 69% of individuals will correctly identify the true news story.

Next we analyze the probability individuals assign to news stories being true or false. As beliefs are continuous variables, our results can be reported at different certainty thresholds. For now, let us say that an individual has a *correct belief* with respect to a true story if she believes it to be true with probability 75% or higher, that she is *uncertain* if she believes it to be true with a probability between 25% and 75%, and that she has a *wrong belief* if she believes it to be true with a probability lower than 25%. This definition means that a (risk-neutral) individual with a correct belief would be willing to accept a bet whereby she wins $1 if the story is true and she loses $3 if the story is false. The converse definition applies to falsehoods: e.g. an individual has the correct belief with respect to a fake story if she believes it to be false with probability 75% or higher.

Our main specification indicates that, on average, 59% of individuals have correct beliefs about the most important real story of the month, 39% are uncertain, and 3% have wrong beliefs. For the second-most important story of the month, the percentages are 44%, 52%, and 4%. For the third-most important story of the month, they are 31%, 62%, and 7%. Instead, only a minority of individuals believe fake news. Looking at fake news that circulated online, on average, 45% of voters have correct beliefs, 45% are uncertain, and 1% have wrong beliefs. The percentages for the synthetic fake news written by our panel of journalists are similar.

To get a benchmark outside of politics, we perform a similar exercise for news about Sports and Entertainment. Sports and Entertainment compete for individuals' attention [Prior, 2005], are fragile to fake news [Pew, 2019], and are sometimes used to illustrate Americans' alleged lack of interest or information about politics. Aided by the panel of journalists, we follow the same protocol to select the three most important news stories of the month about U.S. entertainment and the three most important stories of the month about U.S. sports. Our estimates suggest that Americans are just as informed about national politics as they are about Sports and Entertainment.

At a disaggregate level, the results above already show that a sizeable minority of individuals are quite distant from journalistic truth because they are unable to confidently distinguish real and fake news. A minority of this minority are confident in their incorrect beliefs, while the rest are simply uncertain.

We exploit the structural model to quantify the extent of heterogeneity across stories and individuals. As one would expect, there is significant heterogeneity across news stories, with some stories correctly believed by over 80% of individuals and others by as few as 20%. This variation is mostly absorbed by the uncertainty region: namely, people do not know if certain stories are true or false. The share of people who holds wrong beliefs about real news varies from 1% to 12%. No widely circulated fake story is believed by more than 19% of the population.

Our model contains an individual information level parameter, whose posterior distribution is estimated for each survey subject. On average, an individual in the top-third of the distribution has correct beliefs over 1.67 out of the 3 top stories of the month and an individual in the bottom-third of the distribution has correct beliefs over 1 out of 3 top news story. Similarly, if faced with 1 typical true news story and 1 typical fake news story, an individual in the top-third of the distribution has a probability between 80 and 89% of identifying the true news story whereas the corresponding probability for an individual in the bottom-third of the distribution is between 71 and 79%.

Our model also allows us to quantify the importance of partisan congruence and time passing. We find that partisan congruence between a news story and an individual matters in determining information levels. An individual is about 2% more likely to select a true news story if the news story reflects favorably on his or her preferred political party rather than unfavorably. The same individual is about 7% more likely to hold correct beliefs over congruent true news stories than non-congruent true news stories. Similarly, time passing matters: every month that passes reduces by 7% the probability that an individual successfully completes a news quiz and by 14-18% the probability that he or she holds correct beliefs over a true news story.

We also investigate information inequality by socioeconomic groups (defined by age, gender, race, and income). This is an important policy question because political accountability theory predicts that less informed segments are treated worse by officials with re-election concerns [Strömberg, 2004] and evidence for this effect has been found in a number of contexts [see e.g., Snyder and Strömberg, 2010]. We find evidence of large inequalities across socioeconomic groups. On average, individuals in the best-informed group (wealthy white men aged 47 and more) are 23% more likely to identify a true news story when faced with 1 true news story and 1 fake news story compared to individuals belonging to the least-informed group (low-income minority young women).

They are also 53% more likely to hold correct beliefs over true news stories.[4]

Overall, we find that socioeconomic factors matter much more than partisanship in determining information levels. For example, consider a typical partisan individual (either Republican or Democrat) and a typical news story. The model predicts that varying the gender, race, age, or income of this individual leads to changes in the probability that the individuals holds correct beliefs about the news story that are roughly 3 times larger than the effect of varying the story's favorability towards the Republican or the Democratic Party. Varying all 4 socioeconomic factors jointly leads to changes that are about 18 times larger.

As most of the news stories in our quizzes are related to the Donald Trump presidency, we perform an external validity exercise by looking at news stories about the Democratic Party primaries in the run-up to the 2020 presidential elections. Democratic primary voters are more likely to be young, female, minority and low-income compared to presidential election voters [Kamarck and Podkul, 2018]. We again document large information inequalities. Looking at the same two groups as above, we find that wealthy white men aged 47 and more are 47% more likely to hold correct beliefs over the true news stories compared to average low-income minority young women.

Taken together, our results indicate that the starkest pattern about the ability of U.S. voters to identify major news stories is not a generalized death of truth, or polarization along ideological lines, or the triumph of fake news. The starkest pattern appears to be inequality along socioeconomic lines. In the conclusion we discuss how this finding casts a doubt on some of the policies that have been proposed to combat the death of truth.

Our results are subject to important caveats, which we discuss in depth in the paper. In particular, we only measure people's factual knowledge of important news stories, but not how they interpret them. Also, while we find that many individuals today are able to confidently identify true and fake news stories, we are unable to measure whether information precision has decreased or increased in recent decades. In the conclusion we also suggest possible avenues for future research.

The rest of the paper is structured as follows. Section 2 reviews the news-generating

[4] As noted by Prior [2014], text surveys may exaggerate information inequalities by omitting visual clues (e.g., by not including pictures of actors mentioned in the news and included in our surveys).

process and the survey design. Section 3 describes the model as well as our estimation approach. Section 4 reports our main results. Section 5 presents various extensions of our analysis as well as robustness checks. Section 6 concludes.

1.1 Literature Review

There exists a sizeable literature on voters' information about political news that spans at least three decades. A partial list includes Price and Zaller [1993], Delli Carpini and Keeter [1996], Prior and Lupia [2008], and the references therein. Polling organizations regularly report survey results on voter information [e.g., Pew, 2017, Eurobarometer, 2017].[5] Media outlets sometimes feature news quizzes (e.g., the New York Times's News Quiz). Our survey differs from the literature in a number of important ways.

First and foremost, to the best of our knowledge all scholarly and commercial surveys are subject to the same limitation: the stories subjects are quizzed on are selected (and usually written) by the researcher and the pollster according to an unspecified criterion. This observation – which should not be taken as a criticism of a literature that typically has different research objectives – makes it difficult to use the results to answer questions about general news information levels.

We illustrate this challenge with reference to some of the existing studies. Let us begin with the pioneering work in this area, Price and Zaller [1993], which measures recall of 15 news stories (see Figures F.1 and F.2 in Online Appendix F). Restricting attention to political news, the percentage of people recalling different stories varies from 14% for a congressional debate on catastrophic health insurance to 65% recalling a trip by Bush to Europe. This large variance affects the answer to the question posed above: "If a major political event occurs, what percentage of Americans become aware of it?" If a researcher chooses stories like the congressional debate on catastrophic health insurance, the percentage is low. If the same researcher chooses stories like Bush's trip to Europe, the answer is much more encouraging. The percentage might be

[5] In turn, this literature is situated within a larger literature on voter information not just of news stories but also of other political facts. For instance, Delli Carpini and Keeter [1996] also studies information about institutions and processes (e.g., "What is the purpose of NATO?" or "Name the three branches of government"), information about key actors (e.g., "What is the name of the Secretary of State?"), information about statistical facts ("what is the percentage of population below the poverty line?"), knowledge of geography, and knowledge of social and political history. However, the present paper focuses exclusively on information about news stories.

even higher, or lower, with other political stories that Price and Zaller [1993] did not include in their survey.

This problem is even stronger in Delli Carpini and Keeter [1996]'s questions about domestic news. The percentage of American who are aware of a specific news story varies from "What is the conclusion of the Meese porno report?" (12%) to "What is the steel dispute about?" (96%). Large gaps are also present in contemporary state-of-the-art surveys. Table 2 reports the results of a recent Pews News IQ survey (see Figure F.3 in Online Appendix F): only 37% answer a question about the unemployment rate correctly while 86% of respondents answer a question about the Zika virus correctly.[6]

In a nutshell, the problem is that the set of possible news stories on which subjects can be tested is basically unbounded, and the outcome of the test depends on which news stories are selected. Without some discipline on the selection criterion, we can get extremely low response rates, extremely high response rates, and everything in between. This problem cannot be addressed after the survey is run, because we do not know what other news stories the researcher considered but did not include, and we do not know whether and how those stories differed from the ones that ended up in the survey.

The news selection issue does not affect just the interpretability of the *absolute* value of the results. It also affects *comparisons* across different groups. What can we conclude if Democrats are more likely than Republicans to identify news stories selected by a particular researcher or pollster with a criterion we do not know? Instead in our setting, that finding would be interpreted as the average Democrat being informationally closer to news that mainstream journalists consider important. There would still be a crucial subjective component (maybe there is a liberal media bias), but one that can now be ascribed to a specific set of actors, can be further investigated, and can be compared to other similarly measured subjective biases.

Our work also differs from all existing scholarly contributions because it situates news in a well-defined time horizon (news of the month) and it is repeated over time. Some commercial news quizzes, like the New York Times's, are repeated over time but, besides suffering from the item selection problem discussed above, the sample of

[6]While our paper focuses on information about one category of political facts only, news stories, our methodological critique applies to all categories of topics. For instance, within the "Knowledge of Institutions and Processes" category identified by Delli Carpini and Keeter [1996], 96% of respondents know that the U.S. is a member of the UN while only 2% can name two Fifth Amendment Rights.

respondents is not representative and they are not incentivized.

2 Design

We review the protocol employed to generate the true and fake news stories we insert in our survey quizzes and describe the information collected through the surveys.

2.1 News Generating Process

We design a protocol to identify, each month, the 3 most important news stories about the U.S. Federal Government according to mainstream media. First, we rely on Reuters' publicly-available wire stories about U.S. national politics to approximate the universe of relevant mainstream news stories.[7] Each wire story is composed of a headline, a brief summary, a picture, and a longer article. There are approximately 80 wire stories a week about U.S. national politics.

Second, we employ a panel of 3 professional journalists recruited through the Columbia School of Journalism.[8] In the course of the project, we have worked with 4 journalists in total. All journalists (three women and one man) are U.S. citizens in their late 20s who are currently working for mainstream media companies. To avoid recency effects, each week, each journalist is asked to select the 5 most important wire stories of the week according to him or her.[9] Specifically, journalists are provided with each wire story's headline, brief summary, and url to the longer article. Because multiple wire stories can deal with the same underlying event or "meta story", we ask the journalists to select only one wire story per meta story. In their weekly selection, we rely on journalists' subjective assessment of whether two Reuters wire stories deal with the same underlying event. At the end of every month, we take the four/five previous weeks' selected

[7] Reuters' wires dedicated to U.S. national politics can be found at https://www.reuters.com/news/archive/politicsNews.

[8] An alternative to our 'human' protocol would be an algorithmic approach, perhaps based on rankings produced by news aggregators such as Google News. However, any such approach would also ultimately build on the subjective views of the users whose data generates the ranking, and it would be less transparent as neither the ranking algorithm nor the users' characteristics are known. If we chose stories on that basis, we would not exactly know whose subjective judgment we are relying upon.

[9] Although we give discretion to our jury members in selecting the most important stories ("choose the stories you would cover as an editor..."), we ask them to adopt U.S.-centered criteria of importance.

wire stories and filter out the wire stories that do not cover the Federal Government (by far, most stories deal with the Federal Government).[10] We select a journalist to pool the remaining wire stories into their relevant meta stories (since different weeks' wire stories can deal with the same underlying event). We then present each meta story and associated wire stories to our panel and ask them to select and rank the five most important meta stories of the month. The choices are aggregated to produce the top three stories of the month. Once the three stories are selected, a short statement about each story is written (e.g., *The U.S Senate acquitted Trump of impeachment charges*).[11]

Our main instrument to estimate individuals' information about the news consists of asking them to select 3 out of 6 news stories. Three of these news stories correspond to the 3 true news stories described in the previous paragraph. The remaining 3 news stories are false short statements about the Federal Government. We employed two methods to general the fake news stories. In the first method, we rely on fact-checking website snopes.com to provide us with 3 fake news about the Federal Government that they "fact-checked" during the four weeks that preceded any survey. Journalists at snopes.com identify which rumours and questionable claims to fact-check and classify each as either false, mostly false, mixture, mostly true, or true. We asked snopes.com to provide us exclusively with rumours and headlines classified as false. In the second method, we relied on our panel of journalists to produce three plausible but false short statements. Among other pre-specified rules, journalists were instructed to write false statements of roughly equal length as the true statements, and in the same journalistic style.[12] We refer to these stories as "synthetic" or "well-crafted" fake news.[13]

[10]We adopt the U.S. definition of the "Federal Government" as being composed of the legislative, executive, and judicial branches. During our time period, the few stories that do not cover the Federal Government mostly deal with the presidential primaries. In Section 5, we replicate our analysis by focusing on the Democratic Party presidential primaries.

[11]Often, the statement associated with a meta story is simply one of the underlying wire stories' headline (or a slight modification). Journalists were asked to write primarily in the past tense and to avoid using numbers and figures.

[12]We also instructed the panel to avoid writing negations of events that really took place, to avoid writing statements that could be perceived as related to the real statements, to avoid using numbers and figures, and to primarily use past tenses.

[13]The fake news we inserted in our quizzes are unambiguously false. Although it is a natural starting point for our analysis, we note that in practice fake news can be particularly misleading when they include some elements of truth.

Democratic Party Presidential Primaries. In 5 surveys, we inserted quizzes that covered news about the Democratic Party primaries exclusively. The method employed to generate the 6 stories included in the news quizzes is identical to that described above. To generate the fake stories, we relied exclusively on synthetic fake news.

Sports and Entertainment. In 2 surveys we inserted quizzes that included news exclusively about the world of sports and news quizzes that included news exclusively about the world of entertainment. We relied on our panel of journalists to select the 3 most important news stories of the month for each topic and to produce 3 synthetic fake news. The protocol we employed was identical to that described above, except that the journalists' selection was conducted at the monthly level. Because of Reuters' scant coverage of these topics, we relied on the Associated Press for news about sports and msn.com for news about entertainment.

2.2 Survey Design

This paper mainly exploits data gathered from 8 online surveys we conducted through polling company YouGov. The first survey took place in June 2019 and the last survey in February 2021. For each survey, YouGov enrolled a representative sample of the U.S. citizen adult population.[14] All surveys were administered to about 1,000 individuals. We instructed YouGov to avoid enrolling individuals who participated in prior editions of the survey. This restriction was partially lifted from the 3rd survey onward. Overall, 8251 individuals participated in our 8 surveys. YouGov provides a wide array of background information concerning each respondent (demographics, income, party affiliation, etc.), where the information is collected months before our surveys. Our survey took respondents on average 5-6 minutes to complete. Participants received about $1.9 on average (paid via gift cards) in exchange for completing the survey. Payments

[14] To construct the sample, YouGov employs a two-step procedure. In the first step, a random sample is drawn from the population (using either Census information or the American Community Survey). This sample is referred to as the target sample. In the second step, a matching technique is utilized to match each member of the target sample with members of YouGov's pool of respondents. For further details, see https://smpa.gwu.edu/sites/g/files/zaxdzs2046/f/downloads/YG˙Matching˙and˙weighting˙basic˙description.pdf.

(a) Socioeconomic Characteristics

Statistic	Yougov	ACS 2018
Median Age	50	47.00
% Female	0.53	0.51
% White	0.69	0.73
% Black	0.11	0.13
% 4yr College Degree	0.31	0.31
% Married	0.47	0.48
% Family Inc <30k	0.29	0.17
% Family Inc 30k - 60k	0.19	0.23

(b) Party Affiliations

Party Affiliation	YouGov	Pew 2018
% Democrat	45	48
% Republican	34	39
% Independent	16	7
% Other	4	6

Table 1: Survey Participants Characteristics

included a $0.50 show up fee and bonuses worth $1 for each quiz correctly answered.[15]

Table 1a provides basic descriptive statistics regarding the socioeconomic characteristics of our survey participants. It also reports the corresponding statistics for the population of U.S. adult citizens according to the 2018 American Community Survey of the Census Bureau (ACS).[16] All dimensions appear broadly aligned with the general population, with the exception of family income. Table 1b reports information on the party affiliation of our survey respondents, and compares it with the statistics provided by Pew [2018].[17] We pool the respondents who report that they "Lean Democrat" ("Lean Republican") with the respondents who support the Democratic Party (Republican Party). The proportions are roughly comparable, with the exception of Independents who appear somewhat over-represented in the YouGov sample. As a robustness check, we ran two surveys on samples of respondents recruited through M-Turk. We provide greater detail in the relevant extension.

Our survey was composed of two main parts: (i) a series of questions about media consumption habits and (ii) a series of questions about recent political news. Online

[15]The advantage of incentivizing accurate responses is that it limits partisan behavior on the part of the respondents when completing the surveys. It is however of course possible that the same individuals, when not incentivized to hold accurate beliefs, base their decisions on partisan beliefs.

[16]To obtain the 2018 ACS go to https://www.census.gov/programs-surveys/acs.

[17]YouGov asks respondents to select one option among "Strong Democrat", "Not very strong Democrat", "Lean Democrat", "Independent", "Lean Republican", "Not very strong Republican", "Not sure", "Don't know". About 4% of respondents report either "Not Sure" or "Don't Know". We pool these respondents with the respondents who report being "Independent".

Appendix E presents the language used in all of our survey questions.

2.2.1 Media Consumption Habits

Respondents reported whether they had acquired information about national politics during the previous 7 days, and whether they acquired it online, by watching television, by listening to the radio, and/or by reading a print newspaper. We further asked respondents to report the news sources they relied on (e.g., CNN). Finally, respondents were asked to report the amount of time they dedicated to national politics. Table A.1 in Online Appendix A.1 reports summary statistics.

2.2.2 News Quizzes

All surveys included 1 or 2 quizzes about current news stories (less than 4 weeks old). In a number of surveys, we also included 5-8-week-old and 9-12-week-old news quizzes. Overall, we included 21 distinct news quizzes in our 8 surveys. Our average respondent took 2.38 news quizzes. Each quiz was composed of 6 short statements. Survey respondents were told the list contained exactly 3 true statements and 3 false statements. Respondents were asked to select which 3, to the best of their ability, were the true statements. News quizzes either included only synthetic news or only actual fake news (i.e., we never mixed synthetic and actual fake news). When surveys included both types of quizzes, the quizzes included the same 3 true news stories. To avoid individuals from obtaining information elsewhere, respondents were given 60 seconds to make their selection.[18] We offered $1 (paid via a giftcard) to all respondents who selected all 3 true statements. All survey respondents were revealed the correct answers once they took the quiz. Presumably because of the 60-second limit, 19% of respondents ended up selecting a number of statements different from 3. The vast majority of these respondents selected strictly fewer than 3 statements. We exclude these respondents from our main analysis. In Online Appendix C.4, we re-estimate the model by including respondents who selected fewer than 3 statements.

Why did we tell respondents that exactly 3 statements are true? Telling respondents that 3 statements are true *on average* without recruiting much larger samples of

[18] Imposing a time limit may lead us to under-estimate respondents' information levels [see e.g., Bago et al., 2020].

respondents would make our model estimation much more noisy, especially of the news story-level parameters (see below). Any given statement would be read by significantly fewer respondents and even fewer respondents would face the exact same 6 statements. Finally, including 3 true statements on average would not be practical: our panel of journalists would have to write a much longer list of real and false statements for us to sample from. For a discussion of alternative quiz designs see Online Appendix D.

Table 2 provides descriptive statistics at the statement level. On average, true news stories about the Federal Government were selected by 75% of respondents. Breaking down stories by their ranking according to our journalists, on average, the most important news stories were selected by 84% of respondents, the 2nd most important news stories by 77% of respondents, and the third most important news stories by 64% of respondents. By contrast, synthetic fake news were selected by 25% of respondents and fake news by 22% of respondents.[19],[20]

Partisan Score. Respondents – after having completed the news quiz and being told which statements were true and false – were asked how favorably, in their opinion, each true statement reflected on the Republican Party. Similarly, for each false statement, respondents were asked how favorably, in their opinion, the statement would have reflected on the Republican Party had it been true. Respondents were allowed to select one option among "very unfavorable" (score of -2), "unfavorable" (score of -1), "neither unfavorable nor favorable" (score of 0), "'favorable" (score of 1), and "very favorable" (score of 2). Across all quizzes about Federal politics, the average true statement had an average partisan score of -0.06 (standard deviation: 0.45). The average synthetic fake news had an average partisan score of -0.09 (standard deviation: 0.44). Lastly, the average actual fake news had an average partisan score of -0.01 (standard deviation: 0.25). For each statement j, we construct the continuous variable $b_j \in [-\infty, \infty]$

[19]Table A.2 in Online Appendix A.2 provides descriptive statistics at the quiz level, distinguishing between quizzes about the Federal Government, quizzes about Sports and Entertainment, and quizzes about the Democratic Party presidential primaries.

[20]Pennycook et al. [2021] administer a series of surveys, in which social media users are asked to assess the accuracy of true and false headlines. The authors show that partisan congruence matters more than accuracy in determining users' propensity to share news with their networks, despite the fact that users are able to correctly identify most headlines as true or false. Reassuringly, some raw statistics are similar to ours (e.g., the share of true headlines identified as true). However, making a more direct comparison is difficult because of their focus on social media users and because the criteria they rely on to select the news stories is not reported.

	Mean	St. Dev.	Min	Max	Quizzes	N
Share of true news stories selected	0.75	0.15	0.36	0.95	8	14796
Share of 1st true news stories selected	0.84	0.14	0.62	0.95	8	4932
Share of 2nd true news stories selected	0.77	0.1	0.66	0.9	8	4932
Share of 3rd true news stories selected	0.64	0.16	0.36	0.84	8	4932
Share of synthetic fake news selected	0.25	0.15	0.07	0.63	8	14796
Share of Snopes fake news selected	0.22	0.11	0.08	0.42	3	4308

Table 2: Story Level Summary Statistics

by rescaling the average respondent's score to give the resulting variable a standard deviation equal to 1.

3 Model

We develop our model in two steps. We first consider an agent who is asked to pick the statement that is most likely to be true out of a set of statements and we show that, under standard assumptions, the problem corresponds to a familiar parameterized discrete choice problem. Second, we apply this framework to the survey instrument we are using to arrive at the econometric model that we will be using in the rest of the paper. We also clarify the link between our model and the existing literature.

3.1 A Discrete Choice Model

Suppose agent i is trying to determine whether statement j is true ($\tau_j = 1$) or false ($\tau_j = 0$). Based on her information I_{ij}, she forms a belief about the truth of the statement $\Pr(\tau_j = 1|I_{ij})$. In what follows we use a monotonic transformation of this belief, its log-likelihood ratio: $z_{ij} = \ln(\Pr(\tau_j = 1|I_{ij})/\Pr(\tau_j = 0|I_{ij}))$. The log-likelihood ratio z_{ij} is a random variable which we assume can be written as:

$$z_{ij} := \underbrace{(2\tau_j - 1)\,\hat{\gamma}_j \theta_i \delta^t + \alpha b_j p_i}_{\mu_{ij}} + \lambda + \varepsilon_{ij}, \tag{1}$$

where ε_{ij} has a standard Gumbel CDF. The distribution of z_{ij} depends on the agent's information precision $\theta_i \geq 0$, on the statement's salience $\hat{\gamma}_j \geq 0$, on the number of months t since the story was written via the decay parameter δ, and on the effect of partisan congruence $\alpha b_j p_i$ (where b_j is statement j's partisanship, $p_i \in \{-1, 0, 1\}$ is agent i's partisanship, and α captures the strength of this partisan congruence effect).[21] λ is a free parameter to be determined later.

Suppose now that agent i is trying to determine whether statement j belonging to a set J of statements is the most likely to be true. Assuming that ε_{ij} is iid across statements, the probability agent i selects statement j as the most likely to be true among the set J of statements is:[22]

$$\pi_{ij} = \frac{e^{\mu_{ij}}}{\sum_{k \in J} e^{\mu_{ik}}} = \frac{e^{(2\tau_j - 1)\hat{\gamma}_j \theta_i \delta^t + \alpha b_j p_i}}{\sum_{k \in J} e^{(2\tau_k - 1)\hat{\gamma}_k \theta_i \delta^t + \alpha b_k p_i}}. \tag{2}$$

If j is a true (false) statement, π_{ij} is increasing (decreasing) in j's salience $\hat{\gamma}_j$. If $\theta_i > 0$, as the statement becomes infinitely easy ($\hat{\gamma}_j \to \infty$), the probability tends to 1 if the statement is true and to zero if it is false. Similarly, if statement j is the easiest (i.e., if $\hat{\gamma}_j > \hat{\gamma}_k\ \forall k \neq j$), then the probability increases with agent i's information precision θ_i if the statement is true (and tends to 1 as $\theta_i \to \infty$) and decreases if it is false (and tends to 0 as $\theta_i \to \infty$). Finally, if $\alpha > 0$, the probability of selection π_{ij} is higher if the statements is congruent (i.e., $b_j p_i > 0$) than if it is either neutral (i.e., $b_j p_i = 0$) or

[21] Recall that we interpret $b_j \in \mathbb{R}$ as the partisanship of the news story: a high (low) b_j denotes a story that reflects favorably (unfavorably) on the Republican Party. Similarly, $p_i \in \{-1, 0, 1\}$ denotes agent i's partisanship, where $p_i = 1$ ($p_i = -1$) means that agent i identifies with the Republican Party (Democratic Party) and $p_i = 0$ means that agent i identifies as Independent. The term $b_j p_i$ captures the congruence ($b_j p_i > 0$) or incongruence ($b_j p_i < 0$) between an individual's partisanship and a news story's partisanship. The parameter α measures the strength of the effect of partisan congruence.

[22] The expression above holds under the assumption that the random variable ε_{ij} is independent across statements. In practical terms, this means that the statements are not related in ways that make their plausibility value correlated. An obvious violation occurs when two statements refer to related stories "President Trump visited France" and "President Trump met with President Macron." We believe the independence condition is satisfied in practice within every round as both the true stories and the fake stories are designed to belong to distinct meta-stories (see Section 2).

incongruent (i.e., $b_j p_i < 0$), independently of whether the statement is true or false.

We can use the same expressions to determine the probability that agent i's belief in the truth of statement j is above a certain confidence threshold h, denoted $\rho_{ij}(h)$. Specifically, turning to the log-likelihood ratio z_{ij} and setting $H = \ln\left(\frac{h}{1-h}\right)$, we have:

$$\rho_{ij}(h) = \Pr[z_{ij} \geq H \mid \mu_{ij} + \lambda] = 1 - e^{-e^{\mu_{ij}+\lambda-H}}. \tag{3}$$

Unlike expression (2), $\rho_{ij}(h)$ contains the parameter λ. To calibrate its value, consider a story with $\hat{\gamma} = 0$ and $b = 0$. This is a story over which an agent has no information and holds a neutral prior. We assume an agent assigns a 0.5 probability to such a story being true. This implies a value $\lambda = \ln(\ln 2) \simeq -0.36651$, which we set throughout. As discussed below, assuming that subjects assign a 50% probability to a story over which they have no information is natural in our setting because they know that exactly 50% of the stories are true.

3.2 Econometric Model

Expression (2) states the probability that individual i believes statement j to be the most likely to be true among a set J of statements. In our survey quizzes, respondents read 6 statements (ordered randomly) and they are given the additional information that exactly 3 statements are true and 3 statements are false. They are rewarded if they successfully select these 3 true statements. The solution to the respondents' problem involves the iterated application of expression (2). The intuition for this result is that being told that exactly 3 statements are true may change the posterior probabilities attached to the 6 statements but does not change the rank order of those probabilities.

To see this, assume each respondent maximizes the probability of receiving the monetary reward. Let $T \equiv (\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6) \in \{0, 1\}^6$ be the set of all possible 'truth vectors' over the 6 statements before being told that exactly 3 are true. Let $q_{ij} = \Pr(\tau_j = 1)$ be the posterior probability that subject i assigns to statement j being true before learning that exactly 3 statements are true and 3 are false. Note that as news stories are not directly related, their truth values are independent: if $j \neq k$, $\Pr(\tau_j = 1, \tau_k = 1) = q_{ij} q_{ik}$. Suppose now that subject i is told that exactly 3 statements are true and let T_3 be the set of all truth vectors in T such that exactly 3 statements are true. If the subject selects statements j', j'', j''' as true and k', k'', and k''' as false, the subjective probability she wins the reward is equal to:

$$\frac{q_{ij'}q_{ij''}q_{ij'''}\left(1-q_{ik'}\right)\left(1-q_{ik''}\right)\left(1-q_{ik'''}\right)}{\Pr\left(T_3\right)}. \tag{4}$$

This probability is maximized if and only if the agent chooses the statements with the three highest $q'_{ij}s$. This is equivalent to picking the statement with the highest q out of 6, then the statement with the highest q out of the remaining 5, and finally the statement with the highest q out of the remaining 4.

Given our logit specification (see (2)), the probability of selecting statements $\{j, j', j''\}$ in this exact order is given by:

$$\pi_{ij\in J}\cdot\pi_{ij'\in J\backslash\{j\}}\cdot\pi_{ij''\in J\backslash\{j,j'\}}. \tag{5}$$

Let $S\left(j',j'',j'''\right)$ denote the set of all six possible permutations of statements j', j'', and j''' and s a typical element of $S\left(j',j'',j'''\right)$. Let also π_{is} denote the probability that i selects the statements j', j'', and j''' in the exact order s (see (5)). The total probability of selecting the three statements j', j'', and j''' is equal to $\Pi\left(j',j'',j'''\right)\equiv\sum_{s\in S\left(j',j'',j'''\right)}\pi_{is}$. Our survey instrument generates draws over Π.

3.3 Literature Discussion

Our model is related to Item Response Theory (IRT), a set of statistical models that are used to analyze test results with the objective of inferring the difficulty of the test questions and the traits of the test takers [Van der Linden and Hambleton, 1997]. The model has been applied to analyze data from voter information surveys by, for instance, Baek and Wojcieszak [2009] and Anderson et al. [2010].

In standard IRT applications such as the Rasch model [Rasch, 1960], the researcher can rank alternatives a priori (usually because an answer can only be right or wrong). Here, instead we cannot a priori rank different statement bundles that contain different subsets of true statements. Suppose that A, B, and C are true statements and D, E, and F are false statements: it is not ex ante clear whether choosing, say, (A, B, D) is better than choosing (A, C, E). We are closest to an extension of IRT called Nominal

Response Model (NRM), developed by Bock [1972], which allows items to be ranked in a partially unknown manner [see also Anderson et al., 2010].

One modification is necessary. We cannot use NRM directly because we are interested in measuring two factors: the underlying skill of our respondent (the precision of their information) and the effect of partisan congruence. The latter effect is not salient in educational testing where only skill is measured. We therefore augment Bock [1972] by developing a model where individuals have two traits, skill and partisanship, and news stories have two characteristics, difficulty and partisanship.

3.4 Estimation

In what follows, let $\gamma_j \equiv (2\tau_j - 1)\,\hat{\gamma}_j$. True statements have a positive associated salience parameter γ_j and false statements have a negative associated salience parameter γ_j. We wish to use our survey data to estimate the values of the main model's parameters: the information level θ_i of every subject, the salience γ_j of every (true or false) news story, the decay parameter δ, and the strength of partisan congruence α. For the vector of θ_i's we have only one news quiz per individual (on any given topic) but individuals are drawn from the same distribution. We therefore adopt a hierarchical Bayesian approach, where every individual is endowed with a θ_i drawn from a population distribution, which is in turn drawn from a hyperprior distribution [see Chapter 5.1 in Gelman et al., 2014]. We will estimate a (noisy) posterior distribution of θ_i for each individual and a (precise) posterior distribution for the population. We estimate the model by Bayesian methods, specifically the No U-Turn Sampler [Hoffman and Gelman, 2014] implemented in Numpyro [Phan et al., 2019].[23]

Main Model. In the main analysis, we assume that all individuals' precision parameter θ_i are drawn from a common prior distribution. Specifically, we specify a common prior distribution $\theta_i \sim N(1, \sigma^2)$ with hyperprior $\sigma \sim \exp(\frac{1}{4})$. We also specify that $\gamma_j \sim |Y|$ if $\tau_j = 1$ and $\gamma_j \sim -|Y|$ if $\tau_j = 0$, where $Y \sim N(0, 1)$. Building on the economic model, we thus impose that the γ_j's associated with the true (resp. false) statements are nonnegative (resp. nonpositive). The remaining prior distributions are

[23] For a review of the algorithm and an analysis of its performance on categorical data see Sacher et al. [2021].

specified as $\alpha \sim N(0,1)$ and $\delta \sim N(1,1)$.

Hierarchical Model. In an extension in which we look at heterogeneity across individuals, we estimate a hierarchical model in which individuals' precision parameter θ_i are drawn from a prior distribution whose mean μ_i depend on pre-specified individual characteristics (e.g., age, gender, etc.). Specifically, let X denote a $K \times 1$ vector of covariates X_k. We specify that each $\theta_i \sim N(\mu_i, \sigma^2)$ where (i) $\sigma \sim \exp(\frac{1}{4})$ and (ii) $\mu_i = 1 + \sum_{k=1}^{K} \eta_k X_k$, with $\eta_k \sim N(0,1)$. We subtract sample averages to each individual characteristics, to ensure that an average individual has $\mu_i = 1$.

A common issue with this family of models [e.g., Bock, 1972, and see discussion of the literature above] is that there always exist two additional degrees of freedom. First, one can add a constant to all the γ_j's without modifying $\Pi(j', j'', j''')$. The constraints we impose that ensure that the γ's associated with the true (false) statements are nonnegative (nonpositive) potentially address this concern by acting as lower- and upper-bounds on the values the γ_j's can take. Second, θ_i and γ_j are identified only through their product. One can multiply all θ's by a constant and divide the γ's by the same constant without modifying $\Pi(j', j'', j''')$. We address this challenge by assuming that the mean of the prior distribution of θ_i is equal to 1 in the main model and, similarly, that the "average individual" has a $\mu = 1$ in the hierarchical model. Although this normalization to an arbitrary scale means that θ and γ have no literal meaning, it will not affect the interpretation of our results where we rely on the functions π_{ij} and $\rho_{ij}(h)$ instead. Only the product of θ_i and γ_j matters for π_{ij} and $\rho_{ij}(h)$.

4 Analysis

4.1 Overview

We begin by estimating the main model using the 8 quizzes with synthetic fake news exclusively. This involves estimating the joint posterior distribution of:

- 4,932 individual-level θ_i's: If we integrate across individuals we obtain an aggregate distribution of theta with mean close to 1 and a standard deviation equal to 0.48 (σ has a posterior mean equal to 0.48).

- 48 story-level γ_j's: (i) Conditional on $\tau_j = 1$, the marginal posterior distribution of γ_j has mean 0.92 and standard deviation 0.8; (ii) Conditional on $\tau_j = 0$, the marginal posterior distribution of γ_j has mean -1.15 and standard deviation 0.95.

- The population parameter α: The marginal posterior distribution of α has a mean equal to 0.03, a standard deviation equal to 0.01, a 5th percentile value equal to 0.01, and a 95th percentile value equal to 0.06.

- The population parameter δ: The marginal posterior distribution of δ has a mean equal to 0.69, a standard deviation equal to 0.03, a 5th percentile value equal to 0.63, and a 95th percentile value equal to 0.74.

Because the magnitude of these estimates are not easily interpreted, in what follows we present our main findings by relying on the functions π_{ij} (the probability that individual i selects statement j from a set of statements) and $\rho_{ij}(h)$ (the probability that individual i believes statement j to be true with probability h or higher).

4.2 Aggregate Information Levels

We begin by analyzing individuals' ability to distinguish true from fake news stories. We imagine that individuals are faced with one true statement and one false statement and that they are incentivized to select the statement they believe is the most likely to be true.[24] For simplicity, we assume that these two stories are neutral ($b = 0$) and less than a month old ($t = 0$). The probability that individual i with information precision θ_i selects the true statement when faced with a true statement j with salience γ_j and a false statement j' with salience $\gamma_{j'}$ is equal to $\pi_{ij} = \frac{e^{\theta\gamma_j}}{e^{\theta\gamma_j}+e^{\theta\gamma_{j'}}}$. Let $F_s(\theta, \gamma)$ denote the joint posterior distribution of θ and γ associated with survey $s = 1, \cdots, S$. One can then compute $\overline{\pi} := \frac{1}{S}\sum_{s=1}^{S}\int \pi_{ij} dF_s(\theta, \gamma)$, whose empirical analog is given by $\frac{1}{N}\sum_n \frac{1}{S}\sum_{s=1}^{S}\pi_{ij}(\theta_n, \gamma_n)$ (where N is the number of draws from the joint posterior distribution of θ and γ).

The first row in Table 3 reports the predicted probability $\overline{\pi}$ that, on average, individuals select the true news story when faced with one true and one false statement. The

[24] Recall that in our quizzes respondents must select 3 statements out of 6. Using our parameter estimates to predict individuals' performance in a counterfactual quiz with only 1 true and 1 false statements simplifies the interpretation of the probability that individuals successfully complete the quiz. In Section 4.3 we show that the model's predicted probabilities fit the data well.

table distinguishes between hypothetical quizzes that include the first, second, third, or any top 3 true news story of the month.[25] On average, individuals are 85% likely to select the first news story of the month, 81% likely to select the second news story of the month, and 76% likely to select the third news story of the month. Overall, they are 81% likely to select a top 3 news story of the month.[26] The second row in Table 3 reports the predicted probability that individuals successfully complete a more difficult version of the quiz, in which only 1 out of 4 statements is true. Again, we predict that, on average, most individuals are able to successfully complete the quiz, with probabilities of success ranging from 57% to 71% depending on the true news story's ranking. These numbers suggest that the vast majority of individuals are able to distinguish typical true news stories from fake news. We also note that the news stories our journalists think are more important are more likely to be selected by respondents.

	First	Second	Third	All
$\overline{\pi}\,(\text{true} \mid 1\text{ true}, 1\text{ false})$	0.85	0.81	0.76	0.81
$\overline{\pi}\,(\text{true} \mid 1\text{ true}, 3\text{ false})$	0.71	0.64	0.57	0.64

Table 3: Average Probability of Selecting True Statement

Note: The first row reports the average probability that individuals select the first, second, and third true news story of the month (as well as any true news story ranked third or higher) when faced with one true news story and one synthetic fake story. The second row reports the average probability that individuals select the first, second, and third true news story of the month (as well as any true news story ranked third or higher) when faced with one true news story and three synthetic fake stories.

We now analyze the probability individuals assign to news stories being true. The predicted probability that individual i with information precision θ_i assigns a probability equal to or higher than h to statement j being true is equal to $\rho_{ij}(h)$ (see (3)). We compute $\overline{\rho}(h) := \frac{1}{S}\sum_{s=1}^{S}\int \rho_{ij}(h)\, dF_s(\theta,\gamma)$. The function $\overline{\rho}(h)$ gives the probability

[25] More precisely, we suppose that the salience parameter γ of a true news story is drawn from the marginal posterior distribution of γ conditional on a given rank (first, second, or third news story of the month). By contrast, fake news are not ranked by importance.

[26] Individuals' ability to distinguish between true and fake news stories is also suggested by the raw data: in our average quiz with 3 true and 3 fake news stories, true news stories are selected by 75% of survey respondents and fake news by 22-25% of respondents (see Table 2). Note that these raw figures are lower than those reported in Table 3. This difference is explained by the fact that the probability that a given typical statement is selected by the average individual is higher in a quiz with 2 statements than in a quiz with 6 statements.

that, on average, respondents believe news stories to be true with probability h or higher. We continue to suppose that both the true and the fake news stories are neutral and less than a month old.

Table 4 reports the probability $\overline{\rho}(h)$ for various confidence intervals, by distinguishing between the first, second, and third true news stories of the month and synthetic fake news. To report our results in a way that is easier to comprehend, it is useful to focus on a particular level of confidence h. In what follows, let us say that an individual has a *correct belief* about a true story if she believes it to be true with probability 0.75 or higher, that she is *uncertain* if she believes it to be true with a probability between 0.25 and 0.75, and that she has a *wrong belief* if she believes it to be true with a probability lower than 0.25. The converse definition applies to fake news: for instance, an individual has the correct belief about a fake story if she believes it to be true with a probability 0.25 or lower. Accordingly, the top panel of Table 4 reports the corresponding figures. On average, the model predicts that individuals hold correct beliefs about the first news story of the month with probability 0.59. Similarly, on average, individuals are uncertain with probability 0.39 and they hold wrong beliefs with probability 0.03. These numbers change as we move from the first to the second and third true news stories of the month. For example, on average, the probability that individuals hold correct beliefs about the second and third news stories of the month fall to 0.44 and 0.31, respectively. Turning to fake news, on average, individuals hold correct beliefs about our synthetic fake news with probability 0.45. Individuals are uncertain with probability 0.44 and they hold wrong beliefs with probability 0.11.

Naturally, saying that an individual holds "correct beliefs" about a true (respectively, fake) news story if she assigns a probability at least as high as 0.75 (respectively, no higher than 0.25) to the story being true is arbitrary. The second and third panels of Table 4 report similar figures for alternative thresholds. For example, in the second panel, we report that, on average, 68% of individuals attribute 2 to 1 odds to the first story of the month being true. The corresponding figures for the second and third news stories of the month are 56% and 43%, respectively. Similarly, on average, individuals are 56% likely to attribute 2 to 1 odds to synthetic fake news being false. Last, the third panel of Table 4 reports the likelihood that, on average, individuals attribute a probability of truth greater than or equal to $h = 0.5, 0.6, 0.7, 0.8, 0.9$ to the first, second, and third news story of the month and to our synthetic fake news.

Confidence	True Story			Fake Story
	First	Second	Third	
0 - 0.25	0.03	0.04	0.07	0.45
0.25 - 0.75	0.39	0.52	0.62	0.44
0.75 - 1	0.59	0.44	0.31	0.11
0 - 0.33	0.07	0.09	0.15	0.56
0.33 - 0.66	0.25	0.35	0.42	0.29
0.66 - 1	0.68	0.56	0.43	0.16
0.5 - 1	0.82	0.75	0.63	0.27
0.6 - 1	0.74	0.64	0.5	0.19
0.7 - 1	0.64	0.51	0.37	0.13
0.8 - 1	0.52	0.37	0.25	0.08
0.9 - 1	0.35	0.2	0.13	0.04

Table 4: Average Probability of Holding Correct Beliefs $\overline{\rho}\left(h\right)$

Note: The table reports the average probability $\overline{\rho}\left(h\right)$ that respondents hold correct beliefs about the first, second, and third true news story of the month as well as about synthetic fake news.

An alternative approach to expressing information levels consists of computing the expected number of true news stories – among the top 3 stories of the month – over which individuals hold correct beliefs. We rank individuals by the mean of their associated posterior distribution of information precision θ_i and report results for individuals belonging to the bottom-third, middle-third, and top-third of the distribution. Table 5a reports the probability that, on average, members of these three groups hold correct beliefs about the first, second, and third news stories of the month.[27] Using these numbers, one computes that, on average,– of the top 3 news stories of the month – individuals in the bottom-third of the distribution have correct beliefs about 1 news story, individuals in the middle-third have correct beliefs about 1.35 news stories, and individuals in the top-third have correct beliefs over 1.67 news stories. Table 5b returns to the hypothetical quiz with 1 true and 1 false statement. The table reports, for each tier, the average probability that individuals select the true statement. Individuals in the top-third of the distribution are about 12-13% more likely to successfully complete quizzes relative to individuals in the bottom-third.

[27] As above, we suppose these true stories to be neutral and to be less than a month old.

(a) Average Information Levels

	Information tier		
	Lower	Middle	Higher
First Story	0.42	0.61	0.74
Second Story	0.33	0.44	0.55
Third Story	0.25	0.3	0.38

(b) Average Probability of Selection

	Information tier		
	Lower	Middle	Higher
First Story	0.79	0.87	0.89
Second Story	0.75	0.83	0.85
Third Story	0.71	0.78	0.8

Table 5: Heterogeneity across Information Tiers

Note: The left table ranks individuals by their information precision and reports, for each information tier, the average probability $\overline{\rho}\,(0.75)$ that individuals hold correct beliefs about the average first, second, and third true news story of the month. The right table ranks individuals by their information precision and reports, for each information tier, the average probability $\overline{\pi}$ that individuals select the true news story when faced with 1 true and 1 fake news story, again distinguishing between the first, second, and third true news story of the month.

We conclude this subsection by reporting the marginal posterior distribution of θ in Figure 1. One somewhat striking feature of $F(\theta)$ is its relatively low mass close to zero. Our estimates suggest that very few individuals have little ability to discern the truth. This finding is easily explained by some basic patterns in the raw data. Across all quizzes, around 5% of respondents selected 0 true statements and only 15% selected 1 true statement. By way of comparison, an uninformed individual (with no partisan prior), with no choice but to randomize, chooses 1.5 correct statements on average.[28] The same individual has a probability equal to 50% to select either 0 or 1 true statements. The theta distribution that fits the data cannot place a large weight on individuals that have little ability to discern the truth.

Our results so far are somewhat reassuring. A sizable share of individuals are able to discern true from false statements, and to do so confidently. Only a small share of individuals hold wrong beliefs about true and fake news. Nonetheless, a significant share of individuals are uncertain whether news stories are true or false and there seems to exist significant heterogeneity across individuals in terms of information precision.

[28]Moreover, because each individual completes only one or two quizzes, the variance of the distribution $F_i(\theta)$ is relatively large, so that the common prior assumption tends to pull all individuals upward. Further, the restriction to respondents who selected exactly 3 statements may also in part explain the relatively small mass around 0 (see Online Appendix C.4).

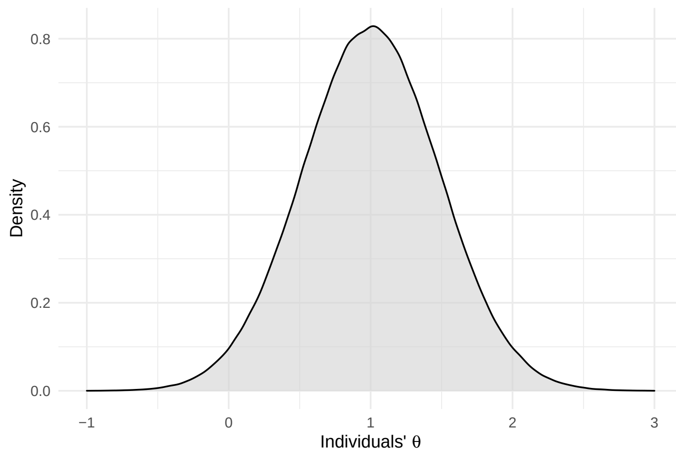


Figure 1: The Posterior Distribution of Information Precision θ

4.3 Heterogeneity across News Stories

Next we explore heterogeneity across news stories. Table 12 in the appendix lists all 24 *true news stories* that were included in our quizzes. Similarly, Table 13 lists all 24 *synthetic fake news*. For each statement, the tables report the share of survey respondents who selected the statement when completing the quiz ("raw mean"), the partisan score b given by the average respondent, the mean of the posterior distribution of γ_j, the predicted share of respondents who – according to our model's estimates – will select the statement when completing the quiz, as well as the probability $\overline{\rho}(h)$ that respondents will assign probability h to statement j being true.[29]

As expected, there exists significant heterogeneity across news stories. Some statements were selected by virtually all our respondents and others were selected only by a tiny share of respondents. Similarly, there also exists significant heterogeneity across news stories in their partisan score: many stories reflected favorably on the Republican Party, many stories reflected favorably on the Democratic Party, and many tended to reflect neither favorably nor unfavorably on either party.[30] Further, recall that the

[29]Unlike the previous subsection in which we assumed stories to be neutral in their partisanship, in this exercise we incorporate the extent to which each news story reflects favorably on the Democratic versus Republican Party.

[30]Tables 12 and 13 report the unscaled values of b_j. Recall that a score of -2 (resp. 2) means that news story j reflects very unfavorably (resp. very favorably) on the Republican Party, a score of -1 (resp. 1) that it reflects unfavorably (resp. favorably), and a score of 0 that it reflects neither unfavorably nor favorably on the Republican Party. Interpreting the magnitude of the partisan score b

parameter γ_j captures how responsive the likelihood of selecting statement j is to information precision θ. What the tables suggest is that some true statements are much more easily detectable as true by informed respondents than others. Similarly, some false statements are much more easily detectable as false by informed respondents than others. Next, the tables report, for each statement, the main model's predicted share of respondents who select it when completing the quiz (computed taking into account the characteristics of the remaining 5 statements that were included in the same quiz). As suggested by the numbers, our model approximates the actual data well, irrespective of whether a statement is chosen by few or many respondents.

Finally, there exists significant heterogeneity across news stories in terms of respondents' beliefs. Looking at real news, the probability that, on average, individuals hold correct beliefs ranges from 21% to 83%. By contrast, the probability that individuals hold wrong beliefs ranges from 1% to 12%. For example, respondents have an 83% probability of holding correct beliefs about the (true) story "*The U.S. Senate acquitted Trump of Impeachment Charges*." By contrast, they hold correct beliefs about the (true) story "*Supreme Court granted a request by President Trump's administration to fully enforce a new rule that would curtail asylum applications by immigrants at the U.S.-Mexico border*" only with probability 42% (despite 69% of our sample selecting the statement when completing the quiz). This last news story – with its relatively large difference between the share of respondents who select the statement and share of respondents who hold correct beliefs – illustrates how our structural approach takes into account the various properties of all the news stories included in the quiz when measuring respondents' beliefs about each single news story. For synthetic fake news, the probability that, on average, individuals hold correct beliefs ranges from 14% to 81%. The probability that they hold wrong beliefs goes from 3% to 20%. Consistently with our earlier findings whereby a large share of individuals are able to discern true from false statements confidently, we find that even the most widely believed fake news is less likely to be believed as true than the least believed real news.

by looking at the raw numbers is difficult. In Section 4.5, we rely on the model to assess the importance of news stories' partisan score in determining individuals' information about political news.

4.4 Fake News

The results presented thus far relied on estimating the main model using the quizzes with synthetic fake news exclusively. We briefly present the results we obtain when we instead estimate the model using the quizzes with actual fake news exclusively. We included quizzes with actual fake news in three surveys. Two of these surveys were run by YouGov and one was run using M-Turk. We use all three samples in what follows.

Table 6 replicates Table 3 by reporting the predicted probability that individuals successfully complete quizzes that contain 1 true and 1 false statement (first row) and quizzes that contain 1 true and 3 false statements (second row). In line with our results with synthetic fake news, the model predicts that, on average, the vast majority of individuals are able to successfully distinguish true from fake news stories.

	First	Second	Third	All
$\overline{\pi}\,(\text{true} \mid 1\ \text{true}, 1\ \text{false})$	0.86	0.85	0.75	0.84
$\overline{\pi}\,(\text{true} \mid 1\ \text{true}, 3\ \text{false})$	0.72	0.69	0.55	0.69

Table 6: Average Probability of Selecting True Statement

Note: The first row reports the average probability that individuals select the first, second, and third true news story of the month (as well as any true news story ranked third or higher) when faced with one true news story and one fake story. The second row reports the average probability that individuals select the first, second, and third true news story of the month (as well as any true news story ranked third or higher) when faced with one true news story and three fake stories.

Further, Table 7 replicates Table 4 by reporting the average probability $\overline{\rho}\,(h)$ that individuals hold correct beliefs for different confidence levels. Inspecting both tables reveals that the aggregate information levels we estimate are very similar independently of the method used to generate fake news.[31] Moreover, the average probability that respondents hold wrong beliefs about fake news is similar for actual and synthetic fake news, suggesting either that our journalists are able to write equally plausible fake news as those that circulate online and/or that very few respondents were actually exposed to the fake news selected by snopes.com.

[31] Table C.8 in Online Appendix C.5 replicates Table 4 by restricting attention to news quizzes with synthetic fakes news that were ran concurrently with the news quizzes with actual fake news. The figures reported in Table C.8 are similar to those reported in Table 7.

Confidence	True Story			Fake Story
	First	Second	Third	
0 - 0.25	0.02	0.02	0.07	0.45
0.25 - 0.75	0.39	0.45	0.64	0.45
0.75 - 1	0.59	0.53	0.28	0.1
0 - 0.33	0.06	0.05	0.16	0.56
0.33 - 0.66	0.26	0.29	0.44	0.29
0.66 - 1	0.68	0.66	0.4	0.15
0.5 - 1	0.82	0.84	0.61	0.26
0.6 - 1	0.73	0.73	0.47	0.19
0.7 - 1	0.64	0.6	0.35	0.13
0.8 - 1	0.53	0.45	0.22	0.08
0.9 - 1	0.39	0.25	0.11	0.04

Table 7: Average Probability of Holding Correct Beliefs $\overline{\rho}\,(h)$

Note: The table reports the average probability $\overline{\rho}\,(h)$ that respondents hold correct beliefs about the first, second, and third true news story of the month as well as about fake news.

Table 14 lists all 9 fake news that were included in the quizzes. Exactly as with the synthetic fake news, no fake news was believed to be true by more than 20% of respondents.

4.5 Effect of Partisan Congruence and Time Passing

Do individuals exhibit a tendency to believe news stories that reflect most favorably on their preferred political party?[32] If so, to what extent? When focusing on the most important political news of the month, how accurate is Obama's statement that voters live in different political information universes and that they do not share a common baseline of facts? The model assumes that all individuals are possibly biased along partisan lines, and that the extent of the bias is identical across individuals. Individuals can hold partisan beliefs for a variety of reasons, including motivated beliefs [e.g., Bénabou and Tirole, 2002, 2006] or selective exposure to news [see, for instance,

[32]Throughout, we rely on the bipartisan nature of American politics to assume that a story that reflects favorably on the Republican party reflects unfavorably on the Democratic Party. Similarly, we assume that a story that "neither reflects favorably nor unfavorably" on the Republican Party does not reflect either favorably or unfavorably on the Democratic Party either.

Sunstein, 2018, on echo chambers and social media].[33] Individuals can also (rationally) rely on their partisan priors when they are unsure about the accuracy of competing news sources, as in our news quizzes with three true and three false statements.

The model also allows time passing to matter in determining individuals' information about the news, via the decay parameter δ. Intuitively, time may matter if individuals have limited memory and older news stories receive less media coverage.

Table 8 considers a hypothetical quiz with 1 partisan true news story and 1 neutral fake news story. We rely on the model's estimates based on the news quizzes with synthetic fake news. Results when instead using the quizzes with actual fake news are similar and omitted for brevity. The first three columns of Table 8 report the function $\overline{\pi}$, for various percentiles in the distribution of the true news stories' associated partisan score b_j (10th, 25th, 50th, 75th, and 90th), by distinguishing between Republicans and Democrats and by considering news stories that are less than four weeks old.[34] On average, respondents are roughly 2% more likely to select the true news story if it reflects very favorably on their preferred party compared to if it reflects very unfavorably on their preferred party. The last two columns of Table 8 repeat the same exercise by assuming that one month and two months have elapsed since both the true and the fake news stories were written. Time plays a significant role: looking for instance at neutral stories, every month that passes reduces the probability that individuals select the true statement by roughly 7%.

Next we look at the probability that individuals hold correct beliefs about the news. The first three columns of Table 9 report the function $\overline{\rho}(h)$, for various confidence intervals and various percentiles in the distribution of b_j (10th, 25th, 50th, 75th, and 90th), by distinguishing between Republicans and Democrats and by considering true news stories that are less than four weeks old.[35] As news stories reflect less favorably on the Republican Party, the share of Republicans who hold correct beliefs over these stories falls. The effect is symmetric for Democrats. On average, individuals are roughly

[33] On echo chambers and, more generally, individuals' tendency to adopt unbalanced news diets see also Gentzkow and Shapiro [2011] and Flaxman et al. [2016].

[34] We suppose that the true and false news stories have salience parameters γ drawn from the corresponding estimated marginal posterior distributions of γ_j. We thus disregard any possible co-dependence between γ_j and b_j.

[35] We again suppose that the true news story has a salience parameters γ drawn from the marginal posterior distribution of γ_j.

Story's partisanship		Months Delayed (a) $t = 0$	(b) $t = 1$	(c) $t = 2$
		Probability of Selection		
Strongly Pro-Republican	Republican	0.8144	0.7623	0.7090
	Democrat	0.8006	0.7451	0.6887
Moderately Pro-Republican	Republican	0.8129	0.7601	0.7062
	Democrat	0.8032	0.7482	0.6923
Neutral	Republican	0.8091	0.7556	0.7010
	Democrat	0.8072	0.7532	0.6980
Moderately Pro-Democrat	Republican	0.8052	0.7503	0.6944
	Democrat	0.8119	0.7592	0.7052
Strongly Pro-Democrat	Republican	0.8015	0.7460	0.6896
	Democrat	0.8149	0.7629	0.7096

Table 8: Effect of Partisan Congruence and Time Passing on Probability of Selecting True Statement $\overline{\pi}$

Note: The table reports the average probability $\overline{\pi}$ that a supporter of a given political party selects the true statement when faced with 1 true and 1 false statement by varying the favorability toward the Republican Party of the true statement (i.e., by setting the true statement's partisan score b equal to the 10th, 25th, 50th, 75th, and 90th percentiles of the distribution of b_j) and assuming a neutral false statement. The table also reports the corresponding average probability $\overline{\pi}$ when the news stories are one-month-old and two-month-old.

7% more likely to hold correct beliefs about a true news story that reflects favorably on their preferred political party.

We now turn to the effect played by time passing in determining individuals' beliefs about real news. The last two columns of Tables 9 report, for various percentiles in the distribution of b_j, the average probability with which Republicans and Democrats hold correct beliefs about true news stories as a function of the number of months that have elapsed. Time has a rather sizable effect on the probability of holding correct beliefs about true news stories: looking again at neutral stories, every month that passes lowers this probability by 14 to 18%.

Table 9 also allow us to investigate whether the effect played by partisan congruence grows larger for older news stories [e.g., Zimmermann, 2020]. We find evidence of such a phenomenon. For example, while on average a Republican is 7% more likely to hold correct beliefs about a news story that reflects very favorably on the Republican Party (relative to very unfavorably) when the news story is less than a month old, the corresponding figure becomes close to 8% when the news story is between 5 to 8 weeks old and close to 10% when the news story is between 9 to 12 weeks old. These effects are similar for Democrats.

		Months Delayed				
			(a) $t = 0$		(b) $t = 1$	(c) $t = 2$
Story's partisanship			Confidence : $h \in$			
		$(0 - 0.25)$	$(0.25 - 0.75)$		$(0.75 - 1)$	
Strongly Pro-Republican	Republican	0.04	0.50	0.46	0.39	0.33
	Democrat	0.05	0.52	0.43	0.36	0.30
Moderately Pro-Republican	Republican	0.04	0.51	0.45	0.38	0.32
	Democrat	0.05	0.51	0.44	0.36	0.31
Neutral	Republican	0.04	0.51	0.45	0.37	0.32
	Democrat	0.05	0.51	0.44	0.37	0.31
Moderately Pro-Democrat	Republican	0.05	0.51	0.44	0.36	0.31
	Democrat	0.04	0.51	0.45	0.38	0.32
Strongly Pro-Democrat	Republican	0.05	0.52	0.43	0.36	0.30
	Democrat	0.04	0.50	0.46	0.38	0.33

Table 9: Effect of Partisan Congruence and Time Passing on Average Probability of Holding Correct Beliefs $\overline{\rho}(h)$

Note: The table reports the average probability $\overline{\rho}(h)$ that a supporter of a given political party assigns a probability of truth within a given confidence interval to true and less than a month old news stories with varying favorability toward the Republican Party (i.e., by setting the news story's partisan score b equal to the 10th, 25th, 50th, 75th, and 90th percentiles of the distribution of b_j). The table also reports the average probability that a supporter of a given political party holds correct beliefs about the same news stories when these are one-month-old and two-month-old.

4.6 Inequalities

There exists an important literature documenting the relationship between media coverage and citizens' information and, in turn, the relationship between citizens' information and the attention received from politicians. One important channel through which this accountability channel operates is voting. If voters are informed about the policies and actions implemented by politicians, the latter have greater incentives to cater to voters' preferences to increase their odds of reelection. Investigating how information levels vary across socioeconomic groups is therefore of interest: as politicians are likely aware of the link between information and voting, they have incentives to skew their policies towards the better informed voters.

To illustrate some of these dynamics, in Online Appendix B we develop a simple model of retrospective voting inspired by Strömberg [2001], Prat and Strömberg [2013], and Matějka and Tabellini [2017]. In the model, various groups of voters differ in their policy preferences $u_g(\cdot)$, their size s_g, and information levels $\bar{\rho}_g$ (the share of informed individuals in group g). We show that an incumbent politician seeking reelection has incentives to allocate weights equal to $\frac{\bar{\rho}_g}{\bar{\rho}} s_g$ on the various groups of voters, where $\bar{\rho}$ denotes the average voter's level of information. By contrast, a utilitarian social planner would allocate weights equal to s_g. The incumbent politician thus places greater weight on the better informed groups of voters.

In this section, we quantify the extent of information inequalities across socioeconomic groups. We estimate the hierarchical model described in Section 3, in which individuals' information precision θ_i are drawn from a prior distribution whose mean μ_i depend on individual characteristics. We restrict our attention to news quizzes about Federal politics with synthetic fake news exclusively. Table 10 reports, for 16 socioeconomic groups, the average predicted probability $\overline{\pi}$ that members of a particular group successfully complete a hypothetical quiz containing 1 true and 1 false statement as well as the average probability $\overline{\rho}(h)$ that they hold correct beliefs about true news stories. These figures are computed by estimating the hierarchical model in which μ_i depends on age, gender, race (white versus minority), and income (above/below the median).[36]

Our results suggest significant differences across socioeconomic groups. To take an extreme example, on average, minority, female voters aged 46 or less with a below-median income have a 71% probability of successfully completing the hypothetical quiz and a 34% probability of holding correct beliefs about true news stories. By contrast, on average, white, male voters aged 47 or more with an above-median income have an 87% probability of successfully completing a quiz (i.e., a 23% difference) and a 52%

[36] Also, we suppose that the true and false news stories are less than a month old and have salience parameters γ drawn from the corresponding estimated marginal posterior distributions and a partisan score b equal to zero. We thus consider recent news stories that are typical in terms of their salience and neutral in their partisanship.

probability of holding correct beliefs (i.e., a 53% difference).

	Age $\geq$ 47	Female	White	Income 60k+	$\overline{\pi}$ (true | 1 true, 1 false)	$h < 0.25$	$h \in (0.25, 0.75)$	$h > 0.75$
1					0.75	0.06	0.57	0.38
2				x	0.80	0.05	0.53	0.42
3			x		0.81	0.04	0.52	0.44
4			x	x	0.83	0.04	0.49	0.47
5		x			0.71	0.07	0.59	0.34
6		x		x	0.75	0.06	0.57	0.37
7		x	x		0.76	0.05	0.56	0.38
8		x	x	x	0.80	0.05	0.53	0.43
9	x				0.80	0.04	0.52	0.43
10	x			x	0.84	0.04	0.48	0.48
11	x		x		0.84	0.04	0.48	0.48
12	x		x	x	0.87	0.03	0.45	0.52
13	x	x			0.77	0.05	0.56	0.39
14	x	x		x	0.81	0.04	0.52	0.44
15	x	x	x		0.81	0.04	0.52	0.44
16	x	x	x	x	0.84	0.04	0.48	0.48

Table 10: Information Levels across Socioeconomic Groups

Note: The table reports, for 16 socioeconomic groups, the average probability $\overline{\pi}$ that an individual belonging to a given group selects the true statement when faced with one true and one false statement and the average probability $\overline{\rho}(h)$ that an individual belonging to a given group assigns a probability of truth within a given interval of confidence to a typical true news story.

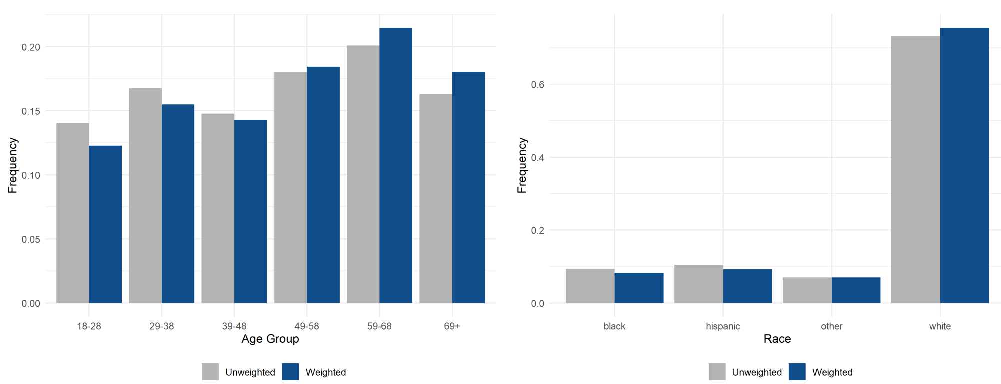
Next, we estimate different versions of the hierarchical model in which we progressively add individual characteristics. We report the mean, the 5th percentile value, and the 95th percentile value of the marginal posterior distributions of the coefficients associated with each characteristic. In what follows, we refer to the mean of the marginal posterior distribution of a coefficient simply as "coefficient." Column (1) in Table 15 (see Appendix A) allows μ_i to depend on the respondent's age, income, college education, gender, and race. Age is the most important characteristic with a coefficient equal to 0.25. To quantify the importance of this effect, note that the information precision θ_i of an average individual is a Normally distributed random variable with mean $\mu_i = 1$ and standard deviation equal to roughly 0.48. The effect of age therefore corresponds to roughly a half of a standard deviation positive shift in the prior distribution of information precision. Further, the coefficients associated with college education and income are equal to 0.16 and 0.13 respectively. The coefficient associated with gender is -0.21.[37] Similarly, the coefficients associated with being African-American and Hispanic are equal to -0.22 and -0.2 respectively. Column (2) adds affiliations with the Democratic and the Republican Parties (the excluded category are Independents) and Column (3) adds a variable measuring whether individual i feels either strongly republican or strongly democrat.[38] Focusing on column (3), the coefficients associated with being a Democrat and a Republican are equal to 0.19 and 0.2 and the coefficient associated with being strongly partisan is equal to 0.17, suggesting (i) that supporters of the Republican and Democratic Parties are associated with higher information levels relative to Independents and (ii) that strength of partisanship is positively associated with information precision. The inclusion of these additional characteristics does not affect much the coefficients associated with the 5 socioeconomic factors (which the exception of the coefficient associated with being African-American, which shifts to the left). Next, Columns (4) and (5) add media consumption habits. In both columns the number of news outlets and time usage (in hours) are positively and strongly associated with information precision. In particular, an additional hour of news consumption corresponds to a 138-148% of a standard deviation increase in the prior distribution of information precision.[39] Also, including media consumption habits virtually removes any association between strength of partisanship and information precision. Finally, Column (6) adds News Interest as an individual characteristic. Self-reported interested in politics has been highlighted by previous work as an important predictor of news literacy [e.g., Prior, 2007]. Our results are consistent: we find that the coefficient associated with general interest in politics is equal to 0.21.

[37] Lizotte and Sidman [2009] suggests that most of the gap in political knowledge between men and women can be explained by differences in risk aversion because women are less likely to venture a guess when they are not certain. However, our methodology is not affected by that potential issue because all subjects must pick three statements.

[38] Recall from Table 1b that only a tiny share of survey respondents report that they do not identify with a party. For this reason, looking at the effect of supporting *any* party is of limited interest.

[39] The median individual in our sample reports spending 240 minutes a week consuming national news and the average individual reports spending 432 minutes (see Online Appendix A.1). These numbers are in line with other measures of news media consumption. For instance, in 2010 the Pew Research Center reported that the average American spent 70 minutes a day consuming the news [Pew, 2010]. Nevertheless, self-reported measures of news consumption are notoriously exaggerated and we thus interpret the coefficients associated with news consumption with caution [see, e.g., Prior, 2009, Guess, 2015]. In Online Appendix C.6 we replicate some of our main results on two samples of U.S. voters recruited through MTurk. MTurk survey participants report spending significantly less time consuming the news compared to YouGov survey participants.

We return to our theoretical framework to illustrate the relevance of our findings from a political economy angle. In Figure 2, the grey bars correspond to the size of various age groups in our sample. By contrast, the blue bars represent the actual weights an incumbent seeking reelection would allocate these various groups, say when designing a policy that affects voters of different ages differently. Consistent with our results above, the incumbent will behave as if voters aged 49 or more make up close to 58% of all voters, even though they make up less than 54% of all voters in reality. Similarly, the incumbent will behave as if whites make up close to 75.5% of the population (in contrast to their actual share of 73%). Comparing these numbers with current U.S. demographic trends helps to assess magnitudes.[40] For instance, the incumbent behaving as if the population of whites is 2-3 percentage points larger than what it actually is is roughly equivalent to saying that the incumbent behaves with a 10 year lag with respect to the demographic composition of the U.S. population.



(a) By Age Groups

(b) By Race

Figure 2: Inequalities in Information about the News

Note: Grey bars correspond to the size of various age groups in our sample. Blue bars correspond to the weights an incumbent seeking reelection would allocate these various groups.

[40] See, for instance, https://www.brookings.edu/blog/the-avenue/2018/03/14/the-us-will-become-minority-white-in-2045-census-projects/.

4.7 Taking stock: Ideological Polarization vs Socio-economic inequality

Our analysis found that both ideological polarization and socio-economic inequality explain the information patterns we observe. However, a cursory look at the estimated coefficients seems to indicate that the role played by socio-economic inequality is much larger. This section confirms this intuition by providing a comparison of the two factors. Ideological polarization could affect voter information about the news in three ways. First, it could be that Republicans are systematically more informed than Democrats or viceversa. However, we find no evidence for such effect. Second, it could be that more extreme voters are less informed than less extreme voters. However, we actually find that the opposite is true. Third, it could be that polarization takes the form suggested by many commentators including Obama: Voters on different sides of the political spectrum have different information. We do find evidence for this partisan congruence effect, and the rest of this section will compare the magnitude of this effect with that of socioeconomic factors.

To perform this comparison, we consider a counterfactual world where there is no partisan congruence between individuals and news stories. As we will argue that the magnitude of polarization is much lower than that of inequality, we make a best-case assumption for the no-polarization world: in this ideal world, the ability of a voter to identify any real news story is the same as the ability of an actual voter to identify congruent news stories. We measure the average increase in $\rho(h)$ that we obtain by moving from our polarized world to this ideal non-polarized world as well as the average increase in $\rho(h)$ that we would obtain if voters in low-information socioeconomic groups had the same ability to identify news stories as voters in high-information socioeconomic groups.

Table 11 focuses on and pools supporters of either the Republican or the Democratic Party. The first row reports the average probability of holding correct beliefs over true news stories, $\overline{\rho}(0.75)$, assuming that these stories are politically congruent. This corresponds to the ideal non-polarized world described above, in which individuals' ability to identify real news stories is the same as their ability to identify congruent news stories. Specifically, we assume that the news stories paired with Democrats lie on the 25th percentile in the distribution of b_j and the news stories paired with Republicans lie on the 75th percentile in the distribution of b_j. We thus assume that the favorability of these news stories towards individuals' preferred political party corresponds to the median level of favorability observed in the data.[41] The first row also distinguishes between various socioeconomic factors as well as membership to the least and the most-informed socioeconomic groups according to Table 10 (i.e., groups 5 and 12). The second row performs a similar exercise, instead assuming median levels of political *incongruence*–in other words, the second row imagines a worst-case polarized world. Comparing both rows therefore provides an upper-bound on the increase in $\overline{\rho}\,(0.75)$ obtained when moving from a polarized world to a non-polarized world. Across all columns (i.e., across all subsamples of individuals), the increase when moving from the bottom row to the top row is always lower than 1 percentage point. By contrast, varying each socioeconomic factor individually leads to changes in $\overline{\rho}\,(0.75)$ ranging from 4 to 7 percentage points. Looking at percentage changes, the effect of varying any socioeconomic factor is roughly 3 times larger than the effect of varying a news story's partisan congruence. Strikingly, belonging to the most versus the least-informed socioeconomic group leads to an increase in $\overline{\rho}\,(0.75)$ about 18 times larger than the increase that occurs when a news story is congruent versus non-congruent.

These findings, combined with our earlier results regarding aggregate information levels, suggest that the largest policy concern might not be the death of journalistic truth or its ideological polarization, but rather information inequality along socioeconomic lines. We discuss some policy implications in Section 6. Naturally, our results should not be interpreted as suggesting that political polarization has been exaggerated or that liberals and conservatives do not disagree about important matters such as climate change or the merits of vaccination. Rather, our results indicate that when it comes to factual news information socioeconomic differences appear to play a much larger role than ideological polarization.[42]

[41]Moreover, the values of γ we select are drawn from the marginal posterior distribution of γ_j, that is, we disregard any possible co-dependence between γ's and b's.

[42]In Online Appendix C.1, we perform an alternative exercise to assess the importance of ideological polarization vs socio-economic inequality – specifically, we compute groups of individuals' ability to distinguish true from fake news – and we draw the same conclusion.

	Gender		Race		Age		Family Income		Group	
	Female	Male	Non White	White	< 47	≥ 47	<\$60k	≥\$60k	5	12
Median-polarization news stories										
Congruent	0.422	0.475	0.405	0.466	0.413	0.478	0.424	0.477	0.343	0.533
Non-congruent	0.416	0.467	0.399	0.459	0.406	0.472	0.417	0.47	0.337	0.526

Table 11: Partisan Congruence versus Socioeconomic Factors

Note: The table reports the probability $\overline{\rho}\,(0.75)$ of holding correct beliefs about a true news story whose salience parameter γ is drawn from the marginal posterior distribution of γ. The table restricts attention to supporters of the Republican and the Democratic Party. The top row ("Congruent") reports the average of (i) the average probability $\overline{\rho}\,(0.75)$ that Republicans hold correct beliefs over a news story with an associated partisan score b that lies on the 75th percentile in the distribution of b_j and (ii) the average probability $\overline{\rho}\,(0.75)$ that Democrats hold correct beliefs over a news story with an associated partisan score b that lies on the 25th percentile in the distribution of b_j. The bottom row ("Non-congruent") reports the average of (i) the average probability $\overline{\rho}\,(0.75)$ that Republicans hold correct beliefs over a news story with an associated partisan score b that lies on the 25th percentile in the distribution of b_j and (ii) the average probability $\overline{\rho}\,(0.75)$ that Democrats hold correct beliefs over a news story with an associated partisan score b that lies on the 75th percentile in the distribution of b_j. Columns restrict attention to particular subsamples of individuals.

5 Extensions and Robustness Checks

5.1 Sports and Entertainment

The "post-truth" phenomenon is often discussed with respect to political news. For this reason, we sought a benchmark outside of politics and included news quizzes about Sports and Entertainment in two of our surveys (four quizzes in total). We estimate our main model to obtain the posterior distributions of the various parameters of interest by combining and relying exclusively on the news quizzes about Sports and Entertainment. We assume these news to be neutral in their partisanship and we rely on synthetic fake news only. Online Appendix C.2 presents our main results. U.S. citizens appear just as well informed about Politics as they are about Sports and Entertainment. On average, respondents hold correct beliefs about true news stories covering Sports and Entertainment with probability 45%. The corresponding figure for true news stories about Federal politics is 44% (in those news quizzes ran concurrently with the quizzes about Sports and Entertainment).

5.2 Democratic Party Presidential Primaries

Most of our news quizzes about the Federal Government included news directly related to the Donald Trump presidency. We cannot exclude the possibility that the sizable differences in information levels across socioeconomic groups that we find are somehow driven by Donald Trump's four-year tenure in the White House. In 5 surveys, we included news quizzes devoted exclusively to news about the Democratic Party presidential primaries. Noting that Democratic primary voters are more likely to be young, female, minority and low-income compared to presidential election voters [Kamarck and Podkul, 2018], this extension allows to perform an external validity exercise. Exactly as before, we estimate the main model highlighted in Section 3 to obtain the posterior distributions of the various parameters of interest. The model is estimated using the quizzes about the Democratic Party presidential primaries exclusively, and it relies on synthetic fake news only. Our main findings are reported in Online Appendix C.3, and Tables C.4 and C.5 include all the true news stories and all the synthetic fake news that were used in the news quizzes.

We again find evidence of large information inequalities. Looking at the least and best-informed socioeconomic groups, we find that, on average, white, male voters aged 47 or more with an above-median income are 47% more likely to hold correct beliefs about true news stories covering the Democratic primaries than minority, female voters aged 46 or less with a below-median income. They are also close to 16% more likely to identify a true news story when faced with one true and one fake news story.

5.3 Robustness Checks

Recall that our main analysis excluded the 19% of respondents who selected fewer than 3 statements when completing the news quizzes. If the tendency to select fewer than 3 statements is correlated with information precision, one may worry that excluding these respondents may bias our results. In Online Appendix C.4, we replicate our main analysis by imputing respondents' "missing choices." Specifically, for all the respondents who selected fewer than 3 statements, we choose uniformly at random the missing choices from the remaining unselected items. As expected, we find that the accuracy of respondents' beliefs decreases somewhat: for example, on average, respondents hold correct beliefs about the most important news story of the month with probability 53% (against 59% when not imputing missing choices).

Finally, to alleviate concerns that some of our findings may be driven by YouGov survey participants being particularly news savvy, in Online Appendix C.6 we replicate our main results by relying exclusively on the news quizzes about the Federal Government that we administered through M-Turk. Our main results line up with the analysis performed using the YouGov sample.

6 Concluding Remarks

The current policy debate has focused mostly on the "death of truth", the prevalence of fake news, and the polarization of news diets along partisan lines. Our findings cast doubt on the policy solutions, mentioned in the introduction, that have been proposed to counter fake news and ideological polarization. Are those policies – especially the most drastic ones, like re-thinking the scope of the First Amendment – appropriate? Much more evidence is needed to diagnose the problem we are trying to solve before we take such extreme steps.[43] Although we are only able to measure the most basic knowledge of the news, our results suggest that the vast majority of voters are able to confidently identify as true the political news mainstream journalists consider to be the most important.

Instead, our findings indicate that the starkest pattern about the ability of voters to identify major news stories is not the generalized death of truth or its ideological polarization but rather its unequal distribution along socioeconomic lines. A sizeable minority of U.S. voters have very low information levels and they are more likely to be found among women, minorities, the young, and the poor. In this sense, our work is related to an old literature in sociology that documents strong heterogeneity in political attitudes across socioeconomic groups [see, e.g., Gaxie, 1978].

Our paper thus begs a fundamental question about voter information: Where do these large inequalities come from? One can advance three possible hypotheses. First, it could be that the we chose a political topic – domestic policy – that is much more

[43] Caution is needed also because outside the U.S. a number of regimes have invoked the death-of-truth narrative to pass laws against fake news that amount to censorship. In 2019, Russia passed a law to block social media that spread factually inaccurate information (as determined by the state). Turkey passed a similar law in 2020. On the role of information in autocracies see [Guriev and Treisman, 2019]

interesting to certain segments. However, this explanation is not promising because we find the same inequality patterns when we investigate news on the Democratic primaries. Second, it could be an economic-driven access story. Certain segments of the population have less disposable income to spend and are therefore less likely to consume costly news sources. However, this is unlikely to be the whole explanation because inequality patterns survive after controlling for income.

The most likely explanation is that the journalistic truth – at least in its current form – is more appealing to certain socioeconomic groups than to others. In turn, this explanation may be due to many factors. One factor that deserves further scrutiny is that news producers tend to belong to the high-information groups: The newsroom staff at major national newspapers is overwhelmingly male [WMC, 2019] and white [Arana, 2018]. Nonetheless, much more research is needed to understand the causes of these inequalities and to analyze the merits of policies that are aimed at equalizing access to news sources [see, for instance, Cagé, 2016, for proposals to change the way news media is funded].

References

Hunt Allcott and Matthew Gentzkow. Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31(2):211–236, 2017.

Hunt Allcott, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. The Welfare Effects of Social Media. *American Economic Review*, 110(3):629–76, 2020.

Jennifer Allen, Baird Howland, Markus Mobius, David Rothschild, and Duncan J. Watts. Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances*, 6(14), 2020a.

Jennifer N. L. Allen, Antonio A. Arechar, Gordon Pennycook, and David G. Rand. Scaling up Fact-checking Using the Wisdom of Crowds. *Working Paper*, 2020b.

Carolyn J. Anderson, Jay Verkuilen, and Buddy L. Peyton. Modeling Polytomous Item Responses Using Simultaneously Estimated Multinomial Logistic Regression Models. *Journal of Educational and Behavioral Statistics*, 35(4):422–452, 2010.

Gabriel Arana. Decades of Failure. *Columbia Journalism Review*, 2018.

Eric Arias, Horacio Larreguy, John Marshall, and Pablo Querubin. Does the Content and Mode of Delivery of Information Matter for Political Accountability? Evidence from a Field Experiment in Mexico. *Working Paper*, 2018.

Eric Arias, Horacio Larreguy, John Marshall, and Pablo Querubin. Priors Rule: When Do Malfeasance Revelations Help or Hurt Incumbent Parties. *Working Paper*, 2019.

Young Min Baek and Magdalena E. Wojcieszak. Don't expect too much! learning from late-night comedy and knowledge item difficulty. *Communication Research*, 36(6): 783–809, 2009.

Julian Baggini. *A Short History of Truth: Consolations for a Post-Truth World*. Quercus Publishing, 2017.

Bence Bago, David G. Rand, and Gordon Pennycook. Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. 149(8):1608–1613, 2020.

James Ball. *Post-Truth: How Bullshit Conquered the World*. Biteback Publishing, 2018.

Abhijit Banerjee, Selvan Kumar, Rohini Pande, and Felix Su. Do Informed Voters Make Better Choices? Experimental Evidence from Urban India. *Working Paper*, 2012.

Oscar Barrera Rodriguez, Sergei M. Guriev, Emeric Henry, and Ekaterina Zhuravskaya. Facts, Alternative Facts, and Fact Checking in Times of Post-Truth Politics. *CEPR Working Paper*, 2018.

Roland Bénabou and Jean Tirole. Self-Confidence and Personal Motivation. *Quarterly Journal of Economics*, 117, 2002.

Roland Bénabou and Jean Tirole. Belief in a Just World and Redistributive Politics. *Quarterly Journal of Economics*, 121(2):699–746, 2006.

Darrell R. Bock. Estimating Item Parameters and Latent Ability When Responses Are Scored in Two or More Nominal Categories. *Psychometrika*, 37:29–51, 1972.

Lee C. Bollinger and Geoffrey R. Stone. *The Free Speech Century*. Oxford University Press, 2018.

John G. Bullock, Alan S. Gerber, Seth J. Hill, and Gregory A. Huber. Partisan Bias in Factual Beliefs about Politics. *Quarterly Journal of Political Science*, 10, 2015.

Julia Cagé. *Saving the Media*. Belknap Press, 2016.

Yuyu Chen and David Y. Yang. The Impact of Media Censorship: 1984 or Brave New World? *American Economic Review*, 109(6):2294–2332, 2019.

Michael d'Ancona. *Post-Truth: The New War on Truth and How to Fight Back*. Ebury Press, 2017.

Stefano DellaVigna and Ethan Kaplan. The Fox News Effect: Media Bias and Voting. *The Quarterly Journal of Economics*, 122(3):1187–1234, 2007.

Michael X. Delli Carpini and Scott Keeter. *What Americans Know about Politics and Why It Matters*. Yale University Press, 1996.

Thomas Eisensee and David Strömberg. News Droughts, News Floods, and U. S. Disaster Relief. *The Quarterly Journal of Economics*, 122(2):693–728, 2007.

Ruben Enikolopov, Maria Petrova, and Ekaterina Zhuravskaya. Media and Political Persuasion: Evidence from Russia. *American Economic Review*, 101(7):3253–85, 2011.

Eurobarometer. *Standard Eurobarometer 87*. European Commision, 2017.

Claudio Ferraz and Frederico Finan. Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes. *The Quarterly Journal of Economics*, 123(2):703–745, 2008.

Seth Flaxman, Sharad Goel, and Justin M. Rao. Filter Bubbles, Echo Chambers, and Online News Consumption. *Public Opinion Quarterly*, 80(S1):298–320, 2016.

Daniel Gaxie. *Le cens caché: Inégalités culturelles et ségrégation politique*. Editions du Seuil, Paris, 1978.

Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman & Hall, 2014.

Matthew Gentzkow and Jesse M. Shapiro. Ideological Segregation Online and Offline. *The Quarterly Journal of Economics*, 126(4):1799–1839, 2011.

Matthew Gentzkow, Jesse M. Shapiro, and Daniel F. Stone. Media bias in the marketplace: theory. In Simon Anderson, Joel Waldfogel, and David Stromberg, editors, *Handbook of Media Economics*, chapter 2. North Holland, 2015.

Alan S. Gerber, Dean Karlan, and Daniel Bergan. Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions. *American Economic Journal: Applied Economics*, 1(2):35–52, 2009.

James. Gleick. *The Information. A History. A Theory. A Flood.* Pantheon Books, 2011.

Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. Fake news on Twitter during the 2016 U.S. presidential election. 363(6425): 374–378, 2019. doi: 10.1126/science.aau2706.

Andrew M. Guess. Measure for Measure: An Experimental Test of Online Political Media Exposure. *Political Analysis*, 23(1):59–75, 2015.

Andrew M. Guess, Jonathan Nagler, and Joshua Tucker. Less than you think: Prevalence and predictors of fake news dissemination on facebook. *Science Advances*, 5(1), 2019.

Andrew M. Guess, Michael Lerner, Benjamin Lyons, Jacob M. Montgomery, Brendan Nyhan, Jason Reifler, and Neelanjan Sircar. A digital media literacy intervention increases discernment between mainstream and false news in the united states and india. *Proceedings of the National Academy of Sciences*, 117(27):15536–15545, 2020a.

Andrew M. Guess, Dominique Lockett, Benjamin Lyons, Jacob M. Montgomery, Brendan. Nyhan, and Jason Reifler. "fake news" may have limited effects beyond increasing beliefs in false claims. *Harvard Kennedy School (HKS) Misinformation Review*, 2020b.

Andrew M. Guess, Brendan Nyhan, and Jason Reifler. Exposure to untrustworthy websites in the 2016 us election. *Nature Human Behaviour*, 4, 2020c.

Sergei Guriev and Daniel Treisman. Informational autocrats. *Journal of Economic Perspectives*, 33(4):100–127, November 2019.

Matthew D Hoffman and Andrew Gelman. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15 (1):1593–1623, 2014.

Michiko Kakutani. *The Death of Truth: Notes on Falsehood in the Age of Trump*. Tim Duggan Books, 2018.

Elain Kamarck and Alexander R. Podkul. The 2018 Primaries Project: The demographics of primary voters, 2018.

Chad Kendall, Tommaso Nannicini, and Francesco Trebbi. How Do Voters Respond to Information? Evidence from a Randomized Campaign. *American Economic Review*, 105(1):322–53, January 2015.

Brian Knight and Ana Tribin. Opposition Media, State Censorship, and Political Accountability: Evidence from Chavez's Venezuela. *NBER Working Paper 25916*, 2019.

Julien Labonne, Cesi Cruz, and Philip Keefer. Buying Informed Voters: New Effects of Information on Voters and Candidates. *Working Paper*, 2019.

David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, and Jonathan L. Zittrain. The science of fake news. *Science*, 359(6380):1094–1096, 2018. ISSN 0036-8075. doi: 10.1126/science.aao2998.

Daniel J. Levitin. *A Field Guide to Lies: Critical Thinking in the Information Age*. New York, New York: Dutton, 2016.

Mary-Kate Lizotte and Andrew H. Sidman. Explaining the gender gap in political knowledge. *Politics amp; Gender*, 5(2):127–151, 2009.

Gregory J. Martin and Ali Yurukoglu. Bias in Cable News: Persuasion and Polarization. *American Economic Review*, 107(9):2565–2599, 2017.

Filip Matějka and Guido Tabellini. Electoral Competition with Rationally Inattentive Voters. *Working Paper*, 2017.

Robinson Meyer. The Grim Conclusions of the Largest-Ever Study of Fake News. *The Atlantic*, 2018.

Sendhil Mullainathan and Andrei Shleifer. The market for news. *American Economic Review*, 95(4):1031–1053, September 2005.

Brendan Nyhan. Facts and Myths about Misperceptions. *Journal of Economic Perspectives*, 34(3):220–36, 2020a.

Porter E. Reifler J. Nyhan, B. Taking Fact-Checks Literally But Not Seriously? The Effects of Journalistic Fact-Checking on Factual Beliefs and Candidate Favorability. *Political Behavior*, 42, 2020b.

Gordon Pennycook and David G. Rand. Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188:39–50, 2019. doi: 10.31234/osf.io/3n9u8.

Gordon Pennycook, Ziv Epstein, Mohsen Mosleh, Antonio Arechar, Dean Eckles, and David Rand. Understanding and reducing the spread of misinformation online. 2019.

Gordon Pennycook, Jonathon McPhetres, Yunhao Zhang, Jackson G. Lu, and David G. Rand. Fighting covid-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science*, 31(7):770–80, 2020.

Gordon Pennycook, Ziv Epstein, Mohsen Mosleh, Antonio A. Arechar, Dean Eckles, and David G. Rand. Shifting attention to accuracy can reduce misinformation online. *Nature*, 592:590–595, 2021.

Pew. *2010 Media Consumption Survey*. Pew Research Center, 2010.

Pew. *From Brexit to Zika: What Do Americans Know?* Pew Research Center, 2017.

Pew. *Wide Gender Gap, Growing Educational Divide in Voters' Party Identification*. Pew Research Center, 2018.

Pew. *Many Americans Say MadeUp News Is a Critical Problem That Needs To Be Fixed*. Pew Research Center, 2019.

Du Phan, Neeraj Pradhan, and Martin Jankowiak. Composable effects for flexible and accelerated probabilistic programming in numpyro, 2019.

Andrea Prat and David Strömberg. *The Political Economy of Mass Media*, volume 2 of *Econometric Society Monographs*, page 135–187. Cambridge University Press, 2013.

Vincent Price and John Zaller. Who Gets the News? Alternative Measures of News Reception and Their Implications for Research. *The Public Opinion Quarterly*, 57 (2):133–164, 1993.

Markus Prior. News vs. entertainment: How increasing media choice widens gaps in political knowledge and turnout. *American Journal of Political Science*, 49(3):577–592, 2005.

Markus Prior. *Post-Broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections*. Cambridge Studies in Public Opinion and Political Psychology. Cambridge University Press, 2007.

Markus Prior. The immensely inflated news audience: Assessing bias in self-reported news exposure. *Public Opinion Quarterly*, 73(1):130–143, 2009.

Markus Prior. Visual Political Knowledge: A Different Road to Competence? *The Journal of Politics*, 76(1):41–57, 2014.

Markus Prior and Arthur Lupia. Money, Time, and Political Knowledge: Distinguishing Quick Recall and Political Learning Skills. *American Journal of Political Science*, 52 (1):169–183, 2008.

Markus Prior, Gaurav Sood, and Kabir Khanna. You Cannot be Serious: The Impact of Accuracy Incentives on Partisan Bias in Reports of Economic Perceptions. *Quarterly Journal of Political Science*, 10, 2015.

Georg Rasch. *Probabilistic Models for Some Intelligence and Attainment Tests*. Danish Institute for Educational Research, 1960.

Szymon Sacher, Laura Battaglia, and Stephen Hansen. Hamiltonian Monte Carlo for Categorical Data. Manuscript, Imperial College London, 2021.

Anya Schiffrin. *Mis-And Disinformation Online: A Taxonomy Of Solutions*. PhD thesis, Universidad de Navarra, 2020.

James M. Snyder and David Strömberg. Press Coverage and Political Accountability. *Journal of Political Economy*, 118(2):355–408, 2010.

Richard Stengel. *Information Wars. How We Lost the Global Battle Against Disinformation and What We Can Do about It*. Grove Press, 2019.

David Strömberg. Mass Media and Public Policy. *European Economic Review*, 45, 2001.

David Strömberg. Mass Media Competition, Political Competition, and Public Policy. *The Review of Economic Studies*, 71(1):265–284, 2004.

David Strömberg. Media and Politics. *Annual Review of Economics*, 7(1):173–205, 2015.

Cass R. Sunstein. *Republic: Divided democracy in the age of social media*. Princeton University Press, 2018.

Cass R. Sunstein. *Too Much Information: Understanding What You Don't Want to Know*. The MIT Press, 2019.

Zeynep Tufekci. *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. Yale University Press, 2017.

Wim J. Van der Linden and Ronald K. (eds) Hambleton. *Handbook of Modern Item Response Theory*. H. New York, NY: Springer, 1997.

Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. 359(6380):1146–1151, 2018. doi: 10.1126/science.aap9559.

WMC. The Status of Women in the U.S. Media 2019. *WMC Reports*, 2019.

Tim Wu. Is the First Amendement Obsolete. *Michigan Law Review*, 117(3), 2018.

Florian Zimmermann. The Dynamics of Motivated Beliefs. *American Economic Review*, 110(2):337–61, 2020.

A Appendix Tables

					$\overline{\rho}\,(h)$		
Statement	Raw Mean	b	γ	Prob of selecting	$h < 0.25$	$h \in (0.25, 0.75)$	$h > 0.75$
At a closed-door meeting at the White House, top envoy to China delivered evidence of rising Farm Belt frustration over bio-fuel policy.	0.36	-0.35	0.03	0.41	0.12	0.67	0.21
Mitch McConnell avoided White House, citing laxity on masks, COVID-19 precautions	0.48	-0.37	0.06	0.5	0.11	0.67	0.22
Second U.S. presidential debate officially canceled after Trump balked	0.62	-0.64	0.39	0.63	0.05	0.65	0.29
The Trump administration credited cooperation from Mexico and Central American countries in cracking down on migrants.	0.63	0.32	0.21	0.64	0.08	0.67	0.25
Senior U.S. House members vowed to pass major defense bill despite pandemic	0.64	0.1	0.07	0.65	0.11	0.67	0.22
President Trump proposed plan to make U.S. immigration more merit-based.	0.66	0.52	0.34	0.66	0.06	0.65	0.29
U.S. Supreme Court allowed President Trump's 'Remain in Mexico' asylum policy	0.66	0.11	0.47	0.67	0.05	0.63	0.32
In win for President Trump, U.S. Supreme Court made deporting immigrants for crimes easier	0.66	0.27	0.13	0.67	0.09	0.67	0.23
Supreme Court granted a request by President Trump's administration to fully enforce a new rule that would curtail asylum applications by immigrants at the U.S.-Mexico border.	0.69	-0.02	0.82	0.69	0.02	0.56	0.42
Mexico agreed to take more migrants seeking asylum in the United States while they await adjudication of their cases.	0.7	0.42	0.5	0.71	0.04	0.63	0.33
President Trump said he would address national debt if re-elected	0.71	0.43	0.19	0.69	0.08	0.67	0.25
Joe Biden said U.S. coronavirus death toll to probably top 500,000 by end of February	0.72	0.07	0.54	0.73	0.04	0.62	0.34
U.S. Senate Republican leader McConnell said Trump 'provoked' Jan. 6 riot	0.75	-0.44	0.64	0.75	0.03	0.59	0.37
Republican lawmakers in the House of Representatives condemned President Trump's decision to withdraw troops from Syria.	0.75	0.1	0.42	0.71	0.05	0.65	0.3
President Trump notified Congress he is firing the inspector general of U.S. intelligence community	0.78	-0.46	0.85	0.77	0.03	0.55	0.43
The House of Representatives passed legislation seeking to rein in President Trump's ability to deploy U.S. forces to fight abroad	0.85	-0.11	1.19	0.83	0.02	0.45	0.53
Attorney General William Barr said that President Trump's attacks on prosecutors, the judge and jurors in the trial of Roger Stone undermined the Justice Department's work	0.88	-0.47	1.34	0.85	0.01	0.41	0.57
Trump Supreme Court pick Amy Coney Barrett pledged to follow law, not personal views	0.89	0.42	1.56	0.89	0.01	0.33	0.66
Alabama's governor signed a bill to ban nearly all abortions in the state.	0.9	0.35	1.56	0.9	0.01	0.33	0.66
Whistle-blower report complains of White House cover-up on Trump-Ukraine scandal.	0.9	-0.76	1.99	0.9	0.01	0.25	0.74
President Trump declared coronavirus a national emergency	0.93	0.66	2	0.91	0.01	0.26	0.73
Joe Biden sworn in as U.S. president	0.94	-0.84	2.17	0.94	0.01	0.21	0.78
A whistleblower filed a complaint against President Trump, leading to an impeachment inquiry.	0.94	-0.73	1.84	0.93	0.01	0.28	0.71

The U.S Senate acquitted Trump of impeachment charges	0.95	-0.08	2.67	0.94	0.01	0.17	0.82

Table 12: True News Stories

***Note*:** The table reports all the true statements included in our quizzes about the Federal Government. For each statement, it also reports the share of respondents who selected it ('raw mean'), the partisan score b (a positive value indicates that the average respondent felt the statement reflected favorably on the Republican Party, and conversely for a negative value), the parameter γ (i.e., the statement's straightforwardness), the model's predicted share of respondents who select the statement, as well as the model's predicted average probability $\overline{\rho}(h)$ that individuals assign a probability within a given confidence interval to the story being true.

					$\overline{\rho}\,(h)$		
Statement	Raw Mean	b	γ	Prob of selecting	$h < 0.25$	$h \in (0.25, 0.75)$	$h > 0.75$
Biden team's Twitter handle under fire after mistakenly reposting anti-Trump tweets	0.07	0.65	-2.93	0.07	0.82	0.15	0.03
A Tape surfaced of President Trump supporting abortion	0.07	-0.63	-2.58	0.08	0.76	0.2	0.04
President Trump's Tax Returns showed billions given to various charities.	0.09	0.31	-3.11	0.08	0.82	0.15	0.03
President Trump fired coronavirus advisor Dr. Anthony Fauci	0.11	-0.94	-1.89	0.12	0.66	0.28	0.05
Mitt Romney decided to run for president against Trump in the 2020 race after breakout role in impeachment	0.12	-0.15	-1.87	0.11	0.66	0.29	0.05
Mike Pence Revealed Bombshell Allegations in Impeachment Trial	0.13	-0.33	-1.68	0.14	0.64	0.3	0.06
Trump administration to continue to allow U.S. research using fetal tissue from abortions.	0.13	-0.01	-1.75	0.14	0.66	0.29	0.05
President Trump took a week-long break from Campaigning to Deal with Coronavirus Outbreak	0.14	0.42	-0.78	0.19	0.37	0.51	0.11
Kanye West called for special prosecutor if Biden elected	0.16	-0.34	-1.4	0.16	0.58	0.36	0.07
President Trump announced his tax returns will be released by Mid-May	0.17	0.13	-1.65	0.18	0.64	0.31	0.06
ISIS beheaded three Americans in response to Al-Baghdadi's death.	0.17	-0.44	-1.74	0.15	0.65	0.3	0.06
Attorney General Barr released text message from Special Counsel prosecutor Robert Mueller: 'We're taking down Trump.'	0.19	-0.07	-1.27	0.19	0.54	0.38	0.07
Around 20% of IRS stimulus checks bounced	0.19	-0.69	-1.48	0.2	0.6	0.34	0.06
Nancy Pelosi under investigation by Justice Department over alleged insider trading during coronavirus outbreak	0.21	0.15	-1.02	0.21	0.45	0.45	0.09
China blacklists Apple and Microsoft amid escalating trade war.	0.23	-0.36	-0.66	0.24	0.34	0.54	0.12
Agriculture trade group marched in Washington to draw attention to export problems	0.31	-0.37	-0.46	0.32	0.27	0.59	0.14
President Trump tweeted about Black Lives Matters protests taking place in front of Mar-a-Lago	0.37	-0.42	-0.33	0.36	0.23	0.62	0.15
President Trump announces he will resume peace talks with Iran at UN General Assembly	0.37	0.35	-0.13	0.36	0.16	0.65	0.18
Biden in favor of temporarily barring guests from Capitol and other federal buildings	0.38	0.1	-0.32	0.37	0.23	0.61	0.16
China and the United States agreed on a new comprehensive trade deal.	0.41	0.65	-0.15	0.49	0.17	0.65	0.18
U.S. Border Patrol facility admitted to measles outbreak among migrant children in custody.	0.42	-0.07	-0.26	0.41	0.21	0.63	0.17
Vaping case to make its way to Supreme Court.	0.44	0.13	-0.03	0.39	0.13	0.67	0.2
White House to host election night viewing party, Fauci calls it 'potential disaster'	0.48	-0.7	-0.06	0.45	0.14	0.66	0.2
President Trump's campaign saw steep rise in donations after press conferences	0.63	0.36	-0.05	0.6	0.14	0.66	0.2

Table 13: Synthetic Fake News

***Note*:** The table reports all the synthetic false statements included in our quizzes about the Federal Government. For each statement, it also reports the share of respondents who selected it ('raw mean'), the partisan score b (a positive value indicates that the average respondent felt the statement reflected favorably on the Republican Party, and conversely for a negative value), the parameter γ (i.e., the statement's straightforwardness), the model's predicted share of respondents who select the statement, as well as the model's predicted average probability $\overline{\rho}\,(h)$ that individuals assign a probability within a given confidence interval to the story being true.

					$\overline{\rho}\,(h)$		
Statement	Raw Mean	b	γ	Prob of selecting	$h < 0.25$	$h \in (0.25, 0.75)$	$h > 0.75$
"Antifa" arsonists have been setting wildfires raging on the West Coast in September 2020	0.07	0.2	-2.38	0.08	0.75	0.21	0.04
Ruth Bader Ginsburg said that pedophilia was good for children	0.08	-0.02	-2.35	0.08	0.74	0.22	0.04
CNN issued a correction that read, 'Sen. Ted Cruz was seen wearing a pin featuring a QAnon symbol. It was later discovered that this was not a QAnon pin, but a Doritos snack chip stuck to his suit.'	0.15	0.27	-1.38	0.16	0.56	0.37	0.07
As of late January 2021, Donald Trump had started a new U.S. political party called the 'Patriot Party.'	0.17	0.03	-1.22	0.18	0.52	0.4	0.08
U.S. Rep. Marjorie Taylor Greene said 'If English was good enough for Jesus, it's good enough for us.'	0.2	-0.05	-1.06	0.2	0.47	0.44	0.09
Kentucky Attorney General Daniel Cameron is married to U.S. Senator Mitch McConnell's granddaughter	0.2	-0.36	-0.68	0.2	0.35	0.53	0.12
Democratic U.S. presidential nominee Joe Biden said that he grew up in section 8 housing during town hall debate.	0.28	0.03	-0.59	0.28	0.32	0.56	0.13
President Trump said: 'The doctors said they've never seen a body kill the Coronavirus like my body. They tested my DNA and it wasn't DNA. It was USA.'	0.35	-0.49	-0.3	0.35	0.22	0.62	0.16
While speaking about Violent Crime Control and Law Enforcement Act of 1994, Joe Biden referred to Black Americans as 'super-predators.'	0.42	0.31	-0.1	0.41	0.15	0.66	0.19

Table 14: Actual Fake News

***Note*:** The table reports all the actual fake news included in our quizzes about the Federal Government. For each statement, it also reports the share of respondents who selected it ('raw mean'), the partisan score b (a positive value indicates that the average respondent felt the statement reflected favorably on the Republican Party, and conversely for a negative value), the parameter γ (i.e., the statement's straightforwardness), the model's predicted share of respondents who select the statement, as well as the model's predicted average probability $\overline{\rho}\,(h)$ that individuals assign a probability within a given confidence interval to the story being true.

	Dependent variable: μ_i					
	(1)	(2)	(3)	(4)	(5)	(6)
Democrat		0.1969	0.1906	0.1189	0.1105	0.103
		(0.1252 , 0.2752)	(0.1182 , 0.2645)	(0.0266 , 0.2106)	(0.0052 , 0.2055)	(0.0049 , 0.205)
Republican		0.1009	0.2052	0.1531	0.1259	0.1085
		(0.034 , 0.1735)	(0.1227 , 0.288)	(0.0381 , 0.2594)	(0.0124 , 0.2377)	(-0.0054 , 0.2176)
Strong Partisan			0.1687	0.0501	0.0258	-0.0175
			(0.0958 , 0.2412)	(-0.0416 , 0.1367)	(-0.0649 , 0.1145)	(-0.1085 , 0.0752)
News Interest						0.2076
						(0.1315 , 0.2839)
Age $\geq$ 47	0.2512	0.2666	0.2579	0.2095	0.1906	0.1627
	(0.1956 , 0.3075)	(0.2122 , 0.3227)	(0.2007 , 0.3151)	(0.1394 , 0.2785)	(0.1167 , 0.2618)	(0.0878 , 0.2421)
Income $\geq$ 60k	0.1334	0.1393	0.1355	0.0918	0.0749	0.0591
	(0.0735 , 0.1889)	(0.0809 , 0.1963)	(0.0792 , 0.1918)	(0.019 , 0.1677)	(4e-04 , 0.1502)	(-0.0138 , 0.137)
College +	0.1645	0.1533	0.1534	0.1272	0.11	0.1002
	(0.1087 , 0.2247)	(0.0966 , 0.2116)	(0.0978 , 0.2097)	(0.0572 , 0.1947)	(0.0382 , 0.1847)	(0.0304 , 0.1703)
Female	-0.2088	-0.2149	-0.2118	-0.1121	-0.1054	-0.0872
	(-0.2637 , -0.1529)	(-0.2704 , -0.1608)	(-0.2669 , -0.1539)	(-0.1846 , -0.043)	(-0.1785 , -0.0339)	(-0.1558 , -0.0167)
Black	-0.2174	-0.2773	-0.2797	-0.29	-0.2662	-0.2325
	(-0.3045 , -0.1293)	(-0.3703 , -0.1874)	(-0.3684 , -0.1942)	(-0.394 , -0.185)	(-0.3728 , -0.1603)	(-0.341 , -0.122)
Hispanic	-0.1951	-0.2167	-0.2089	-0.181	-0.1588	-0.1573
	(-0.2797 , -0.1135)	(-0.2989 , -0.1386)	(-0.2907 , -0.1233)	(-0.2924 , -0.0735)	(-0.2686 , -0.0515)	(-0.2706 , -0.0448)
Sources 3+				0.077	0.112	0.0791
				(4e-04 , 0.1556)	(0.0068 , 0.214)	(-0.0317 , 0.183)
Total Time (hrs)				0.7155	0.6631	0.5244
				(0.4219 , 1.0108)	(0.3853 , 0.9341)	(0.237 , 0.8361)
N	4138	4138	4138	2358	2333	2304
Extra media controls					X	X

Table 15: Socioeconomic Factors

***Note*:** The dependent variable is the mean μ_i of the prior distribution from which individual i's information precision θ_i is drawn. The table reports the mean, the 5th percentile value, and the 95th percentile value of the marginal posterior distribution of the coefficient associated with each individual characteristic included in the hierarchical model. Strong Partisan is a dummy variable taking value 1 if individual i reports being either a strong republican or a strong democrat. Sources 3+ is a dummy variable taking value 1 if individual i reports relying on 3 or more news media outlets during previous 7 days. Total Time is the number of hours dedicated to consuming national news during previous 7 days reported by individual i. News Interest is a dummy variable taking value 1 if individual i reports being interested in general politics. Extra media controls include: voter registration, Indicators for using tv, print, online and radio as a news source, as well as dummies for 10 biggest news sources interacted with using at least 3 sources. Media consumption questions were not included in every survey.

Le LIEPP (Laboratoire interdisciplinaire d'évaluation des politiques publiques) est un laboratoire d'excellence (Labex) distingué par le jury scientifique international désigné par l'Agence nationale de la recherche (ANR). Il est financé dans le cadre des investissements d'avenir *(ANR-11-LABX-0091, ANR-11-IDEX-0005-02)* ***et de l'IdEx Université de Paris*** *(ANR-18-IDEX-0001).*

www.sciencespo.fr/liepp

A propos de la publication

Procédure de soumission :
Rédigé par un ou plusieurs chercheurs sur un projet en cours, le *Working paper* vise à susciter la discussion scientifique et à faire progresser la connaissance sur le sujet étudié. Il est destiné à être publié dans des revues à comité de lecture (peer review) et à ce titre répond aux exigences académiques. Les textes proposés peuvent être en français ou en anglais. En début de texte doivent figurer : les auteurs et leur affiliation institutionnelle, un résumé et des mots clefs.

Le manuscrit sera adressé à : liepp@sciencespo.fr

Les opinions exprimées dans les articles ou reproduites dans les analyses n'engagent que leurs auteurs.

Sciences Po - LIEPP
27 rue Saint Guillaume
75007 Paris - France
+33(0)1.45.49.83.61
liepp@sciencespo.fr