DISCUSSION PAPER SERIES

DP15222

MEASURING VOTERS' KNOWLEDGE OF POLITICAL NEWS

Charles Angelucci and Andrea Prat

INDUSTRIAL ORGANIZATION

PUBLIC ECONOMICS

ISSN 0265-8003

MEASURING VOTERS' KNOWLEDGE OF POLITICAL NEWS

Charles Angelucci and Andrea Prat

Discussion Paper DP15222
Published 29 August 2020
Submitted 25 August 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

MEASURING VOTERS' KNOWLEDGE OF POLITICAL NEWS

Abstract

We propose a methodology to measure knowledge of news about recent political events that combines a protocol for identifying stories, a quiz to elicit knowledge, and the estimation of a model of individual knowledge that includes difficulty, partisanship, and memory decay. We focus on news about the Federal Government in a monthly sample of 1,000 US voters repeated 11 times. People in the most informed tercile are 97% more likely than people in the bottom tercile to know the main story of the month. We document large inequalities across socioeconomic groups, with the best-informed group over 14 percentage points more likely to know the typical story compared to the least-informed group. Voters are 10-30% less likely to know stories unfavorable to their political party. Also, each month passing lowers the probability of knowing a story by 3-4 percentage points. We repeat our study on news about the Democratic Party primaries.

JEL Classification: N/A

Keywords: N/A

Charles Angelucci - cangeluc@mit.edu
MIT Sloan

Andrea Prat - andrea.prat@columbia.edu
Columbia University and CEPR

Measuring Voters' Knowledge of Political News*

Charles Angelucci[†] Andrea Prat[‡]
MIT Sloan Columbia University

July 2020

Abstract

We propose a methodology to measure knowledge of news about recent political events that combines a protocol for identifying stories, a quiz to elicit knowledge, and the estimation of a model of individual knowledge that includes difficulty, partisanship, and memory decay. We focus on news about the Federal Government in a monthly sample of 1,000 US voters repeated 11 times. People in the most informed tercile are 97% more likely than people in the bottom tercile to know the main story of the month. We document large inequalities across socioeconomic groups, with the best-informed group over 14 percentage points more likely to know the typical story compared to the least-informed group. Voters are 10-30% less likely to know stories unfavorable to their political party. Also, each month passing lowers the probability of knowing a story by 3-4 percentage points. We repeat our study on news about the Democratic Party primaries.

Keywords: media, knowledge, inequality

JEL Classification Numbers: L82, D72, D90

*We thank Roland Bénabou, Alessandro Bonatti, Laurent Bouton, Julia Cagé, Teresa Esteban Casanelles, Simone Galperti, Leo Goldman, Emeric Henry, Amit Khandelwal, José L. Montiel Olea, Suresh Naidu, Jacopo Perego, Jett Pettus, Vincent Pons, Miklos Sarvary, Andrey Simonov, Sevgi Yuksel, and seminar audiences at Columbia University, CSEF, the University of Exeter, Harvard University, the NYC Media Seminar, 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.

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

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

1 Introduction

The media plays an important role in providing citizens with the information they need to keep government accountable. Informed citizens are aware of what the government does and are thus in a position to punish or reward the incumbent at the next election. The central role played by the media in maintaining government accountability is well-documented by a growing body of literature in political economy. For example, in the US, Snyder and Strömberg [2010] find that political districts with greater media coverage elect representatives who work harder to promote their constituents' interests. Similarly, in Uganda, Reinikka and Svensson [2005] document that schools in areas with greater newspaper coverage are better run. This logic applies to new media: Gavazza et al. [2018] show that the expansion of broadband internet in the UK crowded out local news and reduced local public spending.[1,2]

A government that is aware of the link between information and voting behavior is also more likely to cater to the better-informed voters. This proposition has received empirical support: for example, Strömberg [2004] shows that US counties with higher radio ownership received greater federal funding during the New Deal. The logic can be formalized in a simple model of retrospective voting [Strömberg, 2001, Prat and Strömberg, 2013]. An incumbent politician knows that voters care about her policies. If different social groups have different levels of information, better informed groups will be more responsive to the incumbent's behavior and the latter will design policies that cater to them.[3] Inequalities in information are likely to exacerbate other types of inequalities [Delli Carpini and Keeter, 1996].

Voters' knowledge of political news is, therefore, a key ingredient of many political economy models. Those theories do not just consider average knowledge but also how knowledge is distributed across topics and voters. Indeed, there exists a sizeable body of work that measures voter knowledge, with some of it focusing on news knowledge.

Polling organizations regularly report survey results on voter knowledge [e.g., Pew,

[1]Other papers showing an effect of news coverage on political outcomes include Eisensee and Strömberg [2007], Ferraz and Finan [2008], Gerber et al. [2009], Enikolopov et al. [2011], Banerjee et al. [2012], Kendall et al. [2015], Labonne et al. [2019], Arias et al. [2018], Arias et al. [2019], Knight and Tribin [2019], and Cagé [2020]. See Strömberg [2015] for a survey.

[2]Media bias also affects political outcomes [e.g., DellaVigna and Kaplan, 2007, Gentzkow et al., 2015, Martin and Yurukoglu, 2017].

[3]A model developed in Online Appendix B shows that if $\bar{\rho}_g$ is the average news knowledge level in social group g, a re-election seeking incumbent will choose her behavior as if maximizing a welfare function where each group's weight is proportional not only to its size but also its knowledge level.

2017, Eurobarometer, 2017].[4] On the academic side, the public opinion literature has provided a number of measures for political knowledge. Price and Zaller [1993] measure recall of 16 news stories. Examples of survey questions include: "Do you remember any recent stories about Marine Colonel Oliver North receiving a sentence for his conviction in the Iran-Contra Affair? [If yes:] Do you recall anything about what sentence he received?" and "Do you recall any stories about a U.S. Supreme Court decision this summer on abortion? [If yes:] Do you remember what the court decided?" They find that respondents' background level of political knowledge is the strongest predictor of current news recall across a wide range of topics.

The canonical work in this area is Delli Carpini and Keeter [1996], who collate about 3,700 questions asked in various surveys from 1940 to 1993, with the objective to measure the American public's level of political knowledge. They divide questions into five categories, one of which is domestic politics. In the last year for which they have information (1990), the statements are: "Who will pay for S&L bailout?"; "Why is the Hubel telescope in the news?"; "Did Bush veto a plant closing bill?"; "What is the illiteracy rate in US?"; "What is the percentage of population that is Hispanic/Black/Jewish?".[5]

In recent years, news knowledge has been examined from the perspective of fake news. Some commentators have argued that misinformation spread through social media has played an important role in elections around the world [e.g., Levitin, 2016, Stengel, 2019]. Allcott and Gentzkow [2017] measure consumption and recall of fake news in the 2016 election, and Barrera Rodriguez et al. [2018] investigate the role played by fake news and fact-checking on French voters' beliefs and political preferences. Lazer et al. [2018] discuss the prevalence and impact of the phenomenon and potential interventions. More recently, Allcott et al. [2019] measure the effect of Facebook on news knowledge.[6] To measure knowledge, they include a list of 15 true and false statements and ask respondents to select which, in their opinion, are true. The true statements are borrowed from recent articles published in leading US

[4] The American National Election Studies (ANES) also include two questions on political knowledge: 'Which party had most members of congress before the election?' and 'Which party had most members of congress after the election?'.

[5] Prior and Lupia [2008] measure political knowledge by administering surveys that include 14 questions about facts relevant to the 2004 presidential election. They find that typical survey methods (quick, unincentivized questions) likely underestimate voters' true knowledge of politics.

[6] See also Chen and Yang [2019] on the relationship between consumption of uncensored information and knowledge of current events in China.

media outlets. The false statements are either modifications of existing articles from the same sources or recent fake news identified by third-party fact-checkers. Allcott et al. [2019] show that Facebook usage tends to increase knowledge of the news.

While the existing literature has uncovered important patterns about voter information, its analysis of the particular area of knowledge we are interested in – political news – displays three related gaps.

First, any knowledge measurement exercise faces an initial challenge: what set of knowledge items should voters be tested on? As the examples above illustrate, this challenge is hard because the set of possible items is unstructured, heterogeneous, and virtually unbounded. To the best of our knowledge, the existing literature approaches this challenge by letting the researchers select the knowledge items over which survey respondents are quizzed. While this methodology is natural, it has a drawback. Only the researcher knows what universe of knowledge items he or she considered and what criterion he or she used to select within that universe the items that ended up on the survey. This creates problems in terms of interpretability and replicability.

Second, political events have an essential time dimension. The big stories of this month are usually different from the big stories of last month. They may be more or less important, they may favor a different political side, or they may relate to different issues. To get a representative picture of voter knowledge, one should perform the same survey repeatedly over time. However, to the best of our knowledge there is no repeated academic study of this kind, perhaps because the researcher-led selection protocol used so far does not naturally lend itself to replication over time.

Third, to the best of our knowledge the existing literature does not attempt to estimate a microfounded structural model of news knowledge that distinguishes among various factors, such as individual news knowledge, partisan bias, news salience, memory decay, etc. This is probably due to the first two obstacles, as this kind of empirical analysis requires multiple surveys based on a well-defined source of comparable news stories.

This paper contributes to the literature by attempting to overcome these three issues. First, we develop a codified news selection protocol that is outside the control of the researcher. Second, we employ this protocol repeatedly to survey a comparable sample of voters over 11 months. Third, we use the resulting data to estimate a structural model of news knowledge that disentangles the factors mentioned above.

Our news selection process consists of two steps: (i) Selection of the universe of relevant news items. The protocol selects a news source, sets an inclusion criterion, and identifies the set of stories that satisfy that criterion. The researcher has no hand on the content and wording of the stories. (ii) Selection of the knowledge items to be included in the test. The protocol specifies a process to select a subset of (i). This step may rely on the subjective judgment of other agents, but the process must be codified. For (i), this paper uses the set of all Reuters news wires devoted to US national politics. For (ii), we assemble a panel of journalists and ask them to select – within the subset identified in (i) – the three most important stories of the month about the Federal Government. We then conduct surveys to measure US voters' knowledge of these stories.

The importance of a story is clearly a subjective matter, and any attempt to measure importance ultimately relies on someone's judgment. Even an algorithmic approach, such as that used by Google News to rank stories, is ultimately built on the subjective views of its users. The goal of our proposed approach 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 claim their views are representative of mainstream journalism: so our survey measures how much voters know about stories that mainstream journalists think are important. Instead, an algorithmic approach based on, say, Google News would instead be less transparent, as neither the ranking algorithm nor the users' characteristics are known. If we chose stories on the basis of that, we would not exactly know whose subjective judgment we are relying upon.[7]

We exploit the protocol in a number of ways. Chiefly, we repeat the survey for 11 months on 11 different panels of approximately 1,000 US voters. On several occasions, we also included 1- and 2-month-old stories, to measure knowledge decay over time. Finally, we extend the protocol to news about the Democratic Party presidential primaries, chosen among the same set of Reuters news wires about national politics and ranked by the same panel of journalists.

Once news stories about the Federal Government are selected, we measure knowledge in a financially incentivized survey in many ways similar to those used by, for instance, Allcott et al. [2019], Guess [2015], Prior et al. [2015], Bullock et al. [2015], and

[7] An advantage of our approach is that we can try to measure how our selectors differ from the rest of the population in their views of which stories are most important. See section 4.1.

Chen and Yang [2019].[8] Respondents are selected by YouGov, a polling company, to produce a nationally representative sample of US adult citizens. As part of the survey, respondents take multiple quizzes. In each quiz, we present our respondents with 6 items: the 3 most important knowledge items of the month according to our panel of journalists as well as 3 plausible but false statements. Consistent with our approach to real news, we rely on the panel of journalists to create the false statements. The false statements cover the Federal Government and are written in the same journalistic style as the true knowledge items. Survey respondents are given 60 seconds to select the 3 statements which, to the best of their recollection, are true. They receive a monetary reward in case all 3 true knowledge items are chosen.[9]

The survey data is used to estimate the distributions of parameters of a news knowledge model. In our model knowledge is a continuous variable: when a respondent is confronted with a news story (true or false), she assigns a probability of truth between zero and one that depends on (i) features of the story like salience and partisanship (e.g., whether the story reflects favorably on the Republican Party) and (ii) features of the respondent like knowledge and ideology. The respondent uses these assigned 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. While every news story is different and may be harder or easier, the stochastic generating process for both true and fake stories is exogenously given. The main object of interest is the posterior distribution of the respondent-level knowledge parameter, but we also obtain estimates for the salience and partisanship of each story, as well as the effect of time passing on news knowledge.

In our main analysis, we measure voters' knowledge of news stories about the Federal Government. An agent's knowledge of a particular news story is the estimated probability the agent assigns to that story being true. Our findings can therefore be reported at different knowledge levels. If for now we define "knowledge" as attributing a chance equal to at least 75% that a news story is true, according to our estimates

[8]On the role of partisanship 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 knowledge see also Prior and Lupia [2008].

[9]This approach implicitly defines knowledge as awareness of a fact. A deeper notion of knowledge entails understanding that fact. One may be aware that President Trump was impeached without truly understanding what the impeachment process is. One limitation of our approach is that we only attempt to measure this more superficial form of knowledge.

the average voter knows 1.3 of the 3 most important news stories of the month. About 64% of US voters know the most important story of the month, and the share of US voters who know the second and third most important stories of the month falls to 37% and 32%, respectively.

Significant heterogeneity across voters exists. For instance, the average individual in the top-third of the distribution knows roughly 1.9 out of 3 news stories. By contrast, the average individual in the bottom-third of the distribution knows roughly 0.9 news story. Similarly, significant heterogeneity across news stories exists, with some stories known by over 80% of individuals and others by fewer than 20%. Reassuringly, only a tiny share of individuals believes the typical true story to be false. Further, we find that time significantly affects knowledge of political news: we document that one month of time reduces by 3-4 percentage points the share of voters who know a given story. We also find a relatively large effect of partisanship on knowledge, with respondents being 10-30% more likely to know news stories that reflect favorably on their preferred political party. We also measure inequalities in news knowledge across socioeconomic groups (defined by age, gender, race, and income). According to our estimates, the average individual in the best-informed group (wealthy white men aged 47 and more) is about 47% more likely to know the typical news story compared to the average individual in the least-informed group (low-income minority young women).[10]

In an extension, we illustrate the replicability of our methodology by focusing on a different set of knowledge items. In 5 surveys, we rely on our panel of journalists to select the 3 most important stories of the month regarding the Democratic Party presidential primaries.

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

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

2 Design

The key components in our analysis are knowledge quizzes, in which respondents are rewarded if they succeed in choosing the true knowledge items included in a list containing both true and false knowledge items. We review the protocol we have employed to generate the true and false knowledge items. We also describe the information we have 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 US Federal Government according to mainstream media.

Universe of Relevant Knowledge Items. We rely on Reuters' publicly-available wire stories about US national politics to approximate the universe of relevant knowledge items.[11] This choice allows us to focus on essential facts covered by mainstream media. 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 US national politics.

Generating 3 True and 3 False Knowledge Items. We rely on a panel of 3 professional journalists recruited through the Columbia School of Journalism.[12] To avoid recency effects, each week, each journalist is asked to select the 5 most important wire stories of the week according to him/her.[13] 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,

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

[12]We describe the protocol we eventually arrived at that shields the production of news from researcher interventions. As we ran surveys modifications were gradually introduced to remove our involvement in the production of the 3 true and 3 false statements (e.g., in the early surveys we would harmonize the use of past tenses across statements or select the false statements ourselves from the list produced by the panel).

[13]Although we give significant 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 US-centered criteria of importance. All jury members are US citizens.

we take the four/five previous weeks' selected wire stories and filter out the wire stories that do not cover the Federal Government (by far, most stories deal with the Federal Government).[14,15] 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 (we rely on randomization to break eventual ties). 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*).[16]

Our main instrument to estimate voters' knowledge of political news consists of asking them to select 3 out of 6 statements. Three of these statements correspond to the 3 true statements described in the previous paragraph. The remaining 3 statements are false short statements about the Federal Government. We relied on our panel of journalists to produce these plausible but ultimately 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.[17,18]

Why did we rely on a panel of human journalists to identify top stories, rather than use some more "objective" machine learning algorithm? One could for instance select the most clicked stories in aggregators like Google News or the most popular articles on mainstream media like the New York Times, or use some ranking that is based on those numbers. But obviously such approach would rely on subjective

[14] We adopt the US definition of the "Federal Government" as being composed of the legislative, executive, and judicial branches. In the first 5 surveys we did not filter out the news stories that did not cover the Federal Government (the inclusion criterion was simply "national politics"). As we show below, our results are unaffected when we restrict our attention to the last 6 surveys.

[15] During our time period, the few stories that do not cover the Federal Government deal with the presidential primaries. In Section 5, we replicate our analysis by focusing on the Democratic Party presidential primaries.

[16] Often, the story that summarizes 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.

[17] 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.

[18] Notice that we could have relied on fake news that actually circulated online, by for instance using third-party fact-checkers. Although it would be interesting to use our method to quantify the extent to which voters believe in fake news, in this paper we limit ourselves to measuring voters' knowledge of real news.

judgment too, that of Google News users or New York Times readers, who are likely to be different in terms of knowledge, partisanship, and taste from other voters. Note that whatever makes Google News users or New York Times readers more likely to click on a story is likely to affect their knowledge of that story too, thus biasing the rest of the analysis.[19]

2.2 Survey Design

This paper exploits data gathered from 11 online surveys we conducted through polling company YouGov. The first survey took place in December 2018 and the last survey in June 2020.[20] For each survey, we asked YouGov to enroll a representative sample of the US citizen adult population.[21] All surveys were administered to 1,000 individuals, except for one survey which was administered to 1,500 individuals. We instructed YouGov to avoid enrolling individuals who participated in prior editions of the survey. This restriction was lifted from the eighth survey onward. Overall, 7,865 individuals participated in our 11 surveys. YouGov provides a wide array of background information concerning each survey respondent (demographics, income, education, party affiliation, interest in politics, etc.), where the information is collected months before our surveys.[22] 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 included a 50¢ show up fee and bonuses worth $1 for each quiz correctly answered.[23]

Table 1 provides basic descriptive statistics regarding the socioeconomic characteristics of the survey respondents who participated in all 11 surveys. It also reports the

[19] In ongoing work, we investigate whether an algorithm can replicate our panel's choices.

[20] Our analysis exploits all the surveys we have run to date. Funding availabilities determined the number and the sequencing of our surveys. Notice also that our time period does not coincide with a presidential election. Recent research suggests that it is information acquired over long periods of time that determines most voters' beliefs [Le Pennec and Pons, 2019].

[21] 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.

[22] As a robustness check, we replicated one survey on a distinct sample of respondents recruited through M-Turk (see Online Appendix C.4).

[23] Our description of the survey is based on the last 7 surveys we administered. Some modifications were introduced as we conducted more surveys. We highlight these modifications when relevant.

Statistic	YouGov	ACS 2018
Median Age	49.00	47.00
% Female	0.52	0.51
% White	0.69	0.73
% Black	0.11	0.13
% 4yr College Degree	0.30	0.31
% Unemployed	0.07	0.06
% Married	0.48	0.48
% Family Inc <30k	0.28	0.17
% Family Inc 30k - 60k	0.20	0.23

Table 1: Socioeconomic Characteristics

Party Affiliation	YouGov	Pew 2018
% Democrat	45	48
% Republican	35	39
% Independent	16	7
% Other	4	6

Table 2: Party Affiliations

corresponding statistics for the population of US adult citizens according to the 2018 American Community Survey of the Census Bureau (ACS).[24] All dimensions appear broadly aligned with the general population, with the exception of family income.

Table 2 reports information on the party affiliation of our survey respondents, and compares it with the statistics provided by Pew [2018].[25] For the purposes of this paper, 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.

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.[26]

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

[25] 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".

[26] In a number of surveys, we collected information on voting intentions and feelings toward the

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.[27] We use this data to create the dummy variables Television_i, Print_i, Radio_i, Online_i. We also create the discrete variable Media_i, defined as the sum of these 4 dummy variables. For all survey respondents who selected one or more types of media, we further asked them to report the news sources they relied on (e.g., CNN and Facebook). We used this information to create the discrete variable News Sources_i.[28] Finally, survey respondents were asked to report the amount of time they dedicated to getting information about national politics. We used this information to code the variable Time_i. Tables E.1 and E.2 in Online Appendix E present the language used in the corresponding survey questions. Table A.1 in Online Appendix A reports summary statistics.

2.2.2 Knowledge of the News

All surveys included 1 or 2 knowledge quizzes about current news stories (less than 4 weeks old).[29] In a number of surveys, we also included 5-8-week-old and 9-12-week-old knowledge quizzes. Overall, we included 16 distinct knowledge quizzes in our 11 surveys. Our average respondent took 1.86 knowledge 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 correct statements.[30] To avoid individuals from obtaining information elsewhere, respondents were given 60 seconds to make their selection. Whereas no incentives were given during the first survey (in addition to the base compensation), from the second survey onward we offered an extra $1 (paid via a giftcard) to all respondents who selected all three correct statements. All survey respondents were revealed the correct answers once they took the quiz. Tables

main political parties. On a few occasions we also elicited how important respondents felt the various news stories included in our quizzes were. Similarly, we sometimes included a test that aimed at measuring the attention paid to the survey. With the exception of news story importance (which we comment on below), we ended up not analyzing this information.

[27]Media consumption questions were included only in surveys 1-8.

[28]Many news sources are available across media. We consolidated news sources as appropriate.

[29]For an overview of our survey design see Figure E.1 in Online Appendix E.

[30]We discuss alternative quiz designs in Online Appendix D.

E.3-E.13 in Online Appendix E include all knowledge quizzes that we administered through our series of surveys. Table E.14 in Online Appendix E reports how the various quizzes were allocated to the various surveys we administered. Presumably because of the 60-second limit, 19% of respondents ended up selecting a number of statements different from 3.[31] We exclude these respondents from our main analysis. In Online Appendix C.2, we re-estimate the model by including respondents who selected fewer than 3 statements. Across all surveys and quizzes, our average survey respondent selected approximately 2.20 true statements (standard deviation: 0.68).

In the last 7 surveys, we asked our survey respondents to report their feelings towards the 6 statements contained in the quiz they completed. Specifically, for each true statement, respondents were asked how favorably, in their opinion, the 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", "unfavorable", "neither unfavorable nor favorable", "'favorable", and "very favorable". We used the resulting information to construct the continuous variable $b_j \in [-\infty, \infty]$ to measure the average respondent's opinion regarding the extent to which statement j reflects favorably on the Republican Party.[32] Across all quizzes, the average true statement has $b = -0.03$ (standard deviation: 0.21), that is, the average survey respondent felt that the average true statement reflected slightly unfavorably on the Republican Party. Similarly, across all surveys and quizzes, the average false statement received a score of $b = -0.04$ (standard deviation: 0.21). Tables E.15 and E.16 in Online Appendix E present the language used in the corresponding survey questions.

3 Model

We develop our model in three steps. We first formulate the basic general problem an agent faces when she is trying to assign a probability of truth to a statement, which is a standard application of Bayesian binary hypothesis testing. In the second step

[31] The vast majority of these respondents selected strictly fewer than 3 statements.

[32] To construct it, we first map the answers such that "neither unfavorable nor favorable" is represented by 0, "very unfavorable" is represented by -1 and "very favorable" is represented by 1. Then, for each statement, we take the average of this measure across respondents, and rescale the resulting variable to have a standard deviation equal to 1.

we 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. Finally, we apply this theoretical 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. In the last subsection, we clarify the link between our model and the existing statistical literature.

3.1 The News Knowledge Problem

Suppose agent i is trying to establish the truth of statement j, which we call $q_j \in \{0,1\}$, where 0 represents a false statement and 1 a true statement. The agent observes a signal y_{ij} about the statement. For simplicity, assume the signal is continuously distributed and has full support on the real line. The signal's conditional distribution depends on q_j, on the agent's knowledge precision θ_i, on the statement's characteristics γ_j (e.g., straightforwardness, salience, or familiarity), and on the number of months t since the story was written: $f\left[y_{ij}|q_j,\theta_i,\gamma_j,t\right]$. The agent is also endowed with a prior probability that the statement is true, which depends on the statement's partisanship b_j and on the agent's party affiliation p_i: $g\left[q_j=1|p_i,b_j\right]$. The agent's posterior probability that the statement is true, $\Pr\left[q_j=1|y_{ij}\right]$, is given by:

$$\frac{f\left[y_{ij}|q_j=1,\theta_i,\gamma_j,t\right]g\left[q_j=1\mid p_i,b_j\right]}{f\left[y_{ij}|q_j=1,\theta_i,\gamma_j,t\right]g\left[q_j=1|p_i,b_j\right]+f\left[y_{ij}|q_j=0,\theta_i,\gamma_j,t\right]g\left[q_j=0|p_i,b_j\right]}.$$

Suppose we wish to know whether the agent believes the statement is true with at least probability $h\in(0,1)$. The relevant condition is:

$$\frac{f\left[y_{ij}|q_j=1,\theta_i,\gamma_j,t\right]}{f\left[y_{ij}|q_j=0,\theta_i,\gamma_j,t\right]}\geq\frac{g\left[q_j=0|p_i,b_j\right]}{g\left[q_j=1|p_i,b_j\right]}\frac{h}{1-h}, \tag{1}$$

or:

$$\begin{aligned}&\ln f\left[y_{ij}|q_j=1,\theta_i,\gamma_j,t\right]-\ln f\left[y_{ij}|q_j=0,\theta_i,\gamma_j,t\right]\\ \geq\ &\ln g\left[q_j=0|p_i,b_j\right]-\ln g\left[q_j=1|p_i,b_j\right]+H,\end{aligned}$$

where $H = \ln(h/(1-h))$. The left-hand side of the inequality is a function of the random variable y_{ij}. As y_{ij} is in turn distributed according to $f\left[y_{ij}|q_j,\theta_i,\gamma_j,t\right]$, we can write the left-hand side as x_{ij}, a real-valued random variable distributed according to some $\tilde{f}\left[x_{ij}|q_j,\theta_i,\gamma_j,t\right]$. The first part of the right-hand side is a deterministic function of p_i and b_j, which we write as $\tilde{g}\left[p_i,b_j\right]$. Thus, the agent assigns at least probability h to statement j being true if:

$$x_{ij} \geq \tilde{g}\left[p_i, b_j\right] + H. \tag{2}$$

Let $\tilde{F}$ be the cumulative distribution function of $\tilde{f}$. For any level h, the probability that the agent assigns at least probability h to statement j is $1 - \tilde{F}\left[\tilde{g}\left[p_i,b_j\right] + H\right]$. This expression is a characterization of the agent's belief in the truth of statement j in terms of the threshold h and the underlying parameters p_i and b_j.

3.2 A Discrete Choice Model

We now make a number of functional form assumptions that lead to a tractable and familiar logit specification. Assume that the random variable on the left-hand side of (2) can be written as:

$$x_{ij} = \left(2q_j - 1\right)\gamma_j\theta_i\delta^t + \lambda - \varepsilon_{ij},$$

where ε_{ij} has a standard Gumbel CDF and λ is a free parameter to be determined later. Recall that we interpret $\theta_i \geq 0$ as agent i's knowledge precision and $\gamma_j \geq 0$ as the straightforwardness (the opposite of difficulty) of the news story. The parameter $\delta \geq 0$ captures the effect of time passing, with $t = 0, 1, \cdots$. Also, assume the prior term can be written as

$$\tilde{g}\left[p_i, b_j\right] = -\alpha b_j p_i.$$

Again, recall that we interpret $b_j \in (-1, 1)$ 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 party affiliation, 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 tendency of voters to believe statements that agree with their ideology and the parameter $\alpha \geq 0$ measures the strength of this partisan effect.[33]

[33] In Online Appendix C.1, we allow for $\tilde{g}$ to depend on time t and re-estimate the model.

This formulation is equivalent to the agent assigning to statement j a *plausibility value*

$$z_{ij} := x_{i,j} - \tilde{g}\left[p_i, b_j, t\right] = (2q-1)\,\gamma_j \theta_i \delta^t + \lambda + \alpha b_j p_i - \varepsilon_{ij},$$

where z_{ij} has a Gumbel distribution with location parameter $(2q-1)\,\gamma_j \theta_i \delta^t + \lambda + \alpha b_j p_i$.

Now suppose the agent is given a set J of statements. What is the probability that the agent assigns to statement j the highest probability of truth? This is similar to a standard logit discrete choice model. Each statement is characterized by its truth q_j, its straightforwardness γ_j, and its partisanship b_j. All statements are t-month old. The error term is i.i.d. across statements. The statement with the highest probability of truth is the statement with the highest associated plausibility value z_j.

Proposition 1 *The probability agent i believes statement j is the most likely to be true among the set J of statements is*

$$\begin{aligned}\pi_{ij} &= \frac{\exp\left(\left(2q_j - 1\right)\gamma_j \theta_i \delta^t + \lambda + \alpha b_j p_i\right)}{\sum_{k\in J} \exp\left(\left(2q_k - 1\right)\gamma_k \theta_i \delta^t + \lambda + \alpha b_k p_i\right)} \\ &= \frac{\exp\left(\left(2q_j - 1\right)\gamma_j \theta_i \delta^t + \alpha b_j p_i\right)}{\sum_{k\in J} \exp\left(\left(2q_k - 1\right)\gamma_k \theta_i \delta^t + \alpha b_k p_i\right)}.\end{aligned} \tag{3}$$

The comparative statics of the expressions above are intuitive.[34] If j is a true (false) statement, π_{ij} is increasing (decreasing) in i's knowledge precision θ_i and j's straightforwardness γ_j. As the agent becomes infinitely knowledgeable ($\theta_i \to \infty$) or the statement becomes infinitely easy ($\gamma_j \to \infty$), the probability tends to 1 if the statement is true and to zero if it is false.

We define $\rho_{ij}(h)$ as the probability that statement j passes a hypothesis test with threshold h:

$$z_{ij} \geq H = \ln \frac{h}{1-h}. \tag{4}$$

[34]The expression above holds under the assumption that the random variable ε_{ij} is independent across the six 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 Emmanuel 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).

For any value of H we can compute the probability that z_{ij} is greater than H:

$$\Pr\left[z_{ij} \geq H \mid (2q-1)\,\gamma_j \theta_i \delta^t + \lambda + \alpha b_j p_i\right] = 1 - e^{-e^{(2q-1)\gamma_j\theta_i\delta^t+\lambda+\alpha b_j p_i - H}}. \tag{5}$$

To calibrate λ, consider a story with $\gamma = 0$ and $b = 0$. This is a story over which the agent's knowledge precision θ and prior are of no use. We assume the agent believes such a story with probability 0.5. This implies that if we set $h = 0.5$ – and hence $H = 0$ – we have:

$$\Pr\left[z_{ij} \geq 0 \mid 0\right] = 1 - e^{-e^{\lambda}} = \frac{1}{2}, \tag{6}$$

which holds if and only if $\lambda = \ln(\ln 2) \simeq -0.36651$. In what follows, we set $\lambda = \ln(\ln 2)$. Therefore, the probability that the agent assigns at least probability h to statement j being true is equal to:

$$\rho_{ij}(h) = 1 - e^{-e^{(2q-1)\gamma_j\theta_i\delta^t+\ln(\ln 2)+\alpha b_j p_i - \ln\frac{h}{1-h}}} \tag{7}$$

Below, we often rely on expression (7) to convey our empirical findings.

3.3 Econometric Model

In our survey quizzes, respondents are given 6 statements (ordered randomly). They are told that exactly 3 statements are true and they receive \$1 if they successfully select these 3 true statements. This creates some mechanical correlation between answers. For instance, if I think that 1 statement is true and I know that only 3 statements are true, then I must be too pessimistic about the other statements. This mechanical correlation is fully incorporated in the estimation procedure. Intuitively, the information that exactly 3 statements are true does not affect the optimal strategy of the respondent: to pick the 3 statements that are most likely to be true (i.e., the 3 statements with the highest plausibility values). For the purposes of our estimation exercise, we rely on the probability of selecting any 3 statements $\{j, j', j''\}$ for all possible orderings of the plausibility values associated to the statements j, j', and j''. Given our logit specification (see (3)), the probability of selecting statements $\{j, j', j''\}$ in this exact order is given by: $\pi_{ij\in J} \cdot \pi_{ij'\in J\setminus\{j\}} \cdot \pi_{ij''\in J\setminus\{j,j'\}}$.

Our objective is to estimate, for each respondent i, a posterior distribution of knowledge precision $\theta_i \in \mathbb{R}$ and, for each statement j (whether true or false), posterior distributions of $\gamma_j \in \mathbb{R}$.[35] In addition, we estimate the posterior distributions of population parameters $\delta \in \mathbb{R}^+$ and $\alpha \in \mathbb{R}$.[36]

In what follows let $g \in G$ denote a socioeconomic group, where groups are defined as intersections of 4 demographic characteristics: Age (below/above median), Gender, Family Income (below/above median), and race (white and minority).

We estimate the model by Bayesian methods, specifically Hamiltonian Monte Carlo [Hoffman and Gelman, 2014] implemented in Stan [Carpenter et al., 2017]. To that end, we specify common prior distributions $\theta_i \sim N(\mu_g, \sigma^2)$ and $\gamma_j \sim N(0, 1)$, with hyperpriors $\mu_g \sim N(0, 10)$, $\forall g$, and $\sigma \sim \exp(\frac{1}{4})$.[37,38] The remaining prior distributions are specified as $\alpha \sim N(0, 10)$ and $\delta \sim N(1, 1)$.

The key identifying assumption is that the processes that generate the γ's and the θ's are stochastically independent.[39] While some months our panel of journalists selects real and false stories that are easier or harder and YouGov selects better or worse respondents (though that is less likely, given our sample size), what is required is that these two sources of variations are not systematically correlated.

We propose a three-step procedure to estimate the parameters of the model. In step 1, we arbitrarily fix $\theta_i = 1, \quad \forall i \in I$, and estimate the remaining parameters. In step 2, we fix γ_j to equal its posterior mean from Step 1 and estimate μ_g, $\forall g$, and σ. Finally, in Step 3, we fix μ_g, $\forall g$, and σ at their posterior means from Step 2 and

[35]Hereafter, the variable γ_j corresponds to the term $\left(2q_j - 1\right)\gamma_j$ in the economic model. As a result, we expect γ_j to be positive (resp. negative) when $q_j = 1$ (resp. $q_j = 0$).

[36]A common problem with this family of models [e.g., Bock, 1972, and see discussion of the literature below] is that θ and γ are identified through their product, so that there always exists one additional degree of freedom. This problem is solved by "anchoring" one of the two variables to some arbitrary scale. Consistent with our Bayesian approach, in our analysis the anchoring is achieved by assuming that γ is distributed according to a standard normal.

[37]Following Bock [1972] we impose the restriction that $\sum_{j=1}^{6} \gamma_j = 0$ by fixing $\gamma_6 = -\sum_{j=1}^{6} \gamma_j$. In the absence of this restriction, one could add any constant to all the γ's without affecting the probability of selecting a given statement j. Our empirical findings are very similar if we remove this constraint.

[38]An alternative approach consists of assuming a common mean for the prior distribution of θ across all individuals and groups. Such an approach would be rather conservative when quantifying knowledge differences across groups. Given the limited data available at the individual level, the posterior distributions of individual knowledge θ_i have a relatively large variance. As a direct consequence, the common prior assumption would pull individual estimates toward the mean. Nevertheless, results under this alternative approach are very similar to those with group-level means, with the exception of our results on inequalities (with smaller differences across groups).

[39]This type of mutual dependence between questions is obviously different from the mechanical dependence discussed above.

reestimate $(\theta_i)_{i\in I}$, γ_j, α, and δ.[40]

We conclude with some final remarks. In the last 7 surveys, we separately asked the respondents to report how favorable to the Republican Party they felt the true and false statements were. We can thus directly use this information to create the variable b_j (the average respondent's opinion about the favorability of statement j toward the Republican Party). Because we did not ask these questions in the first 4 surveys, we first estimate our model by relying on an alternative measure of news story partisanship. Specifically, for each statement we compute the difference between the share of Republicans and Democrats who selected that statement and normalize that variable to have a standard deviation equal to 1. Although this approach suffers from a possible reverse causality problem, we first use it in our main analysis. Later on, we will restrict our attention to the last 7 surveys and rely exclusively on the separately-observed measure of b_j.

3.4 Literature Discussion

The model we develop here is loosely 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]. However, we face two important differences with standard approaches in this literature.

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.

However, we cannot use any of the IRT models, including NRM, directly because of one important difference. The objective of all IRT tests is to measure the underlying skill of test takers. Instead, we are interested in measuring two factors: the underlying skill of our respondent (the precision of their signal) and the effect of partisanship.

[40] Alternatively, we could arbitrarily set one group's mean μ_g to be equal to one and estimate all remaining parameters simultaneously. Our findings are unaffected under this alternative method.

The latter effect is well-known to be important in political knowledge but it is not salient in educational testing. We therefore must augment Bock [1972] by developing a model where individuals have two traits, skill and ideology, and news stories have two characteristics, difficulty and partisanship. The combination of ideology and partisanship determines response rates in a non-monotonic way: it increases or decreases the probability that a person chooses a given statement depending on the congruence between the person's ideology and the statement's partisanship.

4 Analysis

4.1 Knowledge of the News

Within our framework, the probability that individual i with knowledge precision θ_i assigns a probability equal to or higher than h to statement j being true is equal to $\rho_{ij}(h)$ (see (7)). Our first results shed light on the average voter's knowledge of political news. For each statement j and individual i, and for any confidence level h, we can compute the posterior distribution of $\rho_{i,j}(h)$ as well as its average. Let $F(\theta)$ represent the posterior distribution of θ in the sample. One can then compute $\overline{\rho}(h) := \int_{\theta \in \mathbf{R}} \rho(h|\theta) dF(\theta)$, whose empirical analog is given by $\frac{1}{IN} \sum_i \sum_n \rho(h|\theta_{i,n})$ (where I is the number of individuals and N is the number of draws from the posterior distribution of θ_i).[41] Figure 1 plots $\overline{\rho}(h)$ for all values of $h \in [0,1]$, by distinguishing between the top 3 stories of the month. Recall that the ranking of news stories by importance is provided by our panel of journalists. Even within a given rank (say, first story of the month), however, the properties of the news stories –as captured by γ_j – may vary from one month to the next. To address this issue, within each rank, we take the median of the means of the posterior distributions of γ_j across stories. We also suppose this fictitious typical story to be neutral in its partisanship (i.e., we set $b = 0$).

Table 3 reports the average voter's knowledge $\overline{\rho}(h)$ of the typical first, second, and third news story of the month about the Federal Government, for various intervals of confidence.[42] To report our results in a way that is easier to comprehend, it is

[41] We refer to the average voter for simplicity. Formally speaking, though, we compute the average probability that a voter selected at random according to $F(\theta)$ assigns probability h or higher to a statement being true.

[42] As for Figure 1, to compute the typical story, we take the median of the means of the posterior distributions of γ_j across stories that belong to the same rank (1st, 2nd, or 3rd). We also suppose this fictitious typical story to be neutral in its partisanship ($b = 0$). Results are similar if we take

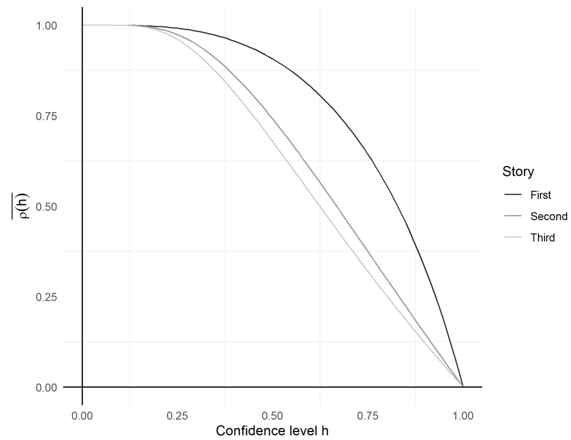


Figure 1: Average Voter's Knowledge of the News

Note: The figure reports the average voter's probability $\overline{\rho}(h)$ of assigning a probability equal to or higher than h to the typical first, second, and third story of the month being true.

useful to focus on a particular level of confidence h. In what follows, we say that an individual knows a (true) statement if he/she assigns a probability $h \geq 0.75$ to the statement being true. Similarly, we say that an individual is uncertain about the truth of a news story if she assigns a probability of truth between 0.25 and 0.75, and that she believes the story to be false if she assigns a probability of truth lower than 0.25. Accordingly, the top panel of Table 3 reports the corresponding figures. For the first news story of the month, the probability that the average voter knows the story is equal to 64%. Similarly, the probability that the average voter is uncertain (i.e., $h \in (0.25, 0.75)$) is equal to 35%, and the probability that the average voter believes the story to be false is 1%. These numbers change as we move from the first to the second and third stories of the month. For example, the probability that the average voter knows the second and third typical story falls to 37% and 32%, respectively. Reassuringly, the ranking of news stories by our panel of journalists is reflected in voters' knowledge of these stories.

Naturally, saying that a voter "knows" a news story if she assigns a probability at least as high as 0.75 to the story being true is arbitrary. The second and third panels of Table 3 report similar figures for alternative definitions of knowledge. For example, in the second panel, we report that the average voter is 77% likely to attribute 2 to 1 odds to the first story of the month being true. The corresponding figures for the

the average of the means of the posterior distributions of γ_j instead of the median.

second and third news stories of the month are 51% and 45%, respectively. Last, the third panel of Table 3 reports the likelihood that the average voter attributes a probability greater than or equal to $h = 0.5, 0.6, 0.7, 0.8, 0.9$ to the first, second, and third news stories of the month being true.

Confidence	First story	Second story	Third story
0 - 0.25	0.01	0.03	0.04
0.25 - 0.75	0.35	0.6	0.64
0.75 - 1	0.64	0.37	0.32
0 - 0.33	0.02	0.07	0.11
0.33 - 0.66	0.21	0.41	0.45
0.66 - 1	0.77	0.51	0.45
0.5 - 1	0.91	0.74	0.68
0.6 - 1	0.83	0.6	0.53
0.7 - 1	0.72	0.45	0.39
0.8 - 1	0.56	0.3	0.25
0.9 - 1	0.33	0.15	0.12

Table 3: Average Voter's Knowledge of the News $\overline{\rho}(h)$

Note: The table reports the average voter's probability $\overline{\rho}(h)$ of assigning a probability within a given interval of confidence to the typical first, second, and third story of the month being true.

An alternative approach to expressing voters' knowledge of political news consists of computing the expected number of news stories – among the top 3 stories of the month – known by voters. In addition to being directly interpretable, this way of measuring knowledge is also particularly amenable to quantifying differences across voters. In what follows, we rank individuals by the mean of their associated posterior distribution of knowledge precision θ_i and report results for the average individual belonging to the bottom-third, middle-third, and top-third of the knowledge distribution. In particular, Table 4 reports the probability that the average member of these three groups knows the average first, second, and third news story of the month.[43] Using these numbers, one computes that – of the top 3 news stories of the month – the average voter in the bottom-third of the distribution knows approximately 0.9 story, the average voter in the middle-third knows approximately 1.3 stories, and the average

[43]By average story we mean a story whose associated parameter γ corresponds to the average value of the mean of the posterior distributions of γ_j across relevant stories. We again suppose this typical story to be neutral in its partisanship ($b = 0$).

voter in the top-third knows close to 1.7 stories.

	Knowledge tier		
	Lower	Middle	Higher
First story	0.385	0.598	0.759
Second story	0.302	0.419	0.551
Third story	0.248	0.3	0.367

Table 4: Average Voter's Knowledge of the News, by Knowledge Tier

Note: The table ranks individuals by their knowledge precision and reports, for each knowledge tier, the average voter's probability $\overline{\rho}(h)$ of assigning a probability equal to or higher than 0.75 to the typical first, second, and third story of the month being true.

We conclude this subsection by reporting the posterior distribution of θ that we recover (see Figure 2). 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, fewer than 1% of respondents selected 0 true statements and only 14% 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.[44] The same individual has a probability equal to 0.05 to choose 0 true statements and a probability equal to 0.45 to choose one true statement. The theta distribution that fits the data cannot place a large weight on individuals that have little ability to discern the truth.

4.2 Heterogeneity across News Stories

Next we explore various dimensions of heterogeneity across news stories. Table 15 lists all *true* statements that were included in our quizzes. Similarly, Table 16 lists all *false* statements that were included in our quizzes. For each statement, the tables report the date, the share of survey respondents who selected the statement when completing the quiz ("raw mean"), the mean of the posterior distribution of γ_j, the predicted share of respondents who – according to our model's estimates – will select

[44]Moreover, because each individual completes only but a few 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.

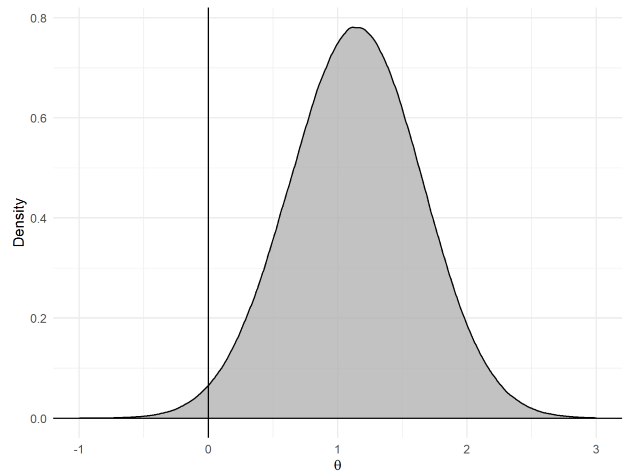


Figure 2: The Posterior Distribution of Knowledge Precision θ

the statement when completing the quiz, as well as the probability that the average voter assigns probability h to statement j being true.

As suggested by the tables, there exists significant heterogeneity across news stories (within both the true and false statements). Some statements were selected by virtually all our survey respondents and others were selected only by a tiny share of respondents. Recall that the parameter γ_j captures how responsive the likelihood of selecting statement j is to knowledge θ. What the tables suggest is that some true statements are much more easily detectable as true by knowledgeable respondents than others. Similarly, some false statements are much more easily detectable as false by knowledgeable respondents than others.[45]

Next, the tables report, for each statement, the predicted probability that the average voter selects 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, the tables suggest that there exists significant heterogeneity across news stories regarding respondents' knowledge. For example, the average voter has a 82% probability of knowing the (true) story "*The US Senate acquitted Trump of Impeachment Charges*." By contrast, it knows the (true) story "*Supreme Court granted a request by President Trump's administration to fully enforce a new rule that*

[45]For 7 statements out of 66, being more knowledgeable was seemingly a disadvantage.

would curtail asylum applications by immigrants at the U.S.-Mexico border*" only with probability 32%, despite 70% of our sample selecting the statement when completing the quiz. This last news story – with its high share of selections – illustrates how our structural approach takes into account the various properties of all the knowledge items included in the quiz to identify voters' actual knowledge of each single item.

Reassuringly, none of the false statements we included in our quizzes are widely believed to be true. In fact, the vast majority of our false statements are believed to be true by fewer than 20% of respondents and none are believed to be true by more than 33% of respondents.

4.3 Effect of Time

In our framework, the probability that a voter knows a story also depends on the number of months that have elapsed since the story was written. Specifically, time affects voters' beliefs through the population parameter δ (see (7)). Figure 3 plots its posterior distribution. It is tightly estimated away from 1, suggesting an effect of time passing on voters' knowledge of the news.

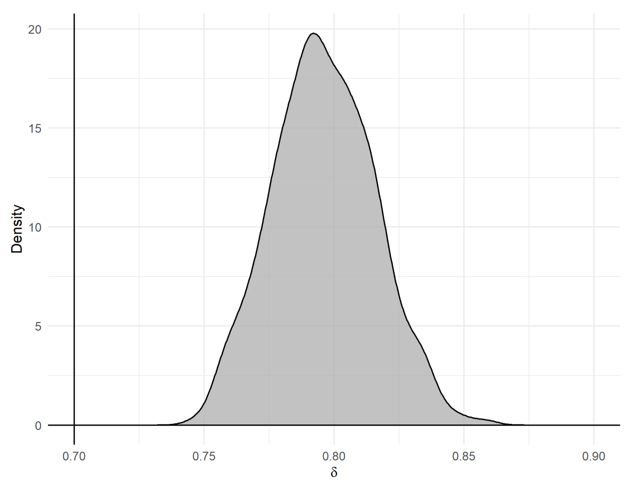


Figure 3: The Posterior Distribution of Time Decay δ

Table 5 reports the probability $\overline{\rho}(h)$ that the average voter attaches various levels of confidence h to the typical news story being true as a function of the number of months that have elapsed.[46] The probability that a voter chosen at random attributes

[46]By typical story we mean a story whose associated parameter γ corresponds to the median of the means of the posterior distributions of γ_j across all the true news stories. To isolate the effect

a probability equal to or higher than 75% that a typical story is true is 38% when the story is less than 4 weeks old, but the corresponding figure falls to 34% when the story is between 5 and 8 weeks old, and to 31% when the story is between 9 and 12 weeks old. In other words, time has a rather sizable effect on the probability of knowing a story. Although determining the exact underlying mechanism is beyond the purview of this paper, the effect of limited memory and motivated beliefs in combination with decreasing coverage are likely drivers of our findings [e.g., Zimmermann, 2020].

	Time Passed (Months)		
Confidence	0	1	2
0 - 0.5	0.02	0.03	0.04
0.25 - 0.75	0.6	0.63	0.65
0.75 - 1	0.38	0.34	0.31

Table 5: Effect of Time Passing on $\overline{\rho}\,(h)$

Note: The table reports the average voter's probability $\overline{\rho}\,(h)$ of assigning a probability within a given confidence interval to the typical news story being true, when the story is 1-4-week-old, 5-8-week-old, and 9-12-week-old.

4.4 Effect of Partisanship

The model allows for multiple dimensions of heterogeneity across news stories. One dimension of particular interest is the extent to which a story reflects favorably on the Republican Party: Is voters' knowledge of political news skewed towards those stories that reflect most favorably on their preferred political party [e.g., Bénabou and Tirole, 2002, 2006]?[47,48] If so, to what extent? The model we estimate assumes that all voters are possibly biased along partisan lines in their baseline knowledge of the news, and that the extent of the bias is identical across voters.

We elicited respondents' feelings towards the news only from the 5th survey onward (see Section 2). To use all 11 surveys, we must thus proxy stories' partisanship

of δ, we suppose this typical story to be neutral in its partisan content ($b = 0$).

[47]Throughout, we rely on the bipartisan nature of American politics to assume that a story that reflects favorably on the Republican party must reflect 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.

[48]See [Le Yaouanq, 2020] on the relationship between political preferences and beliefs about scientific facts.

differently. We proxy the extent to which a news story reflects favorably on the Republican Party by using the difference between the share of Republican respondents and the share of Democratic respondents who selected the story when completing the quiz. Moreover, we normalize this measure to have a variance equal to 1. We then rank the statements according to their partisanship measure b_j, and select statements within given percentile ranks: the 10th, 25th, 50th, 75th, and 90th percentile. Statements with low (high) values of b_j are likely favorable to the Democratic (Republican) party.

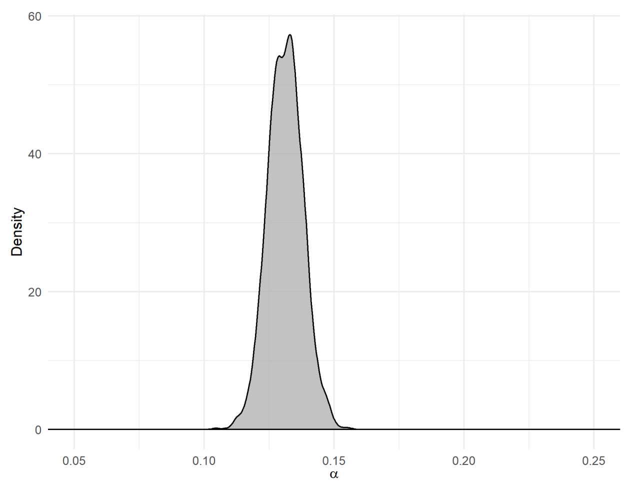


Figure 4: The Posterior Distribution of the Partisan Parameter α

Figure 4 plots the posterior distribution of the population parameter α. The partisan parameter is rather tightly estimated away from zero, suggesting the presence of a partisanship effect. Table 6 reports, for various percentiles in the distribution of b_j, the probability that the average supporter of given party attributes a given probability to a statement being true.

As news stories reflect less favorably on the Republican Party, the share of Republican respondents who attribute a probability of truth greater than or equal to 75% falls. Not surprisingly given that we assumed α to be a population parameter, the effect is symmetric for Democratic respondents. To quantify the magnitude of this effect, we define Partisan Gap as the difference in the average $\rho(h)$ across supporters of a given party, between Republican and Democratic party, normalized by the corresponding value for the Independent respondents. By this metric, for example, supporters of the Republican Party are 30.54% more likely than supporters of the Democratic Party to know a story located on the 90th percentile of the distribution (i.e., a statement that reflects rather positively on the Republican Party). Similarly, Republican respondents are 23.46% less likely to know stories that reflect poorly on the Republican Party (i.e., stories located on the 10th percentile).

Congruence		Confidence		
		$0 - 0.25$	$0.25 - 0.75$	$0.75 - 1$
Strongly Pro-Republican (90th pct)	Republican	0.01	0.55	0.44
	Democrat	0.04	0.63	0.33
	Partisan Gap	-98.9	-13.48	30.54
Moderatly Pro-Republican (75th pct)	Republican	0.02	0.57	0.41
	Democrat	0.03	0.61	0.36
	Partisan Gap	-44.12	-6.23	13.98
Neutral (50th pct)	Republican	0.02	0.59	0.38
	Democrat	0.02	0.59	0.38
	Partisan Gap	2.55	0.24	-0.6
Moderatly Pro-Democrat (25th pct)	Republican	0.03	0.61	0.36
	Democrat	0.02	0.57	0.41
	Partisan Gap	49.44	6.72	-15.24
Strongly Pro-Democrat (10th pct)	Republican	0.04	0.62	0.34
	Democrat	0.01	0.56	0.43
	Partisan Gap	76.36	10.34	-23.46

Table 6: Partisan Knowledge of the News

Note: The table reports the average supporter of a given political party's probability $\overline{\rho}(h)$ of assigning a probability of truth within a given confidence interval to news stories with varying favorability toward the Republican Party. Stories are ranked according to b_j. The table also reports the measure Partisan Gap, defined as the difference in the average $\rho(h)$ across supporters of a given party, between Republican and Democratic party, normalized by the corresponding value for the Independent respondents. We proxy a news story's favorability toward the Republican Party by using the difference between the share of Republican respondents and the share of Democratic respondents who selected the story when completing the quiz.

Table 7 reports the probability, as times passes, that a supporter of a given party attributes a probability of truth equal to or greater than 0.75 to news stories in various percentiles in the distribution of b_j. Time has a large and almost uniform effect on the odds of knowing a story, independently from the partisanship of a news story. Time also has a slight exacerbation effect on the partisanship bias. For news stories that reflect very favorably on the Republican Party, the variable Partisan Gap is equal to 30.54 for 1-4-week-old stories, to 30.97 for 5-8-week-old stories, and to 31.06 for

9-12-week-old stories. Magnitudes are similar for news stories that reflect favorably on the Democratic Party. Alternatively, if we restrict our attention to supporters of the Republican and Democratic Parties, roughly 57% of voters who know a strongly pro-Republican story are Republican when the story is less than a month old. The comparable figure rises to 58% when the story is 2-month old.

Congruence		Delay		
		$t = 0$	$t = 1$	$t = 2$
Strongly Pro-Republican (90th pct)	Republican	0.44	0.39	0.36
	Democrat	0.33	0.29	0.26
	Partisan Gap	30.54	30.97	31.06
Moderatly Pro-Republican (75th pct)	Republican	0.41	0.36	0.33
	Democrat	0.36	0.32	0.29
	Partisan Gap	13.98	13.73	13.4
Neutral (50th pct)	Republican	0.38	0.34	0.3
	Democrat	0.38	0.34	0.31
	Partisan Gap	-0.6	-1.42	-2.1
Moderatly Pro-Democrat (25th pct)	Republican	0.36	0.31	0.28
	Democrat	0.41	0.37	0.34
	Partisan Gap	-15.24	-16.63	-17.67
Strongly Pro-Democrat (10th pct)	Republican	0.34	0.3	0.27
	Democrat	0.43	0.38	0.35
	Partisan Gap	-23.46	-25.19	-26.43

Table 7: Partisan Knowledge of the News as Time Passes

Note: The table reports the average supporter of a given political party's probability $\overline{\rho}(h)$ of assigning a probability of truth equal to or greater than 0.75 to news stories with varying favorability toward the Republican Party as time passes (less than 1 month, between 1 and 2 months, and between 2 and 3 months). Stories are ranked according to b_j. It also reports the measure Partisan Gap as a function of time.

A limitation of the approach highlighted above is that, for each story, we rely on the share of Republicans versus Democrats who selected it to construct its partisanship score b_j. This may mechanically lead the model to find evidence of partisanship in voters' knowledge of the news. To address this problem, we replicate our analysis on the last 7 surveys, using the measure of b_j separately elicited from our survey respondents (see Section 2). Table 8 reports the corresponding results. The magnitude of the congruence effects are smaller but economically significant. For example, the effect of partisanship (being Republican versus Democrat) on the probability of knowing a story that reflects very favorably on a given party is about the same in size as the effect of one month of time on the probability that the average individual knows a typical story (i.e., 4-5 percentage points). For completeness, Table 17 in the Appendix reports our main results regarding the average voter's knowledge of the news when restricting our attention to the last 7 surveys and using the direct measure of b_j. Our main findings appear unaffected.

Congruence		Confidence		
		$0-0.25$	$0.25-0.75$	$0.75-1$
Strongly Pro-Republican (90th pct)	Republican	0.02	0.57	0.41
	Democrat	0.03	0.61	0.36
	Partisan Gap	-44.61	-6.26	14.36
Moderatly Pro-Republican (75th pct)	Republican	0.02	0.58	0.4
	Democrat	0.03	0.61	0.36
	Partisan Gap	-38.18	-5.38	12.33
Neutral (50th pct)	Republican	0.02	0.59	0.39
	Democrat	0.03	0.6	0.37
	Partisan Gap	-12.42	-1.81	4.12
Moderatly Pro-Democrat (25th pct)	Republican	0.03	0.6	0.38
	Democrat	0.02	0.59	0.38
	Partisan Gap	6.35	0.8	-1.9
Strongly Pro-Democrat (10th pct)	Republican	0.03	0.61	0.36
	Democrat	0.02	0.58	0.4
	Partisan Gap	32.78	4.47	-10.35

Table 8: Partisan Knowledge of the News (Second Method)

Note: The table reports the average supporter of a given political party's probability $\overline{\rho}(h)$ of assigning a probability of truth within a given confidence interval to news stories with varying favorability toward the Republican Party. Stories are ranked according to b_j. It also reports the measure Partisan Gap, defined as the difference in the average $\rho(h)$ across supporters of a given party, between Republican and Democratic party, normalized by the corresponding value for the Independent respondents. We restrict our attention to surveys 5-11 for which we are able measure news stories' partisanship directly.

4.5 Inequalities

There exists an important literature documenting the relationship between media coverage and voters' information and, in turn, the relationship between voters' information and the attention received from politicians. One important channel through which this accountability channel operates is voting. If voters are aware of the policies and actions implemented by politicians, the latter have greater incentives to cater to voters' preferences to increase their odds of reelection. Our analysis so far has mostly documented the level of knowledge about political news exhibited by the average voter. Investigating the distribution of knowledge across socioeconomic groups is also 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 knowledge inequalities across socioeconomic groups. Table 9 reports for the 16 socioeconomic groups our model explicitly identifies – the intersections of Age, Gender, Race, and Income (see Section 3), the probability that an average member of a particular group assigns a probability equal to or greater than 0.75 to the typical news story of the month being true.[49] Our results suggest significant differences across groups of voters. To take an extreme example, the average minority, female voter age 47 or less with a below-median income has a 30% probability of knowing the typical news story about the Federal Government. By contrast, the average white, male voter age 48 or more with an above-median income has a 44% probability of knowing the same story.

Next, we explore the explanatory role played by socioeconomic factors in a regression

[49] By typical news story we mean a news story whose associated parameter γ is the median of the means of the posterior distributions of γ_j across all true news stories. We also suppose this typical news story to be neutral (i.e., we set $b = 0$).

	Age > 47	Female	White	Income 60k+	$\rho < 0.25$	$\rho \in (0.25, 0.75)$	$\rho > 0.75$
1					0.05	0.64	0.31
2				x	0.03	0.62	0.34
3			x		0.03	0.61	0.37
4			x	x	0.02	0.59	0.39
5		x			0.05	0.65	0.30
6		x		x	0.05	0.64	0.32
7		x	x		0.04	0.63	0.34
8		x	x	x	0.03	0.62	0.35
9	x				0.03	0.62	0.36
10	x			x	0.02	0.58	0.40
11	x		x		0.02	0.58	0.40
12	x		x	x	0.01	0.55	0.44
13	x	x			0.03	0.62	0.35
14	x	x		x	0.02	0.58	0.40
15	x	x	x		0.02	0.60	0.38
16	x	x	x	x	0.02	0.58	0.40

Table 9: Knowledge of Political News across Socioeconomic Groups

Note: The table reports, for 16 socioeconomic groups, the average member's probability $\overline{\rho}(h)$ of assigning a probability within a given interval of confidence to the typical news story being true.

format.[50,51] Column (1) in Table 18 (see Appendix A) looks at the effects of various socioeconomic factors on the probability $\int \rho_i (0.75)\, dF_i(\theta)$ that voter i knows the second story of the month about the Federal Government.[52] Age is the most important characteristic, with voters aged 47 or more being 5 percentage points more likely to know the typical story. Intuitively, college education and income also positively predict knowledge, by 0.7 and 2.4 percentage points respectively. By contrast, women

[50]Recall that our model allows for different group-level means μ_g across 16 groups defined by Age, Gender, Race, and Income. This approach tends to give these four socioeconomic characteristics greater weight in explaining θ_i (and thus $\rho_i(h)$) compared to other characteristics (e.g., interest in politics, or media usage). For this reason, the coefficients associated to Age, Gender, Race, and Income are relatively large in our regression analysis. By and large, not prioritizing these four variables would still lead to larger coefficients on socioeconomic variables, but the differences would diminish noticeably.

[51]In Online Appendix C.3, we employ standard lasso regression methods to shed light on the most important determinants of news knowledge.

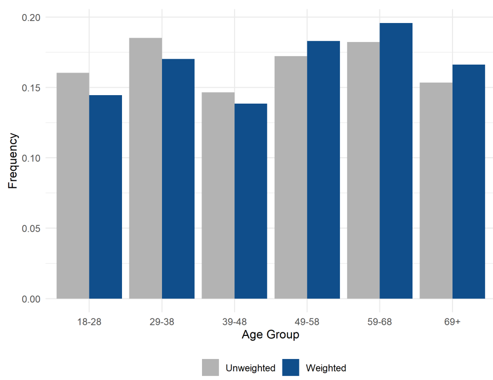
[52]In this exercise, we thus take the second story of the month to be the typical story of the month. Results would be similar if we considered other definitions of a typical story.

and racial minorities are associated with lower knowledge of the news. Women are 2.9 percentage points less likely to know the typical story about the Federal Government. Hispanics and African-Americans are 3.4 and 3.6 percentage points less likely to know the typical news story, respectively.

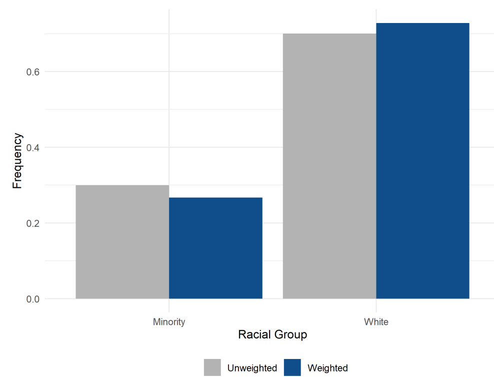
Column (2) adds political affiliations and Column 3 adds general engagement with party politics (partisanship). The coeffcients on political parties are small and they switch sign depending on whether partisanship is included. Partisanship increases the probability of knowing the typical news story by 0.6 percentage points. Table 19 (where Column (1) reproduces Column (3) in Table 18) includes media consumption habits. In both Columns (2) and (3) the number of news outlets and time usage (in minutes) are significantly positively associated with knowledge of the news, and the coefficients on the socioeconomic factors are largely unchanged by the inclusion of these news consumption habits (as well as extra media controls in Column (3)). Finally, Table 20 (which reproduces Table 18's Column (3) and Table 19's Column (3)) adds Political Interest as a control variable. Political Interest has been highlighted by previous work as an important factor in determining knowledge. Our results are consistent: we find that general interest in politics increases the probability of knowing the typical story about the Federal Government by 1.5 percentage points.

Because we measure knowledge of the most important news about the Federal Government from the perspective of mainstream journalists, we cannot rule out the possibility that various socioeconomic groups pay attention to different types of news (e.g., perhaps minorities pay attention to news that affect them more directly and that receive less attention from mainstream media outlets). As mentioned in Footnote 26, in some surveys we asked respondents to report how important they felt the news stories included in our quizzes were. The survey text we used is reported in Table E.16 in Online Appendix E.3.3. With possibly the exception of age, we found no consistent evidence of varying perceived importance of the news according to socioeconomic factors. In addition, in the last survey we administered, we included a knowledge quiz devoted exclusively to political news related to George Floyd's death (see Table E.13 in Online Appendix E.3.2). We estimated a simplified version of our model using this quiz only and found little evidence of lower inequalities in knowledge. These findings (available upon request) are only suggestive: one should not place great confidence on estimates of knowledge differences that use a single knowledge quiz.

We return to our simple theoretical framework to illustrate the relevance of our findings from a political economy angle. In Figure 5, 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 age 49 or more represent close to 54% of voters, even though they represent less than 50% of voters. Similarly, the incumbent will behave as if whites represent close to 73% of the population (in contrast to their actual share of 70%). Comparing these numbers with current US demographic trends helps to assess magnitudes.[53] For instance, the incumbent behaving as if the population of whites is 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 actual demographic composition of the US population.



(a) By Age Groups



(b) By Race

Figure 5: Inequalities in Knowledge of 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.

5 Extensions and Robustness Checks

5.1 Democratic Party Presidential Primaries

We apply our news generating process to select knowledge items pertaining to the Democratic Party presidential primaries. Our objective is twofold. We illustrate the

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

robustness of our method, which can be used to measure voters' knowledge of distinct types of topics. Also, we shed light on voters' knowledge of a key US electoral institution. Exactly as before, we estimate the 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, which were included in surveys 6-10.[54] Even though an individual completed quizzes about both the primaries and the Federal Government, we rely only on his/her performance when completing the quizzes about the primaries to estimate individual parameters.[55,56] Tables (10) and (11) replicate Tables (15) and (16) for our measurement of voters' knowledge of the Democratic primaries. Again, there exists significant heterogeneity across stories. Whereas only 19% of voters knew the story: *Democrats in Presidential debate hint at no swift end to China tariffs*, 82% of them knew the story: *Joe Biden denied alleged sexual assault.*

					$\overline{\rho}$		
Statement	Month	Raw Mean	γ	Prob of selecting	$h < 0.25$	$h \in (0.25, 0.75)$	$h > 0.75$
Democrats in Presidential debate hint at no swift end to China tariffs.	Oct 19	0.45	-0.09	0.44	0.15	0.66	0.19
Joe Biden raised more money than Donald Trump in March	May 20	0.48	-0.18	0.48	0.18	0.64	0.17
Senate Republicans blocked bill condemning Trump over protests	June 20	0.53	0.2	0.54	0.07	0.67	0.25
Democratic groups launched a multi-million digital ad effort to fight President Trump.	Oct/Nov 19	0.6	0.32	0.59	0.05	0.66	0.28
Pentagon ordered remaining active-duty troops to leave the Washington, D.C. area	June 20	0.68	0.56	0.7	0.03	0.61	0.36
Joe Biden announced he will pick a woman to be his vice presidential running mate	Apr 20	0.77	0.84	0.8	0.01	0.52	0.46
In a recent debate, all of the Democratic presidential candidates agreed universal healthcare is a top priority.	Oct 19	0.79	0.76	0.79	0.02	0.55	0.43
Former New York Mayor Michael Bloomberg has been considering whether to run for president.	Oct/Nov 19	0.83	0.98	0.84	0.01	0.47	0.52
Elizabeth Warren catches up with Joe Biden in a national opinion poll.	Oct 19	0.84	0.99	0.86	0.01	0.47	0.52
Two billionaire Democratic presidential hopefuls, Michael Bloomberg and Tom Steyer, collectively spent more in 2019 than the rest of the Democratic candidates combined	Feb 20	0.84	1.02	0.85	0.01	0.46	0.53
Bernie Sanders won New Hampshire's Democratic presidential primary	Feb 20	0.84	1.1	0.87	0.01	0.43	0.56
Elizabeth Warren ended White House bid	Apr 20	0.85	1.22	0.89	0.01	0.38	0.61
President Trump faced condemnation from former leaders of America's armed forces over his approach to civil unrest	June 20	0.88	1.28	0.92	0.01	0.36	0.63
Presidential candidate Elizabeth Warren proposed a Medicare for All plan that she said would not require raising middle-class taxes.	Oct/Nov 19	0.89	1.34	0.92	0.01	0.34	0.65
The Democratic presidential nominating race got off to a chaotic start in Iowa, as the results of the state's caucuses were delayed for hours	Feb 20	0.89	1.37	0.92	0.01	0.33	0.66
Several states postponed Democratic Party primaries amid coronavirus outbreak	Apr 20	0.9	1.43	0.93	0.01	0.31	0.68
Bernie Sanders dropped out of U.S. presidential race	May 20	0.94	1.76	0.98	0.01	0.23	0.76
Joe Biden denied alleged sexual assault	May 20	0.95	2.02	0.99	0.01	0.18	0.82

Table 10: True Statements

Note: The table reports all the true statements included in our quizzes about the Democratic Primaries. For each statement, it also reports the month of the associated survey, the share of respondents who selected it ("raw mean"), the parameter γ (i.e., the statement's straightforwardness), the model's predicted share of respondents who select the statement, as well as the average voter's probability $\overline{\rho}(h)$ of assigning a probability within a given confidence interval to the story being true.

[54] As for the news stories about the Federal Government, we included a knowledge quiz about the Democratic Party primaries in several surveys to produce a representative picture of voter knowledge.

[55] This was true also in our analysis of voters' knowledge of political news covering the Federal Government: we relied exclusively on individuals' performance when completing the quizzes about the Federal Government.

[56] Because we included the quizzes about the primaries in surveys 5 to 9, we are able to separately measure b_j. We therefore present results that rely on our direct measure of b_j only.

					$\overline{\rho}$		
Statement	Month	Raw Mean	γ	Prob of selecting	$h < 0.25$	$h \in (0.25, 0.75)$	$h > 0.75$
Hilary Clinton withdrew endorsement for Joe Biden	May 20	0.09	-2.03	0.06	0.76	0.21	0.04
Joe Biden announced he would not release tax returns	Apr 20	0.09	-1.97	0.04	0.75	0.22	0.04
Pete Buttigieg chose Kamala Harris as his Vice-Presidential pick	Feb 20	0.1	-1.61	0.06	0.68	0.28	0.05
Bernie Sanders admitted to taking Wall Street campaign contributions	Feb 20	0.11	-1.32	0.09	0.61	0.33	0.06
Hillary Clinton endorsed presidential candidate Tulsi Gabbard despite previous spat.	Oct/Nov 19	0.13	-1.77	0.06	0.71	0.25	0.04
President Trump attended George Floyd memorial, despite criticism of protests	June 20	0.15	-1.29	0.11	0.6	0.34	0.06
Black face photo shows up in Joe Biden's past.	Oct 19	0.15	-1.24	0.13	0.58	0.35	0.06
Voting Intentions Poll showed Bloomberg above Biden with white, working class voters.	Oct/Nov 19	0.18	-0.72	0.2	0.4	0.5	0.1
Bernie Sanders ended White House bid	Apr 20	0.18	-0.92	0.14	0.48	0.44	0.08
Joe Biden blamed George Floyd protests for increasing number of coronavirus cases	June 20	0.19	-1	0.16	0.51	0.42	0.08
George Soros refused to donate money to Biden campaign	May 20	0.2	-1.03	0.18	0.52	0.41	0.08
Andrew Yang Endorsed Amy Klobuchar, saying she is 'Most Honest in the Race'	Feb 20	0.21	-0.55	0.2	0.33	0.55	0.12
Kamala Harris ruled out possible role as vice presidential candidate	Apr 20	0.22	-0.6	0.2	0.35	0.53	0.11
Joe Biden announced he would consider Anthony Fauci for Surgeon General	May 20	0.34	-0.54	0.32	0.33	0.55	0.12
Elizabeth Warren plan would slash 70% of mining jobs.	Oct 19	0.37	-0.26	0.37	0.22	0.63	0.16
Pete Buttigieg received a significant donation, pushing him to the front of the fundraising race among all Democratic candidates as of early November.	Oct/Nov 19	0.37	-0.15	0.37	0.18	0.65	0.18
Kamala Harris attacks Cory Booker over Newark's water problem.	Oct 19	0.41	-0.16	0.41	0.18	0.65	0.17
Anthony Fauci warned police against using tear gas, could spread virus particles	June 20	0.57	0.26	0.57	0.06	0.67	0.27

Table 11: False Statements

Note: The table reports all the false statements included in our quizzes about the Democratic Primaries. For each statement, it also reports the month of the associated survey, the share of respondents who selected it ("raw mean"), the parameter γ (i.e., the statement's straightforwardness), the model's predicted share of respondents who select the statement, as well as the average voter's probability $\overline{\rho}(h)$ of assigning a probability within a given confidence interval to the story being true.

Table 12 reports the probability that the average voter knows (for various intervals of confidence h) the typical first, second, and third story of the month about the Democratic Party presidential primaries.[57] As before, the ranking is provided by our panel of journalists. For example, the average voter is 54% likely to assign a probability to the first story of the month being true equal to or greater than 0.75. The corresponding figures for the second and third most important stories of the month are 64% and 37%, respectively.

Confidence	First story	Second story	Third story
0 - 0.25	0.01	0.01	0.03
0.25 - 0.75	0.45	0.35	0.61
0.75 - 1	0.54	0.64	0.37

Table 12: Average Voter's Knowledge of the News about the Democratic Primaries

Note: The table reports the average voter's probability $\overline{\rho}(h)$ of assigning a probability within a given interval of confidence to the typical first, second, and third story of the month about the Democratic Primaries being true.

Next, Table 13 documents the effect of partisanship on the probability of knowing

[57]By typical we mean a story whose associated γ corresponds to the median of the means of the posterior distributions of γ_j across all true stories within a given rank (1st, 2nd, or 3rd).

stories about the Democratic presidential primaries (in various percentiles of the distribution of b_j). Again, we find evidence of partisanship on voters' knowledge of the news, with voters being more likely to know stories that reflect favorably on their preferred party. Last, Table 14 reports the effect of time passing on the probability that voters know the typical news story about the Democratic primaries.[58] As for news on the Federal Government, we find a sizable effect of time, with each month reducing the likelihood that the average voter knows the typical story by about 6-8 percentage points.

Congruence		Confidence		
		$0-0.25$	$0.25-0.75$	$0.75-1$
Strongly Pro-Republican (90th pct)	Republican	0.01	0.44	0.55
	Democrat	0.01	0.47	0.52
	Partisan Gap	-12.51	-4.88	5.08
Moderatly Pro-Republican (75th pct)	Republican	0.01	0.45	0.54
	Democrat	0.01	0.46	0.53
	Partisan Gap	-6.66	-3.2	3.28
Neutral (50th pct)	Republican	0.01	0.45	0.54
	Democrat	0.01	0.46	0.53
	Partisan Gap	-0.26	-1.35	1.31
Moderatly Pro-Democrat (25th pct)	Republican	0.01	0.46	0.53
	Democrat	0.01	0.46	0.53
	Partisan Gap	3.33	-0.31	0.2
Strongly Pro-Democrat (10th pct)	Republican	0.01	0.46	0.53
	Democrat	0.01	0.45	0.54
	Partisan Gap	8.95	1.31	-1.53

Table 13: Partisan Knowledge of the News – Democratic Primaries

Note: The table reports the average supporter of a given political party's probability $\overline{\rho}(h)$ of assigning a probability of truth within a given confidence interval to news stories with varying favorability toward the Republican Party. Stories are ranked according to b_j. It also reports the measure Partisan Gap, defined as the difference in $\overline{\rho}(h)$ across supporters of a given party, between Republican and Democratic party, normalized by the corresponding value for the Independent respondents.

[58] By typical story we mean a story whose associated parameter γ corresponds to the median of the means of the posterior distributions of γ_j across all the true news stories. We also suppose this typical story to be neutral ($b = 0$).

	Time Passed (Months)		
Confidence	0	1	2
0 - 0.5	0.01	0.02	0.02
0.25 - 0.75	0.46	0.53	0.59
0.75 - 1	0.53	0.45	0.39

Table 14: Effect of Time Passing – Democratic Primaries

Note: The table reports the average voter's probability $\overline{\rho}(h)$ of assigning a probability within given a confidence interval to the typical news story being true, when the story is 1-4-week-old, 5-8-week-old, and 9-12-week-old.

5.2 Robustness Checks

We perform a host of further analyses and robustness checks in the Online Appendix. In Online Appendix A, we look into the relationship between various news consumption diets and voters' knowledge of the news. As a robustness check, in Online Appendix C.1 we modify the model to let voters' partisan prior beliefs about news stories depend on the number of months that elapsed since the stories came out. Overall, we find only modest evidence of time playing an important role through this channel.

Further, recall that our main analysis excluded the 19% of respondents who selected fewer than 3 statements when completing the knowledge quizzes. If the tendency to select fewer than 3 statements is correlated with knowledge, one may worry that excluding these respondents may bias our results. In Online Appendix C.2, we replicate our main analysis by imputing respondents' "missing answers." Specifically, for all the respondents who selected fewer than 3 statements, we choose uniformly at random the missing statements from the remaining unselected items. Our main results appear unaffected (e.g., inequalities, effect of time passing, etc.), with the exception of knowledge which decreases across the board.

Last, in Online Appendix C.4 we replicate our 8th survey on a sample of 800 respondents recruited through M-Turk. Our main results line up with the analysis performed using the YouGov sample.

6 Concluding Remarks

This paper develops a new methodology to measure voters' knowledge of political news that combines a protocol for identifying stories, an incentivized quiz to elicit news knowledge, and the estimation of a model of individual knowledge that includes story difficulty, partisanship, and time passing. We apply this method – and repeat it 11 times – to the 3 most important news of the month about the US Federal Government according to mainstream media. We find significant heterogeneity across news stories: some stories are known by more than 80% of voters and others by fewer than 30%. We also find significant heterogeneity across voters in their knowledge of the news. For example, the average individual in the upper tier of the knowledge distribution knows nearly twice as many news stories as the average individual in the bottom tier of the distribution. We also document a sizable effect of time, with each month passing lowering the probability that the average voter knows the typical news story by 4-5 percentage points. Further, we find that voters are significantly more likely to know stories that agree with their partisan preferences. Lastly, we document large inequalities in knowledge of the news across socioeconomic groups. We argue these inequalities likely play a significant role in shaping the policies politicians implement.

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. *Working Paper*, 2019.

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.

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.

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. Media competition, information provision and political participation: Evidence from french local newspapers and elections, 1944–2014. *Journal of Public Economics*, 185, 2020.

Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of statistical software*, 76(1), 2017.

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

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.

Alessandro Gavazza, Mattia Nardotto, and Tommaso Valletti. Internet and Politics: Evidence from U.K. Local Elections and Local Government Policies. *The Review of Economic Studies*, 86(5):2092–2135, 2018.

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.

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

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.

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.

Caroline Le Pennec and Vincent Pons. Vote Choice Formation and Minimal Effects of TV Debates: Evidence from 61 Elections in 9 OECD Countries. *NBER Working Paper 26572*, 2019.

Yves Le Yaouanq. A Model of Ideological Thinking. *Working Paper*, 2020.

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

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.

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.

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. 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.

Ritva Reinikka and Jakob Svensson. Fighting Corruption to Improve Schooling: Evidence from a Newspaper Campaign in Uganda. *Journal of the European Economic Association*, 3(2-3):259–267, 2005. doi: 10.1162/jeea.2005.3.2-3.259.

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. Radio's Impact on Public Spending. *The Quarterly Journal of Economics*, 119(1):189–221, 2004.

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

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

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

A Appendix Tables

Statement	Month	Raw Mean	γ	Prob of selecting	$\overline{\rho}$: $h < 0.25$	$\overline{\rho}$: $h \in (0.25, 0.75)$	$\overline{\rho}$: $h > 0.75$
At a closed-door meeting at the White House, top envoy to China delivered evidence of rising Farm Belt frustration over bio-fuel policy.	Oct 19	0.36	-0.42	0.35	0.28	0.59	0.14
As of May, President Trump's overall popularity had been about the same for more than a year	June 20	0.4	-0.14	0.41	0.17	0.65	0.18
The U.S. Supreme Court gave itself another chance to make a definitive ruling on electoral map disputes	Jan 19	0.41	-0.26	0.4	0.21	0.63	0.16
Vice President Mike Pence visited Nebraska to take stock of the devastation unleashed across the U.S. Midwest by floods.	Apr 19	0.52	0.31	0.59	0.06	0.66	0.28
George Floyd's death sparked Republican and Republican calls in U.S. Congress for action on policing	June 20	0.6	0.28	0.59	0.06	0.67	0.27
The Trump administration credited cooperation from Mexico and Central American countries in cracking down on migrants.	Oct/Nov 19	0.63	0.62	0.67	0.02	0.59	0.38
Senior U.S. House members vowed to pass major defense bill despite pandemic	May 20	0.64	0.41	0.63	0.04	0.65	0.31
U.S. Supreme Court allowed President Trump's 'Remain in Mexico' asylum policy	Apr 20	0.64	0.42	0.67	0.04	0.64	0.32
In win for President Trump, U.S. Supreme Court made deporting immigrants for crimes easier	May 20	0.66	0.52	0.68	0.03	0.62	0.35
President Trump proposed plan to make U.S. immigration more merit-based.	June 19	0.66	0.46	0.67	0.04	0.64	0.33
Trump and Democrats agree to pursue $2 trillion Infrastructure Plan	May 19	0.67	0.46	0.68	0.04	0.63	0.33
U.S. lawmakers to unveil revised criminal justice bill in push for final passage	Dec 18	0.68	0.44	0.67	0.04	0.64	0.32
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.	Oct 19	0.69	0.44	0.7	0.04	0.64	0.32
U.S. Senate hands Trump rebuke on Saudi Arabia	Dec 18	0.7	0.55	0.72	0.03	0.61	0.36
Mexico agreed to take more migrants seeking asylum in the United States while they await adjudication of their cases.	June 19	0.7	0.62	0.73	0.03	0.6	0.38
President Trump said he would address national debt if re-elected	May 20	0.71	0.57	0.7	0.03	0.61	0.36
U.S. Senate returned to Washington amid concerns about coronavirus risk	June 20	0.72	0.6	0.73	0.03	0.6	0.38
Republican lawmakers in the House of Representatives condemned President Trump's decision to withdraw troops from Syria.	Oct/Nov 19	0.75	0.75	0.73	0.02	0.55	0.43
President Trump notified Congress he is firing the inspector general of U.S. intelligence community	Apr 20	0.77	0.75	0.79	0.02	0.55	0.43
Homeland Security Secretary Nielsen resigns amid Trump anger over border	May 19	0.78	0.78	0.8	0.02	0.54	0.44
President Donald Trump vetoed the measure passed by Democrats and Republicans in Congress to end his emergency declaration on building a border wall with Mexico.	Apr 19	0.8	0.49	0.67	0.03	0.63	0.34
Special Counsel Robert Mueller did not find the Trump 2016 campaign knowingly conspired with Russia.	Apr 19	0.82	1.03	0.86	0.01	0.45	0.53
Democratic lawmakers called for further investigation into a revelation that in 2016 Paul Manafort gave polling data to a man linked to Russian intelligence	Jan 19	0.84	0.98	0.87	0.01	0.47	0.51
Rod Rosenstein, U.S. deputy attorney general who appointed Special Counsel Robert Mueller, submits resignation	May 19	0.84	1.1	0.89	0.01	0.43	0.56
The House of Representatives passed legislation seeking to rein in President Trump's ability to deploy U.S. forces to fight abroad	Feb 20	0.84	0.99	0.87	0.01	0.47	0.52
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	Feb 20	0.87	1.17	0.9	0.01	0.4	0.59
Former Trump lawyer Michael Cohen sentenced to three years prison	Dec 18	0.88	1.33	0.93	0.01	0.35	0.65
Alabama's governor signed a bill to ban nearly all abortions in the state.	June 19	0.9	1.47	0.94	0.01	0.3	0.69
Whistle-blower report complains of White House cover-up on Trump-Ukraine scandal.	Oct 19	0.9	1.42	0.95	0.01	0.32	0.67
President Trump declared coronavirus a national emergency	Apr 20	0.92	1.56	0.96	0.01	0.28	0.71
The U.S. Government was partially shut down in fight over Trump's border wall with Mexico	Jan 19	0.94	1.47	0.95	0.01	0.31	0.68
A whistleblower filed a complaint against President Trump, leading to an impeachment inquiry.	Oct/Nov 19	0.94	1.99	0.98	0.01	0.19	0.81
The U.S Senate acquitted Trump of impeachment charges	Feb 20	0.95	2.14	0.99	0.01	0.17	0.82

Table 15: True Statements

Note: The table reports all the true statements included in our quizzes about the Federal Government. For each statement, it also reports the month of the associated survey, the share of respondents who selected it ("raw mean"), the parameter γ (i.e., the statement's straightforwardness), the model's predicted share of respondents who select the statement, as well as the average voter's probability $\overline{\rho}(h)$ of assigning a probability within a given confidence interval to the story being true.

Statement	Month	Raw Mean	γ	Prob of selecting	$\bar{\rho}$ $h<0.25$	$\bar{\rho}$ $h \in (0.25, 0.75)$	$\bar{\rho}$ $h>0.75$
A Tape surfaced of President Trump supporting abortion	Feb 20	0.07	-1.87	0.04	0.72	0.24	0.04
President Trump's Tax Returns showed billions given to various charities.	Oct/Nov 19	0.09	-2.35	0.03	0.8	0.17	0.03
Mitt Romney decided to run for president against Trump in the 2020 race after breakout role in impeachment	Feb 20	0.11	-1.45	0.07	0.64	0.31	0.06
2020 Presidential Candidate Elizabeth Warren took millions in Wall Street campaign contributions.	Apr 19	0.13	-1.33	0.11	0.61	0.33	0.06
Trump administration to continue to allow U.S. research using fetal tissue from abortions.	June 19	0.13	-1.44	0.1	0.64	0.31	0.06
President Trump fired coronavirus advisor Dr. Anthony Fauci	Apr 20	0.14	-1.45	0.1	0.63	0.31	0.06
President Trump took a week-long break from Campaigning to Deal with Coronavirus Outbreak	Feb 20	0.15	-0.98	0.12	0.5	0.42	0.08
Wall Street Journal poll predicts landslide Trump victory in 2020 Presidential Election	June 20	0.16	-1.18	0.14	0.56	0.37	0.07
Trump secures funding for border wall in meeting with top Democrats	Dec 18	0.17	-1.25	0.13	0.58	0.36	0.07
President Trump announced his tax returns will be released by Mid-May	May 20	0.17	-1.04	0.15	0.52	0.41	0.08
ISIS beheaded three Americans in response to Al-Baghdadi's death.	Oct/Nov 19	0.17	-1.24	0.1	0.58	0.36	0.07
Around 20% of IRS stimulus checks bounced	May 20	0.19	-0.92	0.18	0.48	0.44	0.08
Attorney General Barr released text message from Special Counsel prosecutor Robert Mueller: 'We're taking down Trump.'	June 19	0.19	-1.01	0.16	0.51	0.42	0.08
Trump fired Federal Reserve Chairman Jerome Powell for raising interest rates	Jan 19	0.2	-1.18	0.14	0.56	0.37	0.07
Trump releases redacted version of his taxes to Congress	May 19	0.21	-1.01	0.16	0.51	0.42	0.08
Soybean farmers marched on Washington over Chinese tariffs' impacts	Jan 19	0.22	-0.82	0.21	0.44	0.47	0.09
Nancy Pelosi under investigation by Justice Department over alleged insider trading during coronavirus outbreak	Apr 20	0.22	-0.84	0.19	0.44	0.46	0.09
China blacklists Apple and Microsoft amid escalating trade war.	Oct 19	0.23	-0.83	0.23	0.44	0.47	0.09
Saudi Crown Prince To Address Senate In Effort To Clear His Name In Journalist?s Murder	Dec 18	0.25	-0.69	0.23	0.38	0.51	0.11
Clinton Foundation loses nonprofit status	May 19	0.25	-0.63	0.24	0.36	0.53	0.11
The Virginia Bar Association disbars Attorney General Barr for lying to Congress	May 19	0.25	-0.71	0.22	0.39	0.5	0.1
President Donald Trump diverted Puerto Rico aid to fund the border wall with Mexico.	Apr 19	0.29	-0.67	0.23	0.38	0.51	0.11
Agriculture trade group marched in Washington to draw attention to export problems	Apr 20	0.31	-0.44	0.29	0.28	0.58	0.13
Federal Judge rules public funding for Planned Parenthood unconstitutional	Dec 18	0.32	-0.39	0.32	0.26	0.6	0.14
President Trump announces he will resume peace talks with Iran at UN General Assembly.	Oct 19	0.37	-0.39	0.36	0.27	0.59	0.14
Trump Threatened To Raise Border Wall Cost To $7 Billion If Stall By Democrats Continues	Jan 19	0.39	-0.19	0.42	0.19	0.64	0.17
China and the United States agreed on a new comprehensive trade deal.	Oct/Nov 19	0.41	0.24	0.49	0.07	0.67	0.26
U.S. Border Patrol facility admitted to measles outbreak among migrant children in custody.	June 19	0.42	-0.09	0.41	0.15	0.66	0.19
House Republicans Unveil Legislation To Significantly Limit Funding To Planned Parenthood Centers Nationwide.	Apr 19	0.44	0.18	0.53	0.08	0.67	0.25
Vaping case to make its way to Supreme Court.	Oct 19	0.44	-0.22	0.42	0.2	0.63	0.16
Twitter blocked President Trump's account for 24 hours after tweet surfaced as false	June 20	0.46	-0.02	0.46	0.13	0.67	0.2
President Trump's campaign saw steep rise in donations after press conferences	May 20	0.63	0.46	0.65	0.04	0.64	0.33
Joe Biden announced he would publicly name running mate in June	June 20	0.66	0.46	0.67	0.04	0.64	0.33

Table 16: False Statements

Note: The table reports all the false statements included in our quizzes about the Federal Government. For each statement, it also reports the month of the associated survey, the share of respondents who selected it ("raw mean"), the parameter γ (i.e., the statement's straightforwardness), the model's predicted share of respondents who select the statement, as well as the average voter's probability $\bar{\rho}(h)$ of assigning a probability within a given confidence interval to the story being true.

Confidence	First story	Second story	Third story
0 - 0.25	0.01	0.03	0.04
0.25 - 0.75	0.32	0.60	0.64
0.75 - 1	0.67	0.38	0.33

Table 17: Knowledge of the News - Second Approach

Note: The table reports the average voter's probability $\overline{\rho}(h)$ of assigning a probability within a given interval of confidence to the typical first, second, and third story of the month about the Federal Government. Attention is restricted to surveys 5-11 for which we obtain a direct measure of b_j.

	Dependent variable:		
	$\rho_{ij}(0.75)$		
	(1)	(2)	(3)
Democrat		0.004^{***}	-0.001
		(0.001)	(0.001)
Republican		0.0003	-0.004^{***}
		(0.001)	(0.001)
Partisan			0.006^{***}
			(0.001)
Age > 47	0.050^{***}	0.050^{***}	0.049^{***}
	(0.001)	(0.001)	(0.001)
Income > 60k	0.024^{***}	0.024^{***}	0.024^{***}
	(0.001)	(0.001)	(0.001)
College +	0.007^{***}	0.007^{***}	0.007^{***}
	(0.001)	(0.001)	(0.001)
Female	-0.029^{***}	-0.030^{***}	-0.029^{***}
	(0.001)	(0.001)	(0.001)
Black	-0.036^{***}	-0.037^{***}	-0.036^{***}
	(0.001)	(0.001)	(0.001)
Hispanic	-0.034^{***}	-0.034^{***}	-0.034^{***}
	(0.001)	(0.001)	(0.001)
Constant	0.355^{***}	0.354^{***}	0.355^{***}
	(0.001)	(0.001)	(0.001)
Observations	9,785	9,785	9,491
R^2	0.480	0.481	0.482

Notes: The table shows the relationship between the probability of knowing the second story of the month about the Federal Government (across 11 surveys), $\int \rho_i\,(0.75)\,dF_i\,(\theta)$, and various socioeconomic factors. Significance: *p<0.1; **p<0.05; ***p<0.01.

Table 18: Socioeconomic Factors 1/3

	Dependent variable:		
	$\rho_{ij}(0.75)$		
	(1)	(2)	(3)
Democrat	-0.001	-0.001	0.0002
	(0.001)	(0.001)	(0.001)
Republican	-0.004^{***}	-0.003^{***}	-0.002^{*}
	(0.001)	(0.001)	(0.001)
Partisan	0.006^{***}	0.005^{***}	0.004^{***}
	(0.001)	(0.001)	(0.001)
Age > 47	0.049^{***}	0.047^{***}	0.048^{***}
	(0.001)	(0.001)	(0.001)
Income > 60k	0.024^{***}	0.023^{***}	0.023^{***}
	(0.001)	(0.001)	(0.001)
College +	0.007^{***}	0.006^{***}	0.005^{***}
	(0.001)	(0.001)	(0.001)
Female	-0.029^{***}	-0.028^{***}	-0.028^{***}
	(0.001)	(0.001)	(0.001)
Black	-0.036^{***}	-0.036^{***}	-0.034^{***}
	(0.001)	(0.001)	(0.001)
Hispanic	-0.034^{***}	-0.034^{***}	-0.033^{***}
	(0.001)	(0.001)	(0.001)
Sources 3+		0.005^{***}	0.009^{***}
		(0.001)	(0.001)
Total time		0.00001^{***}	0.00001^{***}
		(0.00000)	(0.00000)
Constant	0.355^{***}	0.352^{***}	0.352^{***}
	(0.001)	(0.001)	(0.001)
Extra media controls			X
Observations	9,491	9,491	9,491
R^2	0.482	0.489	0.497

Notes: The table shows the relationship between the probability of knowing the second story of the month about the Federal Government (across 11 surveys), $\int \rho_i (0.75)\, dF_i (\theta)$, and various socioeconomic factors. Significance: *p<0.1; **p<0.05; ***p<0.01. 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.

Table 19: Socioeconomic Factors 2/3

	Dependent variable:		
	$\rho_{ij}(0.75)$		
	(1)	(2)	(3)
Democrat	-0.001	0.0002	0.001
	(0.001)	(0.001)	(0.001)
Republican	-0.004^{***}	-0.002^{*}	-0.001
	(0.001)	(0.001)	(0.001)
Partisan	0.006^{***}	0.004^{***}	0.001
	(0.001)	(0.001)	(0.001)
Poli Interest			0.015^{***}
			(0.001)
Age > 47	0.049^{***}	0.048^{***}	0.045^{***}
	(0.001)	(0.001)	(0.001)
Income > 60k	0.024^{***}	0.023^{***}	0.022^{***}
	(0.001)	(0.001)	(0.001)
College +	0.007^{***}	0.005^{***}	0.004^{***}
	(0.001)	(0.001)	(0.001)
Female	-0.029^{***}	-0.028^{***}	-0.026^{***}
	(0.001)	(0.001)	(0.001)
Black	-0.036^{***}	-0.034^{***}	-0.032^{***}
	(0.001)	(0.001)	(0.001)
Hispanic	-0.034^{***}	-0.033^{***}	-0.032^{***}
	(0.001)	(0.001)	(0.001)
Sources 3+		0.009^{***}	0.007^{***}
		(0.001)	(0.001)
Total time		0.00001^{***}	0.00000^{***}
		(0.00000)	(0.00000)
Constant	0.355^{***}	0.352^{***}	0.347^{***}
	(0.001)	(0.001)	(0.001)
Extra media controls		X	X
Observations	9,491	9,491	9,491
R^2	0.482	0.497	0.512

Notes: The table shows the relationship between the probability of knowing the second story of the month about the Federal Government (across 11 surveys), $\int \rho_i (0.75)\, dF_i(\theta)$, and various socioeconomic factors. Significance: *p<0.1; **p<0.05; ***p<0.01. 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.

Table 20: Socioeconomic Factors 3/3