

Paper_social media

Dan Xu

3/26/2021

Abstract

“Fake news” has become a recurring theme in American politics since the 2016 U.S. election. Many have raised the question of whether social media plays a role in spreading political misinformation and disinformation. This paper is a replication study of two economists Allcott and Gentzkow on fake news propagation in the 2016 U.S. election. Both of the authors examined how deliberately fabricated, as opposed to erroneous or biased stories from well-known or little-known sources spread on social media platforms, such as Facebook. By conducting a survey with 1200 participants a few months after the 2016 U.S. election, the authors found that fake news did not swing the election as people barely remembered it. Therefore, people’s voting behaviors in the 2016 election had not been affected by fake news in social media.

Introduction

Written by Stanford University professor Matthew Gentzkow and NYU professor Hunt Allcott, the original paper aims to address the following questions: Has social media become a major platform for people to receive political information? How do American adults trust information they see on social media? How would such information affect their voting behaviors and their candidate preferences? Fake news is no news. In fact, it has been permeated in American politics for decades. The following graphs illustrate how the political conspiracy theories spread between 1963 and 2010. We can see that the mostly widely circulated one is that the assassination of Martin Luther King was the act of part of a large conspiracy in 1975. 60 percent of the American population believe that it is true, followed by the second mostly widely spread conspiracy of 2003, where the Bush administration was believed to purposely misled the public about evidence that Iraq banned weapons in order to build support for war. If we pay close attention to the conspiracies, it seems that there is no correlation between media and the spread of such conspiracy theories. The graphs above illustrate the political conspiracy theories spread between 1963 and 2010. We can see that the mostly widely circulated one is that the assassination of Martin Luther King was the act of part of a large conspiracy in 1975. 60 percent of the American population believe that it is true, followed by the second mostly widely spread conspiracy of 2003, where the Bush administration was believed to purposely misled the public about evidence that Iraq banned weapons in order to build support for war. If we pay close attention to the conspiracies, it seems that there is no correlation between media and the spread of such conspiracy theories.

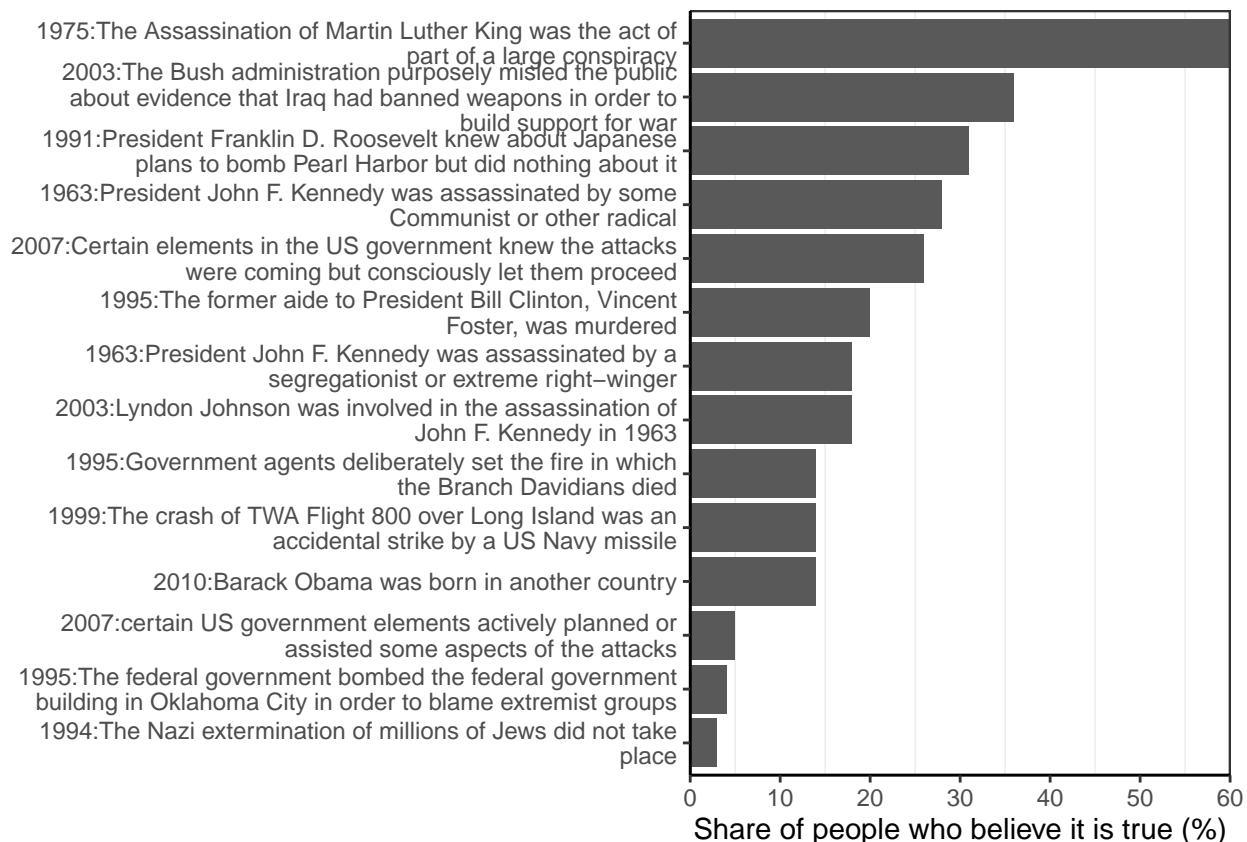
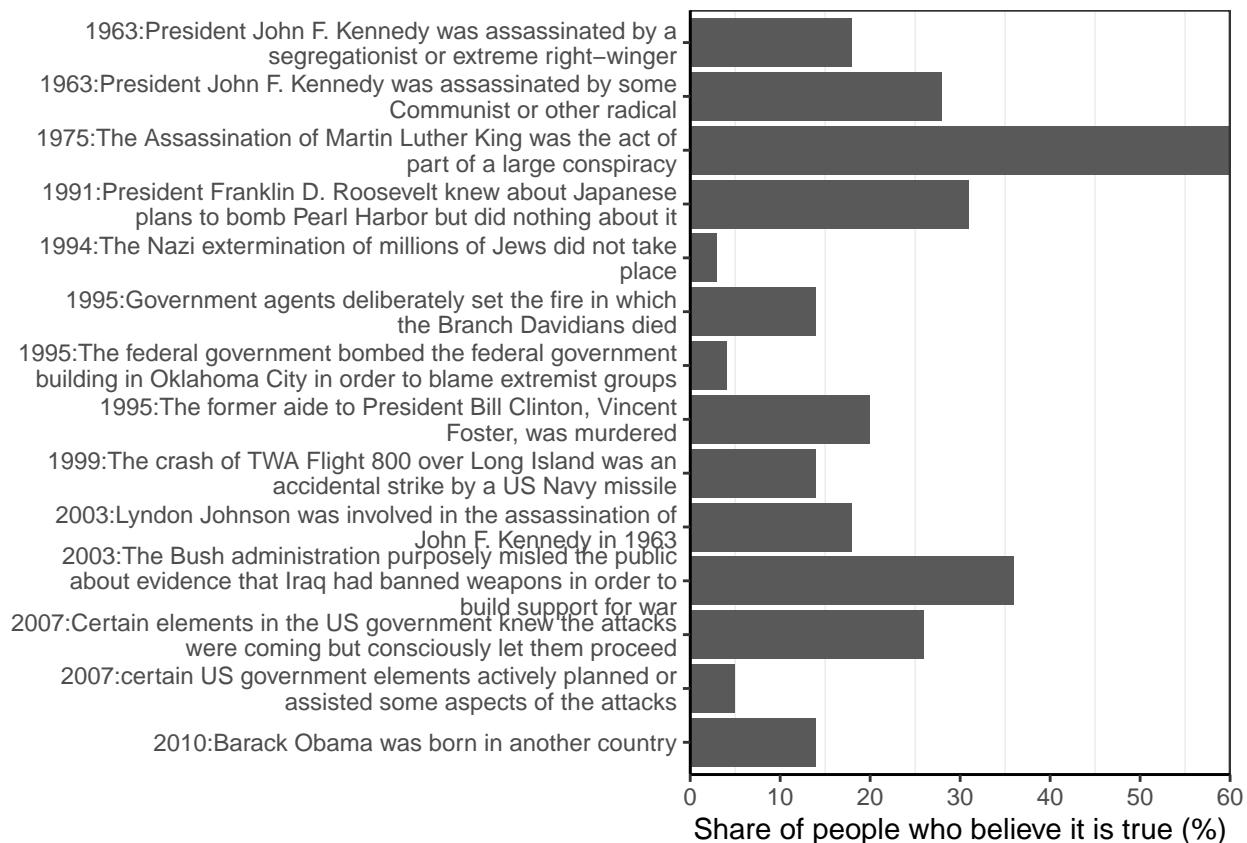


Table 1: Results

Pro	title
Clinton	41
Trump	115

Table 2: Results

Pro	total
Clinton	5518.6
Trump	26156.3

The authors conducted a survey after the presidential election of 2016 to gather data. Specifically, Gentzkow and Allcott ran an online survey of 1,208 US adults recruited from MonkeyMonday in late November, a few weeks after the election. Given that the nature of the recruitment method, the sample size for the survey was not randomly selected. It is indeed biased and the data may not be representative of the population. The main goal of the survey was designed to obtain information about how people rely on social media as an information source to get informed; how exposure to fake news affects people's political behaviors; who are more likely to believe fake news.

Data

Dataset

As the authors put it, there were far more pro-Trump fake news shares — 30.3 million shares for false stories that favored Trump, versus 7.6 million that favored Clinton. My replicated results confirm that there were 41 fake news stories that had been shared on Facebook regarding Clinton, and 115 regarding Trump. However, the total number of fake news stories shared on Facebook was 31674.9 combined; and it was 5518.6 and 26156.3 for Clinton and Trump respectively. The authors would need to specify how they calculated fake news shares on Facebook.

Post - election survey The study was conducted through a post - election survey and a database to understand how people produce, circulate, and receive political fake news. Alongside questions about the respondents' demographics and political affiliations, the survey asked participants for their "most important source of news and information" about the 2016 election. Then, the respondents were presented with 15 news headlines, equally split between pro-Clinton and pro-Trump news, including some deliberately fabricated false stories, fake stories that really existed, and some true ones as well, and placebo headlines — which were untrue and made-up claims that were neither real nor fake. The respondents then were asked to indicate if they "recall seeing this reported or discussed prior to the election?" Following this question, the respondents were asked to evaluate the authenticity of the news stories. The question was "At the time of the election, would your best guess have been that this statement was true?"

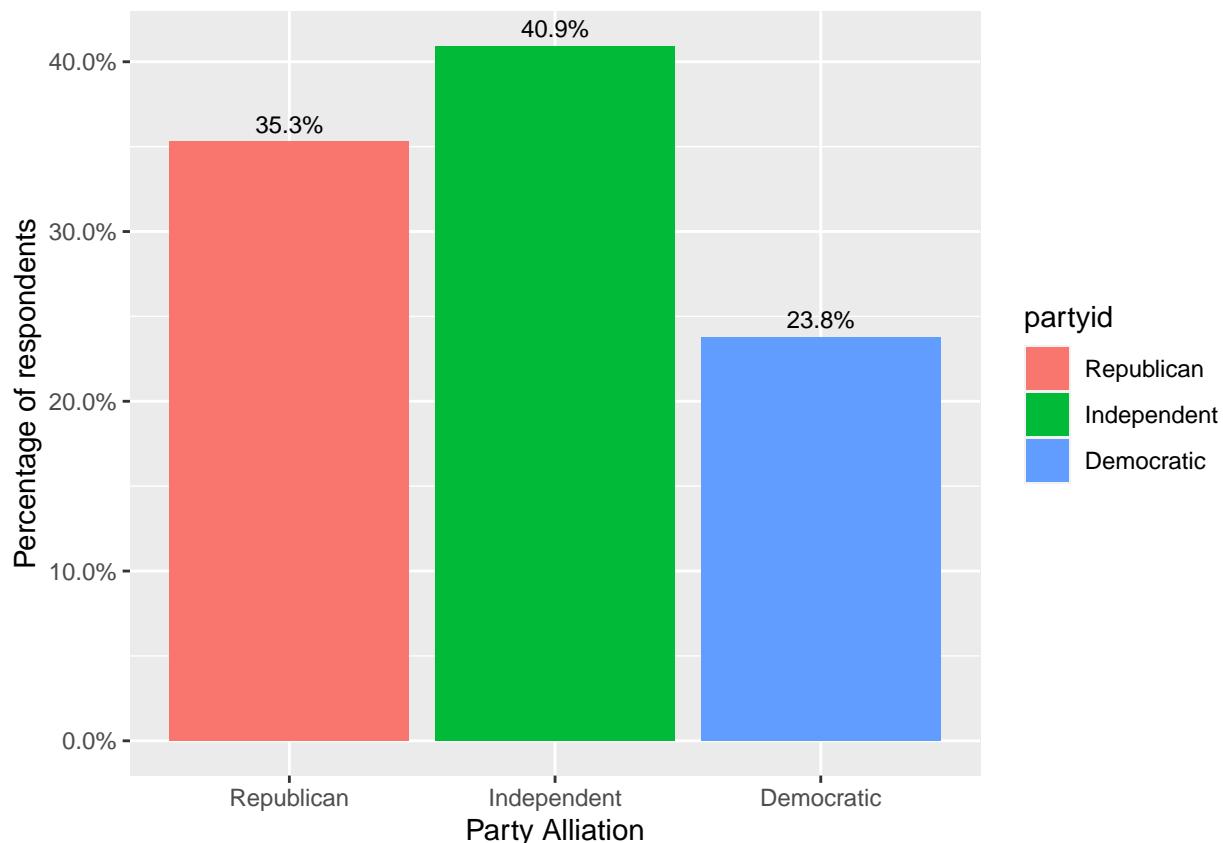
In addition, to understand how political fake news was produced, circulated and received amongst American adults, after acquiring consent from the survey participants to participate and a promise to provide their best answers, the survey asked the respondents to provide their demographic information by answering questions including their party identification/affiliation, who and when they voted for in the 2016 presidential election, and their gender, educational background, race/ethnicity, zipcode. The survey also asked about 2016 election news consumption, including how much time they spent on reading, watching, or listening to election news in general and on social media in particular, and asked them to identify the most important source of news and information about the 2016 election.

According to the authors, the survey was distributed to 1,208 participants. However, there are 1437 observations because some respondents have not completed the survey or have withdrawn from the survey.

Descriptions	Value
Sample size (nrow)	1437
No. of variables (ncol)	63
No. of numeric/integer variables	47
No. of factor variables	0
No. of text variables	16
No. of logical variables	0
No. of identifier variables	0
No. of date variables	0
No. of zero variance variables (uniform)	0
% of variables having complete cases	0% (0)
% of variables having >0% and <50% missing cases	100% (63)
% of variables having >=50% and <90% missing cases	0% (0)
% of variables having >=90% missing cases	0% (0)

Descriptive analysis of the survey sample

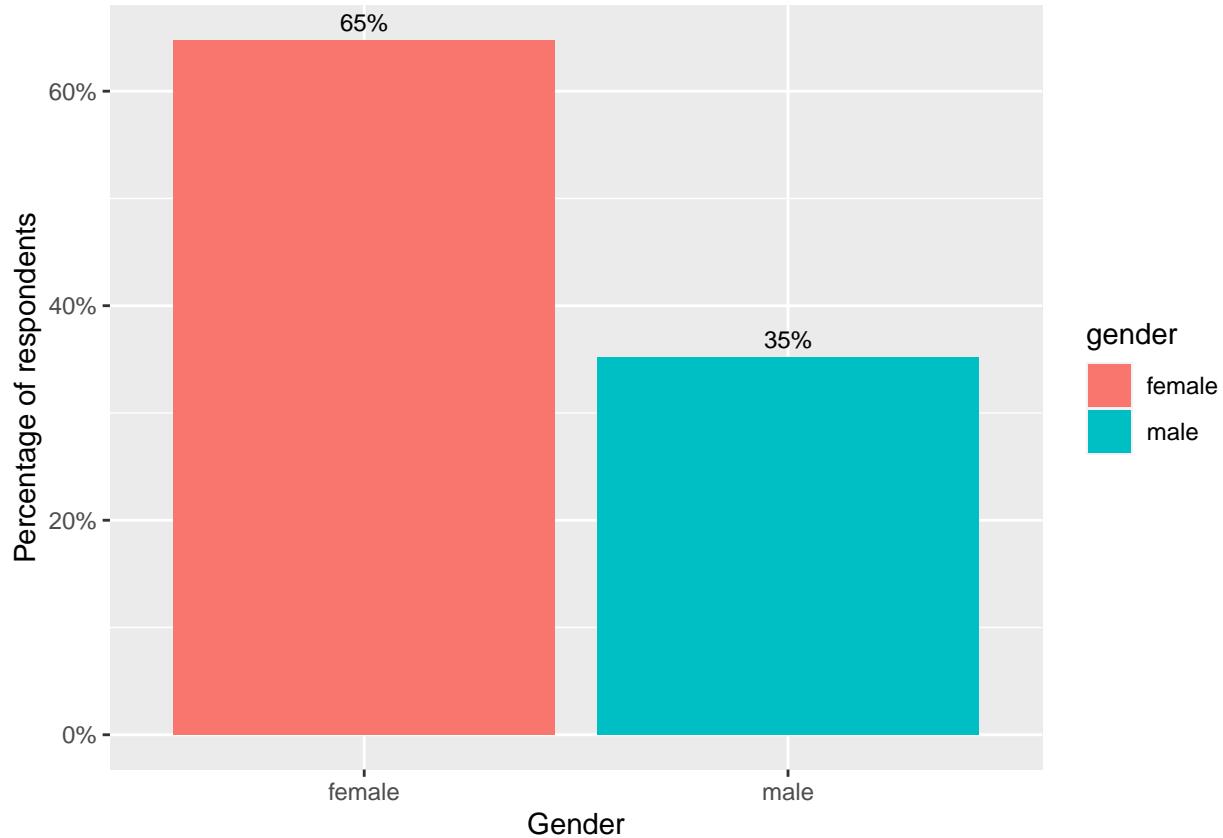
Party affiliation



In terms of party arlliation, the sample size contains 49 percent of respondents who identified as Republicans, and 34 percent of respondents who identified as Democrats. However, the proportion is seemingly inconsistent with the polling results from major U.S. polling agencies. According to Gallup, “37% of registered voters identified as independents, 33% as Democrats and 26% as Republicans” in the year of 2017. The inconsistency is potentially a result of different coding strategies. In the survey, respondents were asked to self identified

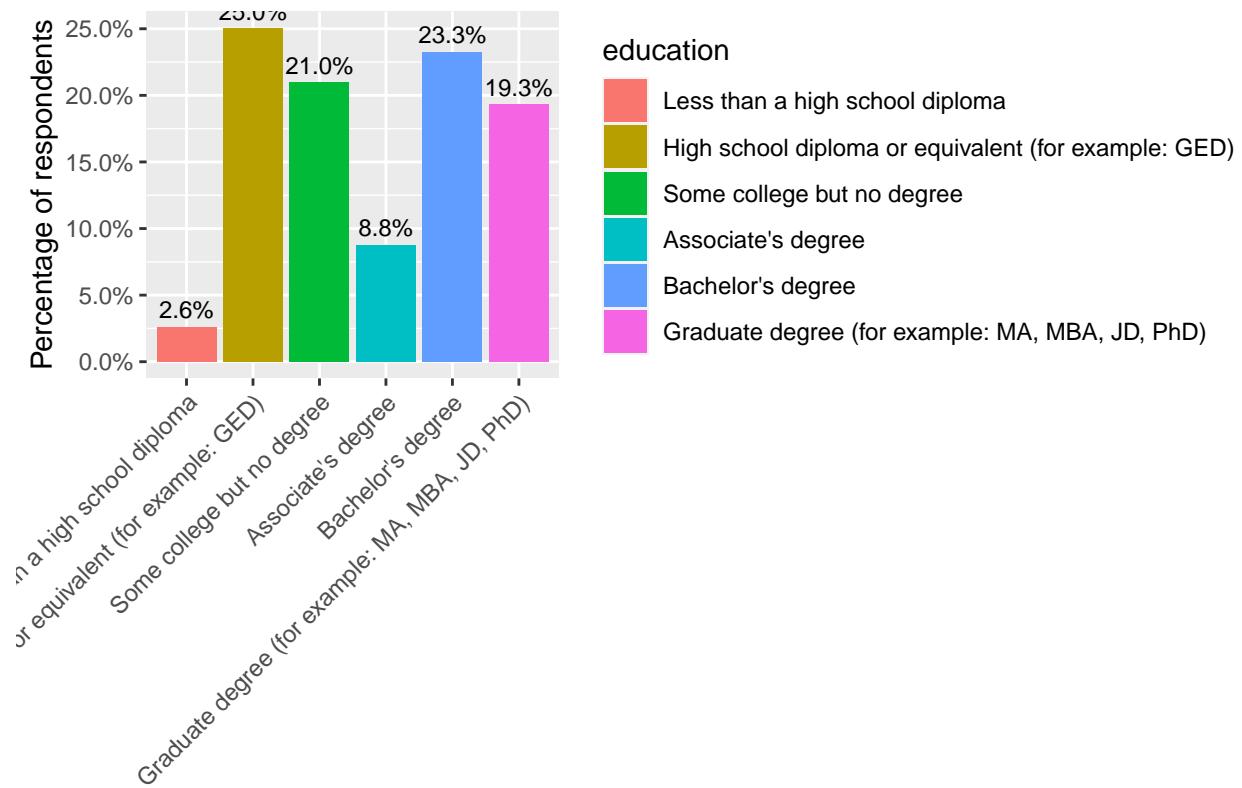
themselves with one of the following categories, namely, “Strongly Republican”, “Weakly Republican”, “Lean toward Republican party”, “Independent”, “Lean toward Democratic party”, “Weakly Democratic”, “Strongly Democratic”, where the “Lean toward Republican party” and “Lean toward Democratic party” categories can also be lumped into the “Independent” category. If we lumped the respondents who lean toward either party, then the results would be different as the independents would be nearly 41% as the Republicans and Democrats being 35.3 percent and 23.8 percent. In this case, we can conclude that the sample size is skewed toward Republicans.

Gender



As the graph above shows, the sample is disproportionately female with nearly 2/3 of the survey respondents were women, and only 1/3 of the respondents were men.

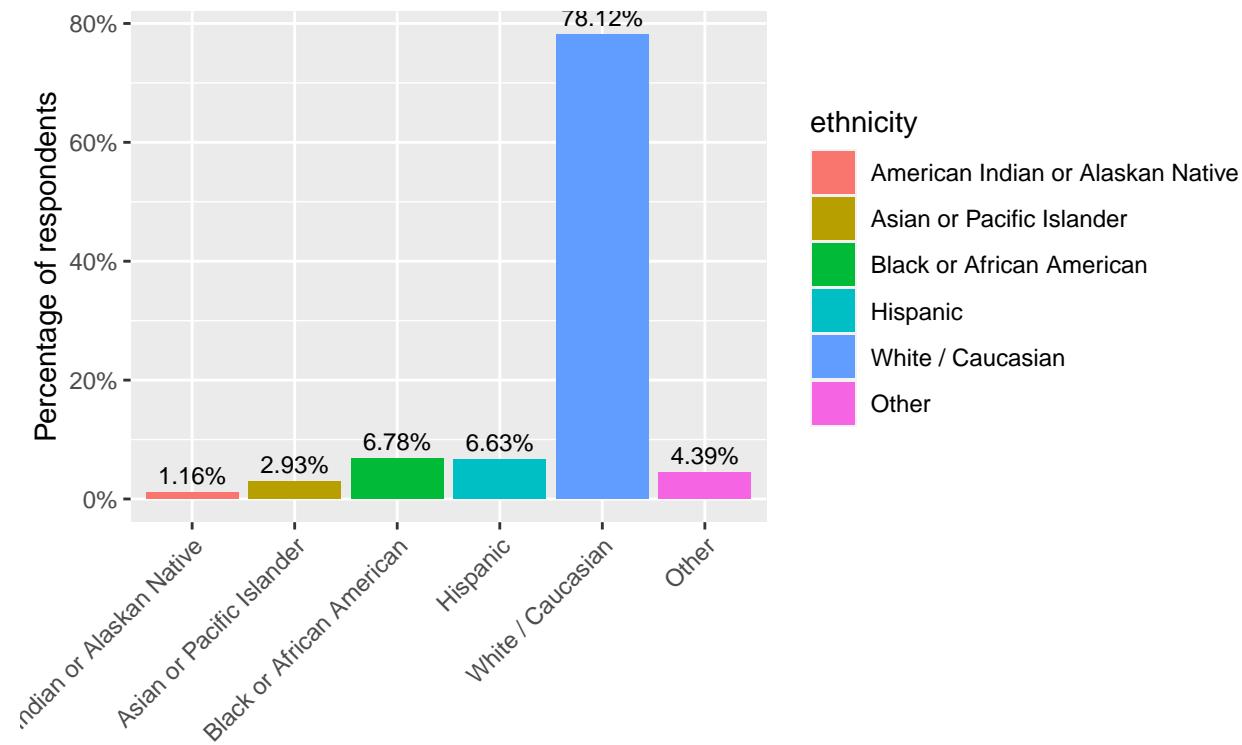
education



Level of education

The sample is also skewed toward highly educated people. As the plot shows, only 27.6 percent of the survey respondents received a high school diploma or less education; while more than 70 percent of the survey participants reported receiving at least some college degree level education, more than 40 percent of them reported holding a bachelor's or higher degree.

Ethnicity



Ethnicity/race

As the graph shows, more than 70% of the study participants were white or Caucasian, suggesting that the sample was not drawn using random sampling techniques.

In summary, the sample was disproportionately white, female, and highly educated, which lacks representativeness and needs to be adjusted with weights.

survey weights

The observations were weighted using weighted means abstracted from national samples for national representativeness. The average weight for education was obtained from the weighted average of American adults who aged 18 and above from 2014 ACS 5-year estimates from United States Census Bureau. Accordingly, respondents who received a bachelor's degree or higher were assigned the weight average of 0.2673650; while other respondents at other levels of education were assigned the average mean of 0.4197737 as the weight. Likewise, for the party identification variable, Respondents who identified with Democrats were assigned the weighted mean of 0.3684771; those who identified themselves as Independents, including Republicans and Democrats who indicated that they leaned toward Independents, were assigned the weighted mean of 0.3429180; For Republican survey participants, they were assigned the average weight of 0.2886049. The weighted means for the party identification were drawn from the survey results of American National Election Studies 2012. For the average of the sample size, 47.1549515 was assigned to the sample as the average age of American adults. In terms of race, the white/Caucasian survey respondents were assigned with a weighted average of 0.6160000, which represents their proportion of the American population. For sex ratio, Men were assigned with the weighted average of 0.4867939.

To sum up, in terms of the survey design, there are a few issues: the survey does not include attention checks. Attention checks are a useful tool to make sure survey respondents are really paying attention, and not

Table 3: Average weights table

Variable	Mean	Notes
Educ_BachelorsOrMore	0.2673650	Weighted average of 18-24 years and 25+ years from 2014 ACS 5-year estimates
Educ_HSOrLess	0.4197737	Weighted average of 18-24 years and 25+ years from 2014 ACS 5-year estimates
Democrat	0.3684771	Weighted results from 2012 ANES, only considering people who were D, R, or I
Independent	0.3429180	Weighted results from 2012 ANES, only considering people who were D, R, or I
Republican	0.2886049	Weighted results from 2012 ANES, only considering people who were D, R, or I
Age	47.1549515	Average age of people 18 or older in the US in 2015
Race_White	0.6160000	https://www.census.gov/quickfacts/table/PST045215/JS-a
Male	0.4867939	Share male of people 18 or older in the US in 2015
Income	76.1625160	Mean income (this is in 2014 inflation-adjusted dollars), times 1.021 for 2.1% inflation (from http://www.usinflationcalculator.com/)
NewsSource_Social	1.2400000	3 is Often, 0 is Never (numbering changed from original Pew coding)
NewsSource_Web	1.5800000	3 is Often, 0 is Never (numbering changed from original Pew coding)

Table 4: Sample weights summary

Variable	Survey sample	U.S. adult population
College graduate	0.44	0.27
High school or less	0.27	0.42
Male	0.35	0.49
Age	45.88	47.15
Caucasian	0.79	0.62
Democrat	0.24	0.37
Republican	0.35	0.29
Web news consumption frequency	2.34	1.58
Social media news consumption frequency	1.88	1.24

simply randomly answer the questions for the sake of completing the survey as quickly as possible. The authors should include such attention check questions to insure the reliability of the survey responses. The income variable is missing in the original survey.

News headlines

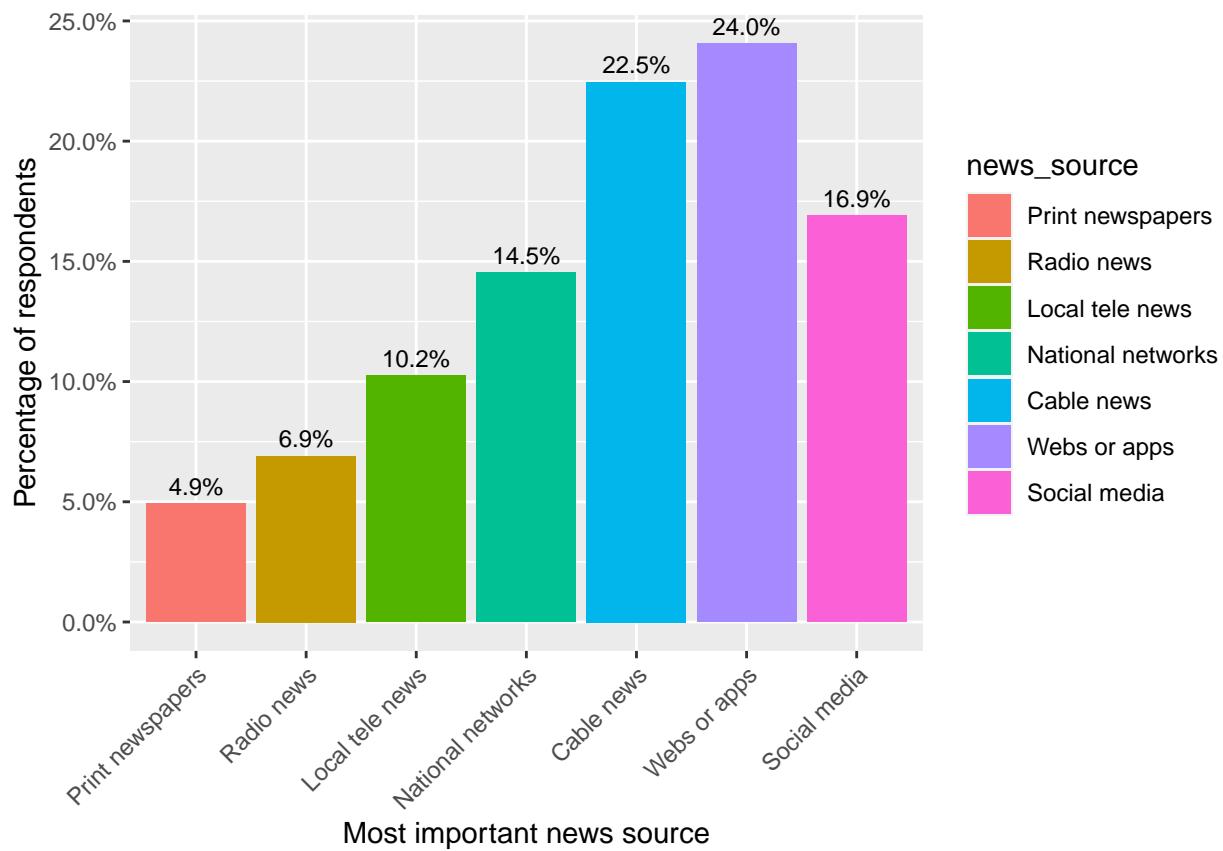
The survey participants were presented with 15 headlines, which were equally split between pro-Clinton and pro-Trump news stories. The authors categorized the stories into “BigFake”, “Small-Fake”, “FakeFake”, “BigTrue”, referring to ambiguously and unambiguously false stories, and unambiguously true ones, and placebo headlines — which were made-up claims that neither true nor fake news outlets had reported. The participants were then asked to report if they had seen it “reported or discussed” during the 2016 U.S. campaign, and how likely they would have been to believe it. False claims included a rumor that the Pope had endorsed Trump for presidency, and that the FBI found evidence of the Clinton Foundation running a pedophile ring. True ones included Trump refusing to confirm that he would concede the election if he lost, as well as Clinton’s comment about Trump supporters belonging in a “basket of deplorables.” Placebos included claims that either Trump or Clinton campaign staff had diverted funds to buy alcohol for expensive parties. Since the authors have not clearly explained the differences amongst these terms, especially between “BigFake” and “FakeFake”, it seems challenging to differentiate them. According to the authors, there were 30 questions ready to be presented to the respondents, but each respondent would only receive a randomly selected 15 of these stories.

Most important news source to get politically informed

As social media has permeated every aspect of our lives, people have started to heavily rely on social media to obtain information. This study tests whether social media has become an important information source from which American adults get news. In this survey, respondents were asked to report what the most important news sources are for them. Surprisingly, the results suggest that only 17 percent of the respondents clearly indicated that they used social media as their primary information source. This finding is consistent with the authors’ findings. In 2016, people still relied on other communication channels to get informed, especially news websites and cable television.

Table 5: Average weights table

Type	Article Text	Article True/False	Pro-Clinton or pro-Trump?
BigFake	Pope Francis endorsed Donald Trump.	FALSE	Trump
BigFake	An FBI agent connected to Hillary Clinton's email disclosures murdered his wife and shot himself.	FALSE	Trump
BigFake	The Clinton Foundation bought \$137 million in illegal arms.	FALSE	Trump
BigFake	Mike Pence said that "Michelle Obama is the most vulgar First Lady we've ever had."	FALSE	Clinton
BigFake	In May 2016, Ireland announced that it was officially accepting Americans requesting political asylum from a Donald Trump presidency.	FALSE	Clinton
BigFake	Celebrity RuPaul said that Donald Trump mistook him for a woman and groped him at a party in 1995.	FALSE	Clinton
SmallFake	At the beginning of November, the FBI uncovered evidence of a pedophile sex ring run under the guise of the Clinton Foundation.	FALSE	Trump
SmallFake	Under Donald Trump's tax plan, it is projected that 51% of single parents would see their taxes go up.	TRUE	Clinton
SmallFake	At a rally a few days before the election, President Obama screamed at a protester who supported Donald Trump.	FALSE	Trump
SmallFake	FBI Director James Comey's October 28th letter about new developments in the investigation of Hillary Clinton's emails went only to Republican members of Congress, and not to Democrats.	FALSE	Clinton
SmallFake	A Republican congressman helped broker a deal for Donald Trump to buy a taxpayer-owned building in New York City.	10	Clinton



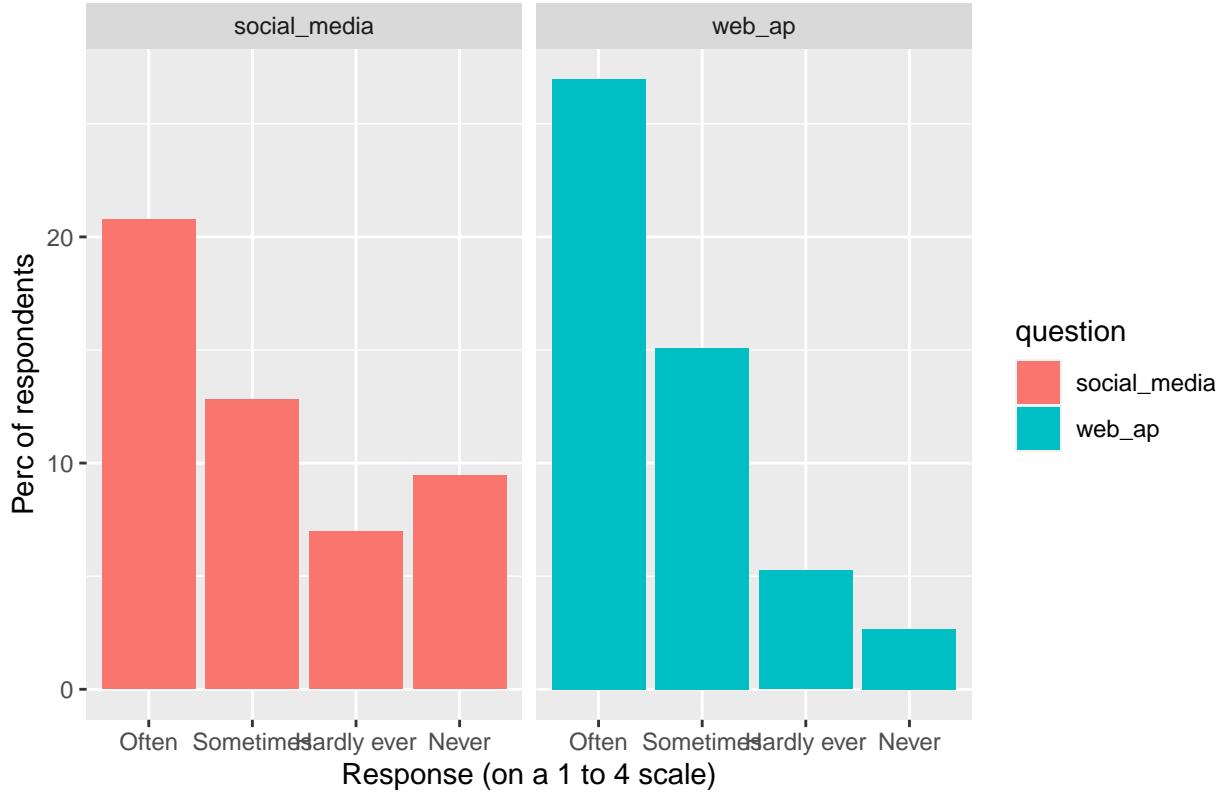
How often do people use social media to get news?

Therefore, people do not use social media as the primary source to get political information. However, it does not mean that social media does not play any role in information consumption. To the contrary, survey results suggest that respondents heavily rely on social media to follow political news. The results show that people heavily rely on social media to get informed. More than 20 percent of the respondents indicated that they often use social media to get news and more than 25 percent of the respondents often turned to websites and apps to get informed.

Table 6: Model results

Rates of seeing and believing fake news relative to placebo fake news	...2	...3	...4	...5
Believability	Recalled seeing	NA	Recalled seeing and believed	NA
News	Fake	Placebo	Fake	Placebo
Share of sample	0.15	0.14000000000000001	0.08	0.08
95% CI	0.17	0.16	0.09	0.1

How often do you get news from social media/web and applications?



Who believed in political fake news?

One of the goals of this study was to test if the share of respondents that was truly exposed and believed the average fake news article. To exam what contributed to the respondents' believability of such political fake news, the authors established a simple linear model as follows to characterize the potential relationship,

$$B_{ij} = aX_i + \alpha + \epsilon$$

In this model, B_{ij} takes 1 when people believe, takes 0 when people do not believe in the story; and takes 0.5 when people are not sure about the veracity of the story. For example, if headline a is true, then B takes value 1 if person i responded “Yes” to “would your best guess have been that this statement was true?”. B_{ij} is the dependent variable and a vector X_i in a linear regression.

This table presents the share people who recall seeing (answered “yes” to the question of “Do you recall seeing this reported or discussed prior to the election?”) but not believe, or recall seeing and believed news

headlines. Columns 1 and 4 include only Fake headlines, columns 2 and 5 include only Placebo headlines, and columns 3 and 6 present differences between the previous two columns. Observations are weighted for national representativeness. Specifically, for the share of people who say they were exposed to the article, we see that Fake and Placebo articles are approximately equally likely to be believed. This is approximately a test of the assumption since all people who recalled seeing Placebo headlines are misremembering, as are almost all people who recalled seeing Fake headlines (for small exposure rates).