

Online Appendix: Social Media and Fake News in the 2016 Election

Hunt Allcott, *New York University and NBER*
Matthew Gentzkow, *Stanford University and NBER*

March 2017

A Data Appendix

A.1 Fake News Database

From Snopes, we scraped all stories dated between August 1st and November 7th, 2016 from www.snopes.com/tag/donald-trump/ and www.snopes.com/tag/hillary-clinton/. From PolitiFact, we scraped all stories dated between August 1st and November 7th, 2016 from www.politifact.com/truth-o-meter/elections/2016/president-united-states/. Most of these stories are fact checks of statements made by presidential candidates, which we drop, but some are fake news headlines. We use fake news headlines that PolitiFact rated as "Pants on Fire" or "False."

We match these articles to data on Facebook shares from BuzzSumo (buzzsumo.com), an online content database that links to the Facebook API and records the number of shares for individual URLs. Individual fake news stories in our database typically occur on multiple URLs—for example, the false story that “the Pope endorsed Donald Trump” was reported independently by a number of different news websites, with different specific URLs for each story. For each story in our fake news database, we searched relevant key words on BuzzSumo, and recorded the number of Facebook shares for every URL that had been shared more than 1000 times. While BuzzSumo does have shares from other social media sites such as Twitter, we do not record shares on these other sites because the number of Facebook shares is orders of magnitude larger. As we carried out these searches in early December 2016, the number of shares includes several post-election weeks, and thus may overstate the number of pre-election shares. We also gather the number of Facebook shares of the fact-check articles from Snopes.¹

¹Some rumors from Snopes were images shared on social media with no specific origin URL, so we do not have Facebook shares of the false “article.” In these cases, we impute the Facebook shares of false articles from the

A.2 Post-Election Survey

Appendix Table 1 presents the news headlines used in the post-election survey, and Appendix Figures 1 and 2 present the share of U.S. adults who recall seeing and who believed each article.

Appendix Table 2 presents summary statistics for the survey sample. We re-weight the sample in column 1 to match population means on all ten variables in column 2, using the entropy weighting procedure of Hainmueller (2012). By construction, the mean weight is one. As diagnostics, the standard deviation of our sample weights is 1.4, the maximum weight is 20.4, 2.3 percent of weights are larger than 5, and 0.25 percent of weights (three observations) are larger than 10. In our unweighted data, Clinton received 15 percentage points more votes than Trump, while in our weighted data, she received 6 percentage points more. The latter margin is statistically indistinguishable from the predictions of most pre-election polls.

Facebook shares of the corresponding Snopes' fact-check articles using a log-log regression, based on the sample of stories for which we have both variables; the R^2 of this regression is 0.17.

Appendix Table 1: News Headlines Used in the Post-Election Survey

(1) Article text	(2) True/false	(3) Article favors
<u>“Big Fake” news headlines covered in New York Times, Wall Street Journal, and BuzzFeed after the election</u>		
Pope Francis endorsed Donald Trump.	FALSE	Trump
An FBI agent connected to Hillary Clinton’s email disclosures murdered his wife and shot himself.	FALSE	Trump
The Clinton Foundation bought \$137 million in illegal arms.	FALSE	Trump
Mike Pence said that “Michelle Obama is the most vulgar First Lady we’ve ever had.”	FALSE	Clinton
In May 2016, Ireland announced that it was officially accepting Americans requesting political asylum from a Donald Trump presidency.	FALSE	Clinton
Celebrity RuPaul said that Donald Trump mistook him for a woman and groped him at a party in 1995.	FALSE	Clinton
<u>“Small Fake” and “Small True” headlines from PolitiFact</u>		
At the beginning of November, the FBI uncovered evidence of a pedophile sex ring run under the guise of the Clinton Foundation.	FALSE	Trump
Under Donald Trump’s tax plan, it is projected that 51% of single parents would see their taxes go up.	TRUE	Clinton
At a rally a few days before the election, President Obama screamed at a protester who supported Donald Trump.	FALSE	Trump
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
A Republican congressman helped broker a deal for Donald Trump to buy a taxpayer-owned building in order to build the Trump International Hotel in Washington, D.C.	FALSE	Clinton
Repeated requests for additional security in Benghazi were routinely denied by Hillary Clinton’s State Department.	TRUE	Trump
<u>“Small Fake” and “Small True” headlines from Snopes, Hillary Clinton tag</u>		
The Clinton campaign secretly paid musicians Beyonce and Jay Z \$62 million to appear at a rally in support of Hillary Clinton.	FALSE	Trump
Hillary Clinton’s first name was spelled with an extra “i” (“Hilliary,” with the word “liar” in the middle) on election ballots printed for use in Lonoke County, Arkansas.	TRUE	Clinton
An email written by Hillary Clinton aide Huma Abedin to her brother revealed that she is a radical Muslim.	FALSE	Trump
<u>“Small Fake” and “Small True” headlines from Snopes, Donald Trump tag</u>		
Donald Trump threatened to deport Puerto Rican Broadway star Lin-Manuel Miranda, not realizing that Puerto Rico is a U.S. territory and Puerto Ricans are U.S. citizens.	FALSE	Clinton
Wikileaks was caught by Newsweek fabricating emails with the intent of damaging Hillary Clinton’s campaign.	FALSE	Clinton
Donald Trump and his campaign donated food and supplies to Hurricane Matthew victims in North Carolina.	TRUE	Trump
<u>“Placebo” headlines that we invented</u>		
Leaked documents reveal that the Clinton campaign planned a scheme to offer to drive Republican voters to the polls but then take them to the wrong place.	FALSE	Trump
Leaked documents reveal that the Trump campaign planned a scheme to offer to drive Democratic voters to the polls but then take them to the wrong place.	FALSE	Clinton
FBI Director James Comey was secretly communicating with Hillary Clinton about when to release results of the FBI investigation into Clinton’s private email server.	FALSE	Trump
FBI Director James Comey was secretly communicating with Donald Trump about when to release results of the FBI investigation into Clinton’s private email server.	FALSE	Clinton
Clinton Foundation staff were found guilty of diverting funds to buy alcohol for expensive parties in the Caribbean.	FALSE	Trump
Trump Foundation staff were found guilty of diverting funds to buy alcohol for expensive parties in the Caribbean.	FALSE	Clinton
<u>“Big True” headlines from the Guardian’s election timeline</u>		
Hillary Clinton said that “you could put half of Trump’s supporters into what I call the basket of deplorables.”	TRUE	Trump
At the 9/11 memorial ceremony, Hillary Clinton stumbled and had to be helped into a van.	TRUE	Trump
At the third presidential debate, Donald Trump refused to say whether he would concede the election if he lost.	TRUE	Clinton
On October 28th, the FBI director alerted members of Congress that it had discovered new emails relevant to its investigation of Hillary Clinton’s personal server.	TRUE	Trump
The musicians Beyonce and Jay Z appeared at a rally in support of Hillary Clinton.	TRUE	Clinton
Two days before the election, the FBI director told Congress that a newer batch of emails linked to Hillary Clinton’s private email server did not change his conclusion that Clinton should face no charges over her handling of classified information.	TRUE	Clinton

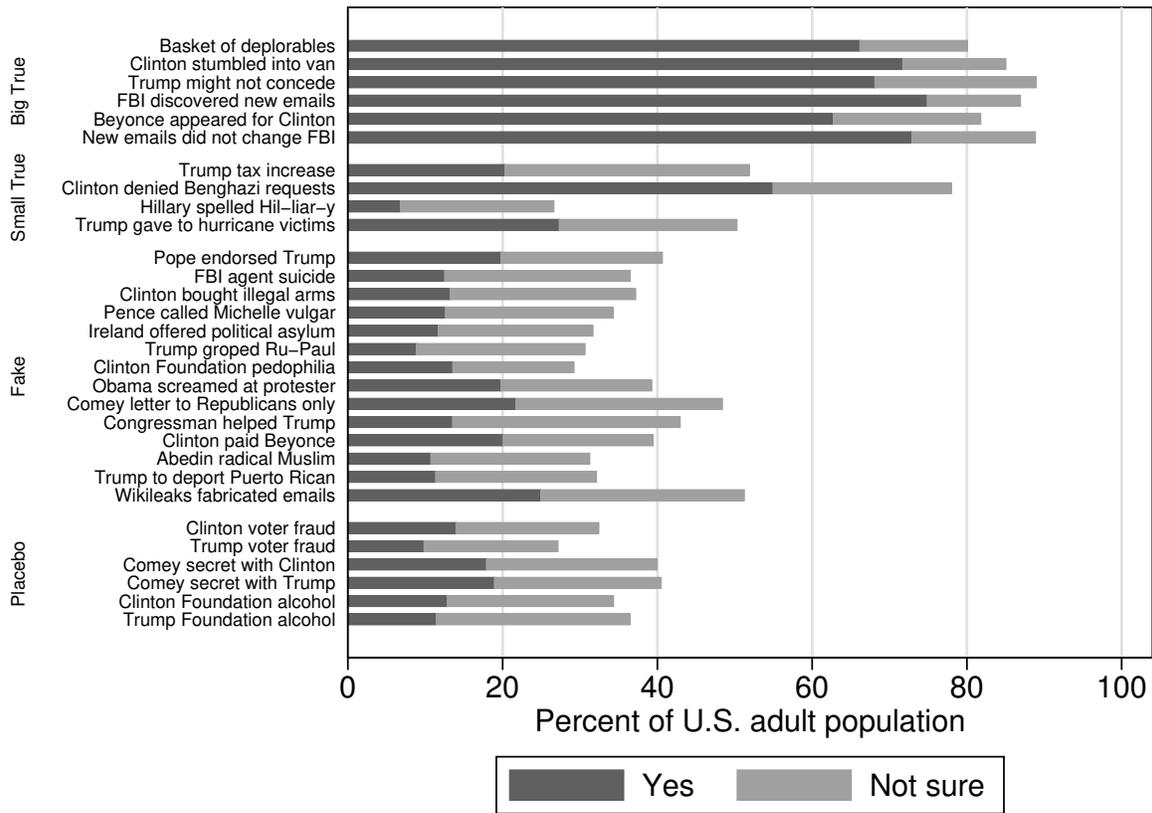
Notes: This table presents the 30 news articles used in the post-election survey. Each respondent received a randomly selected 15 of these stories, stratified to receive three from each of the five major categories listed.

Appendix Table 2: Post-Election Survey Summary Statistics

	(1)	(2)
	Survey sample	U.S. adult population
Household income (000s)	72.73	76.16
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	0.35	0.37
Republican	0.24	0.29
Web news consumption frequency	2.34	1.58
Social media news consumption frequency	1.88	1.24

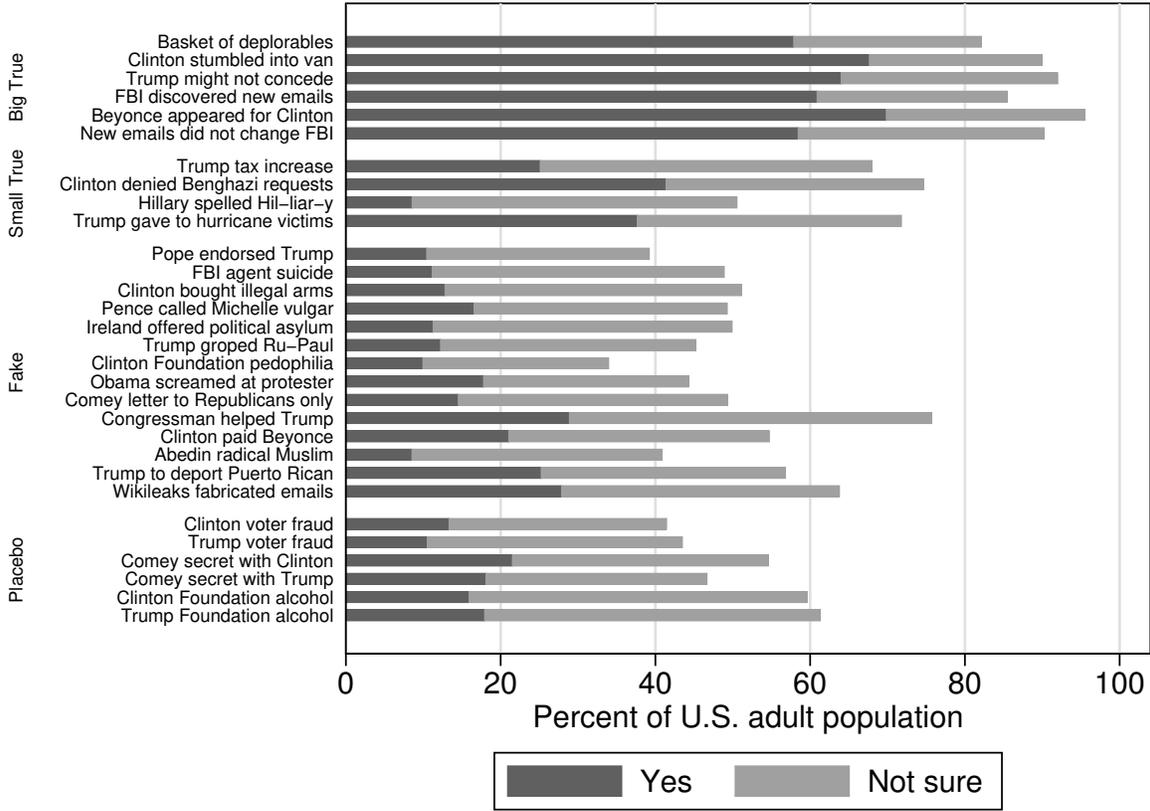
Notes: This table presents demographic data and summary statistics for the post-election survey and the U.S. adult population. News consumption frequency is coded as 3 (often), 2 (sometimes), 1 (rarely), and 0 (never). National average income, education, gender, age, and race are from the U.S. Census and are relevant for the U.S. population aged 18 and over. National party affiliation data are from the American National Election Studies' 2012 Time Series Study. National news consumption frequencies are from the Pew Center (2016b).

Appendix Figure 1: Percent of U.S. adult population that recalled seeing election news, by article



Notes: This figure presents the share of respondents that responded “Yes” and “Not sure” to the question, “Do you recall seeing this reported or discussed before the election,” for each of the 30 headlines listed in table 1. The headline categories written vertically are as defined in Appendix Table 1. Observations are weighted for national representativeness.

Appendix Figure 2: Percent of U.S. adult population that believed election news, by article



Notes: This figure presents the share of respondents that responded “Yes” and “Not sure” to the question, “At the time of the election, would your best guess have been that this statement was true?” for each of the 30 headlines listed in table 1. The headline categories written vertically are as defined in Appendix Table 1 Observations are weighted for national representativeness.

B A simple model of survey response

Using the survey results, we want to know two parameters: the share of population that was truly exposed to the average fake news article in our survey, and the share that was truly exposed and believed the average fake news article. Since the finding of false recall means that true exposure is not directly observed, it is helpful to formalize a simple model of survey response to understand how these two parameters can be inferred.

We assume that the probability that survey respondent i reports seeing (S_{ia}) or believing (B_{ia}) article a is some weakly increasing function G of true exposure $E_{ia} \in \{0, 1\}$ and the plausibility P_{ia} that the respondent assigns to the article. For $Y \in \{S, B\}$, this means that

$$\Pr(Y_{ia} = 1) = G_Y(\beta_Y E_{ia}, \gamma_Y P_{ia}), \quad (1)$$

with $\beta_Y, \gamma_Y \geq 0$. Larger β_S implies better memory, $\beta_B > 0$ if exposure *per se* causes people to believe articles, $\gamma_S > 0$ if respondents consider an article’s plausibility when trying to recall whether they saw it in the media, and $\gamma_B > 0$ simply reflects that more plausible articles are more likely to be believed. We define $M_{ia} \in \{0, 1\}$ as false memory—that is, $M_{ia} = 1$ when $S_{ia} = 1$ but $E_{ia} = 0$.

There are two types of articles, $t \in \{f, p\}$ for Fake and Placebo, and we denote the sets of articles as \mathcal{F} for Fake and \mathcal{P} for Placebo. By construction, the Placebo article exposure rate is zero: $E_{ia} = 0, \forall a \in \mathcal{P}$. Using \mathbb{E} to denote the expectation taken over both individuals and articles, the empirical fact that $\mathbb{E}[S_{ia}|a \in \mathcal{P}] > 0$ demonstrates that $\mathbb{E}[M_{ia}|a \in \mathcal{P}] > 0$. The empirical fact that seeing and believing are correlated for Placebo articles is explained by $\gamma_H, \gamma_B > 0$, i.e. plausibility P_{ia} affects both seeing and believing.

Consider the following two assumptions.

Assumption 1: People do not forget articles if they were actually exposed:

$$S_{ia} = 1 \text{ if } E_{ia} = 1. \quad (2)$$

Assumption 2: For the set of people who misremember seeing articles, plausibility is independent of article type:

$$P_{ia} \perp t, \forall i, a \text{ s.t. } M_{ia} = 1. \quad (3)$$

In essence, Assumption 2 is that Fake and Placebo articles are equally plausible.

We constructed the survey so that these assumptions would be credible. We implemented the survey soon after the election to minimize forgetting and false recall. Assumption 2 is not directly testable because misremembering is unobserved. However, figure 3 shows an approximate test of Assumption 2 if true exposure rates are small. 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 Assumption 2 since all people who recalled seeing Placebo headlines are misremembering, as are almost all people who recalled seeing Fake headlines (for small exposure rates). More broadly, Assumption 2 is likely to hold by design because we wrote the Placebo headlines, and refined them in the pilot, to ensure that they were approximately equally plausible as the Fake headlines.

These two assumptions allow us to infer rates of both true exposure as well as true exposure and believing. Under assumptions 1 and 2, it is straightforward to show that $\mathbb{E}[E_{ia}|a \in \mathcal{F}] = \mathbb{E}[S_{ia}|a \in \mathcal{F}] - \mathbb{E}[S_{ia}|a \in \mathcal{P}]$ and $\mathbb{E}[E_{ia}B_{ia}|a \in \mathcal{F}] = \mathbb{E}[S_{ia}B_{ia}|a \in \mathcal{F}] - \mathbb{E}[S_{ia}B_{ia}|a \in \mathcal{P}]$. In words, subtracting the reported rates for Placebo articles from the reported rates for Fake articles gives the true rates for Fake articles. Intuitively, this is the case because Placebo headlines that are calibrated to be equally-plausible provide a control for false recall.

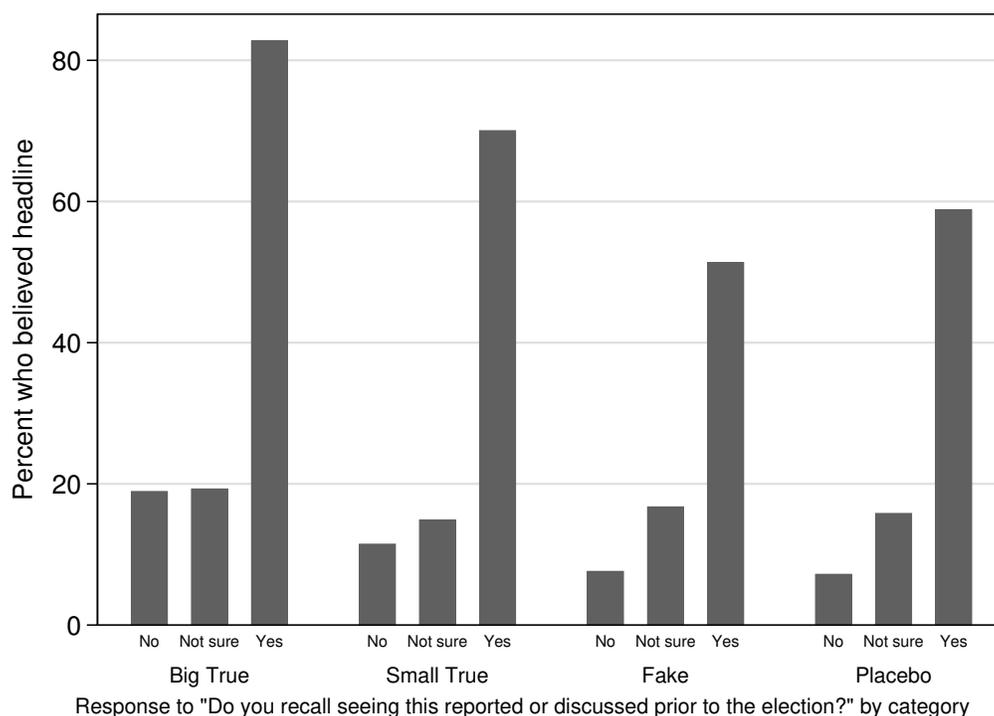
C Additional Figures and Tables

Appendix Table 3: Rates of seeing and believing fake news relative to placebo fake news

	(1)	(2)	(3)	(4)	(5)	(6)
	Fake	Recalled seeing		Recalled seeing and believed		
		Placebo	Fake-Placebo	Fake	Placebo	Fake-Placebo
Share of population	0.153*** (0.009)	0.141*** (0.011)	0.012 (0.009)	0.079*** (0.007)	0.083*** (0.009)	-0.005 (0.007)
N	8,456	3,624	12,080	8,456	3,624	12,080
95 pct confidence bound	.171	.1632	.0288	.0924	.1012	.009

Notes: This table presents the share people who recall seeing (columns 1-3) or recall seeing and believed (columns 4-6) 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. Standard errors are robust and clustered by survey respondent. *, **, ***: statistically significant from zero with 90, 95, and 99 percent confidence, respectively.

Appendix Figure 3: Share who believe news by whether they heard news, by category



Notes: In our post-election survey, we presented 15 headlines. For each headline, the survey asked whether respondents had heard the headline (“Do you recall seeing this reported or discussed before the election?”) and whether they believed it (“At the time of the election, would your best guess have been that this statement was true?”). This figure presents the share of people who believed the headlines in each category, broken down by responses to whether they had heard each headline. Observations are weighted for national representativeness.