

Exposure to untrustworthy websites in the 2016 U.S. election

Andrew M. Guess¹, Brendan Nyhan^{2*} & Jason Reifler³

¹*Department of Politics and Woodrow Wilson School, Princeton University, Princeton, NJ, USA*

²*Department of Government, Dartmouth College, Hanover, NH, USA*

³*Department of Politics, University of Exeter, Exeter, UK*

*Corresponding author: Brendan Nyhan (nyhan@dartmouth.edu)

Abstract

Though commentators frequently warn about “echo chambers,” little is known about the volume or slant of political misinformation people consume online, the effects of social media and fact-checking on exposure, or its effects on behavior. We evaluate these questions for the websites publishing factually dubious content often described as “fake news.” Survey and web traffic data from the 2016 U.S. presidential campaign show that Trump supporters were most likely to visit these websites, which often spread via Facebook. However, these sites made up a small share of people’s information diets on average and were largely consumed by a subset of Americans with strong preferences for pro-attitudinal information. These results suggest that widespread speculation about the prevalence of exposure to untrustworthy websites has been overstated.

Introduction

“Fake news” remains one of the most widely debated aspects of the 2016 U.S. presidential election. Articles from untrustworthy websites that featured factually dubious claims about politics and the campaign were shared by millions of people on Facebook¹. Post-election surveys indicated these claims were often widely believed^{2,3}. Some journalists and researchers have even suggested that “fake news” may be responsible for Donald Trump’s victory⁴⁻⁷.

These developments raise significant democratic concerns about the quality of the information that voters receive. However, little is known scientifically about the consumption of so-called “fake news” from these untrustworthy websites or how it relates to political behavior. In this study, we provide comprehensive individual-level analysis of the correlates and consequences of untrustworthy website exposure in the real world. Our analysis combines a dataset of pre-election survey data and web traffic histories from a national sample of Americans with an extensive list of untrustworthy websites. These data enable us to conduct analyses that are not possible using post-election self-reports of exposure³, aggregate-level data on visits to untrustworthy websites⁸ or behavioral data that lack candidate preference information⁹.

We report five principal findings. First, consistent with theories of selective exposure, people differentially consume false information that reinforces their political views. However, fewer than half of all Americans visited these untrustworthy websites, which represented approximately 6% of people’s online news diet during the study period (95% CI: 5.1%–6.7%). Consumption of news from these sites was instead heavily concentrated among a small subset of people — 62% of the visits we observe came from the 20% of Americans with the most conservative information diets.

Third, we show that Facebook played a central role in spreading content from untrustworthy websites relative to other platforms. Fourth, fact-checks of articles published by these outlets almost never reached their target audience. Finally, we examine whether consumption of factually dubious news affected other forms of political behavior. We find that untrustworthy website consumption does not crowd out consumption of other hard news. However, our results about the relationship between untrustworthy website consumption and both voter turnout and vote choice are statistically imprecise; we can only rule out very large effects.

Theory and expectations

Alarm about “fake news” reflects concerns about rising partisanship and pervasive social media usage, which have raised fears that “echo chambers” and “filter bubbles” could amplify misinformation and shield people from counter-attitudinal information^{10,11}. Relatively little is known, however, about the extent to which selective exposure can distort the factual information that people consume and promote exposure to false or misleading factual claims — a key question for U.S. democracy.

We therefore evaluate the extent to which people engage in selective exposure to untrustworthy websites that reinforce their partisan predispositions during a general election campaign. Studies show that people tend to prefer congenial information, including political news, when given the choice¹²⁻¹⁵. We therefore expect Americans to prefer factually dubious news that favors the candidate they support. However, behavioral data show that only a subset of Americans have heavily skewed media consumption patterns¹⁶⁻¹⁸. We therefore disaggregate the public by the

overall skew in their information diets to observe whether consumption of news from untrustworthy websites mirrors people's broader tendencies toward selective exposure.

We also seek to understand how false and misleading information disseminates online. The speed and reach of social media and the lack of fact-checking make it an ideal vehicle for transmitting misinformation¹⁹. Studies indicate that social media can spread false claims rapidly²⁰ and increase selective exposure to attitude-consistent news and information²¹. These tendencies may be exacerbated by design and platform features such as algorithmic feeds and community structures²². In this study, we test whether social media usage increases exposure to news from untrustworthy websites, a source of misinformation about politics that is attitude-consistent for many partisans. We compare the role that social media plays compared to other web platforms such as Google or webmail.

However, the effects of exposure to this factually dubious content may be attenuated if those who are exposed to it also receive corrective information. Fact-checks are relatively widely read and associated with greater political knowledge²³. Though some studies find that people sometimes resist corrective information in news stories^{24,25}, meta-analyses indicate that exposure to fact-checks and other forms of corrective information generally increase belief accuracy²⁶⁻²⁸. Fact-checks could thus help to counter the pernicious effects of exposure to untrustworthy websites. In the real world, fact-checks may fail to reach the audience that is exposed to the claims they target²⁹. We therefore examine the relationship between exposure to factually dubious news and fact-check consumption. These analyses complement existing experimental fact-checking research by capturing exposure to misinformation and fact-checks in the wild³⁰.

Finally, we examine previous conjectures that exposure to so-called “fake news” affects other types of political behavior — hard news consumption, vote choice, and voter turnout. Most notably, a recent study claims that “false information did have a substantial impact” on the election based on the association in a post-election survey between expressed belief in anti-Hillary Clinton “fake news” and self-reported support for Donald Trump among self-reported supporters of Barack Obama in 2012⁷. However, this research design is correlational and relies on post-election self-reports; it cannot establish causality. Moreover, previous research offers limited support for these conjectures. First, news consumption habits are likely to be ingrained and related to traits such as partisan strength and political interest^{31,32}. Untrustworthy websites are unlikely to displace people’s normal information diets. In addition, the effects of brief exposure to persuasive messages have been found to be small in partisan election campaigns. A meta-analysis of 49 field experiments finds that the average effect of personal and impersonal forms of campaign contact is zero³³. Even the effects of television ads are extremely limited: just one to three people out of 10,000 who are exposed to an additional ad change their vote choice to support the candidate in question by one estimate³⁴. However, exposure could also affect the election by helping to mobilize likely supporters to turn out to vote (or, alternatively, by discouraging likely opponents from voting). Evidence does suggest that television advertising can shift the partisan composition of the electorate³⁴. Exposure to online content that strongly supports a candidate or attacks their opponent could potentially have similar effects.

Results

Total untrustworthy website consumption We estimate that 44.3% of Americans age 18 or older visited an article on an untrustworthy website during our study period, which covered the final weeks of the 2016 election campaign (95% CI: 40.8%–47.7%). In total, articles on these factually dubious websites represented an average of 5.9% of all the articles Americans read on sites focusing on hard news topics during this period (95% CI: 5.1%–6.7%). The content people read on these sites was heavily skewed by partisanship and ideology — articles on untrustworthy conservative websites represented 4.6% of people’s news diets (95% CI: 3.8%–5.3%) compared to only 0.6% for liberal sites (95% CI: 0.5%–0.8%).

Selective exposure to untrustworthy websites We observe stark differences by candidate support and information diet in the frequency and slant of untrustworthy website visits, suggesting powerful selective exposure effects. First, people who indicated in the survey that they supported Trump were far more likely to visit untrustworthy websites — especially those that are conservative and thus very likely pro-Trump — than those who indicated they were Clinton supporters. Among Trump supporters, 56.7% read at least one article from an untrustworthy conservative website (95% CI: 51.7%–61.8%) compared with 27.7% of Clinton supporters (95% CI: 23.1%–32.4%). Consumption of articles from untrustworthy liberal websites was much lower, though also somewhat divided by candidate support. Clinton supporters were modestly more likely to have visited untrustworthy liberal websites (23.0%, 95% CI: 18.8%–27.2%) than Trump supporters (10.7%, 95% CI: 8.1%–13.3%). These differences in the proportions of supporters who were exposed to attitude-consistent untrustworthy news are illustrated in Figure 1.

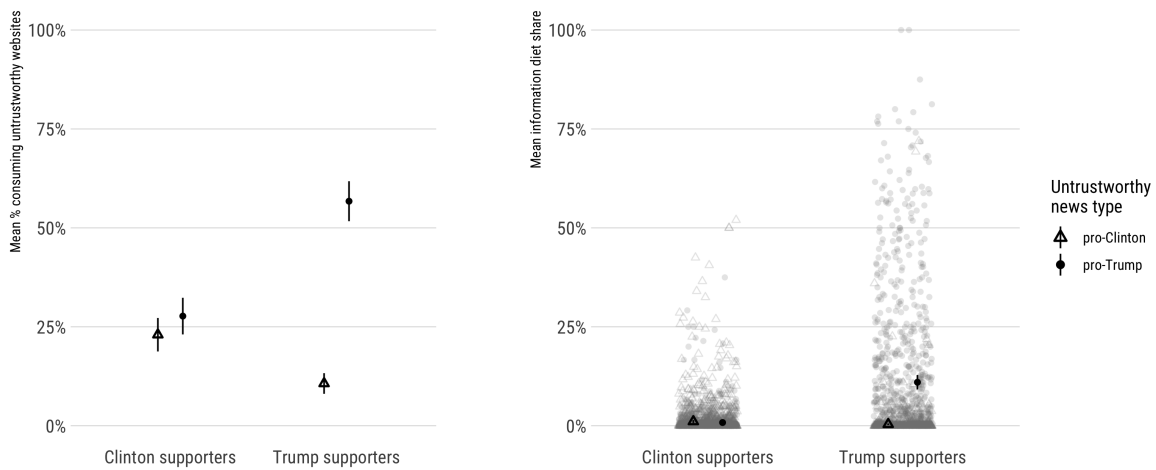


Figure 1: **Selective exposure to untrustworthy websites.** Means and 95% confidence intervals calculated using survey weights for October 7–November 14, 2016 among YouGov Pulse panel members who supported Clinton or Trump ($N = 2,170$ for binary exposure measure; $N = 2,016$ for information diet). The denominator for information consumption includes total exposure to those sites as well as the number of pages visited on websites classified as focusing on hard news topics (excluding Amazon, Twitter, and YouTube). Respondents who did not visit any of these sites are excluded from the information diet graph.

The differences by candidate preference that we observe in untrustworthy consumption are even more pronounced when expressed in terms of the composition of the overall news diets of each group. When we consider pages visited on websites that Bakshy et al.²¹ classified as hard news (excluding Amazon, Twitter, and YouTube) as well as untrustworthy websites that Grinberg et al.³⁵ classify as liberal or conservative, we again observe significant differences in consumption by candidate preference. Untrustworthy conservative websites made up 11.0% of the news diet of Trump supporters (95% CI: 9.2%–12.8%) compared to only 0.8% among Clinton supporters (95% CI: 0.6%–1.1%). The pattern was again reversed but of lesser magnitude for untrustworthy liberal websites: 1.1% of pages on hard news topics for Clinton supporters (95% CI: 0.7%–1.5%) versus 0.4% for Trump supporters (95% CI: 0.2%–0.6%).

The differences we observe in visits to untrustworthy liberal and conservative websites by candidate support are statistically significant in OLS models even after we adjust for standard demographic and political covariates, including a standard scale measuring general political knowledge (Table 1). For both a binary measure of exposure and a measure of the share of the respondent's information diet, Trump supporters were disproportionately more likely to consume untrustworthy conservative news and less likely to consume untrustworthy liberal news relative to Clinton supporters, supporting a selective exposure account. Older Americans (age 60 and older) also consumed more information from untrustworthy websites irrespective of slant conditional on these covariates.

To analyze which specific types of news consumers were most likely to visit untrustworthy websites, we divide users into deciles depending on the slant of their information diet, which we

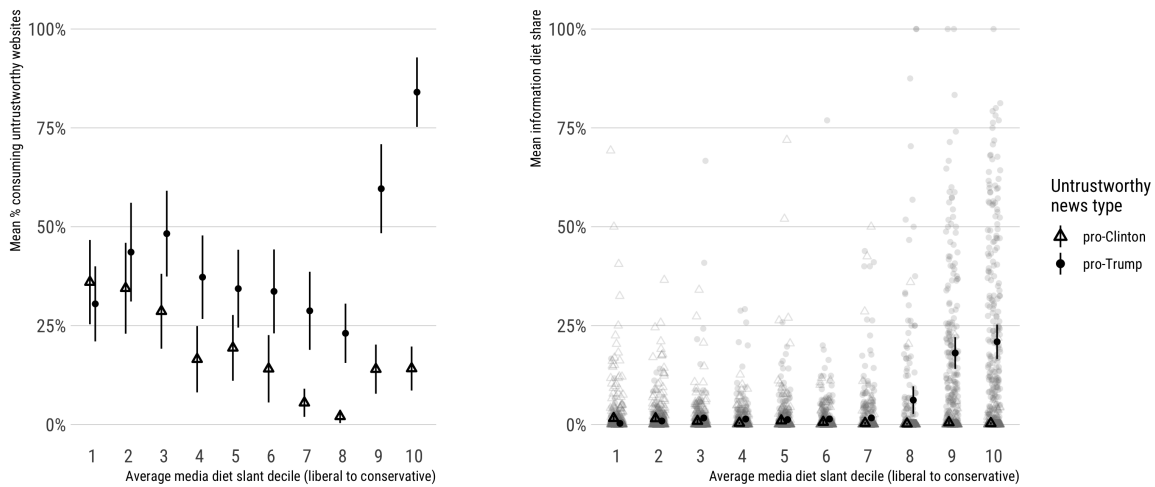


Figure 2: Visits to untrustworthy websites by media diet slant decile. Means and 95% confidence intervals calculated using survey weights for October 7–November 14, 2016 among YouGov Pulse panel members whose average media diet slant could be calculated ($N = 2,422$ for binary exposure measure; $N = 2,334$ for information diet). The denominator for information consumption includes total exposure to those sites as well as the number of pages visited on websites classified as focusing on hard news topics (excluding Amazon, Twitter, and YouTube). Respondents who did not visit any of these sites are excluded from the information diet graph.

compute as the mean slant weighted by pageviews among the websites they visit for which data are available from Bakshy et al.²¹ (which estimate website slant based on differential Facebook sharing by self-identified liberals versus conservatives). Figure 2 shows how consumption of untrustworthy websites varies across these ten deciles, which range from the 10% of respondents who visit the most liberal sites to the 10% who visit the most conservative sites (on average).

The proportion of the sample that visited at least one untrustworthy conservative site ranges inconsistently from 31–48% across the first eight deciles of selective exposure from liberal to conservative, but rises steeply to 59.6% in the second most conservative decile (95% CI: 48.4%–70.9%) and 84.0% in the most conservative decile (95% CI: 75.2%–92.8%). The total amount of untrustworthy news consumption is also vastly greater in the top deciles; news from

untrustworthy conservative websites made up 18.1% of news consumption among the second most conservative decile (95% CI: 14.1%–22.1%) and 20.9% among the 10% of Americans with the most conservative information diets (95% CI: 16.6%–25.3%). These totals, which represent an average of 25 and 91 articles, respectively, are dramatically higher than those observed in the rest of the population (0.3–6.2% across the eight remaining deciles; see Supplementary Table 8 for corresponding statistical results). In total, 62% of all page-level traffic to untrustworthy websites observed in our data during the study period came from the 20% of news consumers with the most conservative information diets.

Engagement with untrustworthy websites Due to the nature of the web consumption data analyzed in this study, we do not have direct measures of whether respondents carefully read articles they visited on untrustworthy websites or believed the claims in those articles. However, auxiliary evidence suggests that respondents engaged with the articles they visited and were vulnerable to believing the claims that they contained.

First, we observe little evidence that most respondents immediately closed articles from untrustworthy websites or otherwise failed to engage with them meaningfully. In fact, they spent more time on pages from these websites than on articles from websites focusing on hard news topics. On average, respondents spent over a minute on articles from untrustworthy websites (64.2 seconds), about two-thirds of a minute on articles from domains that focus on hard news topics (42.1 seconds), and about one quarter of a minute on articles from other sites (24.2 seconds).

In addition, the respondents who visited untrustworthy websites scored lower on a validated measure of cognitive reflection that has been shown to predict greater accuracy in distinguishing

false from real news headlines. People who perform worse on the Cognitive Reflection Test (CRT) are less able to distinguish between false and real headlines^{36,37}. Among the subset of respondents who completed the CRT in separate YouGov surveys, we observe no association between CRT scores and untrustworthy conservative website consumption. But as Figure 3 suggests, Trump supporters who score in the top quintile on the CRT (at least two of the three questions correct) consumed less news from these websites as a share of their information diet than those who got no questions correct ($p=.04$; see Supplementary Table 7). These results suggest that people who got the most news from untrustworthy websites were also more likely to believe it. (Corresponding analyses for untrustworthy liberal websites, which respondents were much less likely to visit, are provided in the Supplementary Information.)

Gateways to untrustworthy websites How do people come to visit an untrustworthy website? Since the election, many have argued that social media, especially Facebook, played an integral role in exposing people to untrustworthy news^{1,3,9}. While we cannot directly observe the referring site or application for the URLs visited by our survey panel, we can indirectly estimate the role Facebook played in two ways.

First, we group respondents who supported either Clinton or Trump into three terciles of observed Facebook usage. The results, which are analyzed statistically in Supplementary Table 10, show a dramatic association between Facebook usage and untrustworthy website exposure. Visits to untrustworthy conservative websites increased from 16.5% among Clinton supporters who do not use Facebook or use it relatively little (95% CI: 8.9%–24.1%) to 24.8% in the middle tercile (95% CI: 17.3%–32.2%) and 45.7% among the Clinton supporters who use Facebook most (95% CI:

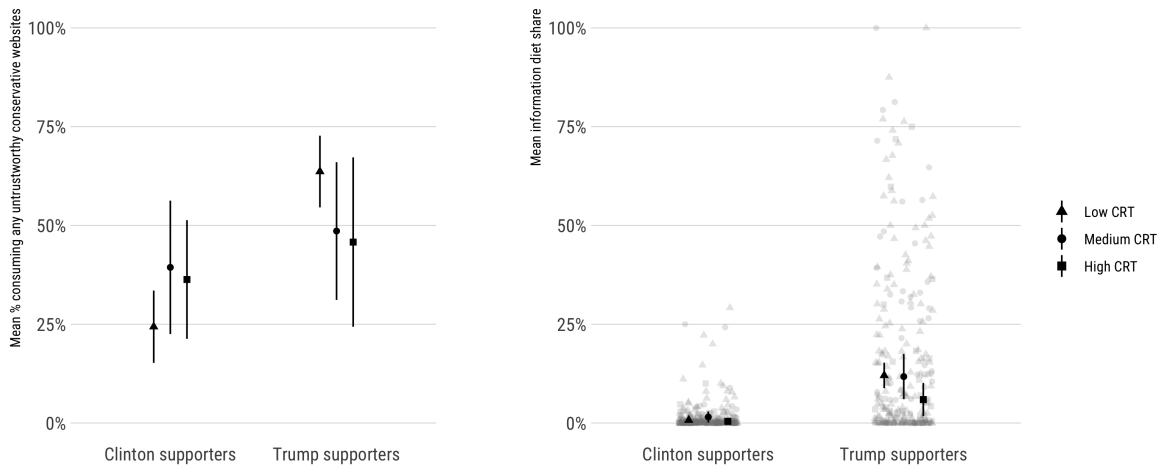


Figure 3: **Consumption of untrustworthy conservative websites by CRT score and candidate preference.** Means and 95% confidence intervals calculated using survey weights for October 7–November 14, 2016 among YouGov Pulse panel members who supported Clinton or Trump ($N = 772$ for binary exposure measure; $N = 711$ for information diet). The denominator for information consumption includes total exposure to those sites as well as the number of pages visited on websites classified as focusing on hard news topics (excluding Amazon, Twitter, and YouTube). Respondents who did not visit any of these sites are excluded from the information diet graph. “Medium” and “high” CRT scores indicate respondents who got one or more than one question correct on the Cognitive Reflection Test (22% and 20%, respectively).

36.2%–55.1%). The increase is similar among Trump supporters, for whom visit rates increased from 39.6% in the lowest third of the Facebook distribution (95% CI: 30.2%–49.0%) to 51.5% in the middle third (95% CI: 42.8%–60.1%) and 74.4% in the upper third (95% CI: 66.7%–82.2%). We observe similar patterns for visits to untrustworthy liberal websites and for exposure levels as a share of respondents' information diets.

Second, following an approach used in prior research, we can make a more direct inference about the role of Facebook by examining the URLs visited by a respondent immediately prior to visiting an untrustworthy website¹⁸. As Figure 4 demonstrates, Facebook was among the three previous sites visited by respondents in the prior thirty seconds for 15.1% of the articles from untrustworthy news websites we observe in our web data. By contrast, Facebook appears in the comparable prior URL set for only 5.9% of articles on websites classified as hard news (excluding Amazon, Twitter, and YouTube). This pattern of differential Facebook visits immediately prior to untrustworthy website visits is not observed for Google (3.3% untrustworthy news versus 6.2% hard news) or Twitter (1.0% untrustworthy versus 1.5% hard news) and exceeds what we observe for webmail providers such as Gmail (9.5% untrustworthy versus 5.4% hard news). Our results demonstrate that Facebook was a key vector of distribution for untrustworthy websites.

The problem of fact-checking mismatch The most prominent journalistic response to “fake news” from untrustworthy websites and other forms of misleading or false information is fact-checking, which has attracted a growing audience in recent years. We found that one in four respondents (25.3%; 95% CI: 22.5%–28.2%) visited a fact-checking article from a national fact-checking website at least once during the study period.

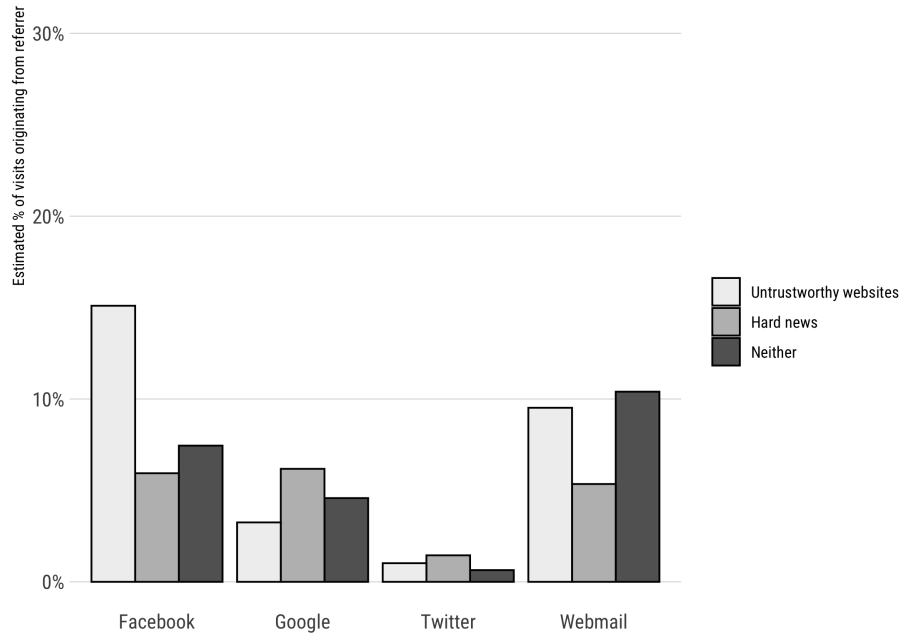


Figure 4: **Referrers to untrustworthy news websites and other sources.** Means and 95% confidence intervals calculated using survey weights for October 7–November 14, 2016 among YouGov Pulse panel members ($N = 2,525$). The denominator for information consumption includes total exposure to those sites as well as the number of pages visited on websites classified as focusing on hard news topics (excluding Amazon, Twitter, and YouTube). Respondents who did not visit any of these sites are excluded from the information diet graph. Facebook, Google, Twitter, or a webmail provider such as Gmail were identified as a referrer if they appeared within the last three URLs visited by the user in the thirty seconds prior to visiting the article.

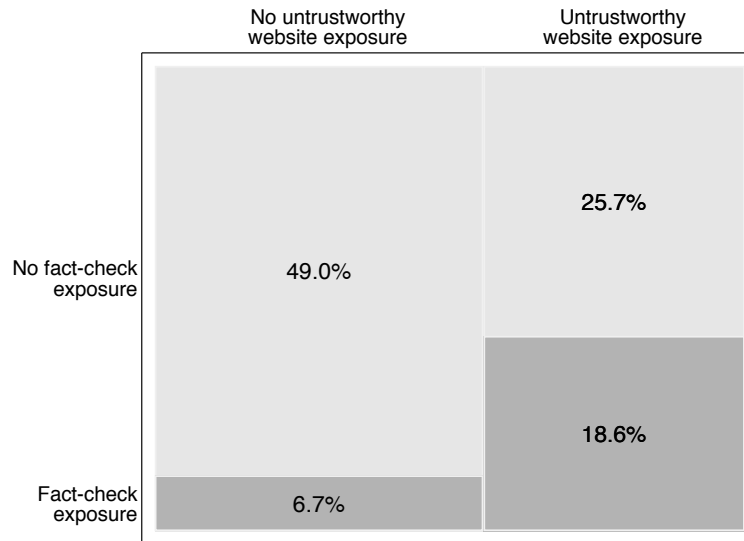


Figure 5: **Fact-check and untrustworthy website visits.** Means and 95% confidence intervals calculated using survey weights for October 7–November 14, 2016 among YouGov Pulse panel members ($N = 2,525$). Fact-check exposure is measured as a visit to PolitiFact, the Washington Post Fact Checker, Factcheck.org, or Snopes.

Recent evidence suggests that this new form of journalism can help inform voters²⁵. However, fact-checking may not effectively reach people who have encountered the false claims it debunks. As Figure 5 illustrates, fewer than half of the 44.3% of Americans who visited an untrustworthy website during the study period also saw any fact-check from one of the dedicated fact-checking websites (18.6%; 95% CI: 16.1%–21.1%).

More specifically, only three of the 111 respondents (2.7%) who read one or more articles from untrustworthy websites that could be matched by Allcott and Gentzkow³ or Grinberg et al.³⁵ to a negative fact-check also read the fact-check debunking the article in question. Searching for additional information more generally also appears to be extremely rare. Google appears among the first three URLs visited in the thirty seconds after a visit for only 2.0% of visits to untrustworthy websites among Clinton/Trump supporters compared to 4.0% for hard news

website visits and 3.9% for all other website visits.

Relationship to news consumption, vote choice, and turnout Observers have suggested that news from untrustworthy websites is displacing hard news consumption³⁸ or that it changed the outcome of the 2016 election⁴⁻⁷. We evaluate both claims.

First, we do not observe evidence that consumption of news from untrustworthy websites crowds out consumption of information from other sources of hard news. Those who consume the most hard news tend to consume the most information from untrustworthy websites (in other words, they appear to be complements, not substitutes). For instance, when we divide the population into terciles by total hard news consumption, we find that the proportion of Trump supporters who visited untrustworthy conservative websites increases from 30.9% in the lowest tercile (95% CI: 23.8%–37.9%) to 67.0% in the middle tercile (95% CI: 58.3%–75.6%) and 79.8% in the high tercile (95% CI: 71.8%–87.7%). This increase is statistically significant (see Supplementary Table 12).

To further verify that news from untrustworthy websites does not crowd out hard news consumption, we compare hard news consumption among respondents for whom online traffic data is also available from a separate February/March 2015 study using a variant of a difference-in-differences approach. Hard news consumption increased substantially from the 2015 study to the 2016 study among those who we know consumed any untrustworthy websites in 2016 but not among those who did not (see Supplementary Table 13). These results are inconsistent with a simple hypothesis that news from untrustworthy websites crowds out hard news consumption.

Other claims concern potential effects on turnout and vote choice. While some have suggested that Donald Trump won the 2016 election because of “fake news”^{4,6}, the most widely cited evidence relies on a post-election survey showing a negative association between belief in false claims about Hillary Clinton and self-reports of having voted for her among self-reported Obama voters in 2012⁷. However, self-reports of vote choice after the election are subject to recall error and bias in self-reporting. More fundamentally, a post-election survey cannot establish that people had even heard these claims before the election. As a result, any such association cannot be interpreted as causal.

Given widespread interest in the claim that “fake news” helped win the election for Trump, we consider whether any relationship exists between prior exposure to untrustworthy conservative websites and two possible outcomes of interest — vote choice in our pre-election survey and a validated measure of voter turnout provided by YouGov, which matched respondents to the TargetSmart voter file. Our results are not sufficiently precise to offer definitive support for or against that claim.

As Table 2 indicates, the relationships between exposure to untrustworthy conservative news and changes in vote choice intention or turnout are imprecisely estimated for both low and high levels of exposure overall (models 1 and 3) and when we distinguish between supporters of Trump and Clinton supporters and respondents who were instead undecided or supported another candidate in July 2016 (models 2 and 4). Equivalence tests reveal that we can only confidently rule out very large effects on vote choice or turnout (ten percentage points or more for Trump support and nine percentage points for turnout; see the Supplementary Information for further details).

Discussion

This paper provides systematic evidence of differential exposure to a key form of false or dubious political information during a real-world election campaign: untrustworthy liberal and conservative websites during the 2016 U.S. presidential election. Our data, which do not rely on post-election survey recall or forced exposure to “fake news” content, indicate that less than half of all Americans visited an untrustworthy website in the weeks before the election and that these websites make up a small percentage of people’s online news diets. However, we find evidence of substantial selective exposure — in particular, Trump supporters differentially consuming news from untrustworthy conservative websites. This tendency was heavily concentrated among a subset of Americans with conservative information diets. In this sense, “echo chambers” are deep (52 articles from untrustworthy conservative websites on average in this subset) but they are also narrow (the group consuming so much of this content represents only 20% of the public).

Our results also provide important evidence about the mechanisms by which factually dubious news disseminates and the effectiveness of responses to it. Specifically, we find that Facebook played an important role in directing people to untrustworthy websites — heavy Facebook users were differentially likely to consume information from these sites, which was often immediately preceded by a visit to Facebook. In addition, we show that fact-checking websites failed to effectively reach visitors to untrustworthy websites — audience overlap was only partial at the domain level and virtually non-existent for the fact-checks that could be matched to specific articles.

Finally, we examine whether visiting untrustworthy websites affects other political behaviors. Our

results indicate that consumption of untrustworthy websites does not decrease hard news consumption. When it comes to turnout and vote choice, however, our results are imprecise and can only rule out very large effect sizes.

Of course, our study only examines visits to untrustworthy websites via web browsers. It would be desirable to observe exposure on mobile devices and social media platforms directly and to measure consumption of forms of hyper-politicized media including hyperpartisan Twitter feeds and Facebook groups, internet forums such as Reddit, more established but often factually dubious websites, memes and other images relevant to political topics, and more traditional media like talk radio and cable news. Future research should also employ designs that allow us to better assess the effects of untrustworthy websites and other forms of misinformation more broadly. It is also important to understand the role played by interpersonal discussion and other forms of indirect communication in exposing people to both misinformation and fact-checking³⁹. In addition, though we find no measurable evidence of effects of consumption of news from untrustworthy websites on news consumption, vote choice, or turnout, further study is necessary to validate these findings and assess how it affects public debate, misperceptions, hostility toward political opponents, and trust in government and journalism. The results presented here also focus on one period of time; the effects of cumulative exposure to websites with factually questionable content and other forms of misinformation deserve future attention⁴⁰. Finally, further investigation is necessary to understand the reasons for the differences we and other scholars observe in visits to untrustworthy news websites by partisanship/ideology and age^{35,41}.

We also acknowledge the difficulty of comprehensively assessing content accuracy at scale.

Despite significant institutional resources, professional fact-checkers can only investigate a small fraction of claims disseminated by media outlets⁴². Likewise, researchers are limited in their ability to evaluate the accuracy of the large quantities of content that they observe respondents consuming in behavioral data. Moreover, even if such an undertaking were feasible, it would not be possible for us to fully resolve the well-known epistemological challenges inherent to the enterprise of judging truth claims^{43,44}. These constraints necessitate the use of site-level quality or accuracy ratings like those employed in this study (i.e., the data we use to construct our list of untrustworthy websites). By using expert, peer-reviewed determinations about the quality of information provided by websites, we sidestep difficult disputes that often arise at a more granular level of analysis. While not perfect, this procedure is transparent, replicable, and successfully identifies sites that publish articles which are most likely to be found to be false by professional fact-checkers.

Nonetheless, these results underscore the importance of directly studying exposure to untrustworthy websites and other dubious information online. As other studies indicate, exposure to these extreme forms of misinformation is concentrated among a subset of Americans who consume this type of content in large quantities⁴⁵. However, these small groups can help propel dubious claims to widespread visibility online, potentially intensifying polarization and negative affect. This pattern represents an important development in political information consumption.

Methods

This study was approved by the New York University Institutional Review Board (Protocol: IRB-FY2017-149) and the Dartmouth Committee for the Protection of Human Subjects (Study:

00029870). Informed consent was obtained from all participants, who also received incentives from the survey company that collected the data.

Data for this study combine responses to an online public opinion survey of a national sample of Americans with online traffic data collected passively from respondents' computers. These data were collected by the survey firm YouGov from members of their Pulse panel who provided informed consent to allow anonymous tracking of their online data. The software tracks web traffic (minus passwords and financial transactions) for all browsers installed on a user's computer. Users provide consent before installing the software and can turn it off or uninstall it at any time. Identifying information is not collected.

Time-stamped URL-level web traffic data was recorded from October 7–November 14, 2016. Survey data was collected on the YouGov survey platform from October 21–31, 2016 (approximately the middle of our online behavioral data collection period). YouGov also appended additional variables to the data on voter turnout (from voter files updated after the 2016 election) and both prior candidate preference and Cognitive Reflection Test scores (from other surveys taken by our respondents). We employ survey weights for the data to accurately represent the adult population of the U.S. Data from the Facebook News Feed is not included due to restrictions on the Facebook API. We also do not analyze mobile traffic data in the main text because it is only available for 19% of respondents ($n = 629$) and does not capture the full URL of each website visited; see the Online Appendix for details on the domain-level traffic patterns we observe in that data.

This sample closely resembles the U.S. population in its demographic characteristics, privacy

attitudes, and voter turnout behavior. Among the 3,251 survey respondents (a sample size that was determined by budget constraints), 52% are female, 68% are non-Hispanic whites, and 29% have a bachelor's degree or higher when survey weights constructed by YouGov are applied to approximate a nationally representative sample.

The data are likely to not be perfectly representative of the U.S. population due to the unusual Pulse panel — people with less than a high school degree are underrepresented and the sample tilts Democratic (42% Clinton versus 33% Trump on a vote intention question that included Gary Johnson, Jill Stein, other, not sure, and probably won't vote as options) — but the participants are diverse and resemble the population on many dimensions.

This study specifically focuses on data from the 2,525 survey respondents for whom page-level online traffic data from laptop or desktop computers are also available.

Table 3 provides a comparison of the demographic composition and political preferences of the full Pulse sample and participants with online traffic data we analyze with the pre-election American National Election Studies (ANES) face-to-face survey, a benchmark study that was also conducted during the general election campaign. The set of respondents for whom we have page-level online traffic data is demographically very similar to the full Pulse sample and closely resembles the composition of the ANES sample. Two exceptions are preferences for Clinton and intention to vote. However, the latter difference is contradicted by turnout data: 56.6% of respondents for whom online traffic data were available were recorded as voting in the 2016 general election according to the TargetSmart voter file data matched to our respondents by YouGov, which corresponds closely to the U.S. voting-age population turnout rate of 54.7% for

2016⁴⁶. Another distinction concerns technology usage. The Pulse sample has somewhat higher levels of home internet access (presumably 100%) compared with the ANES sample (89%). In addition, we note that our sample seems to demonstrate modestly higher levels of Facebook usage than the American public — 88% visited a Facebook URL at least once and 76% did so more than ten times in our sample period compared with 62% of Americans interviewed in the 2016 American National Election Studies face-to-face survey who said they had a Facebook account and had used it in the last month. However, our measures potentially also capture visits to Facebook pages by individuals who do not have an account.

This study considers the relationship between the demographic and attitudinal variables measured in our survey data, the information consumption behavior observed in our web traffic data, and behavioral data on voter turnout by participants. Specifically, we both identify the demographic and attitudinal correlates of consumption of news from untrustworthy websites and analyze the association between this consumption and relevant behaviors (news consumption, voter turnout, and vote choice). We focus specifically in this study on respondents who reported supporting Hillary Clinton or Donald Trump in our survey (76% of our sample) because of our focus on selective exposure by candidate preference.

Studying consumption of untrustworthy websites requires defining which websites frequently publish factually dubious or untrustworthy content. Following previous research^{35,41}, we use a domain-level approach to measurement. Our goal is to analyze consumption of news from untrustworthy sources rather than exposure to false information *per se* — a necessity given the impossibility of assessing the accuracy of all information that people encounter about politics. We specifically seek to identify websites that “lack the news media’s editorial norms and processes for

ensuring the accuracy and credibility of information”⁴⁷. The qualifying untrustworthy websites considered in this study are those classified as “black” (382 websites), “orange” (47 websites), and “red” (61 websites) by Grinberg et al.³⁵ The “black” sites were those previously identified by journalists and fact-checkers as notable publishers of false or misleading content. Grinberg et al. also classified sites as “orange (negligent or deceptive)” or “red (little regard for the truth)” using human annotation of website editorial practices among sites fact-checked by Snopes or that were frequently mentioned in their Twitter data³⁵.

There are of course many lists of untrustworthy online content, but we present robustness tests in the Supplementary Information using two alternate outcome measures that yield results which are highly consistent with those presented in Table 1. First, we show that results are highly similar using an alternative domain-level measure: domains identified by Allcott and Gentzkow³ as frequently publishing dubious information that were created soon before the 2016 election and overwhelmingly supported one of the two major candidates. Second, we show that our results are also highly similar for a measure of exposure to articles from the Grinberg et al.³⁵ set of untrustworthy websites that were specifically fact-checked and found to be false or misleading. Finally, we also validate our preferred classification of untrustworthy websites in the Supplementary Information by merging the Grinberg et al.³⁵ designations with data on professional fact-checks. Reassuringly, we find that over 93% of articles from any of the three untrustworthy categories (black, red, and orange) were determined to be false, a proportion that is substantially higher than what we observe for articles from domains not in these categories.

We thus measure whether people visited one of the qualifying websites (a binary measure) as well as the fraction of people’s diet of news and information that came from these sites (a proportion)

both overall and for those that Grinberg et al.³⁵ could classify as liberal or conservative based on site exposure patterns (see their Supplementary Materials for details). Table S2 provides a list of the untrustworthy sites most frequently visited by respondents in our sample, their quality rating, and their slant classification. As in numerous other studies of this topic^{3,35,41,48}, we find ideological/partisan asymmetries in the set of untrustworthy news websites respondents visited and how much traffic they received. This asymmetry thus does not appear to be a measurement artifact. The reasons we observe such an asymmetry are beyond the scope of this study, however, and should be investigated in future research.

Additionally, we compute two key explanatory measures from web traffic data. First, following Guess (N.d.), we measure the overall ideological slant of respondents' online media consumption (or "information diet") by calculating the average slant of the websites they visit, which are based on differential sharing of websites by self-identified liberals versus conservatives on Facebook²¹. We then divide our sample into ten equally-sized groups (deciles) ranging from the 10% of respondents with the most liberal information diets to the 10% with the most conservative information diets. We choose deciles given the evidence of substantial within-party heterogeneity in selective exposure and other relevant political behaviors⁴⁹. These deciles correspond to meaningful differences in political attitudes and behavior. For instance, 89% of Americans in the most conservative decile of media consumption preferred Trump to Clinton (95% CI: 81.6%–96.8%). This group consumed a median of eight articles from Fox during the study period and the top quartile read between 136 and 1611.

Finally, we measure respondent consumption of "hard news" sites classified by a topic model as focusing on national news, politics, or world affairs²¹. We sum this measure with total

consumption of untrustworthy websites as defined above to construct the denominator for the estimated proportion of people's news and information diet coming from untrustworthy websites.

In our statistical analyses, we use OLS models due to their simplicity, ease of interpretation, and robustness to misspecification⁵⁰, but we demonstrate in the Supplementary Information that the conclusions for our binary exposure measures are consistent if estimated using a probit model.

We have not formally tested whether the assumptions of OLS have been met.

Data availability

Data files necessary to replicate the results in this article are available at the following Dataverse repository: <https://doi.org/10.7910/DVN/YLW1AZ>.

Code availability

R/Stata scripts that replicate the results in this article are available at the following Dataverse repository: <https://doi.org/10.7910/DVN/YLW1AZ>.

References

1. Silverman, C. This analysis shows how fake election news stories outperformed real news on Facebook (2016). BuzzFeed, November 16, 2016. Downloaded December 16, 2016 from https://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook?utm_term=.ohXvLeDzK#.cwwgb7EX0.
2. Silverman, C. & Singer-Vine, J. Most americans who see fake news believe it, new survey says (2016). BuzzFeed, December 6, 2016. Downloaded December 16, 2016 from https://www.buzzfeed.com/craigsilverman/fake-news-survey?utm_term=.lazQnopg3#.tbR2yvrL6.
3. Allcott, H. & Gentzkow, M. Social media and fake news in the 2016 election. *Journal of Economic Perspectives* **31**, 1–28 (2017).
4. Parkinson, H. J. Click and elect: how fake news helped Donald Trump win a real election (2016). *The Guardian*, November 14, 2016. Downloaded September 6, 2018 from <https://www.theguardian.com/commentisfree/2016/nov/14/fake-news-donald-trump-election-alt-right-social-media-tech-companies>.
5. Solon, O. Facebook's failure: did fake news and polarized politics get trump elected? (2016). *The Guardian*, November 10, 2016. Downloaded March 24, 2017 from <https://www.theguardian.com/technology/2016/nov/10/facebook-fake-news-election-conspiracy-theories>.

6. Blake, A. A new study suggests fake news might have won Donald Trump the 2016 election (2018). *Washington Post*, April 3, 2018. Downloaded September 6, 2018 from https://www.washingtonpost.com/news/the-fix/wp/2018/04/03/a-new-study-suggests-fake-news-might-have-won-donald-trump-the-2016-election/?utm_term=.9db09d55f95f.
7. Gunther, R., Beck, P. A. & Nisbet, E. C. “Fake news’ and the defection of 2012 Obama voters in the 2016 presidential election. *Electoral Studies* (2019).
8. Nelson, J. L. & Taneja, H. The small, disloyal fake news audience: The role of audience availability in fake news consumption. *New Media & Society* (2018).
9. Fourney, A., Racz, M. Z., Ranade, G., Mobius, M. & Horvitz, E. Geographic and temporal trends in fake news consumption during the 2016 US presidential election. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2071–2074 (2017).
10. Sunstein, C. R. *Republic.com* (Princeton University Press, 2001).
11. Pariser, E. *The filter bubble: How the new personalized web is changing what we read and how we think* (Penguin, 2011).
12. Stroud, N. J. Media use and political predispositions: Revisiting the concept of selective exposure. *Political Behavior* **30**, 341–366 (2008).
13. Hart, W. *et al.* Feeling validated versus being correct: a meta-analysis of selective exposure to information. *Psychological Bulletin* **135**, 555 (2009).

14. Iyengar, S. & Hahn, K. S. Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication* **59**, 19–39 (2009).
15. Iyengar, S., Hahn, K. S., Krosnick, J. A. & Walker, J. Selective exposure to campaign communication: The role of anticipated agreement and issue public membership. *Journal of Politics* **70**, 186–200 (2008).
16. Gentzkow, M. & Shapiro, J. M. Ideological segregation online and offline. *Quarterly Journal of Economics* **126**, 1799–1839 (2011).
17. Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A. & Bonneau, R. Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science* **26**, 1531–1542 (2015).
18. Flaxman, S. R., Goel, S. & Rao, J. M. Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly* **80**, 298–320 (2016).
19. Swire, B. & Ecker, U. K. Misinformation and its correction: Cognitive mechanisms and recommendations for mass communication. In Southwell, B. G., Thorson, E. A. & Sheble, L. (eds.) *Misinformation and Mass Audiences*, 195–211 (University of Texas Press, 2018).
20. Vosoughi, S., Roy, D. & Aral, S. The spread of true and false news online. *Science* **359**, 1146–1151 (2018).
21. Bakshy, E., Messing, S. & Adamic, L. A. Exposure to ideologically diverse news and opinion on facebook. *Science* **348**, 1130–1132 (2015).

22. Helmsley, J. The role of middle-level gatekeepers in the propagation and longevity of misinformation. In Southwell, B. G., Thorson, E. A. & Sheble, L. (eds.) *Misinformation and Mass Audiences*, 263–273 (University of Texas Press, 2018).
23. Gottfried, J. A., Hardy, B. W., Winneg, K. M. & Jamieson, K. H. Did fact checking matter in the 2012 presidential campaign? *American Behavioral Scientist* **57**, 1558–1567 (2013).
24. Nyhan, B. & Reifler, J. When corrections fail: The persistence of political misperceptions. *Political Behavior* **32**, 303–330 (2010).
25. Flynn, D., Nyhan, B. & Reifler, J. The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Advances in Political Psychology* **38**, 127–150 (2017).
26. Chan, M.-p. S., Jones, C. R., Hall Jamieson, K. & Albarracin, D. Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological science* **28**, 1531–1546 (2017).
27. Walter, N. & Murphy, S. T. How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs* **85**, 423–441 (2018).
28. Walter, N., Cohen, J., Holbert, R. L. & Morag, Y. Fact-checking: A meta-analysis of what works and for whom. *Political Communication* (2019).
29. Shin, J. & Thorson, K. Partisan selective sharing: The biased diffusion of fact-checking messages on social media. *Journal of Communication* **67**, 233–255 (2017).

30. Weeks, B. E. Emotions, partisanship, and misperceptions: How anger and anxiety moderate the effect of partisan bias on susceptibility to political misinformation. *Journal of Communication* **65**, 699–719 (2015).
31. Prior, M. *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections* (Cambridge University Press, 2007).
32. Diddi, A. & LaRose, R. Getting hooked on news: Uses and gratifications and the formation of news habits among college students in an internet environment. *Journal of Broadcasting & Electronic Media* **50**, 193–210 (2006).
33. Kalla, J. L. & Broockman, D. E. The minimal persuasive effects of campaign contact in general elections: Evidence from 49 field experiments. *American Political Science Review* **112**, 148–166 (2018).
34. Spenkuch, J. L. & Toniatti, D. Political advertising and election results. *The Quarterly Journal of Economics* (2018).
35. Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B. & Lazer, D. Fake news on twitter during the 2016 us presidential election. *Science* **363**, 374–378 (2019).
36. Pennycook, G. & Rand, D. G. Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition* (2018).
37. Frederick, S. Cognitive reflection and decision making. *Journal of Economic Perspectives* **19**, 25–42 (2005).

38. Bacevich, A. J. The real news we ignore at our peril (2018). *The American Conservative*, January 11, 2018.
39. Van den Putte, B., Yzer, M., Southwell, B. G., de Bruijn, G.-J. & Willemsen, M. C. Interpersonal communication as an indirect pathway for the effect of antismoking media content on smoking cessation. *Journal of Health Communication* **16**, 470–485 (2011).
40. Southwell, B. G. & Thorson, E. A. The prevalence, consequence, and remedy of misinformation in mass media systems. *Journal of Communication* **65**, 589–595 (2015).
41. Guess, A., Nagler, J. & Tucker, J. Less than you think: Prevalence and predictors of fake news dissemination on facebook. *Science Advances* **5**, eaau4586 (2019).
42. Graves, L. *Deciding what's true: The rise of political fact-checking in American journalism* (Columbia University Press, 2016).
43. Uscinski, J. E. & Butler, R. W. The epistemology of fact checking. *Critical Review* **25**, 162–180 (2013).
44. Lim, C. Checking how fact-checkers check. *Research & Politics* **5**, 2053168018786848 (2018).
45. Benkler, Y., Faris, R., Roberts, H. & Zuckerman, E. Study: Breitbart-led right-wing media ecosystem altered broader media agenda (2017). *Columbia Journalism Review*, March 3, 2017. Downloaded March 14, 2017 from <http://www.cjr.org/analysis/breitbart-media-trump-harvard-study.php>.

46. McDonald, M. P. 2016 November General Election Turnout Rates (2016). United States Election Project, September 5, 2018. Downloaded September 12, 2018 from <http://www.electproject.org/2016g>.
47. Lazer, D. M. *et al.* The science of fake news. *Science* **359**, 1094–1096 (2018).
48. Budak, C. What happened? The spread of fake news publisher content during the 2016 us presidential election. In *The World Wide Web Conference*, 139–150 (ACM, 2019).
49. Guess, A., Lyons, B., Nyhan, B. & Reifler, J. Avoiding the echo chamber about echo chambers: Why selective exposure to like-minded political news is less prevalent than you think (2018). Knight Foundation report, February 12, 2018. Downloaded October 23, 2018 from https://kf-site-production.s3.amazonaws.com/media_elements/files/000/000/133/original/Topos_KF_White-Paper_Nyhan_V1.pdf.
50. Angrist, J. D. & Pischke, J.-S. *Mostly Harmless Econometrics: An empiricist's companion* (Princeton University Press, 2009).

Acknowledgements

This project received funding support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 682758). We are grateful to the Poynter Institute, Knight Foundation, and American Press Institute for funding support; the funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We thank Dan Kahan and Craig Silverman for sharing data; Samantha Luks and Marissa Shih at YouGov for assistance with survey administration; and Kevin Arceneaux, Yochai Benkler, David Ciuk, Travis Coan, Lorien Jasny, Dan Kahan, David Lazer, John Leahy, Thomas Leeper, Adam Seth Levine, Ben Lyons, Cecilia Mo, Simon Munzert, and Spencer Piston for providing comments and feedback. We are also grateful to Brillian Bao, Jenna Barancik, Angela Cai, Jack Davidson, Kathryn Fuhs, Jose Burnes Garza, Guy Green, Jessica Lu, Annie Ma, Henry Parkhurst, Sarah Petroni, Morgan Sandhu, Priya Sankar, Amy Sun, Jacob Sweetow, Andrew Wolff, and Alexandra Woodruff for research assistance.

Author contributions

A.M.G., B.N., and J.R. designed the study, conducted the analysis, and drafted and revised the manuscript.

Competing interests

The authors declare no competing interests.

Tables

Table 1: Who chooses to visit untrustworthy news websites (behavioral data)

	<i>Untrustworthy conservative sites</i>						<i>Untrustworthy liberal sites</i>					
	Binary			% of info diet			Binary			% of info diet		
	<i>b</i>	s.e.	<i>p</i>	95% CI	<i>b</i>	s.e.	<i>p</i>	95% CI	<i>b</i>	s.e.	<i>p</i>	95% CI
Trump supporter	0.25	0.04	0.00	0.18, 0.32	0.09	0.01	0.00	0.07, 0.11	-0.17	0.03	0.00	-0.22, -0.12
Political knowledge	0.02	0.01	0.05	-0.00, 0.04	0.00	0.00	0.83	-0.00, 0.00	0.02	0.01	0.00	0.01, 0.03
Political interest	0.08	0.03	0.01	0.02, 0.13	0.02	0.01	0.00	0.01, 0.03	0.06	0.02	0.00	0.03, 0.09
College	0.02	0.03	0.55	-0.05, 0.08	-0.02	0.01	0.05	-0.03, -0.00	-0.00	0.03	0.98	-0.05, 0.05
Female	0.00	0.04	0.95	-0.07, 0.07	0.02	0.01	0.06	-0.00, 0.04	0.03	0.03	0.30	-0.02, 0.08
Nonwhite	-0.05	0.04	0.22	-0.14, 0.03	-0.01	0.01	0.35	-0.03, 0.01	-0.11	0.03	0.00	-0.18, -0.05
Age 30–44	0.03	0.08	0.69	-0.12, 0.18	0.00	0.01	0.64	-0.02, 0.02	0.04	0.05	0.34	-0.05, 0.13
Age 45–59	0.09	0.07	0.20	-0.05, 0.23	0.02	0.01	0.11	-0.00, 0.04	0.06	0.05	0.16	-0.03, 0.15
Age 60+	0.10	0.07	0.13	-0.03, 0.23	0.04	0.01	0.00	0.02, 0.06	0.07	0.04	0.10	-0.01, 0.16
Constant	-0.11	0.12	0.33	-0.34, 0.12	-0.08	0.02	0.00	-0.12, -0.04	-0.06	0.07	0.41	-0.19, 0.08
R ²	0.14				0.17				0.10			0.03
N	2167				2167				2167			2014

OLS models with survey weights (*p*-values two-sided). Online traffic statistics cover the October 7–November 14, 2016 period among YouGov Pulse panel members. The denominator for information consumption includes total exposure to those sites as well as the number of pages visited on websites classified as focusing on hard news topics (excluding Amazon, Twitter, and YouTube). Respondents who did not visit any of these sites are excluded from the information diet models.

Table 2: Correlates of Trump support and voter turnout in the 2016 election

	Trump support				Voter turnout			
	<i>b</i>	s.e.	<i>p</i>	95% CI	<i>b</i>	s.e.	<i>p</i>	95% CI
Clinton supporter (July)	-0.19	0.04	0.00	-0.26, -0.12	-0.05	0.04	0.25	-0.14, 0.04
Trump supporter (July)	0.69	0.04	0.00	0.61, 0.78	-0.06	0.04	0.12	-0.13, 0.02
Untrustworthy conservative website exposure (binary)	0.05	0.03	0.08	-0.01, 0.11	0.04	0.03	0.23	-0.02, 0.10
Liberal information diet	-0.05	0.03	0.09	-0.10, 0.01	0.04	0.04	0.33	-0.04, 0.12
Conservative information diet	0.01	0.02	0.78	-0.04, 0.05	-0.03	0.04	0.43	-0.10, 0.04
Political knowledge	-0.01	0.01	0.28	-0.02, 0.01	-0.01	0.01	0.36	-0.03, 0.01
Political interest	0.04	0.02	0.01	0.01, 0.07	0.02	0.02	0.47	-0.03, 0.06
College	-0.06	0.03	0.03	-0.11, -0.01	-0.02	0.03	0.44	-0.08, 0.03
Female	-0.01	0.02	0.54	-0.06, 0.03	-0.00	0.03	0.93	-0.06, 0.06
Nonwhite	-0.01	0.03	0.69	-0.08, 0.05	-0.02	0.04	0.59	-0.10, 0.05
Age 30–44	0.07	0.05	0.11	-0.02, 0.16	0.06	0.05	0.26	-0.04, 0.15
Age 45–59	0.08	0.04	0.04	0.00, 0.15	0.06	0.05	0.29	-0.05, 0.16
Age 60+	0.12	0.04	0.00	0.04, 0.21	0.03	0.04	0.46	-0.05, 0.12
Constant	0.07	0.09	0.42	-0.10, 0.25	0.24	0.08	0.00	0.08, 0.40
Controls for past turnout	No				Yes			
R^2	0.77				0.53			
N	1715				1715			

OLS models with survey weights (*p*-values two-sided). Online traffic statistics for October 7–21, 2016 among YouGov Pulse panel members. Trump support was measured in a survey conducted October 21–31, 2016. YouGov matched validated vote data from TargetSmart to survey respondents. “Controls for past turnout” are separate indicators for voting in the 2012 presidential primaries, the 2016 presidential primaries, and the 2012 general election (see the Supplementary Information for full results).

Table 3: Demographics of respondents

	ANES FTF	Full Pulse sample	Laptop/desktop data available	Mobile data available
Candidate preference				
Trump	0.32	0.33	0.34	0.27
Clinton	0.36	0.42	0.42	0.47
Other/DK/won't vote	0.32	0.24	0.24	0.25
Age				
18-29	0.21	0.19	0.16	0.27
30-44	0.22	0.25	0.22	0.34
45-59	0.29	0.28	0.30	0.25
60+	0.28	0.28	0.32	0.13
Race				
White	0.69	0.68	0.69	0.65
Black	0.12	0.12	0.12	0.11
Hispanic	0.12	0.13	0.12	0.16
Asian	0.03	0.03	0.02	0.03
Sex				
Male	0.47	0.48	0.48	0.45
Female	0.51	0.52	0.52	0.55
N	1181	3251	2525	660

Respondents are participants in the 2016 American National Election Studies pre-election face-to-face study (ANES FTF) and YouGov Pulse panel members. The columns of YouGov Pulse data are not mutually exclusive — the third and fourth columns represent differing subsets of the full Pulse sample. Estimates calculated using survey weights.