SOCIAL SCIENCES

Less than you think: Prevalence and predictors of fake news dissemination on Facebook

Andrew Guess[1]*, Jonathan Nagler[2], Joshua Tucker[2]

Copyright © 2019 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

So-called "fake news" has renewed concerns about the prevalence and effects of misinformation in political campaigns. Given the potential for widespread dissemination of this material, we examine the individual-level characteristics associated with sharing false articles during the 2016 U.S. presidential campaign. To do so, we uniquely link an original survey with respondents' sharing activity as recorded in Facebook profile data. First and foremost, we find that sharing this content was a relatively rare activity. Conservatives were more likely to share articles from fake news domains, which in 2016 were largely pro-Trump in orientation, than liberals or moderates. We also find a strong age effect, which persists after controlling for partisanship and ideology: On average, users over 65 shared nearly seven times as many articles from fake news domains as the youngest age group.

INTRODUCTION

One of the most discussed phenomena in the aftermath of the 2016 U.S. presidential election was the spread and possible influence of "fake news"—false or misleading content intentionally dressed up to look like news articles, often for the purpose of generating ad revenue. Scholars and commentators have raised concerns about the implications of fake news for the quality of democratic discourse, as well as the prevalence of misinformation more generally (*1*). Some have gone so far as to assert that such content had a persuasive impact that could have affected the election outcome, although the best evidence suggests that these claims are farfetched (*2*). While evidence is growing on the prevalence (*3*), believability (*2*), and resistance to corrections (*4*, *5*) of fake news during the 2016 campaign, less is known about the mechanisms behind its spread (*6*). Some of the earliest journalistic accounts of fake news highlighted its popularity on social media, especially Facebook (*7*). Visits to Facebook appear to be much more common than other platforms before visits to fake news articles in web consumption data, suggesting a powerful role for the social network (*3*), but what is the role of social transmission—in particular, social sharing—in the spread of this pernicious form of false political content? Here, we provide important new evidence complementing the small but growing body of literature on the fake news phenomenon.

Data and method

Our approach allows us to provide a comprehensive observational portrait of the individual-level characteristics related to posting articles from fake news–spreading domains to friends on social media. We link a representative online survey (N = 3500) to behavioral data on respondents' Facebook sharing history during the campaign, avoiding known biases in self-reports of online activity (*8*, *9*). Posts containing links to external websites are cross-referenced against lists of fake news publishers built by journalists and academics. Here, we mainly use measures constructed by reference to the list by Silverman (*7*), but in the Supplementary Materials, we show that the main results hold when alternate lists are used, such as that used by peer-reviewed studies (*2*).

Overall, sharing articles from fake news domains was a rare activity. We find some evidence that the most conservative users were more likely to share this content—the vast majority of which was pro-Trump in orientation—than were other Facebook users, although this is sensitive to coding and based on a small number of respondents. Our most robust finding is that the oldest Americans, especially those over 65, were more likely to share fake news to their Facebook friends. This is true even when holding other characteristics—including education, ideology, and partisanship—constant. No other demographic characteristic seems to have a consistent effect on sharing fake news, making our age finding that much more notable.

RESULTS

It is important to be clear about how rare this behavior is on social platforms: The vast majority of Facebook users in our data did not share any articles from fake news domains in 2016 at all (Fig. 1), and as the left panel shows, this is not because people generally do not share links: While 3.4% of respondents for whom we have Facebook profile data shared 10 or fewer links of any kind, 310 (26.1%) respondents shared 10 to 100 links during the period of data collection and 729 (61.3%) respondents shared 100 to 1000 links. Sharing of stories from fake news domains is a much rarer event than sharing links overall. The right panel of Fig. 1 reveals a large spike at 0, with a long tail that goes as far as 50 shares for a single Facebook user, and we see in Table 1 that over 90% of our respondents shared no stories from fake news domains. According to our main measure of fake news content, 8.5% of respondents for whom we have linked Facebook data shared at least one such article to their friends. Again referencing Fig. 1, among those who shared fake news to their friends, more were Republicans, both in absolute (38 Republican versus 17 Democratic respondents) and in relative (18.1% of Republicans versus 3.5% of Democrats in our sample) terms.

We further explore the factors that explain the variation in fake news sharing behavior. As shown in Fig. 2A, Republicans in our sample shared more stories from fake news domains than Democrats; moreover, self-described independents on average shared roughly as many as Republicans (0.506 and 0.480, respectively). A similar pattern is evident for ideology (Fig. 2C): Conservatives, especially those

[1]Department of Politics and Woodrow Wilson School, Princeton University, Fisher Hall, Princeton, NJ 08544, USA. [2]Wilf Family Department of Politics and Social Media and Political Participation (SMaPP) Lab, New York University, 19 West 4th Street, New York, NY 10012, USA.
*Corresponding author. Email: aguess@princeton.edu

Downloaded from http://advances.sciencemag.org/ on March 3, 2020

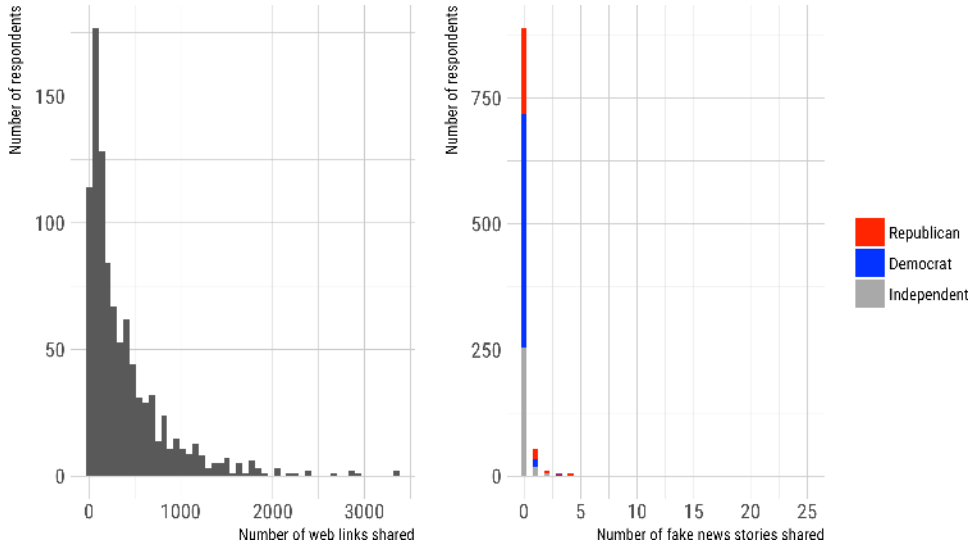

Fig. 1. Distribution of total and fake news shares. (**Left**) Histogram of the total number of links to articles on the web shared by respondents in the sample who identified as Democrats, Republicans, or independents. (**Right**) Stacked histogram of the number of fake news articles shared by respondents who identified as Democrats, Republicans, or independents using the measure derived from (*7*).

Table 1. Distribution of fake news shares.

0	1	2	3	4	5–10	11–50
1090 (91.5%)	63 (5.3%)	12 (1.0%)	8 (<1.0%)	5 (<1.0%)	9 (<1.0%)	4 (<1.0%)

identifying as "very conservative," shared the most articles from fake news domains. On average, a conservative respondent shared 0.75 such stories [95% confidence interval (CI), 0.537 to 0.969], and a very conservative respondent shared 1.0 (95% CI, 0.775 to 1.225). This is consistent with the pro-Trump slant of most fake news articles produced during the 2016 campaign, and of the tendency of respondents to share articles they agree with, and thus might not represent a greater tendency of conservatives to share fake news than liberals conditional on being exposed to it (*3*).

Figure 2D shows that, if anything, those who share the most content in general were less likely to share articles from fake news–spreading domains to their friends. Thus, it is not the case that what explains fake news sharing is simply that some respondents "will share anything." These data are consistent with the hypothesis that people who share many links are more familiar with what they are seeing and are able to distinguish fake news from real news. (We note that we have no measure as to whether or not respondents know that what they are sharing is fake news.) Turning to a key demographic characteristic of respondents, a notable finding in Fig. 2B is the clear association between age group and the average number of articles from fake news domains shared on Facebook. Those over 65 shared an average of 0.75 fake news articles (95% CI, 0.515 to 0.977), more than twice as many as those in the second-oldest age group (0.26 articles; 95% CI, 0.206 to 0.314). Of course, age is correlated with other characteristics, including political predispositions. Thus, we turn to a multivariate analysis to examine the marginal impact of individual characteristics.

Table 2 shows that the age effect remains statistically significant when controlling for ideology and other demographic attributes. The association is also robust to controlling for party, as the various alternative specifications provided in the Supplementary Materials illustrate. In column 2, the coefficient on "Age: over 65" implies that being in the oldest age group was associated with sharing nearly seven times as many articles from fake news domains on Facebook as those in the youngest age group, or about 2.3 times as many as those in the next-oldest age group, holding the effect of ideology, education, and the total number of web links shared constant ($e^{1.9} \approx 6.69$, $e^{1.9-1.079} \approx 2.27$). This association is also found in the specifications using the alternate peer-reviewed measure (*2*) as a dependent variable in columns 3 and 4, with those over 65 sharing between three and four times as many fake news links as those in the youngest age group.

Aside from the overall rarity of the practice, our most robust and consistent finding is that older Americans were more likely to share articles from fake news domains. This relationship holds even when we condition on other factors, such as education, party affiliation, ideological self-placement, and overall posting activity. It is robust to a wide range of strategies for measuring fake news (see Materials and Methods). Further, none of the other demographics variables in our model—sex, race, education, and income—have anywhere close to a robust predictive effect on sharing fake news. We subject our findings to a battery of robustness tests in the Supplementary Materials. Among them, we show that model specification, other predictors such as political knowledge, and distributional assumptions about the dependent variable do not appear to be driving our results (tables S1 to S8 and S13). Last, we show in table S14 that, when we try to explain patterns of hard news sharing behavior using the same approach, the predictors are more varied and do not include age.

Downloaded from http://advances.sciencemag.org/ on March 3, 2020

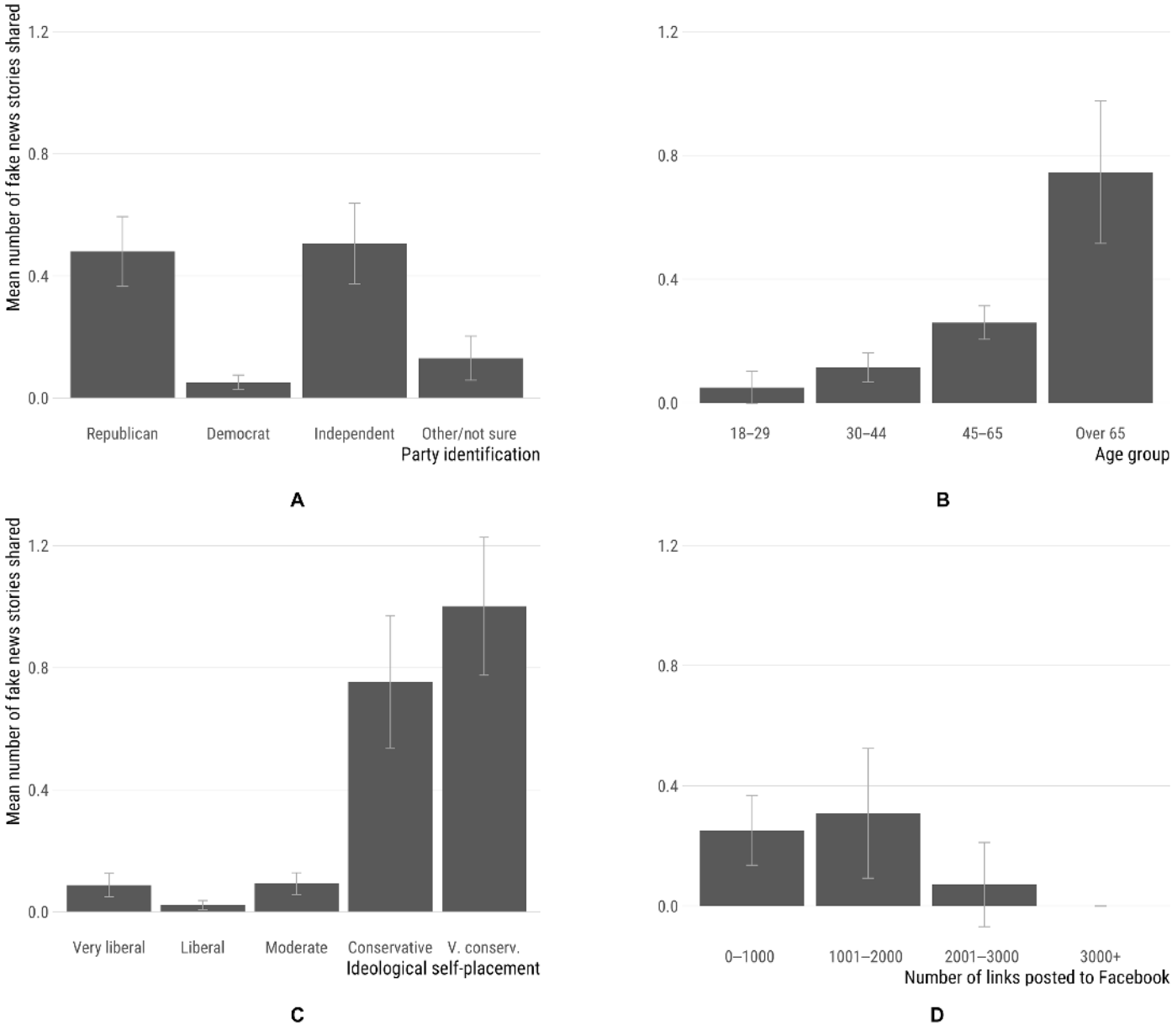

Fig. 2. Average number of fake news shares (and 95% CIs) using the list of domains derived from (*7*). (**A**) Party identification, (**B**) age group, (**C**) ideological self-placement, and (**D**) overall number of Facebook wall posts. Proportions adjusted to account for sample-matching weights derived from the third wave of the SMaPP YouGov panel survey.

DISCUSSION

Using unique behavioral data on Facebook activity linked to individual-level survey data, we find, first, that sharing fake news was quite rare during the 2016 U.S. election campaign. This is important context given the prominence of fake news in post-election narratives about the role of social media disinformation campaigns. Aside from the relatively low prevalence, we document that both ideology and age were associated with that sharing activity. Given the overwhelming pro-Trump orientation in both the supply and consumption of fake news during that period, including via social pathways on Facebook (*3*), the finding that more conservative respondents were more likely to share articles from fake news–spreading domains is perhaps expected. More puzzling is the independent role of age: Holding constant ideology, party identification, or both, respondents in each age category were more likely to share fake news than respondents in the next-youngest group, and the gap in the rate of fake news sharing between those in our oldest category (over 65) and youngest category is large and notable.

These findings pose a challenge and an opportunity for social scientists. Political scientists tend to favor explanations based on stable, deeply held partisan or ideological predispositions (*10*, *11*). The predictive power of demographic traits evaporates when subjected to multiple regression analyses that control for other characteristics correlated with those demographics. Yet, when an empirical relationship

Downloaded from http://advances.sciencemag.org/ on March 3, 2020

Table 2. Determinants of fake news sharing on Facebook. Quasi-Poisson models with YouGov's sample-matching weights applied. Dependent variables are counts of fake news articles shared using measures derived from (*7*) (columns 1 and 2) and (*2*) (columns 3 and 4). The reference category for ideology is "Not sure." "Number of links shared" refers to the number of Facebook posts by each respondent that includes a link to an external URL. A&G, Allcott and Gentzkow.

	Number of stories shared		Number of stories shared (A&G)	
	(1)	(2)	(3)	(4)
Very liberal	0.487	0.387	1.634*	1.485*
	(1.238)	(1.209)	(0.876)	(0.800)
Liberal	−1.127	−1.141	0.873	0.812
	(1.439)	(1.404)	(0.886)	(0.809)
Moderate	0.333	0.392	0.748	0.824
	(1.186)	(1.157)	(0.875)	(0.799)
Conservative	2.187*	2.248**	1.736**	1.800**
	(1.155)	(1.128)	(0.868)	(0.794)
Very conservative	2.366**	2.297**	2.231**	2.087***
	(1.158)	(1.132)	(0.869)	(0.795)
Age: 30–44	0.772	0.742	0.253	0.172
	(0.811)	(0.791)	(0.390)	(0.356)
Age: 45–65	1.136	1.079	0.602*	0.488
	(0.765)	(0.746)	(0.359)	(0.328)
Age: over 65	2.052***	1.900**	1.389***	1.152***
	(0.766)	(0.750)	(0.362)	(0.333)
Female	−0.114	0.008	−0.329**	−0.199
	(0.217)	(0.219)	(0.155)	(0.146)
Black	−0.880	−0.806	−0.609	−0.536
	(0.754)	(0.736)	(0.400)	(0.366)
Education	−0.085	−0.091	−0.021	−0.021
	(0.081)	(0.081)	(0.055)	(0.052)
Income	−0.007	−0.007	0.003	0.003
	(0.008)	(0.008)	(0.004)	(0.003)
Number of links shared		0.001***		0.001***
		(0.0002)		(0.0001)
Constant	−3.416**	−3.635***	−1.201	−1.502*
	(1.379)	(1.348)	(0.931)	(0.851)
N	1041	1040	1041	1040

*$P < 0.1$ **$P < 0.05$ ***$P < 0.01$.

such as the one documented here emerges, we are challenged to view demographic traits not as controls to be ignored but as central explanatory factors above and beyond the constructs standard in the literature (*12*). This is especially the case with age, as the largest generation in America enters retirement at a time of sweeping demographic and technological change. Below, we suggest possible avenues for further research incorporating insights from multiple disciplines.

Given the general lack of attention paid to the oldest generations in the study of political behavior thus far, more research is needed to better understand and contextualize the interaction of age and online political content. Two potential explanations warrant further investigation. First, following research in sociology and media studies, it is possible that an entire cohort of Americans, now in their 60s and beyond, lacks the level of digital media literacy necessary to reliably determine the trustworthiness of news encountered online (*13*, *14*). There is a well-established research literature on media literacy and its importance for navigating new media technologies (*15*). Building on existing work (*16*, *17*), researchers should further develop competency-based measures of digital media literacy that encompass the kinds of skills needed to identify and avoid dubious content designed to maximize engagement. Research on age and digital media literacy often focuses on youth skills acquisition and the divide between "digital natives" and "digital immigrants" (*18*), but our results suggest renewed focus on the oldest age cohorts.

Within this cohort, lower levels of digital literacy could be compounded by the tendency to use social endorsements as credibility

Downloaded from http://advances.sciencemag.org/ on March 3, 2020

cues (*19*). If true, this would imply a growing impact as more Americans from older age groups join online social communities. A second possibility, drawn from cognitive and social psychology, suggests a general effect of aging on memory. Under this account, memory deteriorates with age in a way that particularly undermines resistance to "illusions of truth" and other effects related to belief persistence and the availability heuristic, especially in relation to source cues (*20*–*22*). The severity of these effects would theoretically increase with the complexity of the information environment and the prevalence of misinformation.

We cannot definitively rule out the possibility that there is an omitted variable biasing our estimates, although we have included controls for many individual-level characteristics theoretically related to acceptance of misinformation and willingness to share content online. Even if our models are correctly specified, we use observational data that cannot provide causal evidence on the determinants of fake news sharing. This study takes advantage of a novel and powerful new dataset combining survey responses and digital trace data that overcomes well-known biases in sample selection and self-reports of online behavior (*8*, *9*). However, we are still limited in our ability to collect these data unobtrusively. Despite our high response rate, half of our respondents with Facebook accounts opted not to share their profile data with us. Any inferences are therefore limited insofar as the likelihood of sharing data is correlated with other characteristics of interest.

In addition, while our approach allows for enhanced measurement of online sharing behavior, we lack data on the composition of respondents' Facebook News Feeds. It is possible, for instance, that very conservative Facebook users were exposed to more fake news articles in their networks and that the patterns we observe are not due to differential willingness to believe or share false content across the political spectrum. While other evidence suggests the limits of the "echo chambers" narrative (*23*), we cannot rule out this possibility. Similarly, it may be the case that the composition of older Facebook users' News Feeds differs in systematically important ways from that of younger users; while we lack the data in the present work to test this proposition, future research with access to these data could prove illuminating. These concerns aside, the evidence we have presented is strongly suggestive of an emerging relationship between not only ideological affiliation but also age and the still-rare tendency to spread misinformation to one's friends on social media platforms. If the association with age holds in future studies, there are a whole host of questions about the causes, mechanisms, and ramifications of the relationship that researchers should explore. First, how much of the effect can be attributed to lack of digital or media literacy as opposed to explanations rooted in cognitive abilities and memory? Answering this question will require developing measures of digital literacy that can be behaviorally validated. Second, what is the role of the (currently) unobserved News Feed and social network environment on people's tendency to see, believe, and spread dubious content? How are consumption and spreading, if at all, related? How does social trust over networks mediate the relationship between age and the sharing of misinformation?

Last, if media literacy is a key explanatory factor, then what are the interventions that could effectively increase people's ability to discern information quality in a complex, high-choice media environment replete with contradictory social and political cues? Both theory and existing curricula could serve as the basis of rigorously controlled evaluations, both online and in the classroom, which could then help to inform educational efforts targeted at people in different age groups and with varying levels of technological skill. These efforts leave open the possibility that simple interventions, perhaps even built into online social environments, could reduce the spread of misinformation by those most vulnerable to deceptive content. Developing these innovations would be further aided by increased

Table 3. Comparison of samples. FB, Facebook.

	Full sample	Sample 2*	Sample 3†	*P*‡	Sample 4§
% Democrat	31	32	40	0.17	40
Mean ideology (five-point)	2.98	2.89	2.76	0.01	2.75
% Vote intention (Clinton)	36	37	47	0.07	47
% Voted in 2016 general	0.59	0.59	0.63	0.01	0.63
% Knowledge (0–2)	2.05	2.04	2.13	0.03	2.13
Mean age	51	49	49	0.16	49
% High school or less	23	20	22	0.17	22
Self-reported					
% Post to FB several times/day		26	28	0.28	28
% Look at FB often		65	67	0.42	68
N	3500	2711	1331		1191

*Column 2 summarizes characteristics of respondents who said in the survey that they have a Facebook account (i.e., they selected "Facebook" from the list of response options to the question "Do you have accounts on any of the following social media sites?"). †Column 3 subsets to respondents (regardless of their answer in the previous question) who consented to share Facebook profile information with the researchers. ‡*P* values are computed from *t* tests of the difference in means between the sample of respondents who reported having a Facebook account and those who consented to provide access to their profile data. §The final column subsets to those who shared any Facebook data at all that we were able to link back to the survey.

Downloaded from http://advances.sciencemag.org/ on March 3, 2020

cooperation between academic researchers and the platforms themselves (*24*).

MATERIALS AND METHODS

Survey data

We designed and conducted a panel survey (fielded by online polling firm YouGov) during the 2016 U.S. presidential election to understand how social media use affects the ways that people learn about politics during a campaign. In addition to including a rich battery of individual-level covariates describing media use and social media use, we were able to match many respondents to data on their actual Facebook behavior (see below). The survey had three waves. Wave 1 was fielded 9 April to 1 May 2016 (3500 respondents), wave 2 was fielded 9 September to 9 October 2016 (2635 respondents), and wave 3 was fielded 25 October to 7 November 2016 (2628 respondents).

Facebook profile data

We were able to obtain private Facebook profile data from a substantial subset of our survey respondents. Starting 16 November 2016, YouGov e-mailed all respondents with a request to temporarily share information from their Facebook profiles with us. That request read, in part: "We are interested in the news people have read and shared on Facebook this year. To save time and to get the most accurate information, with your permission we can learn this directly from Facebook. Facebook has agreed to help this way. Of course, we would keep this information confidential, just like everything else you tell us."

Those who consented to do so were asked to provide access to a Facebook web application and to specifically check which types of information they were willing to share with us: fields from their public profile, including religious and political views; their own timeline posts, including external links; and "likes" of pages. We did not have access to the content of people's News Feeds or information about their friends. Respondents read a privacy statement that informed them that they could deactivate the application at any time and that we would not share any personally identifying information. The app provided access for up to 2 months after respondents who chose to share data agreed to do so. This data collection was approved by the New York University Institutional Review Board (IRB-12-9058 and IRB-FY2017–150).

Of 3500 initial respondents in wave 1, 1331 (38%) agreed to share Facebook profile data with us. The proportion rises (49.1%) when we consider that only 2711 of our respondents said that they use Facebook at all. We were successfully able to link profile data from 1191 of survey respondents, leaving us with approximately 44% of Facebook users in our sample. (See the "Sample details" section for a comparison of sample characteristics on various demographic and behavioral dimensions. Linked respondents were somewhat more knowledgeable and engaged in politics than those who did not share data.) For the purposes of this analysis, we parsed the raw Facebook profile data and identified the domains of any links posted by respondents to their own timelines.

We did not have access to posts that respondents deleted before consenting to temporarily share their data with us. It is theoretically possible that some respondents posted fake news articles to their profiles and then deleted them before we had the opportunity to collect the data. To the extent that this activity reflects second-guessing or awareness of how fake news posting is perceived by social connections, we interpret the sharing data that we were able to gather as genuine—posts that our respondents did not feel compelled to remove at a later time. There may be an additional concern that some types of people were more likely to delete fake news articles that they posted, leading us to biased inferences. However, to the extent that these characteristics are negatively correlated with the characteristics that predict posting in the first place, such deletion activity (which is likely very rare) should reduce noise in the data that would otherwise be generated by split-second sharing decisions that are immediately retracted.

Defining fake news

The term fake news can be used to refer to a variety of different phenomena. Here, we largely adopted the use suggested in (*25*) of knowingly false or misleading content created largely for the purpose of generating ad revenue. Given the difficulty of establishing a commonly accepted ground-truth standard for what constitutes fake news, our approach was to build on the work of both journalists and academics who worked to document the prevalence of this content over the course of the 2016 election campaign. In particular, we used a list of fake news domains assembled by Craig Silverman of BuzzFeed News, the primary journalist covering the phenomenon as it developed (*7*). As a robustness check, we constructed alternate measures using a list curated by Allcott and Gentzkow (*2*), who combined multiple sources across the political spectrum (including some used by Silverman) to generate a list of fake news stories specifically debunked by fact-checking organizations.

The Silverman list is based on the most-shared web domains during the election campaign as determined by the analytics service BuzzSumo. Silverman and his team followed up their initial results with in-depth reporting to confirm whether a domain appeared to have the hallmark features of a fake news site: lacking a contact page, featuring a high proportion of syndicated content, being relatively new, etc. We took this list and removed all domains classified as "hard news" via the supervised learning technique used by Bakshy *et al.* (*23*) to focus specifically on fake news domains rather than the more contested category of "hyperpartisan" sites (such as Breitbart). (The authors used section identifiers in article URLs shared on Facebook that are associated with hard news—"world," "usnews," etc.—to train a machine learning classifier on text features. They ultimately produced a list of 495 domains with both mainstream and partisan websites that produce and engage with current affairs.) The resulting list contains 21 mostly pro-Trump domains, including well-known purveyors such as abcnews.com.co, the Denver Guardian, and Ending the Fed. In analyses using this list, we counted any article from one of these domains as a fake news share. (See below for details on these coding procedures and a list of domains in what we refer to as our main BuzzFeed-based list.)

The Allcott and Gentzkow list begins with 948 fact checks of false stories from the campaign. We retrieved the domains of the publishers originating the claims and again removed all hard news domains as described above. Then, we coded any article from this set of domains as a fake news article. For robustness, in table S9, we used only exact URL matches to any of the 948 entries in the Allcott and Gentzkow list as a more restrictive definition of fake news, but one that does not require assuming that every article from a "fake news domain" should be coded as fake news. Since the list contains the researchers' manual coding of the slant of each article, we also presented models analyzing pro-Trump and pro-Clinton fake news sharing activity only.

Downloaded from http://advances.sciencemag.org/ on March 3, 2020

Additional lists

In addition to these primary measures, we report (below) analyses using three supplementary collections of fake news articles produced after the election. Two lists were also produced by Silverman and his team at BuzzFeed (*26*), and the third is a crowdsourced effort headed by Melissa Zimdars of Merrimack College. Our key results are essentially invariant to whatever measure of fake news we use.

Modeling strategy and covariates

We aggregated all shares to the individual respondent level so that our dependent variables are counts (i.e., number of fake news stories shared). To account for this feature of the data, as well as the highly skewed distribution of the counts, we primarily used Poisson or quasi-Poisson regressions to model the determinants of Facebook sharing behavior. We conducted dispersion tests on the count data and used quasi-Poisson models if the null hypothesis of no dispersion is rejected. Below, we included negative binomial and Ordinary Least Squares (OLS) regressions to show that our results are generally not sensitive to model choice. All models applied weights from YouGov to adjust for selection into the sample. We specifically used sample-matching weights produced for the third wave of the survey, which was closest to the Facebook encouragement sent to respondents (*27*). (Results also do not appear to be sensitive to the use of weights.)

We included a mix of relevant sociodemographic and political variables as predictors. These include age (reference category, 18 to 29), race, gender, family income, and educational attainment. In all models, we included either five-point ideological self-placement, three-point party identification, or both. Since these variables were correlated ($r = 0.31$), we addressed possible multicollinearity via transparency—we provided our main results all three ways. (In all models, the reference case for party identification and ideology is "Not sure." Specifications including additional racial/ethnic categories are statistically and substantively unchanged; available from the authors.) Last, we included a measure of the total number of wall posts including a URL. This is intended to capture the overall level of respondents' Facebook link-sharing activity regardless of political content or verifiability.

Details on main BuzzFeed-based list

Our BuzzFeed-based list began with 30 domains identified by that site's reporting as purveyors of intentionally false election-related stories generating the most Facebook engagement. To do this, the journalists, led by Craig Silverman, used keywords and existing lists combined with the analytics service BuzzSumo. To ensure that our analysis stayed clear of websites that could be construed as partisan or hyperpartisan (rather than intentionally or systematically factually inaccurate), we additionally filtered out domains identified as hard news by a supervised learning classifier developed by Bakshy *et al.* (*23*). The nearly 500 hard news domains encompass a wide range of news and opinion websites, both mainstream and niche. The classifier was trained on the text features of roughly 7 million web pages shared on Facebook over a 6-month period by U.S. users, with training labels for hard and soft news generated using bootstrapped keyword searches on the URLs. Once matches to this list of hard news domains were removed (for example, Breitbart.com), we were left with 21 domains, shown below.

(1) usanewsflash.com
(2) abcnews.com.co
(3) denverguardian.com
(4) rickwells.us
(5) truepundit.com
(6) redstatewatcher.com
(7) worldpoliticus.com
(8) subjectpolitics.com
(9) conservativestate.com
(10) conservativedailypost.com
(11) libertywritersnews.com
(12) worldnewsdailyreport.com
(13) endingthefed.com
(14) donaldtrumpnews.co.
(15) yesimright.com
(16) burrardstreetjournal.com
(17) bizstandardnews.com
(18) everynewshere.com
(19) departed.co.
(20) americanmilitarynews.com
(21) tmzhiphop.com

Sample details

Table 3 reports raw proportions of characteristics and self-reported behaviors across various sample definitions. Knowledge ranged from 0 to 4 and was constructed from a grid of questions about the majority party in the House and Senate, in addition to questions about whether the uninsured rate and earnings had increased over the course of 2016. Voter turnout was verified by our survey provider, which matches individual respondents to the TargetSmart voter file.

A potential concern about sample-selection bias is that those who consented to share Facebook data were different from the rest of the sample along some dimension that is also related to our outcome of interest (fake news sharing behavior). Table 3 suggests that, at least on closely related observable characteristics, the subgroup for which we have profile data is a valid cross section of the overall sample. In particular, frequent self-reported Facebook sharing activity is roughly indistinguishable between those who report having a Facebook account and those who provided access ($P = 0.28$). The samples are also comparable on age, frequency of looking at Facebook, and vote intention. Those who shared data were slightly more liberal on average ($P = 0.01$), but we controlled for this in our models and we expected differences between the samples to arise due to chance alone. Last, it may not be surprising that those who provided access to profile data were also more likely to participate in elections, as measured by verified voter turnout in the 2016 general election. We see this somewhat heightened political engagement in the Facebook subsample as important to note, and we accounted for the effects of this difference when we controlled for overall posting activity in our models.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/5/1/eaau4586/DC1

Tables S1–S13. Determinants of fake news sharing on Facebook (alternate specification).

Table S14. Determinants of hard news sharing on Facebook.

Fig. S1. Average number of fake news articles shared by age group (with 95% confidence intervals), using the URL-level measure derived from (*2*).

REFERENCES AND NOTES

1. D. M. J. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, M. Schudson, S. A. Sloman, C. R. Sunstein, E. A. Thorson, D. J. Watts, J. L. Zittrain, The science of fake news. *Science* **359**, 1094–1096 (2018).

Downloaded from http://advances.sciencemag.org/ on March 3, 2020

2. H. Allcott, M. Gentzkow, Social media and fake news in the 2016 election. *J. Econ. Perspect.* **31**, 211–236 (2017).
3. A. Guess, B. Nyhan, J. Reifler, "Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 U.S. presidential campaign" (2016); http://www-personal.umich.edu/~bnyhan/fake-news-2016.pdf.
4. G. Pennycook, T. D. Cannon, D. G. Rand, Prior exposure increases perceived accuracy of fake news. *J. Exp. Psychol. Gen.*, in press.
5. E. Porter, T. J. Wood, D. Kirby, Sex trafficking, Russian infiltration, birth certificates, and pedophilia: A survey experiment correcting fake news. *J. Exp. Polit. Sci.* **5**, 159–164 (2018).
6. S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online. *Science* **359**, 1146–1151 (2018).
7. C. Silverman, "This analysis shows how fake election news stories outperformed real news on facebook" (BuzzFeed, 2016); www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook.
8. A. M. Guess, Measure for measure: An experimental test of online political media exposure. *Polit. Anal.* **23**, 59–75 (2015).
9. A. Guess, K. Munger, J. Nagler, J. Tucker, How accurate are survey responses on social media and politics? *Polit. Commun.* 1–18 (2018).
10. A. Campbell, P. E. Converse, W. E. Miller, *The American Voter* (Wiley, 1960).
11. D. Green, B. Palmquist, E. Schickler, *Partisan Hearts and Minds* (Yale Univ. Press, 2002).
12. C. A. Klofstad, The sociological bases of political preferences and behavior, in *The Future of Political Science: 100 Perspectives* (Routledge, 2009).
13. B. Schäffer, The digital literacy of seniors. *Res. Comp. Int. Educ.* **2**, 29–42 (2007).
14. B. B. Neves, F. Amaro, Too old for technology? How the elderly of lisbon use and perceive ICT. *J. Community Inform.* **8**, 1–12 (2012).
15. S. Livingstone, Media literacy and the challenge of new information and communication technologies. *Commun. Rev.* **7**, 3–14 (2004).
16. E. Hargittai, Survey measures of web-oriented digital literacy. *Soc. Sci. Comput. Rev.* **23**, 371–379 (2005).
17. E. Hargittai, An update on survey measures of web-oriented digital literacy. *Soc. Sci. Comput. Rev.* **27**, 130–137 (2009).
18. S. Bennett, K. Maton, L. Kervin, The 'digital natives' debate: A critical review of the evidence. *Br. J. Educ. Technol.* **39**, 775–786 (2008).
19. S. Messing, S. J. Westwood, Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. *Commun. Res.* **41**, 1042–1063 (2012).
20. E. L. Glisky, S. R. Rubin, P. S. R. Davidson, Source memory in older adults: An encoding or retrieval problem? *J. Exp. Psychol. Learn. Mem. Cogn.* **27**, 1131–1146 (2001).
21. I. Skurnik, C. Yoon, D. C. Park, N. Schwarz, How warnings about false claims become recommendations. *J. Consum. Res.* **31**, 713–724 (2005).
22. B. Swire, U. K. H. Ecker, S. Lewandowsky, The role of familiarity in correcting inaccurate information. *J. Exp. Psychol. Learn. Mem. Cogn.* **43**, 1948–1961 (2017).
23. E. Bakshy, S. Messing, L. A. Adamic, Exposure to ideologically diverse news and opinion on facebook. *Science* **348**, 1130–1132 (2015).
24. G. King, N. Persily, "A new model for industry-academic partnerships" (2018); https://gking.harvard.edu/files/gking/files/partnerships.pdf.
25. J. A. Tucker, A. Guess, P. Barberá, C. Vaccari, A. Siegel, S. Sanovich, D. Stukal, B. Nyhan, *Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature* (William and Flora Hewlett Foundation, 2018).
26. C. Silverman, Here are 50 of the biggest fake news hits on facebook from 2016 (BuzzFeed, 2016); www.buzzfeed.com/craigsilverman/top-fake-news-of-2016.
27. D. Rivers, *Sample Matching: Representative Sampling from Internet Panels* (Polimetrix White Paper Series, 2006).

Acknowledgments: This study has been approved by the New York University Institutional Review Board (IRB-12-9058 and IRB-FY2017-150). We thank S. Luks at YouGov for facilitating survey data collection and J. Ronen for development and technical assistance. **Funding:** This research was supported by the INSPIRE program of the NSF (Award SES-1248077). **Author contributions:** A.G., J.N., and J.T. designed the research. J.N. and J.T. oversaw data collection. A.G. analyzed the data and produced the figures. A.G. drafted the paper, and all authors contributed to revisions. **Competing interests:** J.T. received a small fee from Facebook to compensate him for time spent in organizing a 1-day conference for approximately 30 academic researchers and a dozen Facebook product managers and data scientists that was held at NYU in the summer of 2017 to discuss research related to civic engagement. He did not provide any consulting services or any advice to Facebook as part of this arrangement. He also participated in other unpaid meetings of academics researchers and Facebook product managers and data scientists to discuss topics related to social media and politics. All of these events occurred after the data collection for the current paper. All other authors declare that they have no competing interests. **Data and materials availability:** Counts and survey variables at the respondent level needed to replicate all tables and figures in the paper are available at doi.org/10.7910/DVN/IKDTPZ. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 12 June 2018
Accepted 30 November 2018
Published 9 January 2019
10.1126/sciadv.aau4586

Citation: A. Guess, J. Nagler, J. Tucker, Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Sci. Adv.* **5**, eaau4586 (2019).

Downloaded from http://advances.sciencemag.org/ on March 3, 2020

ScienceAdvances

Less than you think: Prevalence and predictors of fake news dissemination on Facebook

Andrew Guess, Jonathan Nagler and Joshua Tucker

Sci Adv **5** (1), eaau4586.
DOI: 10.1126/sciadv.aau4586

ARTICLE TOOLS http://advances.sciencemag.org/content/5/1/eaau4586

SUPPLEMENTARY MATERIALS http://advances.sciencemag.org/content/suppl/2019/01/07/5.1.eaau4586.DC1

REFERENCES This article cites 16 articles, 3 of which you can access for free
http://advances.sciencemag.org/content/5/1/eaau4586#BIBL

PERMISSIONS http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service

Science Advances (ISSN 2375-2548) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.

Copyright © 2019 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

Downloaded from http://advances.sciencemag.org/ on March 3, 2020