Cracking Open the News Feed: Exploring What U.S. Facebook Users See and Share with Large-Scale Platform Data

Authors' names withheld

Abstract

In this study, we analyze for the first time newly available engagement data covering millions of web links shared on Facebook to describe how and by which categories of U.S. users different types of news are seen and shared on the platform. We focus on articles from "fake news" publishers, credible news sources, purveyors of clickbait, and news specifically about politics, which we identify through a combination of curated lists and supervised classifiers. Our results support recent findings that more fake news is shared by older users and conservatives and that both viewing and sharing patterns suggest a preference for ideologically congenial misinformation. We also find that fake news articles related to politics are more popular among older Americans than other types, while the youngest users share relatively more articles with clickbait headlines. Across the platform, however, articles from credible news sources are shared over 5 times more often and viewed over 7 times more often than articles from untrustworthy sources. These findings offer important context for researchers studying the spread and consumption of information — including misinformation — on social media.

Since the rise of so-called "fake news" outlets on social media was linked to populist political victories, including that of Donald Trump in the United States, intense scholarly and popular interest has focused on the kinds of information — and misinformation — people encounter on their social news feeds, and where it comes from (Lazer et al. 2018; Allcott, Gentzkow and Yu 2019; Grinberg et al. 2019; Guess, Nyhan and Reifler 2020). The stakes are high: If misleading and inflammatory content is widespread, social media present a challenge for democracies that both rely on informed debate and seek to protect free expression.

One of the major responses to these developments has been the refinement of a large research literature focusing on informational, skills-based and other interventions that could help to counteract online misinformation (e.g., Wittenberg and Berinsky 2020; Guess et al. 2020; Pennycook et al. 2020), though scholars continue to debate how to share best practices and collaborate with platforms while maintaining independence (Amazeen et al. 2020). Regardless, any sustainable response — whether from social platforms or government policy — depends not only on credible research on the impact of these strategies, but on a solid descriptive foundation about the ways in which people encounter, share, and interact with online misinformation in the first place. This is critical for assessing the relative costs and benefits of proposed alternatives as well as for effectively tailoring interventions to relevant contexts or contingencies.

Fortunately, the past four years have already yielded some clear answers. In contrast to popular narratives, researchers have consistently found that fake news consumption (through website visits on desktop and mobile devices) represents a small share of most people's news diets, though it is more concentrated among those with the most conservative-leaning media diets (Guess, Nyhan and Reifler 2020; Allen et al. 2020). Similarly, the likelihood of both encountering and sharing links from fake news sources on Twitter is very low (Grinberg et al. 2019). Similar results have been found in terms of sharing links from fake news sources on Facebook (Guess, Nagler and Tucker 2019), but up until now there has not been individual-level data available to researchers working outside the platform on exposure to news (whether untrustworthy or not) within Facebook's News Feed, though web visit data suggests that it was the most important source of referrals to fake news websites during the 2016 U.S. campaign (Guess, Nyhan and Reifler 2020). Thus, in a relatively short amount of time, the accumulated evidence has shed light on key patterns in the consumption and dissemination of online misinformation. (For a review of recent evidence from the United States and elsewhere, see Guess and Lyons 2020.)

Despite these advances, key questions remain about the proliferation of untrustworthy content on social media. While much research to date has focused on individual-level exposure and engagement with "fake" or misleading content, the question of the precise boundaries between so-called "fake news" and adjacent categories has been neglected. This is partially because these distinctions are often subtle (Tucker et al. 2018), and it is difficult to produce reliable classifications from datasets of limited size. Yet some of the outstanding puzzles about online misinformation — such as the relationship between consumption and sharing, and the role of age-related differences in explaining these patterns — may depend on disaggregating these categories of content.

The best evidence to date leverages important advances in the use of digital trace data to understand the correlates of visits to untrustworthy news websites (e.g., Stier et al. 2020). Even so, researchers have been unable to assemble a complete picture of the "fake news" ecosystem due to a conspicuously missing piece: data on what users encounter on social platforms themselves, particularly on Facebook. This inability to observe activity on the Facebook News Feed has placed a sharp limit on research focusing on incidental exposure to fake news via snippets and headlines that may draw users' attention, if not motivate a click (see Anspach and Carlson 2018).

In addition to this major shortcoming, two additional challenges continue to plague existing research. First, most studies rely on conventional-sized samples (for exceptions, see Grinberg et al. 2019 and Allen et al. 2020), which are not ideally suited to accurately estimating the prevalence of habits and behaviors that are rare on average. Second, since the relevant digital trace data are costly and difficult to obtain, the timeline of gathering, analyzing, and publishing data about the role of social media means that the rate of change of target phenomena in the real world is vastly greater than the pace at which scientific evidence about those phenomena accumulates (for a discussion, see Munger 2018).

Until very recently, large-scale evidence that speaks directly to these concerns has not been available for a variety of technical and privacy-related reasons (Tucker et al. 2018). Behavioral data of the kind needed to describe the varied experiences of hundreds of millions of social media users presents staggering new computational, privacy, and ethical challenges for social scientists working with proprietary data (Lazer et al. 2020). Fortunately, the Social Science One partnership[1] has emerged as a way to facilitate independent scholars' access to aggregated Facebook platform data under a differential-privacy framework (King and Persily 2018). In this paper, we analyze the "Condor" dataset of URL-level Facebook activity[2] to precisely quantify both shares and views of different kinds of news content across age and ideological categories according to three dimensions of interest: whether a news source is

[1]See `http://socialscience.one` for more information.

[2]See: `https://socialscience.one/blog/unprecedented-facebook-urls-dataset-now-available-research-through-social-science-one`

untrustworthy or not; whether it is related to politics; and whether it is "clickbait."

We combine this unprecedented access to platform data on URL views and engagements with classification methods to provide new, fine-grained evidence on news consumption and sharing on Facebook. Specifically, we focus on the following research questions:

1. How prevalent is the *sharing* and *consumption* of untrustworthy news in the U.S. Facebook population?
2. What ideological groups are most likely to share and view this "fake news" on Facebook?
3. What age groups are most likely to share and view this content on Facebook?
4. When looking at patterns of consumption and engagement, how does traditional "fake news" differ from "clickbait" or simply news articles that are related to politics?

By providing aggregate numbers of URL shares and views across age brackets and different ideological categories, we are able for the first time to provide answers to these basic descriptive questions concerning the overall quality of the information people encounter or share on their Facebook News Feeds.

Literature, Research Questions and Hypotheses

Our conception of "fake news" relies on definitions used in prior research which emphasize the credibility and trustworthiness of news at the source level as determined by adherence to journalistic norms and practices designed to ensure a commitment to factual accuracy (Lazer et al. 2018; Grinberg et al. 2019). We consider articles from designated "fake news" domains to be instances of "fake news," though since some portion of content produced by these outlets is likely to be factual (Guess, Nyhan and Reifler 2020), the approach can be considered an upper bound. In practice, "fake news" articles can be thought of as a subset of disinformation — factually inaccurate messages intended to be misleading — featuring the trappings of traditional news (Tucker et al. 2018). In this article, we interchangeably refer to "fake news," "low-quality news" and "untrustworthy websites" as opposed to "credible" news. We specify our methods of classifying categories of news, including fake and credible news, in the next section.

A starting point for our investigation is a recent study that analyzed fake news sharing behavior during the 2016 U.S. presidential campaign by merging Facebook profile data obtained via an authenticated application with individual-level survey responses (Guess, Nagler and Tucker 2019). Three main findings stand out from this study. First, sharing activity appeared to be quite rare: over 90% of the sample did not appear to share any fake news content whatsoever, and this was not related to overall sharing behavior. The sharing of misinformation appeared to exhibit a highly skewed distribution with a "long tail" of respondents who made up the bulk of all fake news shares. Second, there was a robust relationship between ideological self-placement and fake news sharing activity, with more conservative-leaning respondents more likely to share dubious content. Third, the older respondents were, the more likely they were to share fake news content. Holding constant other demographic and political features (including ideology, education, and the total number of web links shared), the study found that being over the age of 65 was associated with sharing nearly 7 times as many links to fake news domains as those in the youngest age group or about 2.3 times as many as those in the next-oldest age group. These findings motivate our present focus on two predictors of URL-level viewing and sharing activity on Facebook: age and ideological self-placement.[3]

As suggested, since these sharing patterns are highly skewed, inferences about the kinds of behaviors comprising the "long tail" are limited in conventional samples. Indeed, in the study's dataset a total of 101 respondents out of 1,191 shared any fake news articles to their friends on Facebook — 302 shares in all. Subsequent research has found comparable patterns using a much larger sample of Twitter users matched to the voter file (Grinberg et al. 2019), but questions still remain about the generalizability of these findings to later periods and to the Facebook platform, which is more widely used in the population (Gottfried and Shearer 2016). By encompassing much more of the population of interest in this study, we are able to achieve precise inferences that are not subject to small-N or sample attrition concerns.

In motivating our hypotheses and research questions, we begin with a simple descriptive analysis of the relative prevalence of fake news and credible news among both overall views and shares of URLs on Facebook in the United States. We ask: **Is the total number of views and shares of credible news articles on Facebook greater than the number of views and shares of fake news? (RQ1)**

Following prior research, we next turn to individual-level predictors that speak to debates about the mechanisms underlying fake news engagement on social media. Since these may differ for private (viewing) and relatively more observable (sharing) behavior (Osmundsen

[3]Associations with age and partisanship have more recently been replicated by Osmundsen et al. (N.d.) using surveys linked to Twitter data, with sample size of a comparable magnitude. They also investigate a number of other potential predictors of fake news sharing, such as cognitive reflection (Pennycook and Rand 2019), which we cannot test here.

et al. N.d.), we separately explore each type of outcome.

People may share articles and other forms of content on social media for numerous reasons. One hypothesized purpose is to signal affiliation with one's identity or affinity with a particular cultural or social group (Kahan 2017). In other words, people share content not to inform their friends and social connections but to affirm their membership in an ingroup (and possible opposition to an outgroup). We focus here on ideological affiliation as a potential marker of political identity (Mason 2018). First, we expect fake news sharing to be more common among those at the poles of the ideological continuum (**H1a**). Following research on the asymmetry of fake news production and consumption in 2016, we additionally expect fake news sharing to be more common among conservative users than liberal users (**H1b**). We also expect that fake news is shared predominately by Facebook users that are ideologically congruent to the ideological slant of the news publisher (**H1c**). *Viewing* content on social media feeds may be governed by different mechanisms, though we expect that homophily will tend to produce viewing patterns that reflect the basic sharing dynamics of users' social networks (**H2a–c**).

A second factor proposed as an explanation for fake news sharing and viewing relates to digital literacy, a generalized set of skills and competencies that include the ability to spot online deception and assess the trustworthiness of information sources (Hargittai and Micheli 2019). People with lower levels of digital literacy may not share others' understanding of how to interpret the relationship between content disseminated on one's feed and one's identity, beliefs, and other characteristics. This disconnect could lead to more indiscriminate sharing (and viewing) behavior, especially if having lower digital literacy is associated with being connected to others with similarly low levels (though see Hargittai 2010). Given that older Americans tend to have lower levels of digital literacy, this has been suggested as an explanation for why older Americans have been previously found to be the most likely to share fake news (Guess and Munger 2020).

Of course, age is a blunt category associated with a whole host of period, life-cycle, and individual-level traits, but there have been a number of suggested reasons why older Americans could be more likely to share fake news. In addition to internet skills and digital literacy, these include systematic differences in social trust and social ties; increased tendency toward source amnesia; difficulty detecting lies; or fewer potential concerns with the consequences of being out of the labor market (Brashier and Schacter 2020; Van Bavel et al. 2020; Guess, Nagler and Tucker 2019). Regardless of the particular explanation, based on prior research **we expect both fake news sharing and viewing to be more common among older users than younger users (H3a–b).**

Finally, we explore an additional, under-appreciated possible explanation for differences in engagement with fake news: that specific characteristics of news on social media are relatively more attractive to people in different age categories. Though fake news has been noted for its novelty and ability to surprise (Vosoughi, Roy and Aral 2018), it tends to have characteristics in common with two other basic forms of online content: "clickbait," which describes articles with headlines optimized to entice clicks (often via dubious appeals to curiosity); and, more simply, any content that is related to politics. Clickbait exploits a psychological phenomenon known as the curiosity gap (Loewenstein 1994) and tricks readers into clicking a headline to fill this created knowledge gap. Since some scholars have found evidence that older Americans have a higher preference for clickbait headlines (Munger et al. 2018), it is possible that false clickbait headlines could drive higher levels of fake news sharing among older Americans. Given that older adults are more likely to be interested in politics (Delli Carpini 2000) and vote (Mesch and Coleman 2007; Henn and Foard 2012), and that fake news is often about politics, it is also possible that the political nature of fake news has a similar effect.[4]

In short, a reductive — and perhaps inaccurate — way to describe the kind of online, politically charged fake news studied since 2016 is the intersection of clickbait and political content. Following this logic, we test whether these characteristics of online news are more likely in fake news stories and if fake news that contain these characteristics are more likely to be shared among older Americans. Focusing on URL shares, we expect that fake news is more likely to be classified as clickbait than credible news (**H4a**). We also expect that fake news is more likely to be political than credible news (**H4b**). We then explore which age groups are more likely to share clickbait headlines (**RQ2a**) and political news (**RQ2b**). Finally, as an exploratory analysis we investigate how sharing patterns for different combinations of these news categories vary across age groups (**RQ3**).

Data and Methods

Condor Dataset

The release of the Social Science One (SS1) "Condor" dataset of URLs shared on Facebook allows us to test these relationships in a de-identified, privacy-protecting manner (Messing et al. 2020). The dataset summarizes information about 38 million URLs shared worldwide

[4]We define an online article as political if the central claim of the story relates to government institutions, policy, or a political actor.

more than 100 times publicly on Facebook (between January 1, 2017 and July 31, 2019, inclusive). Each row for U.S. users represents a count of actions (e.g., likes, shares, views, clicks) at the URL-year-month-age-gender-"political page affinity" level, with Gaussian noise added to each count designed according to a differential-privacy framework.[5] Therefore, we observe a noisy value where random noise is drawn from a zero-centered normal distribution with standard deviation σ, where σ is constant for each attribute, such as age category (see Messing et al. 2020 for a detailed explanation). Hence, when calculating the frequency of shares or views within a given category c, we can only construct a noisy estimate of this quantity $\hat{X}_c$. Given that that the noise-injection process is public[6] and we have the number of rows that we sum over to create these counts, we can create confidence intervals around the counts that we present (Eq. 1):

$$\hat{X}_c \pm 2 \times \sqrt{\sum_{r \in c} \sigma^2} \tag{1}$$

We find that the confidence intervals are tight enough that we can still make inferences never before possible about Facebook engagement; in most of our figures, these error bars are barely visible. For every statistic, we report 95% confidence intervals for all of the measures. Adding random noise induces measurement error, which leads to attenuation bias when using statistical methods such as linear or logistic regression (Evans and King 2020), but by only presenting the sums of rows we do not suffer from this problem.

We collect all URLs located in the Condor dataset that were first posted on Facebook between January 1, 2018 and December 31, 2018 (inclusive), filtered in the following additional ways. First, we only include URLs from domains evaluated by NewsGuard, an independent organization that comprehensively evaluates the quality of online information sources with a transparent rating scheme and a staff of full-time journalists.[7] We additionally eliminate all URLs from sites classified by NewsGuard as satire or sites not given a numeric rating, including YouTube. We opted for NewsGuard given its extensive coverage of credible and low-quality online news producers (it contains ratings for 5,135 domains).[8] We focus strictly

[5]Given our study's focus, we aggregate over the month and gender dimensions. Political page affinity is a five-point integer scale (from -2 to 2) that categorizes users' political affinity based on the pages they follow, adapting the method used by Barberá et al. (2015). We treat this as an estimate of a user's ideological self-placement. The age categories are: 18-24, 25-34, 35-44, 45-54, 55-64, 65+. There is also an "NA" age category, which is excluded from our figures. This represents Facebook users for whom age could not be determined from their profiles.

[6]$\sigma = 14$ for shares and $\sigma = 2,228$ for views (Messing et al. 2020)

[7]NewsGuard scores inform an internet plug-in that surfaces metadata on news sources' credibility in search results pages and social media feeds. More information can be found at `https://www.newsguardtech.com`.

[8]Reassuringly, the NewsGuard list contains most of the fake news publishers identified by Allcott,

on 2018 given that at the time of analysis it was the most recent full year for which we could retrieve data.[9] We further limit URLs to those that received more shares from American Facebook users than users from other countries. The aggregated counts we use also cover American Facebook users only. These steps leave us with 981, 829 URLs.

URL Classification

At the URL level, we create three binary measures: (1) credible vs. fake news source; (2) political topic; (3) clickbait. We also classify URLs according to the estimated political slant of their parent domain.

Credibility. We use NewsGuard's ratings to determine the "credibility" of a domain. NewsGuard employs a team of trained journalists and editors to review and rate news and information websites based on nine criteria assessing basic practices of credibility and transparency. Based on a site's performance on these criteria, it is assigned a red or green rating. A histogram of NewsGuard scores (from 0 to 100) for the majority of online news domains can be found in the Online Appendix in Figure 13. Domains that receive a score of 60 points or higher are deemed credible; otherwise, we code them as "low-quality" (i.e., a publisher of fake news). This is the same threshold that NewsGuard's browser extension uses to determine whether users see a "red" or "green" shield next to a news headline. Following the standard in the literature, we classify URLs on the basis of their parent domains' credibility rating.

Political news. We determine each news URL's relevance to politics using a supervised machine classifier. To create a training set, three independent coders identified a random sample of headlines and blurbs as political or not (expressed as a binary yes or no).[10] Given that we are comparing credible and non-credible sources, this sample was balanced between the two.[11]

We used a "gold standard" dataset (100 headlines/blurbs from non-credible sources and 400 headlines/blurbs from mainstream sources) to train three undergraduate coders. Once

Gentzkow and Yu (2019); in the Condor dataset, more than 96% of URL shares from domains in the Allcott, Gentzkow and Yu (2019) list are also captured by the NewsGuard list.

[9]NewsGuard's coverage also begins in 2018.

[10]We define an online article as political if the central claim of the story relates to government institutions, policy, or a political actor.

[11]This is especially important given that headlines/blurbs from low-quality and credible sources are likely to rely on different language styles.

they achieved acceptable reliability, they labeled a large balanced set of additional headlines/blurbs. 8,552 headlines/blurbs were labeled in total (3,971 from credible sources and 4,581 from non-credible sources) by independent coders. We monitored their reliability by having two coders label 15% of the headlines/blurbs.[12] For this subset labeled by two coders, there was 86% agreement.[13] We then used these labels to train a random forest classifier to classify the rest of the URLs in our dataset. Using a train/test split of 70-30, this model produced 90% accuracy and an F1 score of 0.90. Additional validation details are provided in Table 3 in the Online Appendix.

Clickbait. To classify URLs as clickbait or not, we employed a pre-trained classifier based on prominent linguistic features of headlines. Specifically, we used a support vector machine (SVM) with radial basis function kernel built by Chakraborty et al. (2016) using a training set of 16,000 clickbait (BuzzFeed, Upworthy, ViralNova, Scoopwhoop, and ViralStories) and 16,000 non-clickbait headlines (from Wikinews articles collected by NewsReader). This model has an accuracy of 93%, a precision of 0.95 and recall of 0.9 on the original authors' set of headlines.

Ideological slant. To compare news sources across the ideological spectrum, we employ two different estimates of source ideology. For credible news, we use media partisanship scores developed by Eady et al. (2020), which are estimated from URLs jointly shared by social media users and news accounts.[14] For fake news sources, we rely on manual classifications by Aslett et al. (2020) covering the still-active low-quality sources identified by Allcott, Gentzkow and Yu (2019). They classified each source as liberal leaning, conservative leaning, or with no clear partisan orientation. Combined with NewsGuard ratings, these slant estimates allow us to disaggregate both quality and low-quality news domains to those from a primarily liberal or conservative viewpoint.[15]

[12]In cases of disagreement, we had a third coder break the tie before adding the result to our final training set.

[13]Table 2 in the Online Appendix provides more detailed inter-rater reliability statistics.

[14]Sites with a score below zero are classified as liberal and those above zero are classified as conservative.

[15]The list of sources with ideological scores (70 low-quality domains, 150 credible domains) can be found in the Online Appendix. Half of the URL shares recorded in the dataset are from domains for which we have an ideological score.

Results

Prevalence of Fake News (RQ1)

Figures 1 and 2 present the number of shares[16] and views of news URLs on Facebook in the United States in 2018, according to whether their domains are classified as credible or low-quality. Of URLs from sources rated by NewsGuard in the Condor dataset, we find that approximately 84% of shares were from credible news domains.[17] The rest (15%) were shared from low-quality news domains. We find that a slightly larger share of views, approximately 89%, was from credible news domains, and, not surprisingly, there are on the order of 100 times more views across categories than shares. Clearly, views and shares of high-quality news far outnumber those of low-quality or untrustworthy news. Though the proportion of credible news that we find is significantly higher than public perceptions — a recent Knight Foundation study estimated that Americans believe 65% of news on social media constitutes misinformation (Knight 2018) — news from low-quality sources still achieves a nontrivial amount of circulation: roughly one out of eight views of at least moderately popular news articles on Facebook are from these sources.[18]

[16] Share counts include reshares.

[17] Calculating percentages using two different sums of differentially private data may induce bias, but work using this dataset has demonstrated that with a sufficient number of views or shares, the bias in this estimator is not practically significant (Buntain et al. 2021). Given that every viewing and sharing estimate we report in this paper is much higher than minimum thresholds set by Buntain et al. (2021), we can confidently report percentages throughout this paper.

[18] It is worth reiterating here that we only have access to news links shared at least 100 times publicly in the Condor dataset.

Figure 1: News URL Shares by Source Credibility. 95% confidence intervals are displayed.

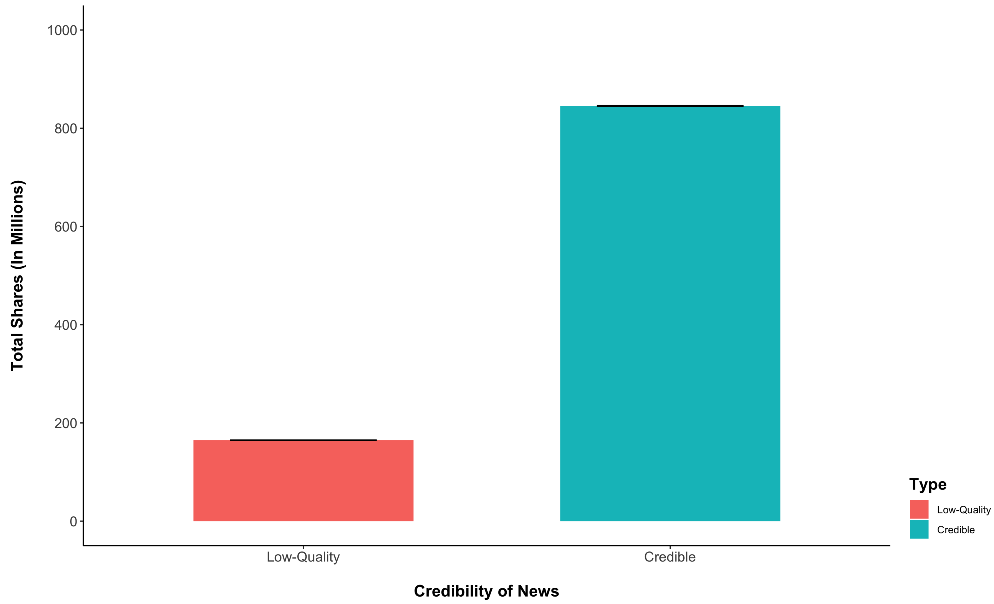
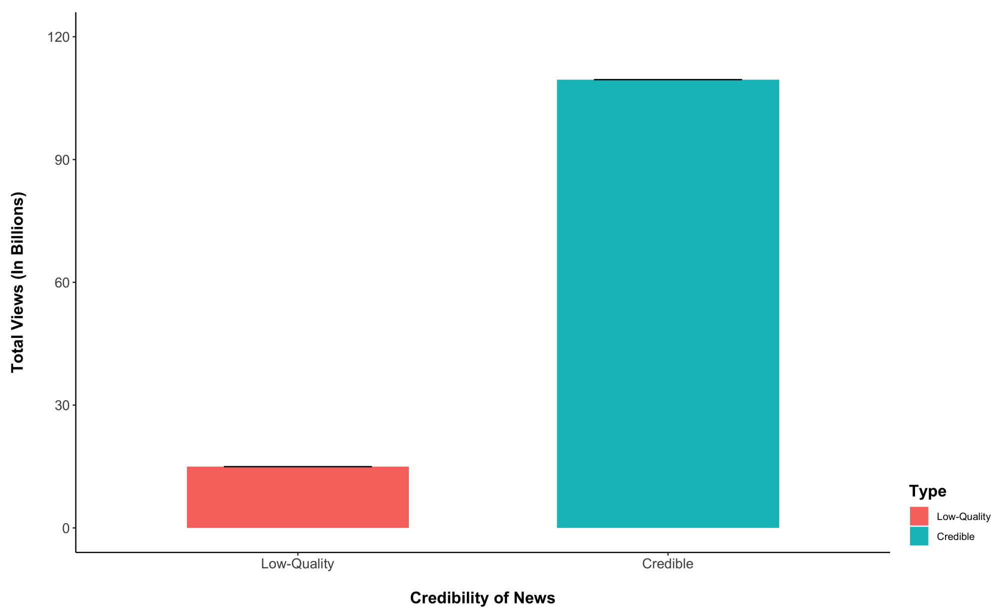


Figure 2: News URL Views by Source Credibility. 95% confidence intervals are displayed.



Ideology and Fake News (H1, H2)

Next, we turn to the distribution of these shares and views by users' inferred ideological category. In Figure 3, looking purely at counts, we see that somewhat more links to articles from fake news domains were shared by conservatives (especially those estimated to be in the most conservative category). This is not driven by more shares overall; the most conservative categories are associated with *fewer* shares from credible news domains. In fact, 27% of news URLs shared by very conservative Facebook users are from low-quality news publishers, but only 9% of news URL shared by very liberal Facebook users are from low-quality news publishers. A similar pattern is visible in Figure 4 for views, though the overall ideology gradient is more slight: 19% of news URLs very conservative Facebook users view are from low-quality news publishers, but only 7% of news URLs very liberal Facebook users view are from low-quality news publishers. For both types of behavior, fake news as a proportion of all news is the greatest for the most conservative category, supporting **H1b/H2b** but not **H1a/H2a**, which predicted more engagement at both ideological poles.

Figure 3: News URL Shares by User Ideology. 95% confidence intervals are displayed.

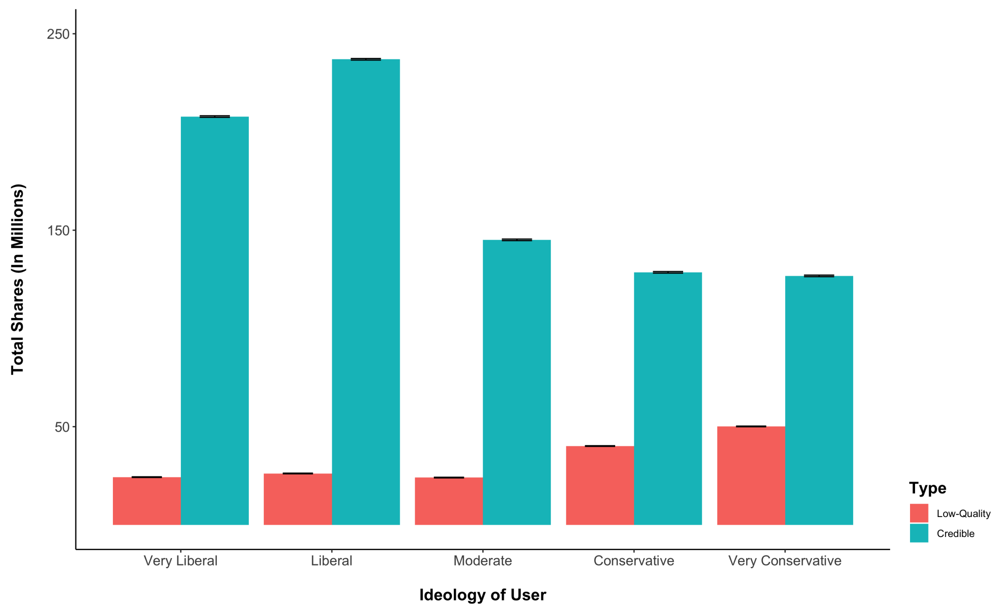
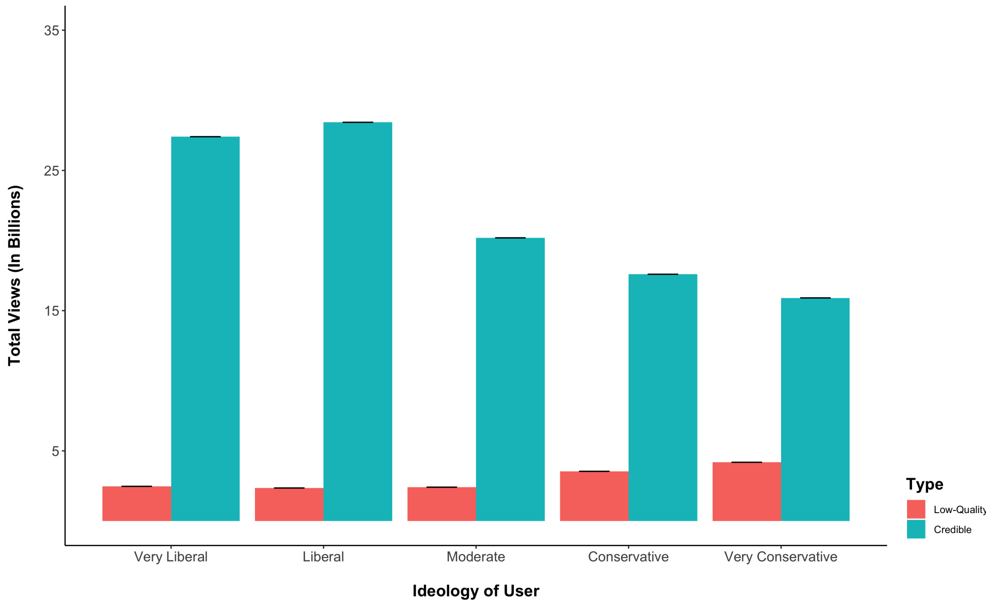


Figure 4: News URL Views by User Ideology. 95% confidence intervals are displayed.



Exploring further, **Hypotheses 1c/2c** predict that Facebook users will view and share more fake news from ideologically congruent sources. Disaggregating our untrustworthy news measure by estimated ideological slant, we find that this is the case for both types of behavior. The number of shares and views of low-quality news by both estimated ideological slant of the source and inferred user ideological category are reported in Figures 5 and 6. As expected, Facebook users appear to mainly view and share ideologically congenial fake news. (We present analogous results for credible news domains in the Online Appendix in Figures **??** and **??**.) Our results, especially among the liberal subgroup, illustrate the advantage of large-scale platform data for studying what is often relatively tail-end behavior: subtle patterns emerge that would be obscured by sampling variability in datasets of conventional size.

Figure 5: Low-Quality News URL Shares by User Ideology and News Slant. 95% confidence intervals are displayed.

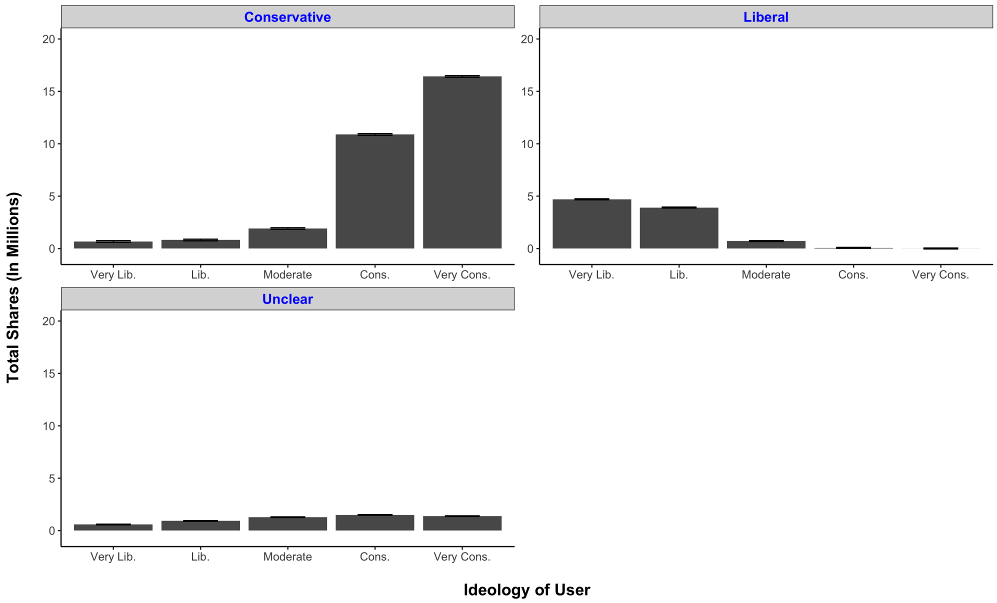
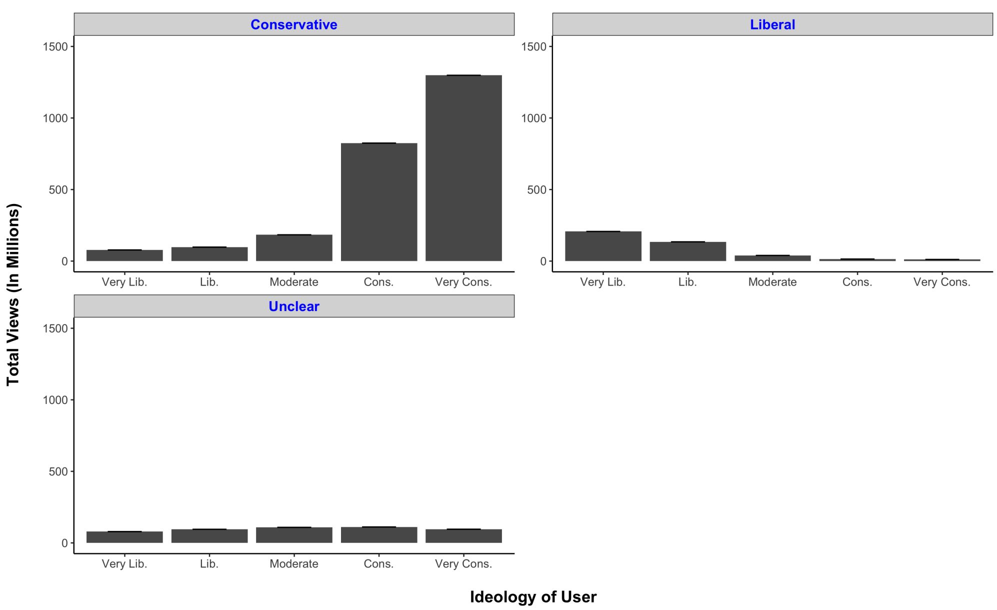


Figure 6: Low-Quality News URL Views by User Ideology and News Slant. 95% confidence intervals are displayed.



Age and Fake News (H3)

In Figures 7 and 8 we present the number of shares and views of URLs from low-quality and credible domains by Facebook users across different age brackets. As previous research suggests, we find that older Americans share more links to articles from fake news domains (**H3a**) than do younger Americans. Taking into account the number of shares from credible news sources, we see that fake news is also greatest in proportional terms among Americans in the oldest age category: 20% of news URLs shared by Facebook users 65 years old or older are from low-quality news domains, but only 10% of news URLs shared by Facebook users from the lowest age bracket (18-24) are from low-quality news domains. We also look at the age gradient within ideological subgroups and find a similar pattern in each, with the exception of moderates (Figure 9). Interestingly, older Americans as a group do not seem to view much more fake news (**H3b**). However, a different pattern emerges when we consider the proportion of news URLs within age categories: 19% of news URLs the oldest Facebook users view are from low-quality news publishers, but only 7% of news URLs the youngest Facebook users view are from low-quality news publishers.

Figure 7: News URL Shares by Age Group. 95% confidence intervals are displayed.

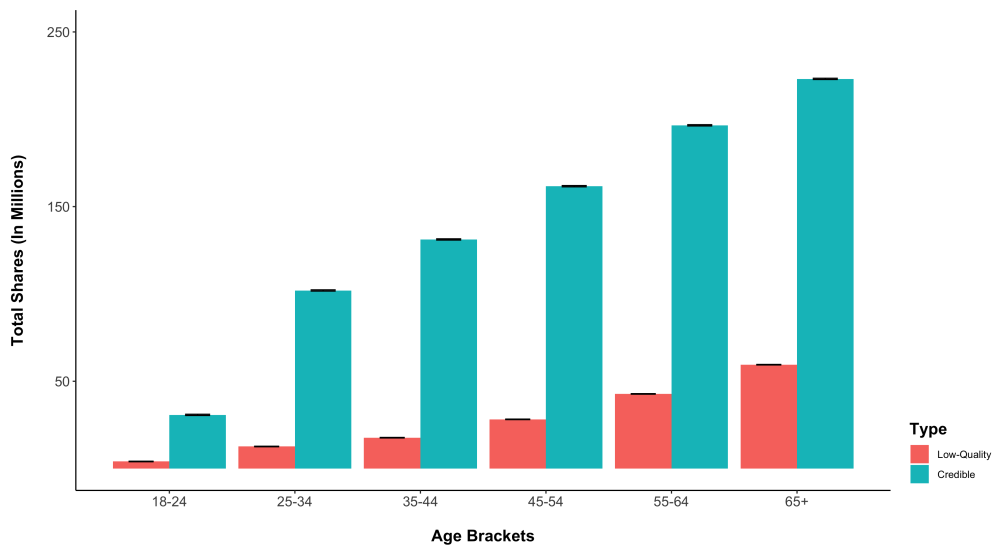


Figure 8: News URL Views by Age Group. 95% confidence intervals are displayed.

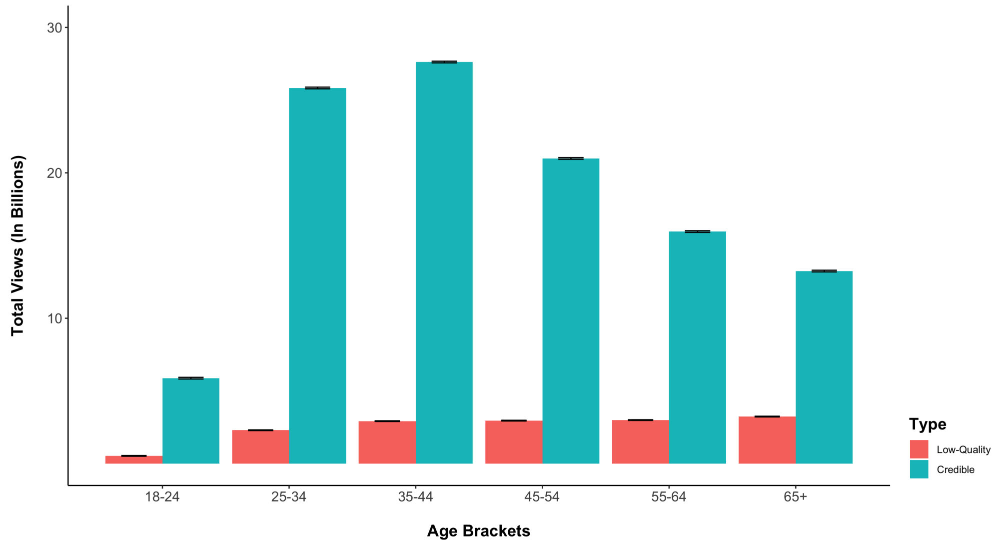
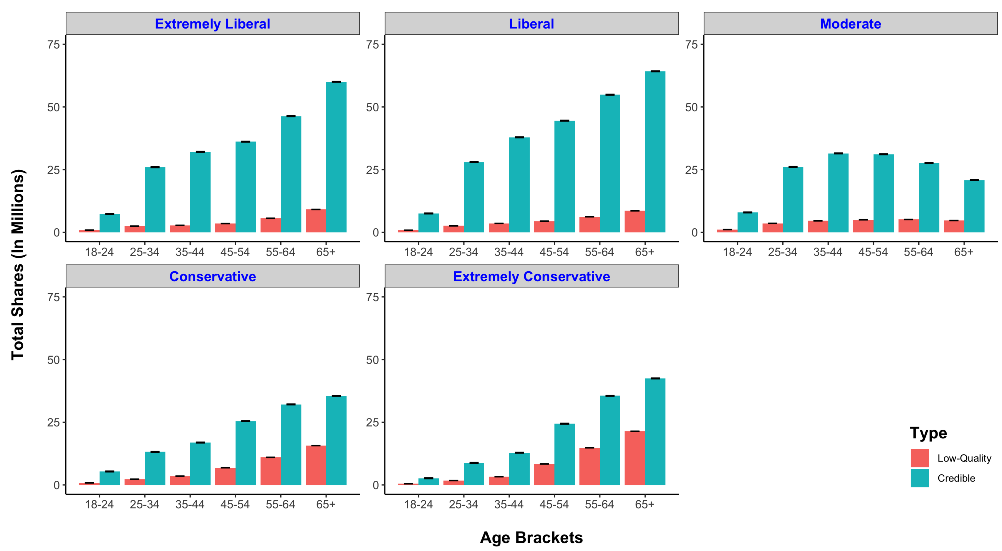


Figure 9: News URL Shares by Ideology and Age Group. 95% confidence intervals are displayed.



Fake News, Clickbait, and Politics (H4, RQ2-3)

Having provided new descriptive evidence on both credible and non-credible news viewing and sharing activity on Facebook across age and ideological groups, we now explore whether similar patterns exist for related but distinct categories — political news and clickbait. First, we look at how these classifications relate to each other. Aggregating over all our URLs, we calculate the proportion classified as political/non-political and clickbait/non-clickbait among both credible and low-quality domains (Table 1). In support of **H4a** and **H4b** we find that URLs from low-quality news domains are both more likely to have clickbait headlines and to have a central claim that is political than URLs from credible news domains, though the latter difference is especially pronounced (37.3% vs. 25.4%). Among low-quality domains, 12.6% are classified as both political and clickbait (compared to 6.3% of credible news domains), which suggests that "fake news" tends to be more than the intersection of clickbait headlines about political topics.

Table 1: Proportion of Links that are Clickbait and Political from Credible and Low-Quality Sources

Credible News Domains		
Clickbait	Political	Percentage of Links
Yes	Yes	6.3%
Yes	No	21.4%
No	Yes	25.4%
No	No	46.9%
Low-Quality News Domains		
Yes	Yes	12.6%
Yes	No	23.9%
No	Yes	37.3%
No	No	26.2%

Given limited overlap across these categories, we investigate whether Facebook users in certain age groups are more likely to share news with clickbait headlines or political news (**RQ2a/RQ2b**). Figure 10 presents the number of shares of URLs with clickbait headlines from all domains in our dataset across age categories. We do not find support for our hypothesis that older users are more likely to share clickbait articles than younger users: in the youngest age bracket (18-24), around 36% of news shares were from URLs that had a clickbait headline, compared to 27% of URLs shared in the oldest age bracket (65 and over). This finding holds when we only look at shares from low-quality news domains as well (see Figure 16 in the Online Appendix).

Figure 11 presents the number of shares of political URLs across age categories. Here we do find support for our hypothesis that older Facebook users are more likely to share political articles than younger users: in the youngest age bracket, around 24% of news shares were from URLs that are political compared to 58% of shares from the oldest age bracket. As before, this finding also holds when we only look at shares from low-quality news domains (see Figure 17 in the Online Appendix).

Figure 10: News URL Shares by Age Category and Clickbait. 95% confidence intervals are displayed.

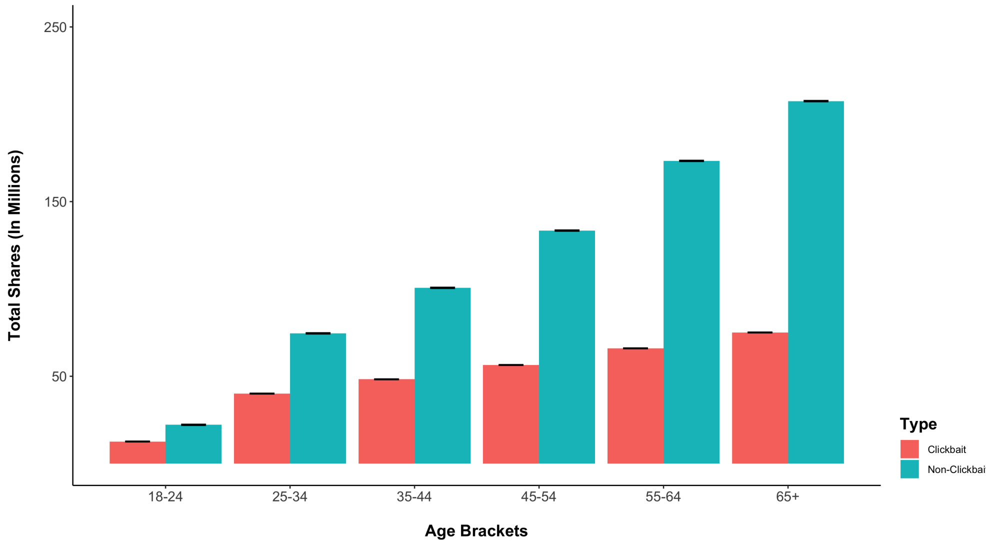
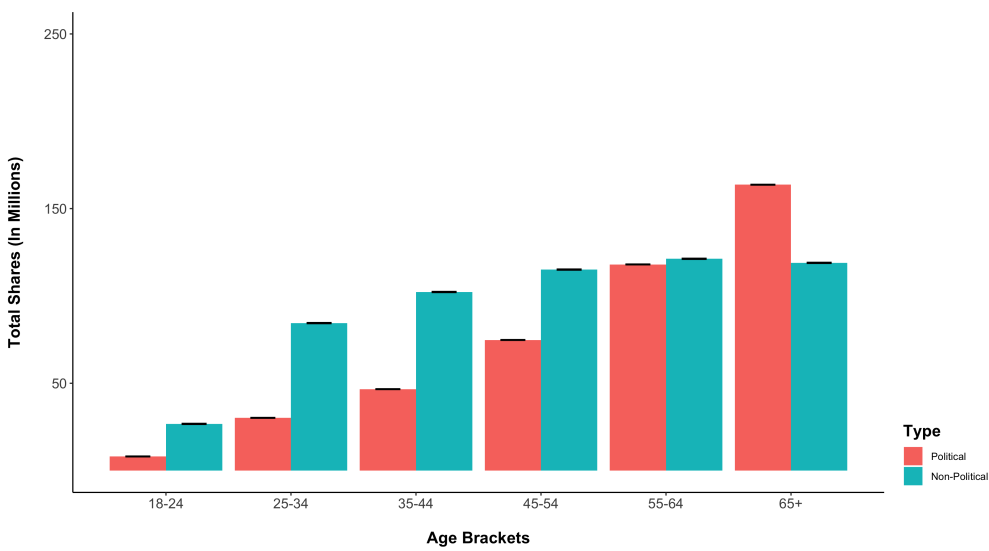


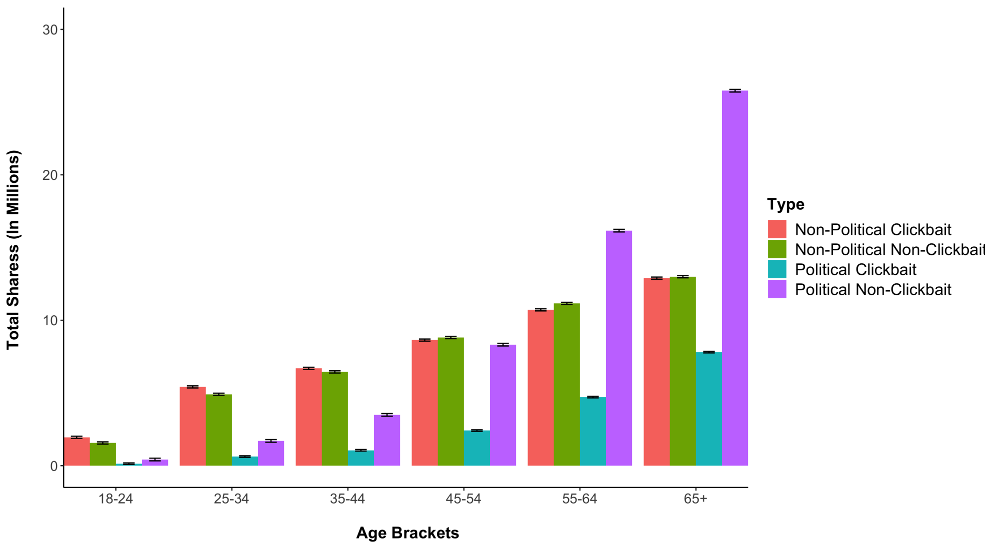
Figure 11: News URL Shares by Age Category and Political Topic. 95% confidence intervals are displayed.



Finally, we explore whether any notable variation emerges *within* the set of fake news URL shares by age category. In other words, do the sharing patterns we uncovered appear to be driven by one of the specific subsets of fake news that we identified in the bottom half of Table 1 above? Figure 12 presents the number of URLs shared from low-quality news domains by age bracket and their clickbait and political classifications. According to the figure, a large percentage of fake news shared by the oldest age bracket is political (around

58%), but only a small percentage of fake news shared by the youngest group of Americans is (around 15%). Clickbait articles only make up around 36% of fake news shared by the oldest Americans, while it makes up around 51% of fake news shared by the youngest Americans. By placing all URLs shared from low-quality news domains into four possible categories of political/non-political × clickbait/non-clickbait we can display the number of URLs shared for each category in Figure 12. As the purple bars in the figure indicate, political *and* non-clickbait fake news make up close to a majority of fake news shares for the oldest age bracket (around 44%), whereas it makes up 11% of fake news shares among the youngest group.

Figure 12: News URL Shares from Low-Quality News Domains. 95% confidence intervals are displayed.



Discussion

Using unprecedented access to a new dataset of large-scale Facebook URL engagement, we document several findings that contribute to our understanding of descriptive patterns in the quality of information Americans view and share in their News Feeds. Focusing on URLs shared in 2018, we confirm findings from earlier research that sharing and consumption of online misinformation on Facebook is relatively rare, with only 12% of news shares — and 16% of news exposures — being linked to what NewsGuard classifies as non-credible news domains (Guess, Nagler and Tucker 2019; Guess, Nyhan and Reifler 2020). We also confirm the existence of a relationship between the conservatism of inferred ideological affinity and both the number and proportion of shares to fake news domains by users in that category. This likely reflects the greater popularity of conservative-leaning fake news overall, though we see among both the most liberal and most conservative user categories a relatively greater number of shares and views of ideologically congenial content from low-quality news domains.

Also in line with earlier work, we find an even stronger relationship between the user age category and the number of shares from low-quality domains, one that is most pronounced within the two most conservative ideological groups — among which older users also share more links to suspect news domains. Strikingly, this pattern is not as clear in the view counts, which suggests that fake news sharing among older Americans may not simply reflect a greater prevalence of low-quality news posts in their feeds (e.g., Grinberg et al. 2019). It is important to note when interpreting these results that the data we analyze are counts aggregated over users, with observations (in this case) at the *URL-age category* level. Thus we cannot strictly make inferences about individual user behavior, especially given prior evidence suggesting that fake news engagement mainly occurs within smaller subgroups than the ones we can identify with the relatively coarse categories made available to researchers in the Condor data release. An alternative explanation for the sharing/viewing discrepancy across age categories, then, is that more fake news sharing *is* conditional on greater exposure on the News Feed — but among a small fraction of users whose feeds are mostly overrun with misinformation and who drive a disproportionate quantity of shares.

Our analyses demonstrate how combining large-scale URL data with machine classification enhances our understanding of news sharing patterns. Looking beyond the traditional credible/untrustworthy dichotomy, we show that clickbait constitutes a larger proportion of news shares for the youngest age category than the oldest. Perhaps less surprisingly, the oldest Americans share more news about politics than any other topic, while political news is relatively rare among the youngest group. The ability to deploy highly accurate supervised learning classifiers trained on large-N platform data enables a more nuanced account of the kinds of content people engage with on social media.

In addition to its necessarily aggregated nature, several other caveats regarding the Condor dataset bear repeating. First, each count we observe is the sum of the true count plus a random number drawn from a known distribution. Though this process adds statistical noise, we are able to report counts with extremely narrow confidence intervals. In addition to this differential-privacy protocol, the URLs themselves are selected for inclusion into the dataset according to a threshold of more than 100 global public shares (Messing et al. 2020). Thus our conclusions are not based on an unbiased sample of the complete universe of web links ever shared by users on Facebook; they are instead limited to inferences based on more popular URLs.[19] Finally, though our estimated counts have the advantage of being straight-

[19]This does not, however, mean that we are not picking up private shares in our analyses, as private shares

forward to interpret and unbiased, they risk misleading conclusions about prevalence that can arise without a reasonable denominator. In most cases, we display counts for both low-quality and credible news, which allows for a natural comparison. Even so, it is important to keep in mind how news exposure on social media compares to other content users may encounter on their feeds. Our data span nearly 280 billion URL views by American users in 2018 of which over 124 billion (44.5%) are from our set of news domains, and nearly 2.1 billion shares of which over 1 billion (48.2%) are from those news domains. These proportions are an order of magnitude greater than other studies have found,[20] which is a striking divergence, though this may partially reflect an imperfect denominator — much of people's News Feeds contain content that does not originate with URLs — in addition to ecological inference issues.

Despite these qualifications, the analyses presented here offer a proof of concept for future research. Though we cannot make inferences about individuals or causal mechanisms, our descriptive findings help to build a foundation of evidence that can inform future research and policy debates. We view this as part of an ongoing, cumulative process: Our evidence will not indefinitely capture the reality of engagement with news URLs on Facebook, and we encourage researchers to build on our work to further replicate and refine our approach as more recent data become available. In addition to refreshing this foundation of evidence, we hope that such efforts will continue to build trust between independent academic researchers and platforms as they work to expand the scope and resolution of future data releases.

are included in observed counts if a given URL was *also* shared publicly more than 100 times.

[20]By one recent estimate, news content accounts for no more than 4.2% of total online consumption (including social media) on average (Allen et al. 2020).

References

Allcott, Hunt, Matthew Gentzkow and Chuan Yu. 2019. "Trends in the diffusion of misinformation on social media." *Research & Politics* 6(2):2053168019848554.

Allen, Jennifer, Baird Howland, Markus Mobius, David Rothschild and Duncan J Watts. 2020. "Evaluating the fake news problem at the scale of the information ecosystem." *Science Advances* 6(14):eaay3539.

Amazeen, Michelle A, Fabrício Benevenuto, Nadia M Brashier, Robert M Bond, Lia C Bozarth, Ceren Budak, Ullrich KH Ecker, Lisa K Fazio, Emilio Ferrara, Andrew J Flanagin et al. 2020. "Tackling misinformation: What researchers could do with social media data." *The Harvard Kennedy School Misinformation Review* 1(8).

Anspach, Nicolas M and Taylor N Carlson. 2018. "What to believe? Social media commentary and belief in misinformation." *Political Behavior* pp. 1–22.

Aslett, Kevin, William Godel, Zeve Sanderson, Nathaniel Persily, Jonathan Nagler, Richard Bonneau and Joshua Tucker. 2020. "Measuring Belief in Fake News in Real-Time.".

Barberá, Pablo, John T Jost, Jonathan Nagler, Joshua A Tucker and Richard Bonneau. 2015. "Tweeting From Left to Right Is Online Political Communication More Than an Echo Chamber?" *Psychological science* .

Brashier, Nadia M and Daniel L Schacter. 2020. "Aging in an Era of Fake News." *Current Directions in Psychological Science* p. 0963721420915872.

Buntain, Cody, Richard Bonneau, Jonathan Nagler and Joshua Tucker. 2021. "A Summary Statistic for Political Ideology of Web Domains Using Differentially Private Engagement Data with Applications to Social Science One.".

Chakraborty, Abhijnan, Bhargavi Paranjape, Sourya Kakarla and Niloy Ganguly. 2016. Stop clickbait: Detecting and preventing clickbaits in online news media. In *2016 ieee/acm international conference on advances in social networks analysis and mining (asonam)*. IEEE pp. 9–16.

Delli Carpini, Michael X. 2000. "Gen. com: Youth, civic engagement, and the new information environment." *Political communication* 17(4):341–349.

Eady, Gregory, Richard Bonneau, Joshua A Tucker and Jonathan Nagler. 2020. "News Sharing on Social Media: Mapping the Ideology of News Media Content, Citizens, and

Politicians.”.
URL: *osf.io/ch8gj*

Evans, Georgina and Gary King. 2020. “Statistically Valid Inferences from Differentially Private Data Releases, with Application to the Facebook URLs Dataset.”.

Gottfried, Jeffrey and Elisa Shearer. 2016. “News Use Across Social Media Platforms 2016.” Pew Research Center, May 26, 2016. Downloaded March 14, 2017 from `http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/`.

Grinberg, Nir, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson and David Lazer. 2019. “Fake news on Twitter during the 2016 US presidential election.” *Science* 363(6425):374–378.

Guess, Andrew, Brendan Nyhan and Jason Reifler. 2020. “Exposure to untrustworthy websites in the 2016 US election.” *Nature Human Behaviour* pp. 1–9.

Guess, Andrew, Jonathan Nagler and Joshua Tucker. 2019. “Less than you think: Prevalence and predictors of fake news dissemination on Facebook.” *Science advances* 5(1):eaau4586.

Guess, Andrew M and Benjamin A Lyons. 2020. “Misinformation, Disinformation, and Online Propaganda.” *Social Media and Democracy: The State of the Field, Prospects for Reform* p. 10.

Guess, Andrew M, Michael Lerner, Benjamin Lyons, Jacob M Montgomery, Brendan Nyhan, Jason Reifler and Neelanjan Sircar. 2020. “A digital media literacy intervention increases discernment between mainstream and false news in the United States and India.” *Proceedings of the National Academy of Sciences* 117(27):15536–15545.

Guess, Andy and Kevin Munger. 2020. “Digital Literacy and Online Political Behavior.”.
URL: *https://osf.io/3ncmk/*

Hargittai, Eszter. 2010. “Digital na (t) ives? Variation in internet skills and uses among members of the “net generation”.” *Sociological inquiry* 80(1):92–113.

Hargittai, Eszter and Marina Micheli. 2019. “Internet skills and why they matter.” *Society and the internet: How networks of information and communication are changing our lives* p. 109.

Henn, Matt and Nick Foard. 2012. “Young people, political participation and trust in Britain.” *Parliamentary affairs* 65(1):47–67.

Kahan, Dan M. 2017. “Misconceptions, misinformation, and the logic of identity-protective cognition.”.

King, Gary and Nate Persily. 2018. “A New Model for Industry-Academic Partnerships.” Current version: GaryKing.org/partnerships.

Knight. 2018. “Americans' Views Of Misinformation In The News And How To Counteract It.”.
URL: *https://knightfoundation.org/reports/americans-views-of-misinformation-in-the-news-and-how-to-counteract-it/*

Lazer, David M. J., Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts and Jonathan L. Zittrain. 2018. “The science of fake news.” *Science* 359(6380):1094–1096.
URL: *http://science.sciencemag.org/content/359/6380/1094*

Lazer, David MJ, Alex Pentland, Duncan J Watts, Sinan Aral, Susan Athey, Noshir Contractor, Deen Freelon, Sandra Gonzalez-Bailon, Gary King, Helen Margetts et al. 2020. “Computational social science: Obstacles and opportunities.” *Science* 369(6507):1060–1062.

Loewenstein, George. 1994. “The psychology of curiosity: A review and reinterpretation.” *Psychological bulletin* 116(1):75.

Mason, Lilliana. 2018. *Uncivil Agreement: How Politics Became Our Identity*. University of Chicago Press.

Mesch, Gustavo S and Stephen Coleman. 2007. New media and new voters: Young people, the Internet and the 2005 UK election campaign. In *Young Citizens in the Digital Age*. Routledge pp. 49–61.

Messing, Solomon, Christina DeGregorio, Bennett Hillenbrand, Gary King, Saurav Mahanti, Zagreb Mukerjee, Chaya Nayak, Nate Persily, Bogdan State and Arjun Wilkins. 2020. “Facebook Privacy-Protected Full URLs Data Set.”.
URL: *https://doi.org/10.7910/DVN/TDOAPG*

Munger, Kevin. 2018. “Temporal Validity in Online Social Science.”.

Munger, Kevin, Mario Luca, Jonathan Nagler and Joshua Tucker. 2018. "The effect of clickbait.".
URL: *https://rubenson. org/wp-content/uploads/2018/09/munger-tpbw18.pdf*

Osmundsen, Mathias, Alexander Bor, Peter Bjerregaard Vahlstrup, Anja Bechmann and Michael Bang Petersen. N.d. "Partisan polarization is the primary psychological motivation behind 'fake news' sharing on Twitter." Working paper available at `https://psyarxiv.com/v45bk/`.

Pennycook, Gordon and David G Rand. 2019. "Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning." *Cognition* 188:39–50.

Pennycook, Gordon, Jonathon McPhetres, Yunhao Zhang and David Rand. 2020. "Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy nudge intervention." *PsyArXiv Preprints* 10.

Stier, Sebastian, Johannes Breuer, Pascal Siegers and Kjerstin Thorson. 2020. "Integrating survey data and digital trace data: key issues in developing an emerging field.".

Tucker, Joshua A., Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal and Brendan Nyhan. 2018. "Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature." *Report Commissioned by the Hewlett Foundation* .
URL: *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3144139*

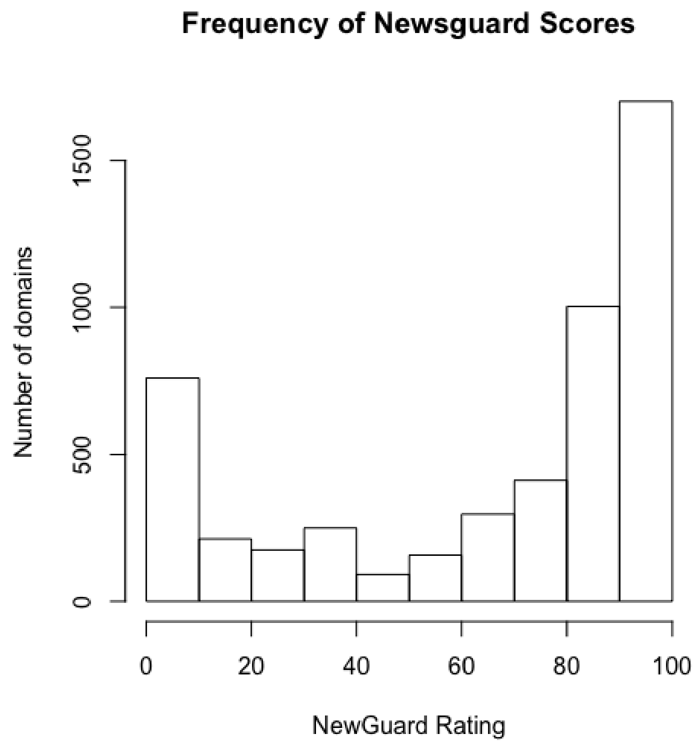
Van Bavel, Jay J, Elizabeth Ann Harris, Philip Pärnamets, Steve Rathje, Kimberly Doell and Joshua A Tucker. 2020. "Political psychology in the digital (mis) information age: A model of news belief and sharing.".

Vosoughi, Soroush, Deb Roy and Sinan Aral. 2018. "The spread of true and false news online." *Science* 359(6380):1146–1151.
URL: *http://science.sciencemag.org/content/359/6380/1146*

Wittenberg, Chloe and Adam J Berinsky. 2020. "Misinformation and its correction." *Social Media and Democracy: The State of the Field, Prospects for Reform* p. 163.

Online Appendix

Figure 13: Histogram of NewsGuard Scores



Coding instructions for research assistants coding URLs as political or not:

For each headline/blurb you will determine if the main subject story informing readers of events that relate to government institutions, policy, or political actors. If the headline or blurb directly references a political figure it is political. The central claim has to fit this definition. Below are some example of headlines that are political and headlines that are not political:

Political Headlines:

(1) For the first time, prosecutors have tied President Trump to a federal crime. But can a sitting president be indicted?
(2) Poll: Freedom Caucus Republican losing to Dem challenger in longtime GOP-held district
(3) China says U.S. should withdraw arrest warrant for Huawei executive
(4) Border Patrol agent kills undocumented woman in Texas
(5) Migrants describe hunger and solitary confinement at for-profit detention center

Not Political Headlines:

(1) American Airlines agent saves two teenage girls from human trafficking scheme
(2) Sharks vs. humans: At 100 million deaths against 6 each year, it's not a fair fight
(3) Puerto Rico says hurricane death toll 20 times higher than first reported
(4) Men Outtalk Women 2 to 1 on the Big Screen, Study Finds
(5) Longwood Area Evacuated, Reported Gas Main Leak: Boston Fire

Note: You can use background knowledge and context to help you make judgements on these articles. If you are unaware of the situation, please do online research to obtain necessary context.

Inter-Rater Reliability Statistics:

Table 2: Inter-Rater Reliability Statistics For The Training Data For The Political Classifier

Coders	Agreement	Cohen Kappa Score	Headlines/Blurbs Coded By Both
01 and 02	0.844	0.699	531
01 and 03	0.874	0.739	509
02 and 03	0.865	0.730	453

Table 3: Precision and Recall of Political Classifier:

Measure	Political	Non-Political	Average
Accuracy	0.90	0.90	0.90
Precision	0.91	0.88	0.90
Recall	0.89	0.91	0.90
F1	0.90	0.90	0.90

Table 4: Credible News Domains in Dataset and Their Ideological Lean

	Domain	Ideological Lean
1	alternet.org	Liberal
2	apnews.com	Liberal
3	atlantablackstar.com	Liberal
4	attitude.co.uk	Unclear
5	axios.com	Liberal
6	bizpacreview.com	Conservative
7	blackamericaweb.com	Liberal
8	blavity.com	Liberal
9	bloomberg.com	Liberal
10	breitbart.com	Conservative
11	businessinsider.com	Liberal
12	buzzfeednews.com	Liberal
13	c-span.org	Conservative
14	campusreform.org	Conservative
15	cbsnews.com	Liberal
16	circa.com	Conservative
17	cjr.org	Liberal
18	cnbc.com	Conservative
19	cnn.com	Liberal
20	cnsnews.com	Conservative
21	commentarymagazine.com	Conservative
22	commondreams.org	Liberal
23	conservativereview.com	Conservative
24	cookpolitical.com	Liberal
25	counterpunch.org	Liberal
26	crooksandliars.com	Liberal
27	csmonitor.com	Liberal
28	dailycaller.com	Conservative
29	dailydot.com	Liberal
30	dailysignal.com	Conservative
31	dailywire.com	Conservative
32	democracynow.org	Liberal
33	dennismichaellynch.com	Unclear
34	economist.com	Liberal
35	elitedaily.com	Liberal
36	endtimeheadlines.org	Conservative
37	express.co.uk	Conservative
38	factcheck.org	Liberal
39	fivethirtyeight.com	Liberal
40	forbes.com	Liberal

Table 5: Credible News Domains in Dataset and Their Ideological Lean

	Domain	Ideological Lean
41	foreignaffairs.com	Liberal
42	foreignpolicy.com	Liberal
43	fortune.com	Liberal
44	foxbusiness.com	Conservative
45	foxnews.com	Conservative
46	freebeacon.com	Conservative
47	gallup.com	Conservative
48	good.is	Liberal
49	governing.com	Liberal
50	harpers.org	Liberal
51	hbr.org	Liberal
52	hotair.com	Conservative
53	huffingtonpost.com	Liberal
54	ibtimes.com	Liberal
55	ijr.com	Conservative
56	ijr.com	Unclear
57	inquisitr.com	Unclear
58	inthesetimes.com	Liberal
59	jacobinmag.com	Liberal
60	jezebel.com	Liberal
61	judicialwatch.org	Conservative
62	justsecurity.org	Liberal
63	lawfareblog.com	Liberal
64	lifezette.com	Conservative
65	mashable.com	Liberal
66	mcclatchydc.com	Liberal
67	mediaite.com	Conservative
68	mediamatters.org	Liberal
69	mic.com	Liberal
70	motherjones.com	Liberal
71	mrctv.org	Conservative
72	msmagazine.com	Liberal
73	msnbc.com	Liberal
74	nationalinterest.org	Conservative
75	nationalreview.com	Conservative
76	nbcnews.com	Liberal
77	newrepublic.com	Liberal
78	newsbusters.org	Conservative
79	newsmax.com	Conservative
80	newsone.com	Liberal

Table 6: Credible News Domains in Dataset and Their Ideological Lean

	Domain	Ideological Lean
81	newsweek.com	Liberal
82	newyorker.com	Liberal
83	nowthisnews.com	Liberal
84	npr.org	Liberal
85	nymag.com	Liberal
86	nypost.com	Conservative
87	nytimes.com	Liberal
88	observer.com	Liberal
89	ozy.com	Liberal
90	pbs.org	Liberal
91	pewresearch.org	Liberal
92	politico.com	Liberal
93	politicususa.com	Liberal
94	politifact.com	Liberal
95	poynter.org	Liberal
96	propublica.org	Liberal
97	prospect.org	Liberal
98	psmag.com	Liberal
99	publicintegrity.org	Liberal
100	qz.com	Liberal
101	rare.us	Conservative
102	rasmussenreports.com	Conservative
103	rawstory.com	Liberal
104	realclearpolitics.com	Conservative
105	reason.com	Conservative
106	reuters.com	Liberal
107	revealnews.org	Liberal
108	rollcall.com	Liberal
109	salon.com	Liberal
110	slate.com	Liberal
111	snopes.com	Liberal
112	spectator.org	Conservative
113	splcenter.org	Liberal
114	splinternews.com	Liberal
115	talkingpointsmemo.com	Liberal
116	taskandpurpose.com	Conservative
117	theamericanconservative.com	Conservative
118	theatlantic.com	Liberal
119	theblaze.com	Conservative
120	theconversation.com	Liberal

Table 7: Credible News Domains in Dataset and Their Ideological Lean

	Domain	Ideological Lean
121	thedailybeast.com	Liberal
122	theepochtimes.com	Conservative
123	theguardian.com	Liberal
124	thehill.com	Liberal
125	theintercept.com	Liberal
126	themarshallproject.org	Liberal
127	thenation.com	Liberal
128	therealnews.com	Liberal
129	theroot.com	Liberal
130	theweek.com	Liberal
131	thinkprogress.org	Liberal
132	time.com	Liberal
133	townhall.com	Conservative
134	tribunist.com	Unclear
135	truth-out.org	Liberal
136	truthdig.com	Liberal
137	univision.com	Liberal
138	upi.com	Conservative
139	usatoday.com	Liberal
140	usnews.com	Liberal
141	vanityfair.com	Liberal
142	vice.com	Liberal
143	voanews.com	Liberal
144	vox.com	Liberal
145	washingtonexaminer.com	Conservative
146	washingtonmonthly.com	Liberal
147	washingtonpost.com	Liberal
148	washingtontimes.com	Conservative
149	weeklystandard.com	Conservative
150	worldpoliticsreview.com	Liberal
151	wsj.com	Conservative

Table 8: Low-Quality News Domains in Dataset and Their Ideological Lean

	Domain	Ideological Lean
1	100percentfedup.com	Conservative
2	ahtribune.com	Conservative
3	americanjournalreview.com	Conservative
4	americanthinker.com	Conservative
5	anonews.co	Unclear
6	awarenessact.com	Unclear
7	bb4sp.com	Conservative
8	beforeitsnews.com	Unclear
9	bipartisanreport.com	Liberal
10	blacknews.com	Liberal
11	clashdaily.com	Conservative
12	collective-evolution.com	Unclear
13	conservativedailypost.com	Conservative
14	conservativefiringline.com	Conservative
15	conservativetribune.com	Conservative
16	dailykos.com	Liberal
17	dcclothesline.com	Conservative
18	disclose.tv	Unclear
19	downtrend.com	Conservative
20	drudgereport.com	Conservative
21	eaglerising.com	Conservative
22	en-volve.com	Conservative
23	frontpagemag.com	Conservative
24	healthnutnews.com	Unclear
25	higherperspectives.com	Unclear
26	huzlers.com	Unclear
27	ilovemyfreedom.org	Conservative
28	infowars.com	Conservative
29	intellihub.com	Conservative
30	joeforamerica.com	Conservative
31	madworldnews.com	Conservative
32	medicalkidnap.com	Unclear
33	mintpressnews.com	Conservative
34	naturalnews.com	Unclear
35	neonnettle.com	Conservative
36	nowtheendbegins.com	Conservative
37	oann.com	Conservative
38	palmerreport.com	Liberal
39	pjmedia.com	Conservative
40	projectveritas.com	Conservative

Table 9: Low-Quality News Domains in Dataset and Their Ideological Lean

	Domain	Ideological Lean
41	projectveritasaction.com	Conservative
42	puppetstringnews.com	Conservative
43	redstate.com	Conservative
44	redstatewatcher.com	Conservative
45	reverbpress.com	Liberal
46	rt.com	Conservative
47	shareblue.com	Liberal
48	stream.org	Conservative
49	teaparty.org	Conservative
50	theconservativetreehouse.com	Conservative
51	thedailybanter.com	Liberal
52	thefederalist.com	Conservative
53	thehornnews.com	Conservative
54	themindunleashed.com	Unclear
55	thenationalmarijuananews.com	Unclear
56	thepoliticalinsider.com	Conservative
57	therightscoop.com	Conservative
58	trueactivist.com	Unclear
59	truepundit.com	Conservative
60	trunews.com	Conservative
61	truthuncensored.net	Unclear
62	twitchy.com	Conservative
63	uschronicle.com	Conservative
64	westernjournal.com	Conservative
65	whatdoesitmean.com	Unclear
66	wnd.com	Conservative
67	worldnewsdailyreport.com	Unclear
68	worldtruth.tv	Unclear
69	yournewswire.com	Conservative
70	zerohedge.com	Unclear

Slant Coding for Low-Quality News Sources

We determine the partisan lean of the low-quality domains by asking three independent coders to determine the partisan perspective of the all low-quality websites located in the Allcott, Gentzkow and Yu (2019) list of fake news publishers (conservative, liberal, and unclear). Coders were asked to use the headlines, the content of the articles, as well as the websites' domain and about pages to make this determination, and to classify websites that had a clear partisan affiliation based on this information. Websites were not classified as liberal or conservative unless at least 50% of their content appeared to have a partisan or political nature. If websites did not meet this threshold they were classified as unclear. If the coders did not unanimously agree, a fourth coder was asked to evaluate the website, and the majority decision was used (split cases were included as active). There was over 75% level agreement among the coders and the Fleiss' Kappa is .705. In total, six domains were placed in the liberal low-quality news stream, fifty domains were placed in the conservative low-quality news stream, and forty-three domains were placed in the unclear low-quality news stream. The prevalence of conservative and unclear low-quality news streams is in line with previous research that provides evidence for the asymmetric production of false/misleading news (Guess, Nyhan and Reifler 2020).

Figure 14: All News URLS Views by Ideology of User. 95 percent confidence intervals are displayed.

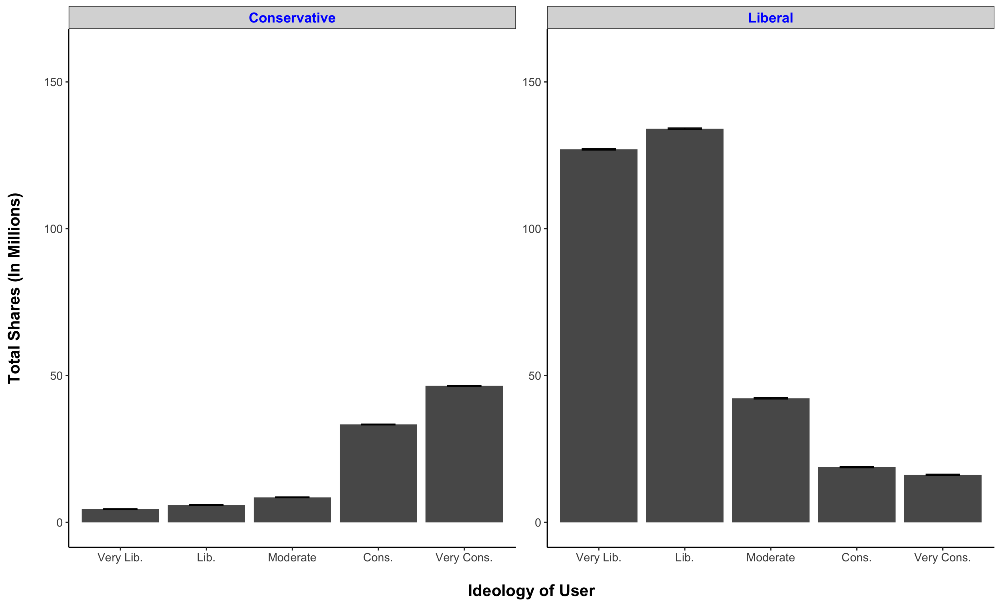


Figure 15: All News URLS Views by Ideology of User. 95 percent confidence intervals are displayed.

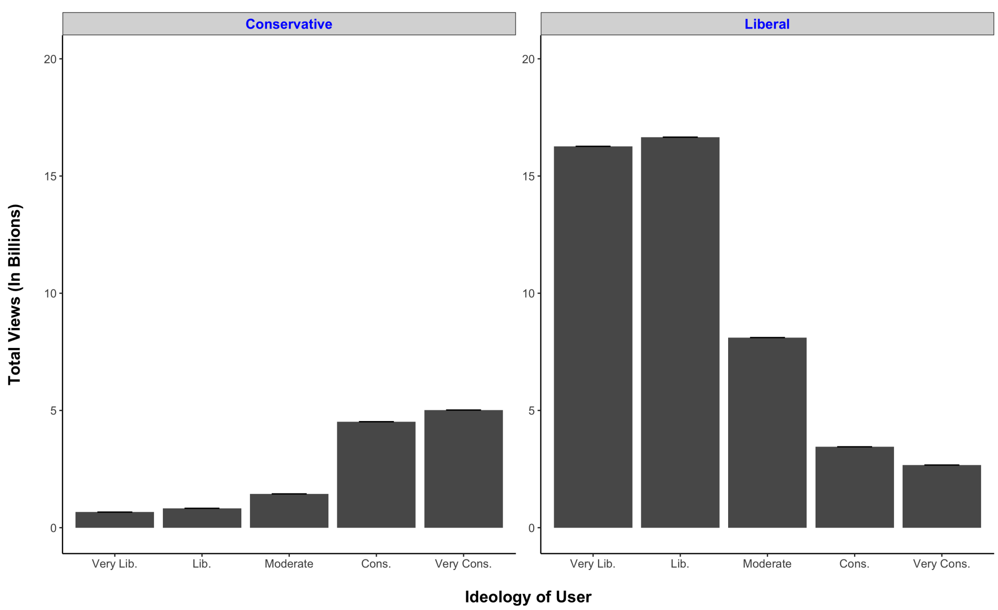


Figure 16: All News URLS Shares by Whether the Article is Clickbait from Low-Quality News Domains. 95 percent confidence intervals are displayed.

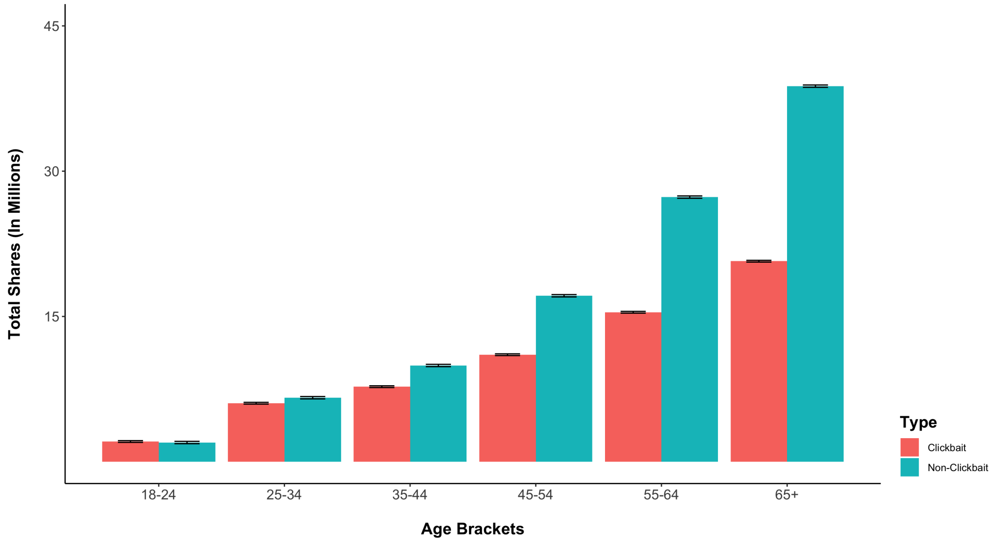


Figure 17: All News URLS Shares by Whether the Article is Political from Low-Quality News Domains. 95 percent confidence intervals are displayed.