

Understanding Online News Behaviors

Frank Bentley
Yahoo/Oath
Sunnyvale, CA
fbentley@oath.com

Katie Quehl
Yahoo/Oath
Sunnyvale, CA
katiequehl@oath.com

Jordan Wirfs-Brock
Yahoo/Oath
Sunnyvale, CA
jordan.wirfs-brock@colorado.edu

Melissa Bica
Yahoo/Oath
Sunnyvale, CA
melissa.bica@colorado.edu

ABSTRACT

The news landscape has been changing dramatically over the past few years. Whereas news once came from a small set of highly edited sources, now people can find news from thousands of news sites online, through a variety of channels such as web search, social media, email newsletters, or direct browsing. We set out to understand how Americans read news online using web browser logs collected from 174 diverse participants. We found that 20% of all news sessions started with a web search, that 16% started from social media, that 61% of news sessions only involved a single news domain, and that 47% of our participants read news from both sides of the political spectrum. We conclude with key implications for online news, social media, and search sites to encourage more balanced news browsing.

CCS CONCEPTS

• **Human-centered computing** → *Empirical studies in HCI*; *Social media*; • **Applied computing** → *Media arts*;

KEYWORDS

News; Web Search; Log Analysis; Polarization; Bias

ACM Reference Format:

Frank Bentley, Katie Quehl, Jordan Wirfs-Brock, and Melissa Bica. 2019. Understanding Online News Behaviors. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019)*, May 4–9, 2019, Glasgow, Scotland UK. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3290605.3300820>

1 INTRODUCTION

Traditionally, if one wanted to read news they could turn to their local newspaper of record. These papers were highly

edited and typically produced daily. While sometimes these sources represented the opinions or political views of the editor, they were generally seen as a trusted source of truth to understand what was happening in the community and the world. From the 1666 publication of the London Gazette to Publick Occurrences, America’s first newspaper in 1690, through the local and regional newspapers that remained strong through most of the 20th century, when it came to reading written news, very little changed.

However, over the past few decades the news landscape in America has changed dramatically. The web and alternative news sources have added many choices to the ways that people can find and consume news. In addition, the collapse of local, daily newspapers [2] and the growing 24-hour cable news cycle have provided additional changes.

There are now thousands of news sources available at any time online, representing a wide variety of political viewpoints and regional, national, or international focus. Now, users can find a single article in isolation, through a web search, social media post, email newsletter, or other direct links from the web or from a friend. Browsing through all of the top stories of the day, as one would read a newspaper, is becoming less common as news arrives in different ways through direct article links [21].

These changing practices highlight a number of questions around how news is consumed in 2018. Although we have conducted hundreds of qualitative interviews about news behavior over the years, we wanted to more deeply understand actual behaviors in news consumption over remembered interactions. We wanted to quantify the role of different referring sites (e.g. social networks, search, email) in prompting news consumption as well as explore temporal and session-based statistics about news consumption on the web — topics that are not well covered in the existing news literature.

Specifically, we had the following research questions:

- (1) How do people get to news articles on the web? What percent are reached through search vs. social media?
- (2) What temporal patterns exist in browsing news on the web? Are there particular types of news that are consumed most often in the morning? On weekends?
- (3) How many different news domains do people visit? Do people focus their attention on one end of the political spectrum? Or do they seek broad viewpoints?

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI 2019, May 4–9, 2019, Glasgow, Scotland UK

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-5970-2/19/05...\$15.00

<https://doi.org/10.1145/3290605.3300820>

To answer these questions, we collected complete web browsing logs from the computers of 174 diverse Americans, covering an average of 138 days of browsing. We then analyzed these logs to identify web sessions that included news, the referring source that led to news articles, the topics of articles that were read, and the political biases of the sources that participants turned to. This data has helped us to more deeply understand news browsing behaviors in 2018.

Our contribution lies in quantifying how participants get to news and the prevalence of different sources. We also show the diversity of bias in the sources that people turn to and temporal patterns throughout the day in the topics that people are reading. This data is important for the design of future news platforms, social networks, search engines, and email newsletters, as well as for the American political process itself.

2 RELATED WORK

Researchers have been studying how people receive news information for quite some time. Modern research into this topic in the 1980s and 90s focused on the role of TV broadcast news versus newspapers in how people receive news. Chaffee and Frank [5] explored the differing roles that TV news and newspapers played in bringing information to different audiences, with newspapers being read by people more actively seeking news and television reaching groups that were more lacking in political information.

Prior [20] pointed to the effects of many of these changes in the news media landscape. Cable news contributes to polarization in political opinions, and viewing habits vary greatly, with some households watching many hours a day but most households not watching at all. He also showed that access to the Internet had “widened gaps in news exposure [and] political knowledge” with some engaging much more than others with online news.

Tewksbury [24] found two different types of news consumers, one group that actively attended to a small number of topics, and a larger group that browsed a wider variety of topics. Looking deeper at the people who do not engage with the news, Eliasoph [9] wrote the stigma Americans have around talking about politics at work or in public has led to diminished awareness of political issues and general apathy towards much of the news and politics in the country.

Online News Consumption

The Internet brought new ways for people to get their news, changing the paradigm of broadcast and print to include on-demand access to individual articles. Goel et al. [15] conducted a study in 2012 analyzing how people browse the web online. Interestingly, in this study, they found that the average user only browsed five news pages per month. We

find this number surprisingly low, but do not have the authors’ definition of a news site to know if it included celebrity, sports, or local news.

Other researchers specifically explored news browsing in more detail. Tewksbury [23] analyzed web browsing logs from 13 specific news sites using data from 2000. They analyzed the specific topics of news that users viewed (e.g. sports, politics, weather, etc.) finding that 54% of users only browsed news in a single topic over the two months of data that they collected.

Purcell et al. [21] conducted a survey of news behaviors, finding that American news habits are based on “foraging and opportunism” and that Americans report visiting between two and five online news sources. Kleppe and Otte [16] studied the news browsing behaviors of young Dutch participants and found the majority of news browsing session started on news home pages. Flaxman et al. [11] explored the news behaviors of participants who more heavily engaged in news, finding that the majority of their browsing started on a news home page and that only 6% of sessions started from social media sites. We expected things to have changed dramatically since their 2013 data, with the rise of social media potentially driving users to a broader set of sites.

Social Media and the Filter Bubble

More recent changes in online news occurred through the development of social media platforms as a place to share and discuss news. This opened up many new ways for people to discover a news article online. Instead of needing to actively choose to go to a news website or explicitly search for a news topic, now news began to passively arrive on social media in posts shared from friends and family or from news sources that a user follows.

In 2013, Weeks and Holbert [25] studied social media news users, finding them to be largely young, technologically literate, and infrequent consumers of newspapers or television news. They also found that political partisans share news on social media more often than moderates, potentially leading to a more biased news experience for those who follow them.

Lottridge and Bentley [18] studied the wider variety of ways that people could share news with each other – including privately in text messaging, semi-publicly on social media, or publicly on sites such as Twitter or Reddit. They found differences in the types of news shared on each platform, with the most polarizing political content shared more publicly. This has clear implications for the types of news people receive if they are tuned into these social sources.

Social media platforms have led to what has been termed the “filter bubble,” [19] where users are said to receive news of a particular political bias due to the accounts that they follow and the friends that they have on social media.

Multiple studies have explored partisanship online and the political leanings of certain news publications. Pew Research has tracked political polarization of both news sources and the general population over time with reports in 2014 [7] and 2017 [8]. The 2017 study shows a large-scale shift in the political leanings of Americans to the left over time.

Various sites such as All Sides¹ and Media Bias/Fact Check² maintain ratings of news sources and their biases. A recent study from the Harvard Berkman Center [10] also explored the political biases of a number of sources based on how they were shared online. Researchers such as Resnick et al. [22] have explored strategies for one to expand their filter bubble to be exposed to articles from more diverse viewpoints.

In addition is the rise of so-called “fake news” — stories that are published but not true. Flinham et al. [12] explored people’s ability to identify “fake news” online and found that one third of their respondents had been fooled by fake news in the past.

This related work raised more questions for us than it answered. Platforms such as Facebook or Twitter have changed the way that news is presented and sorted in the feed over the years, which could also have a large impact on how people get to news from social media sites. We wanted to take a fresh look at how people browse news on their computers, how they get to news articles (e.g. via search, social media, or browsing), and how users engage with multiple news sites within a news session. This would enable us to develop a broad understanding of current news practices on the web in America, as well as how news sites and search engines can adapt to these current practices.

3 METHODS

In order to broadly understand how people are consuming news on the web, we collected a set of complete web browser histories from the computers of a diverse set of Americans. Participants were recruited on Amazon Mechanical Turk. After agreeing to participate, users were directed to an online survey which provided a detailed explanation of the data we were collecting and why, followed by instructions for how to find one’s own browser history file (for either Chrome or Firefox) stored locally on their computer. The files contained a timestamped entry for each webpage viewed for either the last 3 months (Chrome) or since the user first started using the browser (Firefox), in addition to the URL and page title.

We paid participants \$5 for their browser history files, and received valid data from 174 participants, totaling nearly 10 million unique page views over an average of 138 days of history per participant. This amounted to an average \$60/hour

wage and is in line with existing research that shows browsing history being valued at about the price of a Big Mac [4]. These participants are representative of the general US adult population in terms of age (18-72), gender (49% female), and household income (median \$50k), and reside in 39 distinct US states. Previous studies have shown that MTurk samples can be quite accurate when studying technology use in the broader American population [3].

We then created a list of 1,160 unique news domains. We began with the list of most trafficked news sites from the comScore Media Metrix [6] and added the local news sites of major television affiliates (ABC, NBC, CBS, Fox) and local newspapers that are a part of the ten largest media conglomerates. The analysis below will count any page view from one of these domains as a “news page.” Any page in this set with a URL longer than 50 characters was labeled as an “article page.” We manually checked several hundred URLs and found no misclassifications on articles given this approach.

To understand the topics of articles, we ran each article through the Yahoo Content Analysis Service.³ This system returned Yahoo Content Taxonomy (YCT) and Wikipedia tags for each article. YCT is a hierarchical classification system for news content, with top level tags such as “news,” “entertainment,” and “sports,” and lower-level categories such as “politics,” “movies,” and “football.”

To understand the political bias of news sources that participants visited, we utilized the classification from Media Bias/Fact Check.⁴ This was the largest database of sources that we could find, with 1,434 sources listed in five categories from Left Bias (-2), Left-Center Bias (-1), Least Biased (0), Right-Center Bias (1) and Right Bias (2). All domains that users visited were tagged according to this taxonomy.

While other lists of bias exist from sources such as Pew [1], these lists contain far fewer sources. In general, these lists highly agree on their ratings, with sources such as HuffPost and Mother Jones on the far left, National Public Radio (NPR) and the New York Times to the center-left, Reuters and the Associated Press (AP) as unbiased, the New York Post and Forbes as center-right, and Breitbart and Fox News on the far right.

We compared the Media Bias/Fact Check list with a list of 115 sources from a recent paper from the Berkman Center at Harvard [10] and found a 0.77 Pearson correlation with $p < 2.2 \times 10^{-16}$ when comparing bias ratings for all 87 sources that the lists had in common. We also correlated the list from Media Bias/Fact Check with 105 sources listed in the Media Bias Chart from Ad Fontes Media⁵. We found a 0.92 Pearson correlation with $p < 2.2 \times 10^{-16}$ for the 85 sources

¹<https://www.allsides.com/media-bias/media-bias-ratings>

²<https://mediabiasfactcheck.com/>

³<https://developer.yahoo.com/search/content/V2/contentAnalysis.html>

⁴<https://mediabiasfactcheck.com/>

⁵<https://www.adfontesmedia.com>

these lists shared. Finally, we compared the larger list with a set of 200 matching sources from the AllSides Media Bias Ratings.⁶ Here, we found a Pearson correlation of 0.81 with $p < 2.2 * 10^{-16}$. These correlations are strong enough for us to trust the larger partisanship classification of 1,434 sources for our analysis. It is important to have such a large set of sources in order to accurately determine the broader media consumption diets of our participants — as described below our participants visited hundreds of these sources.

We analyzed the web history data in terms of web “sessions.” We used a one-hour-idle session delimiter to break web browsing activity into distinct sessions of interaction. This is standard practice, following guidance in Lalmas et al. [17] and is widely used throughout the Internet industry to segment online behavior.

All research was approved by our institutional processes for conducting work with human subjects and log data. Participants were clearly informed about our institutional identity, the exact data that was being collected, and our data retention policies.

4 FINDINGS

Our dataset contained 9,487,564 total page views from our 174 participants. Using one hour idle session delimiters, there were 43,415 total web browsing sessions in this dataset. In this section, we will explore general news reading behaviors, how users arrived at news articles, temporal patterns, and behaviors within sessions.

To check the representative nature of our dataset, we compared visitation rates with audience data from the ComScore Media Metrix [6] from June 2018 (the month of our study) in the “News/Information” category. This report includes the number of visitors for the each of the top news sites in the United States. We performed a Pearson correlation between the audience numbers for the top 100 sites and the number of users in our sample who visited each site. The correlation was quite strong, at 0.87 ($p < 0.001$), showing that our participants visited the same news sites in similar proportions to the overall US population.

General News Behaviors

Within our dataset, there were 126,298 news page views in 10,026 total sessions (an average of 5.3 news pages in 0.4 sessions per day). Interestingly, 23% of all web browsing sessions contained news page views, even though news pages only accounted for 1.3% of all web page views. For our participants, news was a frequent activity, but not one that often involved deep exploration. There were a total of 83,164 news article views in the dataset, at an average of 3.6 per user per day. However, the median number of articles per day was

Figure 1: Total number of articles viewed (left) and unique news domains visited (right) per participant.

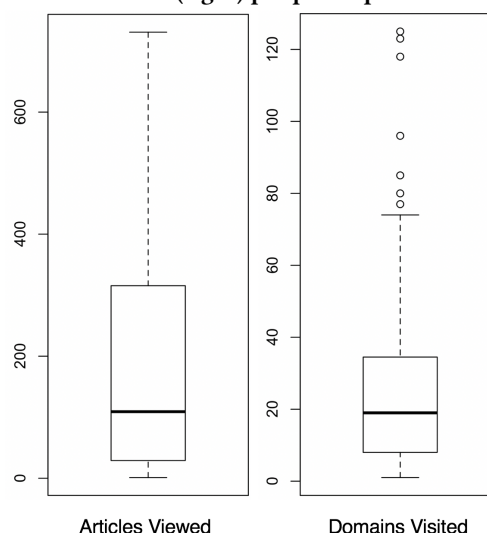
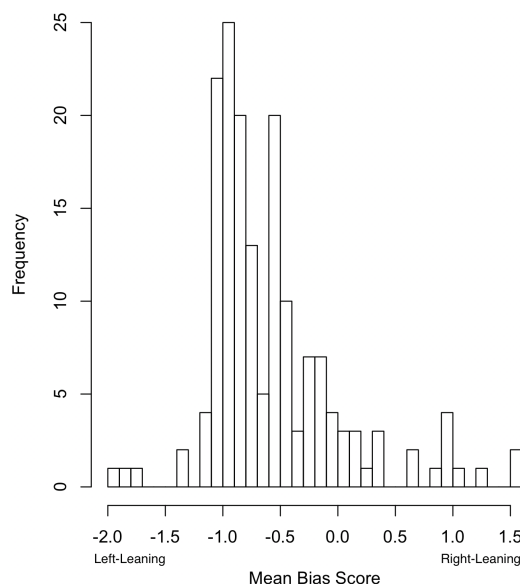


Figure 2: Histogram of mean media bias scores for the articles that each user read.

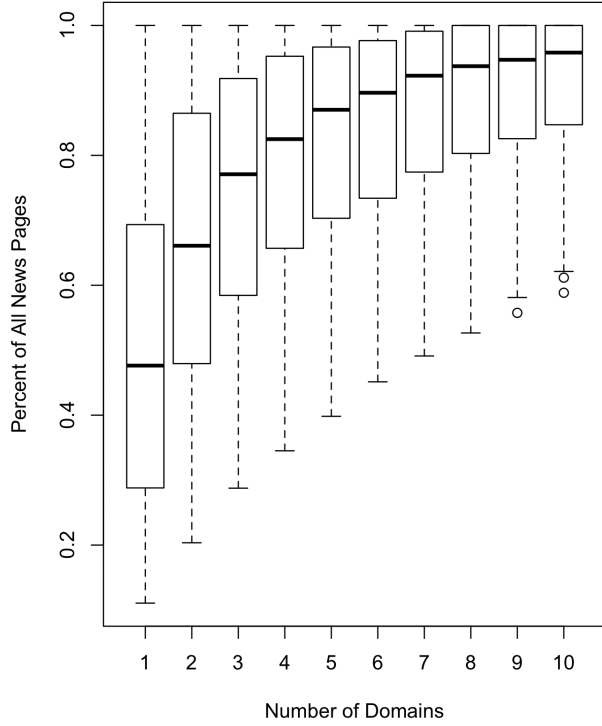


0.8, showing a definite skew in news reading across the population, aligning with findings from Prior [20] about the vast differences in news consumption in America.

Participants visited a mean of 25 different news domains in the average of 138 days of data that we collected, with a median of 19. Figure 1 illustrates the wide range of unique news sources visited by each participant. Our dataset includes visits to a broad range of sources—everything from HuffPost on the left to The Drudge Report on the right. Figure 2 illustrates

⁶<https://www.allsides.com/media-bias/media-bias-ratings>

Figure 3: The percentage of news that a user receives from their top ten sources. The median user receives 47% of their online news from just a single source.

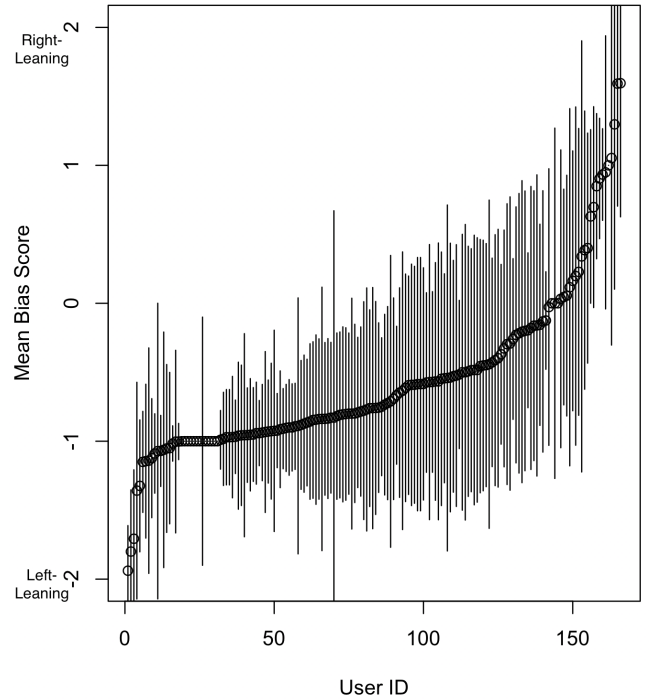


that mean media bias scores for articles that each user read, showing a range of left and right leaning sources, with a majority falling moderately left of center. We find it interesting that every participant visited at least one news site, with 75% of participants viewing more than eight distinct sites and 25% of participants viewing more than 34. One participant visited 120 distinct news sites in their dataset. This is much higher than the two to five sites reported in previous survey-based studies [21] and illustrates why behavioral studies are necessary to gain an accurate ground truth in topics where users might not consider or remember each interaction. In addition, as mentioned in the Introduction, there are more news sites available online today than in the past.

Exploring the domains further, we found that the median participant read 47% of their news articles from a single source. Figure 3 shows the distribution of news read from a user’s top ten news sources. By the second source, the median user has read 66% of all of the news articles that they consume. This increases to 87% by the fifth source and 96% by the tenth source.

Next, we explored the variation of sources visited per participant. Were users in their own “filter bubbles” or did

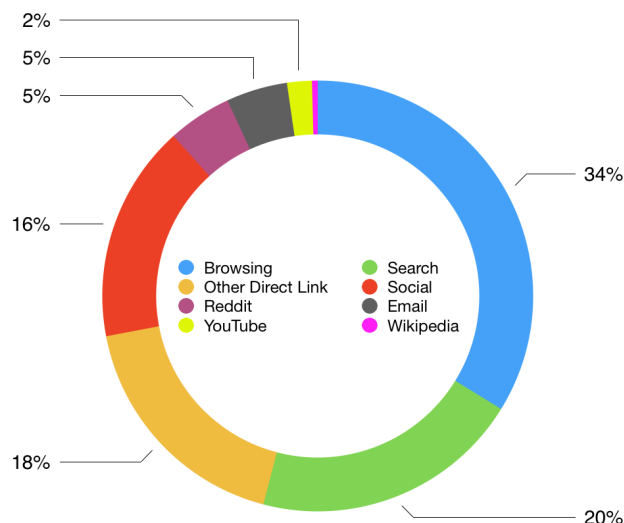
Figure 4: Mean (circle) and standard deviation (lines) of media bias scores for the articles that each user read. Users are arranged by increasing mean score.



they view a variety of sources? Figure 4 shows the mean media bias scores of each participant in our study based on all articles read. Error bars indicate the standard deviation. The majority of participants read articles from sources that leaned to the left – 87% of participants had a mean media bias score less than zero. When taking into account standard deviation, only 48% received the vast majority of their news from left-leaning sources. Only 5% of participants had a ($mean - standarddeviation$) score above zero, with the vast majority of their news coming from right-leaning sources.

The relatively high average standard deviation (0.66) shows that most participants took in news from at least some differing perspectives. In fact, 47% of participants had standard deviation lines that crossed the center line. Perhaps the polarization is not as strong as some would indicate, although we do see the general left-leaning bias in the articles read, which matches similar data from Pew about the American population as a whole [8] and matches behaviors seen in the broader population through the ComScore Media Metrix data [6] where only one of the top ten online news sources was right-leaning (foxnews.com) as of June, 2018. Our data also aligns with the political views of the general population: 38% of our participants viewed at least one article on

Figure 5: The origin of the first article view in a news session. Browsing for news stories represents only 34% of the ways participants landed on an article.



foxnews.com, compared to Donald Trump’s 41% approval rating at the time of the study [14] and 11 percentage points larger than the size of the Republican base (27% in June 2018) [13].

How People Get to News Articles

News sessions can start in many ways. Historically, news was something that one browsed for or received passively. One could read through a newspaper and stop on articles that were interesting or listen to the evening news on the radio or TV absorbing the stories that were shared, but with no control over what was broadcast. The Internet brought many new ways to discover news content. Now, people can browse the news on a news website or portal, similar to browsing a newspaper. But they can also get to articles through social media posts, web searching, email links, and other online sources such as Reddit, YouTube, or Wikipedia.

We were interested in examining the ways that users started a news session. We took the first news article pageview in each news session and found the referring page in the browser logs. Figure 5 shows the different ways that our participants arrived at on the first news article of a session. Although it’s the most frequent single way to get to an article, only 34% of articles were reached through browsing on a news site or portal (such as going to <http://www.nytimes.com> and clicking on a story). Given that this was the original way that news was read, this relatively small percentage represents a massive change to the way that news is consumed.

The remaining two-thirds of the time, users reached a story by a link from another site or shared to them by a friend. Web search represented the largest portion of these direct links, with 20% of news sessions starting from the result of an online search. Links from Social Media sites (e.g. Facebook, Twitter, LinkedIn) led to 16% of news sessions, showing the growing importance of these sites in shaping one’s media diet compared to the 6% observed in 2013 [11]. Other online sites such as Reddit (5%), YouTube (2%) and Wikipedia (0.4%) also led users to the news, in addition to email (including Gmail, Yahoo Mail, and Outlook/Hotmail domains) at 5%. We find the high volume of news preceded by a search at 20% to be quite high, a point we will return to in the discussion.

We observed some variation in how users arrived at news stories of different types. Table 1 shows the breakdown of different referral mechanisms for different types of news content. Finance, Sports, and Travel news had the highest percentage of searches leading to an article. Health and Politics led social referrals and were also at the top for email referrals. Technology, Politics, and Sports led in articles found via browsing a news site. Entertainment, Health, and Finance were the least likely to be found via browsing on a news site.

We will now turn to media bias scores by referring source. The mean absolute value of the bias score from all articles read was 1.02. The most biased articles came from social media (1.07) and email (1.04), with the least biased articles coming from search results (0.92) and Reddit (0.93). We find the lower bias scores from search results to be interesting, and will return to implications for search in the Discussion. We also find the higher bias in email and social media to be interesting, likely because users receive email update newsletters from more biased sources (e.g. the daily “Huff-Post Morning Email”) As has been shown by Weeks et al. [25], partisan users share more frequently on social media.

Temporal Patterns of News Reading

Next, we were interested in the temporal patterns of news reading on the web. Do people not want to wake up to serious news? Or do they catch up on celebrity news in the evening? We used our behavioral data from the logs to confirm or deny these and other statements about news reading. All analysis was conducted in the local time zone of the user, and we segmented all news articles by hour of day and day of week.

Figure 6 shows the number of news sessions observed by hour of day across all of our participants. Interestingly, the highest news consumption occurs in the evening, with the three highest points being from 6–8pm. There is another small peak around noon and one at 7am. Overall, news consumption stays quite high from 6am through 9pm. This is in interesting contrast to historical patterns of news consumption, where one would wake up to the morning paper and

Category	Search	Social	Reddit	Email	YouTube	Wikipedia	Other Link	Browsing
News	23.7%	15.5%	1.4%	6.4%	0.9%	0.7%	33.9%	17.4%
Entertainment	25.1%	15.7%	1.5%	5.0%	0.8%	0.7%	37.2%	13.9%
Technology	17.3%	8.5%	2.5%	0.7%	1.1%	0.5%	41.8%	27.7%
Sports	27.0%	14.3%	1.8%	4.6%	0.4%	0.2%	29.8%	22.0%
Politics	23.1%	16.9%	0.9%	6.9%	0.7%	0.4%	27.1%	24.1%
Travel	28.3%	11.8%	1.2%	5.6%	1.6%	0.3%	33.0%	18.1%
Finance	38.8%	13.3%	0.8%	6.1%	0.0%	0.0%	26.3%	14.6%
Health	27.4%	18.6%	5.2%	5.7%	1.0%	0.2%	27.6%	14.3%

Table 1: How participants arrived at news articles of various types at the start of sessions containing news.

Figure 6: Total news sessions by hour of day.

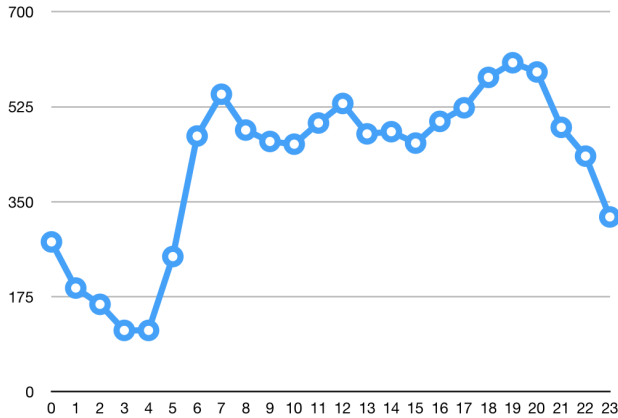
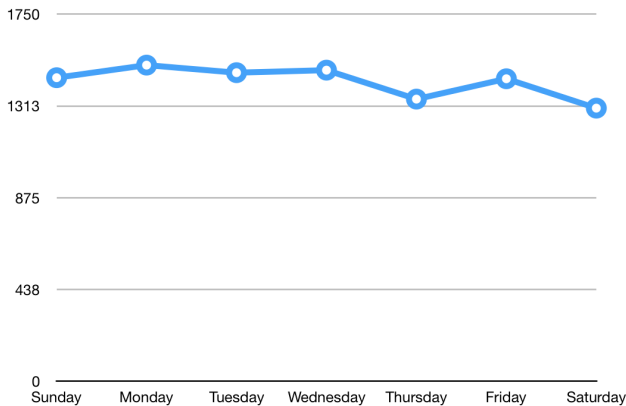
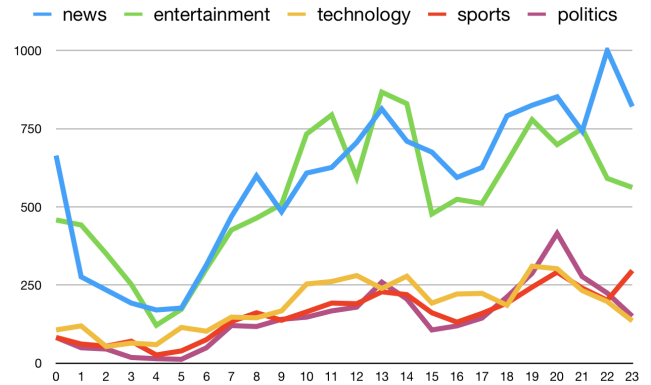


Figure 7: Total news sessions by day of week.



unless they also purchased an evening paper, took care of most of their written news consumption in the morning.

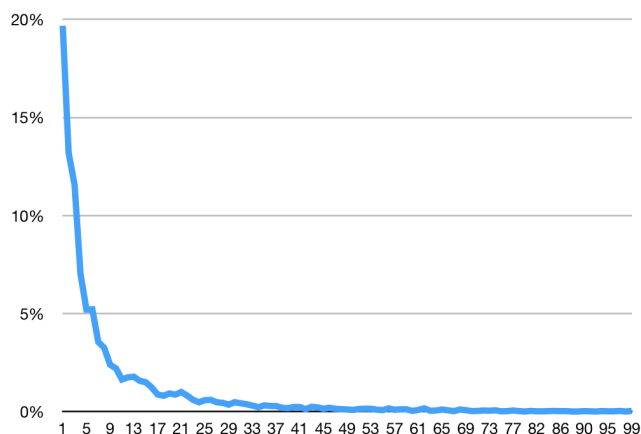
Figure 8: Topics of news articles views by hour of day.



We were also interested in news consumption by day of week. Were there particular days with more news consumption than others? Figure 7 shows that news consumption is fairly flat, with a slight decreasing trend as the week goes on from Monday to Saturday, but only a difference of 15% between the highest (Monday) and lowest (Saturday) days. Although the total number of news sessions remains fairly stable, it is interesting to note that weekends showed 50% fewer total web page views than the highest weekdays (Tuesday and Wednesday), making news a much higher percent of total web browsing on weekends.

We were further interested in the topics of articles that were read throughout the day and by day of week. As discussed in the Methods section, we obtained YCT categories for each article. We then looked at topics by time of day and day of week to see if there were specific temporal dynamics to reading news on particular topics. Figure 8 shows the top five news topics (hard news, entertainment, technology, sports, and politics) by the hour of day that the article was read. At the 8am hour, harder news is the most common type of news read. However, entertainment news surpasses hard news for the 10am, 11am, 1pm, and 2pm hours, likely

Figure 9: Number of news articles viewed in a session. 20% of all news sessions contain just one article that was read.



for a momentary break and distraction during the workday. Entertainment news peaks at 1pm. Hard news then regains the lead from 3pm through midnight with its daily peak at 10pm.

Sports news reading peaks at 8pm and 11pm, just as the evening games are kicking off and ending (the study was conducted during baseball season in America). Political news also peaks at 8pm, surpassing technology and sports reading for the 8pm and 9pm hours whereas technology news surpasses both sports and politics for much of the workday.

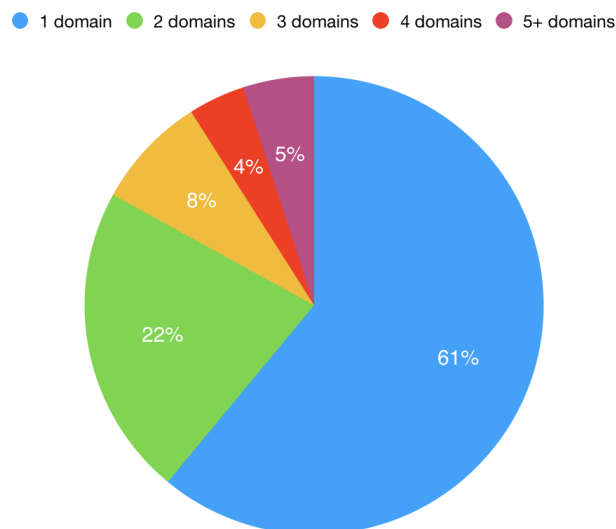
These patterns illustrate the overall popularity of news and entertainment news, which track each other for much of the day. It also shows that entertainment news is not more popular during the morning waking-up hours.

Within-Session Behavior

We now turn to look within sessions to understand how participants consumed multiple news stories within a session. Particularly, we explored if participants were viewing multiple articles on the same topics as well as the diversity of sources within the session.

Figure 9 shows the number of news articles viewed in each news session. Overall, 20% of news sessions involved only a single article, with an additional 13% containing just two articles. 43% of all news sessions had over five articles viewed. Interestingly, the source of the news session had a significant impact on the number of news articles viewed in that session. Sessions started from email links to the news resulted in the most news browsed (at an average of 14 pages). Search and Reddit both led to an average of 12 news pages being viewed, while social links to news led to an average of 11 news pages viewed within that session. All of these are larger than the

Figure 10: Number of distinct news domains browsed within a news session.



average news session of 9.5 pages viewed, given a much smaller number of articles read when directly browsing news sites. This can seem counter-intuitive, but often users arrive on news portals just to browse the headlines or on their way to read their email without clicking on any articles.

The largest number of sessions with news (61%) involved just a single domain, as shown in Figure 10. Yahoo News, The New York Times, CNN, MSN, and BuzzFeed were the most popular sources viewed in sessions with no other news domains. Sessions with two domains comprised 22% of news sessions, with the most common pairing being CNN and Yahoo, occurring 179 times. Both of these sources score as fairly unbiased in the center-left category (-1). The New York Times and Politico were the next most common pairing, occurring 102 times in the dataset. These are also both categorized as center-left. Eight percent of sessions contained three domains, 4% contained four, and 5% of sessions contained five or more news domains.

As the number of domains visited increases, the trend of browsing sites with a similar bias score continues. Overall, the news sites that are browsed together within a session have an average variance of 0.29 points of bias. Only 7.7% of sessions include sites with a bias score variance greater than one. Participants were not frequently going out of their way to find diverse accounts of a story.

Of sessions that contained articles where we could extract Wikipedia entities for people or places mentioned in the article, 36% contained more than one domain browsed. Of these sessions, 70% contained multiple articles with the

same Wikipedia entity. The most common Wikipedia entities where users viewed articles from multiple domains within the same session included: Donald Trump, White House, Melania Trump, Kim Kardashian, Catherine Zeta-Jones, Anthony Bourdain, and Star Trek. This represents a mix of political and celebrity/entertainment news similar to the overall trends across all articles viewed.

In summary, we have seen that 61% of all news sessions only involve a single domain and 20% only involve a single article. When users do branch out, they most often seek sites with similar political bias, however 7.7% of the time they look for a news source with a different viewpoint. 70% of the time that users visit multiple news domains in a session, they read a story on the same topic on multiple sites, representing 25% of all news sessions. This analysis has allowed us to more deeply understand news browsing behaviors and how people get to and explore news on a variety of topics.

5 LIMITATIONS

While this study enabled us to get a deep, behavioral look into current news consumption in the United States, there are some limitations. We are reliant on the Mechanical Turk userbase for our data, which may be more likely to use desktop computers to take their HITs. However, previous research has shown large agreement between the behaviors of this panel and the broader US population [3]. Our sample only contains data from 174 participants. While we have a high degree of diversity in the panel that matches the US population (ages 18-72, 49% female, \$50k median household income, and 39 distinct US states), this may not be enough to capture the very diverse user behaviors across over one thousand news sites. Larger samples should be studied by companies who have access to this much larger browser history data.

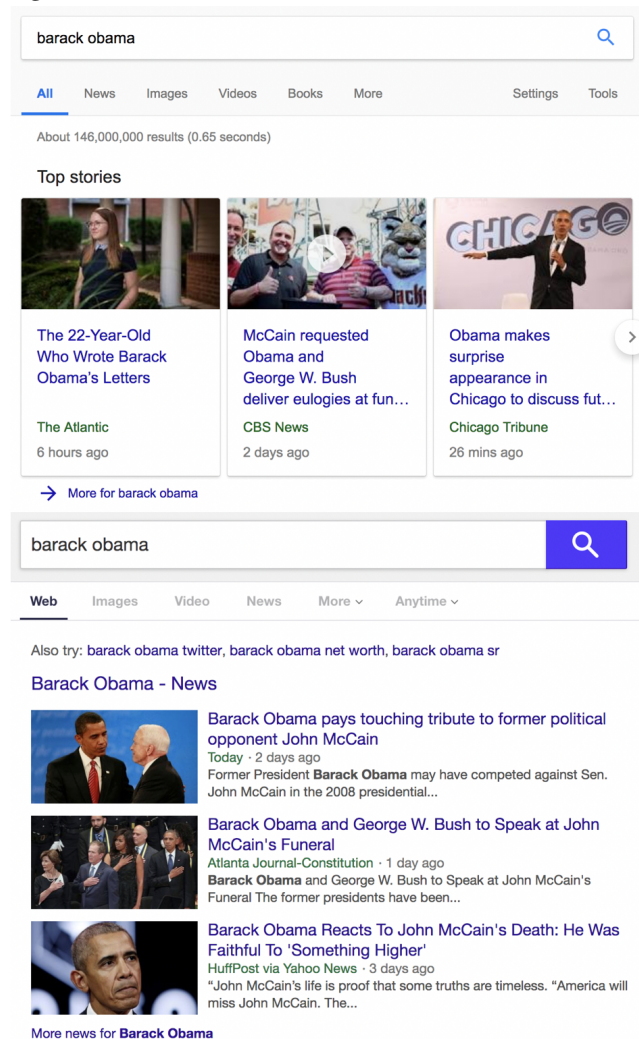
We were unable to capture mobile browsing behavior for several reasons. Most mobile applications (e.g. Facebook, Twitter) use their own embedded browsers, so news page views in these applications are not captured in mobile browsing logs. As it is also not possible to get browsing log data from within dedicated news apps (e.g. The NYT app), we focused on the desktop web for our analysis, where we could get complete logs of user behavior over months of interactions.

Finally, this dataset was collected in late June 2018. With an average of 138 days of data per participant, this means that almost all data was from the Spring. There are likely seasonal differences in use (e.g. sports during the NFL season) and conducting studies with larger log sets or at other times of the year can be useful future work.

6 DISCUSSION

By analyzing the behavioral data of a broad range of Americans and how they browse news on the web, we have been

Figure 11: The current search experience on Google (top) and Yahoo (bottom), highlighting news results in boxes on top of algorithmic search results.



able to more deeply understand news reading behaviors in a way that avoids the bias of self-reporting. We have explored how users get to news articles via search, social media, email, and other sources as well as how they interact within news sessions in terms of the political bias of the news sites visited and topics viewed by time of day.

Through this analysis, several larger themes have emerged, with key implications for the design of news platforms, as well as for the understanding of news consumption in general. These include the importance of search, ways to address temporal browsing preferences, and a reflection on news polarization and filter bubbles.

Search and News

One of the most interesting findings to us was discovering that 20% of all news sessions begin with a web search, as we did not anticipate so much traffic driven this way. We also observed that news sessions that start from search have the most neutral partisan bias of all types of news sessions.

Currently, when searching for news topics or celebrities on Google or Yahoo, a direct display box is shown with top news stories, as shown in Figure 11. Based on the findings of our analysis, we know that users seek multiple stories on the same topic in 25% of news sessions. Search engines can make this experience easier by aggregating news stories on particular topics (e.g. in the Obama example above, showing stories about Senator John McCain together and stories about his trip to Chicago together) so that users can easily find different viewpoints on the same topic.

In addition, search engines can indicate the political bias of sources and explicitly include results that lean left, right, and center so that readers can understand a breadth of opinions on a topic. Boxes might also summarize the different viewpoints based on articles of differing biases to get a quick view of the diversity of angles on a topic. For McCain, a long-serving moderate Republican senator and former POW in Vietnam who recently died, this might include centrist articles praising his service and bipartisanship, left-leaning articles on his support of healthcare and refusal to support torture, and right-leaning articles on times he has strayed from the republican mainstream or defending the White House in not keeping the flag lowered after his death. This can help users to quickly understand complex topics.

Temporal News Browsing

Also interesting to us were the temporal patterns that we observed, highlighting a peak in news browsing in the evening and harder news being more frequent than celebrity in the morning. The peaks mid-day are most often for celebrity and entertainment news.

News sites can explicitly design for these temporal patterns. They can choose the types of news that are editorially selected to be at the top of the page to match the news that people are seeking at different times. Early morning might be a summary of what happened in harder news overnight. Mid-day might include top celebrity or entertainment stories, with more breaking news at the lunch hour. The evening can highlight more sports and other general news. This can help people find the content that they are most open to reading at a given time.

Filter Bubbles

We have observed a large amount of polarization, or a “filter bubble” effect in the news habits of our 174 participants.

Almost half (48%) of participants received the vast majority (mean + standard deviation) of their news from left-biased sources while 5% received the vast majority of their news from right-biased sources. While the US-population overall is leaning farther left [8], the fact that people are not exposed to ideas from different viewpoints can lead to a deeper “bubble” where people are not aware of the ideas of other portions of the population.

One way to help with this problem is to improve search (as described above) to focus on showing a wider variety of results. But search only makes up 20% of all news sessions. News portals, such as yahoo.com, msn.com, or news.google.com can use many of the same techniques suggested for search engines to show a broader diversity of viewpoints for aggregated articles on specific topics. News sites that give rewards points for viewing articles can give additional points for reading stories on different sides of the political spectrum, or can give bonuses for reading stories from mostly unbiased sources to disincentivize reading heavily-biased sources.

We do find it heartening that 7.7% of all news sessions contained multiple news domains with different political bias scores. In addition, 47% of participants had a distribution of news sources that crossed the center. Users are seeking to break out of their bubbles at least a little bit, and this behavior can be encouraged through new features of major news portals.

7 CONCLUSION

This exploration of behavioral data of news browsing, from 174 diverse Americans over an average of 138 days, has highlighted several important trends in news consumption in 2018. Most notably, users are frequently reading news on the web, with 23% of all web browsing sessions containing news. Users are reading from a wide variety of sources with the median user visiting 19 distinct sources. The majority of users, 52%, did not receive significant amounts of news from opposing right or left leaning sources; they were in a filter bubble. We have also seen that how people are getting to news is changing and diverse — 20% of all news sessions came from search, 16% from social media, only 34% from direct browsing.

These findings have implications for the design of search and news portals and helping users to get a more balanced view of the world. We hope future work seeks to replicate these findings in other cultural contexts and in other times of the year. Logs, such as these, can also be used to examine how people explore the news around specific types of events, such as the death of a celebrity or a major political story. The news ecosystem is complex and changing rapidly, and more work that explores actual human behaviors as opposed to idealized answers in surveys or interviews is deeply needed at this time.

REFERENCES

- [1] Rachel Weisel Amy Mitchell. 2014. Political Polarization and Media Habits: From Fox News to Facebook, How Liberals and Conservatives Keep Up with Politics. (2014). <http://assets.pewresearch.org/wp-content/uploads/sites/13/2014/10/Political-Polarization-and-Media-Habits-FINAL-REPORT-7-27-15.pdf>
- [2] Michael Barthel. 2017. Despite subscription surges for largest U.S. newspapers, circulation and revenue fall for industry overall. (2017). <http://www.pewresearch.org/fact-tank/2017/06/01/circulation-and-revenue-fall-for-newspaper-industry/>
- [3] Frank R. Bentley, Nediya Daskalova, and Brooke White. 2017. Comparing the Reliability of Amazon Mechanical Turk and Survey Monkey to Traditional Market Research Surveys. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. ACM, New York, NY, USA, 1092–1099. <https://doi.org/10.1145/3027063.3053335>
- [4] Juan Pablo Carrascal, Christopher Riederer, Vijay Erramilli, Mauro Cherubini, and Rodrigo de Oliveira. 2013. Your Browsing Behavior for a Big Mac: Economics of Personal Information Online. In *Proceedings of the 22Nd International Conference on World Wide Web (WWW '13)*. ACM, New York, NY, USA, 189–200. <https://doi.org/10.1145/2488388.2488406>
- [5] Steven Chaffee and Stacey Frank. 1996. How Americans get political information: Print versus broadcast news. *The Annals of the American Academy of Political and Social Science* 546, 1 (1996), 48–58.
- [6] ComScore. 2018. ComScore Media Metrix report for News/Information. (6 2018). <https://mymetrix.comscore.com>
- [7] Michael Dimock. 2014. Political Polarization in the American Public. (2014). <http://www.people-press.org/2014/06/12/political-polarization-in-the-american-public/>
- [8] Carol Dougherty. 2017. Political Polarization, 1994–2017. (2017). <http://www.people-press.org/interactives/political-polarization-1994-2017/>
- [9] Nina Eliasoph. 1998. *Avoiding politics: How Americans produce apathy in everyday life*. Cambridge University Press.
- [10] Robert Faris, Hal Roberts, Bruce Etling, Nikki Bourassa, Ethan Zuckerman, and Yochai Benkler. 2017. Partisanship, propaganda, and disinformation: Online media and the 2016 US presidential election. (2017).
- [11] Seth Flaxman, Sharad Goel, and Justin M Rao. 2016. Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly* 80, S1 (2016), 298–320.
- [12] Martin Flintham, Christian Karner, Khaled Bachour, Helen Creswick, Neha Gupta, and Stuart Moran. 2018. Falling for Fake News: Investigating the Consumption of News via Social Media. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 376, 10 pages. <https://doi.org/10.1145/3173574.3173950>
- [13] Gallup. 2018. Party Affiliation. (2018). <https://news.gallup.com/poll/15370/party-affiliation.aspx>
- [14] Gallup. 2018. Presidential Approval Ratings - Donald Trump. (2018). <https://news.gallup.com/poll/203198/presidential-approval-ratings-donald-trump.aspx>
- [15] Sharad Goel, Jake M Hofman, and M Irmak Sirer. 2012. Who Does What on the Web: A Large-Scale Study of Browsing Behavior.. In *ICWSM*.
- [16] Martijn Kleppe and Marco Otte. 2017. Analysing and understanding news consumption patterns by tracking online user behaviour with a multimodal research design. *Digital Scholarship in the Humanities* 32, suppl_2 (2017), ii158–ii170.
- [17] Mounia Lalmas, Heather O'Brien, and Elad Yom-Tov. 2014. *Measuring User Engagement*. Morgan & Claypool Publishers. Synthesis Lectures on Information Concepts, Retrieval, and Services.
- [18] Danielle Lottridge and Frank R. Bentley. 2018. Let's Hate Together: How People Share News in Messaging, Social, and Public Networks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 60, 13 pages. <https://doi.org/10.1145/3173574.3173634>
- [19] Eli Pariser. 2011. *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.
- [20] Markus Prior. 2007. *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge University Press.
- [21] Kristen Purcell, Lee Rainie, Amy Mitchell, Tom Rosenstiel, and Kenny Olmstead. 2010. Understanding the participatory news consumer. *Pew Internet and American Life Project* 1 (2010), 19–21.
- [22] Paul Resnick, R. Kelly Garrett, Travis Kriplean, Sean A. Munson, and Natalie Jomini Stroud. 2013. Bursting Your (Filter) Bubble: Strategies for Promoting Diverse Exposure. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion (CSCW '13)*. ACM, New York, NY, USA, 95–100. <https://doi.org/10.1145/2441955.2441981>
- [23] David Tewksbury. 2003. What do Americans really want to know? Tracking the behavior of news readers on the Internet. *Journal of communication* 53, 4 (2003), 694–710.
- [24] David Tewksbury, Michelle L Hals, and Allyson Bibart. 2008. The efficacy of news browsing: The relationship of news consumption style to social and political efficacy. *Journalism & Mass Communication Quarterly* 85, 2 (2008), 257–272.
- [25] Brian E Weeks and R Lance Holbert. 2013. Predicting dissemination of news content in social media: A focus on reception, friending, and partisanship. *Journalism & Mass Communication Quarterly* 90, 2 (2013), 212–232.