Fake News on Facebook and Twitter: Investigating How People (Don't) Investigate

Christine Geeng **Savanna Yee** **Franziska Roesner**
Paul G. Allen School of Computer Science & Engineering
University of Washington
{cgeeng,savannay,franzi}@cs.washington.edu

ABSTRACT
With misinformation proliferating online and more people getting news from social media, it is crucial to understand how people assess and interact with low-credibility posts. This study explores how users react to fake news posts on their Facebook or Twitter feeds, as if posted by someone they follow. We conducted semi-structured interviews with 25 participants who use social media regularly for news, temporarily caused fake news to appear in their feeds with a browser extension unbeknownst to them, and observed as they walked us through their feeds. We found various reasons why people do not investigate low-credibility posts, including taking trusted posters' content at face value, as well as not wanting to spend the extra time. We also document people's investigative methods for determining credibility using both platform affordances and their own ad-hoc strategies. Based on our findings, we present design recommendations for supporting users when investigating low-credibility posts.

Author Keywords
Misinformation; disinformation; fake news; social media; Facebook; Twitter; trust; verification;

CCS Concepts
•**Human-centered computing** → **Social media;** *Empirical studies in collaborative and social computing;* •**Security and privacy** → Social aspects of security and privacy;

INTRODUCTION
While propaganda, conspiracy theories, and hoaxes are not fundamentally new, the recent spread and volume of misinformation disseminated through Facebook, Twitter, and other social media platforms during events like the 2016 United States election has prompted widespread concern over "fake news" online. Social media companies have taken steps to remove misinformation (unintentional false stories) and disinformation (intentional false stories) [43] from their sites, as well as the accounts who spread these stories. However, the speed, ease, and scalability of information spread on social media means that (even automated) content moderation by the platforms cannot always keep up with the problem.

The reality of misinformation on social media begs the question of how people interact with it, whether they believe it, and how they debunk it. To support users in making decisions about the credibility of content they encounter, third parties have created fact-checking databases [28, 75, 78], browser extensions [29, 63], and media literacy initiatives [8, 41, 70]. Facebook and Twitter themselves have made algorithm and user interface (UI) changes to help address this. Meanwhile, researchers have investigated how people assess the credibility of news on social media [33, 44, 49, 81]. However, prior work has typically not studied users' interactions with fake news posted by people they know on their own social media feeds, and companies have given us little public information about how people use the platforms' current design affordances.

To better understand how people investigate misinformation on social media today, and to ultimately inform future design affordances to aid them in this task, we pose the following research questions:

1. How do people interact with misinformation posts on their social media feeds (particularly, Facebook and Twitter)?
2. How do people investigate whether a post is accurate?
3. When people fail to investigate a false post, what are the reasons for this?
4. When people do investigate a post, what are the platform affordances they use, and what are the ad-hoc strategies they use that could inspire future affordances?

We focus specifically on Facebook and Twitter, two popular social media sites that many people use for news consumption [77]—note that we use the term "feed" in this paper to refer generally to both Facebook's News Feed and Twitter's timeline. We conducted an in-person qualitative study that included (a) semi-structured interviews to gain context around people's social media use and prior experience with misinformation and (b) a think-aloud study in which participants scrolled through their own feeds, modified by a browser extension we created to temporarily cause some posts to look as though they contained misinformation.

CHI '20, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
http://dx.doi.org/10.1145/3313831.3376784

Our results show how people interact with "fake news" posts on Facebook and Twitter, including the reasons they ignore posts or choose not to investigate them further, cases where they take false posts at face value, and strategies they use to investigate questionable posts. We find, for instance, that participants may ignore news posts when they are using social media for non-news purposes; that people may choose not to investigate posts due to political burn-out; that people use various heuristics for evaluating the credibility of the news source or account who posted the story; that people use ad-hoc strategies like fact-checking via comments more often than prescribed platform affordances for misinformation; and that despite their best intentions, people sometimes believe and even reshare false posts. Though limited by our participant sample and the specific false posts we showed, our findings contribute to our broader understanding of how people interact with misinformation on social media.

In summary, our contributions include, primarily, a qualitative investigation of how people interact with fake news on their own Facebook and Twitter feeds, surfacing both reasons people fail to investigate posts as well as the (platform-based and ad-hoc) strategies they use when they do investigate, and how these relate to information-processing theories of credibility evaluation. Additionally, based on our findings, we identify areas for future research.

BACKGROUND AND RELATED WORK

Prior work has discussed terminology for referring to the phenomenon of misinformation online, including misinformation (unintentional), disinformation (intentional), fake news, and information pollution [43, 48, 88]. In this paper, we typically use the term "misinformation" to remain agnostic to the intention of the original creator or poster of the false information in question, or use the more colloquial term "fake news".

Social Media, News, and Misinformation

Concerns about misinformation online have been rising in recent years, particularly given its potential impacts on elections [30], public health [7], and public safety [10]. Prior work has studied mis/disinformation campaigns on social media and the broader web, characterizing the content, the actors (including humans and bots), and how it spreads [4, 57, 76, 79]. The spread of misinformation on social media is of particular concern, considering that 67% of Americans as of 2017 get at least some of their news from social media (with Facebook, Youtube, and Twitter making up the largest share at 45%, 18%, and 11% respectively [77]). Moreover, prior work has shown that fact-checking content is shared on Twitter significantly less than the original false article [75].

How People Interact With Misinformation

In this work, we focus on how people interact with misinformation they encounter on Facebook and Twitter. Our work adds to related literature on how people consume and share misinformation online—for example, that fake news consumption and sharing during the 2016 U.S. election was associated with (older) age and (more conservative) voting behaviors [35,39]—as well as the strategies people use to evaluate potential fake news. Flintham et al. [33] suggest that people evaluate the trustworthiness of posts on Facebook or Twitter based on the source, content, or who shared the post, though prior work also suggests that people take the trustworthiness of the source less into account than they think [57], less than trustworthiness of the poster [81], or less when they are not as motivated [44]. Lee et al. [49] also explored the effect of poster expertise on people's assessment of tweet credibility.

More generally, researchers have studied how people process information with different motivations [9, 15, 67], how people use cues as shortcuts for judging credibility when not highly-motivated [31, 34, 61, 80], and frameworks for correcting different types of information misperceptions [50]. At its root, misinformation can be hard to combat due to taking advantage of human cognitive biases [58], including the backfire effect [51, 64]—though other work contests the prevalence of the backfire effect [92] and notes that a tipping point for correcting misconceptions exist [69].

Prior academic work has not studied how people interact with potential false information in the context of their *own* social media feeds [33, 44, 81], without adding followers for the purposes of the study [49]. Thus, our study investigates people's strategies in a more ecologically-valid setting for both Facebook and Twitter—sometimes corroborating prior findings or theories, and sometimes providing new perspectives.

Mitigations for Social Media Misinformation

Platform Moderation

One approach to combating misinformation on social media platforms is behavior-based and content-based detection and moderation. For example, Twitter and Facebook remove accounts that manipulate the platform and display inauthentic behavior [38, 71, 74]. They also both demote posts on their feeds that have been flagged manually or detected to be spam or clickbait [22, 24, 85]. One challenge with platform-based solutions is that they may require changes in underlying business practices that incentivize problematic content on social media platforms [37]. Outside of platforms, research tools also exist to detect and track misinformation or bot related behavior and accounts on Twitter [1, 65, 75].

Supporting Users

Other solutions to misinformation aim to engage and support users in evaluating content and identifying falsehoods. This includes media literacy and education [8, 41] (e.g., a game to imbue psychological resistance against fake news [70]), professional and research fact-checking services and platforms (e.g., Snopes [78], PolitiFact [68], Factcheck [28], and Hoaxy [75]), and user interface designs [36] or browser extensions [29, 63] to convey credibility information to people.

Facebook and Twitter, the sites we focus on in this study, both provide a variety of platform affordances related to misinformation. For example, Facebook users can report a post or user for spreading false news. Facebook also provides an information ("i") button giving details about the source website of an article [25], provides context about why ads are shown to users [26] (although research has shown this context may be too vague [3]), and warns users by showing related articles (including a fact-checking article) before they share

Facebook

Title	Type	Summary *(All are false claims.)*
Lettuce	meme[1]	Lettuce killed more Americans than undocumented immigrants last year [52].
CA Bill	article	CA Democrats Introduce LGBTQ Bill that would protect pedophiles [12].
Dishwasher	image	Dishwashers are a safe place to store valuable documents during hurricanes [47].
NZ Fox	meme	The government of New Zealand pulled Fox News off the air [53].
Church	image	A church sign reads "Adultery is a sin. You can't have your Kate and Edith too" [19].
Billionaires	meme	Rep. Alexandria Ocasio-Cortez said the existence of billionaires was wrong [54].
Eggs	image	A photograph shows Bernie Sanders being arrested for throwing eggs at civil rights protesters [55].
Sydney Storm	image	A photograph shows a large storm over Sydney, Australia [18].
E. coli	article	Toronto is under a boil water advisory after dangerous E.coli bacteria found in the water [56].

Twitter

Lettuce	text	Lettuce killed more Americans than undocumented immigrants last year [52].
NZ Fox	article	The government of New Zealand pulled Fox News off the air [53].
Texas	article	A convicted criminal was an illegal immigrant [66].
Dog	video	Photographs show a large, 450-pound dog [17].
Abortion Barbie	image	A photograph shows a toy product Abortion Barbie [13].
Daylight Savings	article	AOC opposes Daylight Savings Time [16].
Anti-vax	article	A Harvard study proved that "unvaccinated children pose no risk" to other kids [46].

Table 1. Summary of false post information, paraphrased from Snopes.com. The titles are shorthand used in the rest of the paper.
[1]A meme is "an amusing or interesting item (such as a captioned picture or video) that is spread widely online especially through social media" [59].

something that is known to be false news. Facebook has iterated on the design and timing of this warning over the last several years [20, 21, 23] due in part to concerns about the backfire effect (though recent work has called this effect into question [92]). The use of "related stories" for misinformation correction has been supported by non-Facebook research [6].

Twitter has fewer misinformation-specific affordances. Twitter allows user to report tweets (e.g., if something "isn't credible") or accounts (e.g., for being suspicious or impersonating someone else). It has also added a prompt directing users to a credible public health source if users search keywords related to vaccines [86], similar to Facebook [27]. Twitter also annotates "verified" (i.e., authentic) accounts with blue checkmark badges (as does Facebook), though these badges do not indicate anything about the credibility or accuracy of the account's posts. Indeed, Vaidya et al. found that Twitter users do not confuse account verification, as indicated by the blue badge, with post credibility [87].

Despite this wide range of intended solutions, there is a lack of public research on how people use these platform affordances to investigate potential fake news posts. To address this, we study how people react to misinformation on their native news feeds, how and when they take content at face value, and how they behave when skeptical. We consider not just the affordances designed for fake news, but also other ways users make use of the platform.

METHODS

We conducted semi-structured interviews about participants' social media use, and then conducted a think-aloud session as participants scrolled through their *own* feeds, in which a browser extension we developed modified certain posts to look like misinformation posts during the study. The study was conducted in-person either in a user study lab or at a cafe. Researchers audio-recorded the interviews and took notes, and participants were compensated with a $30 gift card. We focused on Twitter and Facebook, two social media sites which were primary traffic sources for fake news during 2016 [35]. Our study was considered exempt by our institution's IRB; because we recognize that IRB review is necessary but not sufficient for ethical research, we continued to conduct our study as we would have given continued review (e.g., submitting modifications to the protocol to our IRB). We discuss ethical considerations throughout this section.

Recruitment and Participants

We posted recruitment flyers around a major university campus, as well as public libraries and cafes across the city. We also advertised to the city's local AARP chapter (to reach older adults), as well as neighborhood Facebook groups. Given that older people shared fake news most often during the 2016 election [39], we sought to sample a range of ages.

Table 2 summarizes our participants. We recruited people who used Twitter and Facebook daily or weekly for various news (except for P11); most participants used social media for other reasons as well (e.g., communicating or keeping up with people, entertainment). Because the study required our browser extension, we primarily recruited people who use social media on a laptop or desktop, though most participants used phones or tablets as well. About one-third of our participants are students from a large public U.S. university. Most participants responded to a question about their political orientation by stating they were left-leaning.

Misinformation Browser Extension

While prior work has primarily observed participants interacting with misinformation from researcher-created profiles, we wanted participants to interact with posts on their own feeds for enhanced validity. To do this while also controlling what they encountered, we built a Chrome browser extension to

Figure 1. Example tweet. Participants would see this as liked or retweeted by someone they follow.

show misinformation posts to participants. Our extension temporarily modified the content of random posts on participants' feeds, making them look as if they contained misinformation, on the client-side in the current browser. During the study, posts with our content appeared on Facebook as if posted by a friend of the participant, within a Group, or as a sponsored ad. On Twitter, posts appeared either as a direct tweet, like, or retweet by someone the participant follows. We did not control for what types of posts were randomly modified.

Though the false posts appeared to participants while the extension was active, these posts did *not* exist in their real feeds, and this content could not be liked or shared. In other words, there was no possibility for participants to accidentally share misinformation via our modifications. If a participant attempted to share or like a modified post on Facebook, no request to Facebook was actually made. On Twitter, if someone attempted to like or retweet a modified posts, in practice they liked or retweeted the real post in their own feed that had been modified by our extension. We consider the risk here to be similar to accidentally retweeting something on one's own feed, something that can happen under normal circumstances. In practice, only one participant retweeted one of the posts our extension modified during the study, and we helped them reverse this action during the debriefing phase (described below).

For the misinformation posts we showed, we used social media posts and articles that were debunked by Snopes [78], a reputable fact-checking site. Many posts were platform-specific, so the selection for Facebook and Twitter was not identical. These posts (summarized in Table 1) occurred within the past few years, and covered three categories identified in prior work: humorous fakes, serious fabrications, and large-scale hoaxes [72]. We included a variety of topics including health, politics (appealing to both left- and right-leaning viewpoints), and miscellaneous like weather; Figure 1 and Figure 2 show examples. Screenshots of all posts can be found in the Supplementary Materials. Prior to recruitment and throughout the study period, we tested the extension on our laptops to ensure it would only make the cosmetic changes we intended.

On Facebook, some of our false posts showed up as Sponsored, allowing us to observe participants' reactions in the context of advertising on their feeds. (We note that ads have been used in disinformation campaigns [73, 83].) On Twitter, modified posts only showed up as non-sponsored tweets.

Consent Procedure

For enhanced validity, we designed the study to avoid prompting participants to think about misinformation before the debrief. We initially deceived participants about the study's purpose, describing it only as investigating how people interact with different types of posts on their feeds, from communication to entertainment to news. During the consent procedure,

Figure 2. Example Facebook post. Participants would see this as posted by a person or group they follow.

we stated that our browser extension would visually modify Facebook and Twitter (which it actually did) and would keep a count of the participant's likes and shares (which it did not; we used this as misdirection to avoid participants focusing on possible visual modifications). As is standard ethical practice, and because changed posts or other news feed content could be upsetting to someone, participants were told they could discontinue the study at any point and still receive compensation.

Interview and Social Media Feed Procedures

We started with a semi-structured interview, asking about what social media platform people use, whose content they see on it, and what they use it for.

Then participants either logged into their social media accounts on our laptop, which had the browser extension installed, or we installed the browser extension onto their computer's browser. We did not store login information, and we logged them out (or uninstalled the extension if it was installed on their computer) at the conclusion of the study. We asked participants to scroll through their feed while thinking aloud about their reactions to various posts, e.g., why they interacted with it, why they skipped it, etc. We asked participants to keep scrolling until they had seen all possible inserted posts. Each participant saw a majority of 9 Facebook or 7 Twitter posts modified, as they scrolled through their feeds within around 15 minutes. After this, we explicitly asked participants about their experiences with fake news posts prior to the study.

Due to technical difficulties, not all modified posts showed up on everyone's feed. Some participants could not complete the

feed-scrolling portion at all (P5 and P7 had a new Twitter UI that was incompatible with our extension, and we met P23 in a place with poor WiFi). Some participants normally use Twitter in a different way than the procedures used in the study, e.g., P21 normally uses Tweetdeck (a modified interface with no ads), and P15 and P18 do not normally view their own feeds but rather use search or view other accounts' pages.

Debriefing

Finally, we disclosed the true purpose of the study and explained that we had modified posts in their feeds to look like they contained misinformation. To minimize potential loss of self-esteem by participants who were fooled by our modified posts, we normalized these reactions by emphasizing that misinterpreting false news is common and that identifying it is challenging, and that their participation in the study was helpful towards addressing this issue. In one case, P9 had retweeted the real post underlying our fake post, so we helped them undo this action (within 10 minutes). No participants showed signs of distress during the study or debriefing, most responded neutrally, and some self-reflected (one with disappointment) on their ability to detect fake news.

To ensure that participants knew which posts had appeared due to the study, we showed them screenshots of all of our misinformation posts. We aimed to help participants avoid believing the false information itself as well as to clarify that their friends or followees had not actually posted it. The debriefing occurred immediately after the interview, so participants did not have any opportunity to share our false posts with anyone else online or offline between the study and the debriefing.

Data Analysis

Audio recordings of the interviews were transcribed and annotated with hand-written notes. We then followed an iterative coding process to analyze the data. To construct the codebook, two researchers read several transcripts to code inductively, developing a set of themes pertaining to our research questions. After iteratively comparing and refining codes to develop the final codebook, each researcher coded half of the interviews and then double-checked the other's coding for consistency.

RESULTS

We now describe our findings about how people interact with potential misinformation on their own social media feeds. We surface reasons why people may not deeply read or investigate posts, as well as the strategies they use when they do choose to investigate.

Interactions with Misinformation (and Other) Posts

We describe how participants interact with posts as they scroll through their feed, focusing in particular on reactions to the fake posts we showed them, but often contextualize our findings using observations about how they interact with posts *in general* (any of which, in practice, could be misinformation).

Skipping or Ignoring Posts

Before someone can assess the credibility of a social media post, they must first pay attention to it. We observed participants simply scrolling past many posts on their feeds, including our false posts, without fully reading them.

Participant	Age	Platform
1	18-24	Facebook
2	18-24	Facebook
3	18-24	Twitter
4	25-34	Facebook
5	18-24	Twitter*
6	18-24	Facebook
7	18-24	Twitter*
8	18-24	Twitter
9	55-64	Twitter
10	25-34	Facebook
11	45-54	Facebook
12	25-34	Facebook
13	25-34	Twitter
14	45-54	Facebook
15	25-34	Twitter
16	35-44	Facebook
17	45-54	Facebook
18	45-54	Twitter
19	45-54	Facebook
20	25-34	Facebook
21	65-74	Twitter
22	45-54	Facebook
23	65-74	Facebook*
24	25-34	Facebook
25	25-34	Twitter

Table 2. Participant ages and the platform on which we conducted the think-aloud session. *Due to technical difficulties, we could not complete the feed-scrolling portion of the study with these participants.

One reason that participants ignored posts was because they would take too much time to fully engage with (long text or videos). In contrast, shorter posts and memes often caught participants' attention. For example, P10 skipped the E. coli article, but read and laughed at the short Lettuce meme, explaining: "If it was something funny like a meme or something, then I'd probably care about it, but it's just words. So a little bit less interested." Of course, different people have different preferences. P16 ignores memes that don't "grab her right away" but is more interested in personal posts written by people she knows (what she called "high-quality" content).

In addition to preferring short posts, some participants were also drawn to posts with significant community engagement (likes or shares). For example, P13 skipped through many tweets, including our fake articles, but read the Lettuce tweet because, "It got so many re-tweets and likes....Maybe part of it is the fact that it's one sentence, it's not like there's multiple paragraphs like this tweet below it. It's not like it's a video....Like if it's just a sentence and it's getting this much engagement maybe there's a reason why people are reading it." P18 also mentioned that posts with over 10,000 likes or retweets will jump out at him.

Participants discussed making quick decisions about whether they found a post interesting or relevant enough to fully read. For example, when encountering our false Dishwasher post, P10, who does not live in Florida, stated, "I read the word Florida and I stopped reading. I was like, Okay that's not

important." Similarly, P20 stated, "So I'd read the headline, [and] unless it's something super interesting to me... I generally skip over. And people that I don't care about so much, I'll generally skip over without reading."

Another common reason—as with P20 above—for choosing to ignore a post was that participants identified it as an ad. We found that many participants were aware of the ads on their feed and explicitly ignored them, and sometimes told Facebook to "Hide this ad". For example, P6 hid two of our false posts which showed up as sponsored. (In contrast, some people did pay attention to ads if they were interested in the content: for example P9 liked a political candidate's promoted tweet because she "support[s] it" and thinks "that liking things makes it pop up more often in other peoples' feeds".)

When we debriefed participants at the end of the interview, pointing out the fake news posts that we had "inserted" into their feeds, we found that participants did not always remember the posts that they had skipped. For example, P3 later had no recollection of the Anti-vax tweet, which she skipped while scrolling. (In contrast, P24 skipped the Lettuce and E. coli Facebook posts, but later remembered the general ideas.) As we discuss further in the Discussion, this finding raises the question: to what extent do people remember the fake news posts that they ignore, and to what extent might these posts nevertheless affect their perception of the topic?

Taking Content at Face Value

We often observed participants taking the content of posts at face value and not voicing any skepticism about whether it was true. For example, P14 reacted to a close friend appearing to post our false E. coli article by saying, "I would definitely click on that and read the article" (not to investigate its claims but to learn the news).

Often the root cause of this trust seemed to be trust in the person who posted the content, and/or confirmation bias on the part of the participants when the post aligned with their political views. For example, when P9 saw a public figure she trusts appear to retweet our false NZ Fox post, she stated, "I'm actually going to retweet that because it's something I wholeheartedly support." P3 also accepted this false post at face value, trusting the celebrity who posted it: "Okay, this is a news article. But, it's from a celebrity I follow. I think I would click into this... I think it's good when celebrities post articles that reflect my political beliefs because I think that if they have the platform, they should use it for good... [though] obviously, I don't get all of my political insights from [them]."

Sharing or Liking Posts

In a few cases, participants directly attempted to share or "like" the posts we modified. (Again, they could not actually share the false content during the study). P9 was the only participant to reshare a post, retweeting the NZ Fox article; P21 stated she would email her friends the NZ Fox article; several participants "liked" modified posts. We note that prior work has found that fake news was not reshared on Facebook during the 2016 election as much as commonly thought [39]. Nevertheless, even people who do not actively share or "like" the false posts they take at face value may share their content outside of the platform (e.g., via email or conversation) or incorporate them into their worldviews.

Skepticism About Content

In other cases, people voiced skepticism about posts. For example, some were skeptical of potentially manipulated images. While several participants scrolled past the Sydney Storm Facebook post without much thought, P22 correctly noted, "Well it looks photoshopped... I've seen similar things before." Sometimes skepticism about the content was compounded by skepticism about the source. For example, P5 did not believe the 450-pound dog video on Twitter, saying that "the type of breed of dog that it was showing doesn't grow that large. I mean possibly it could happen, right, but I think there would be much more news about it if that was actually true, and I can't remember if the Dodo [the video creator] is actually a real news outlet or not." Sometimes this skepticism was sufficient for participants, who then ignored the post; in other cases, this prompted them to investigate further, using strategies we discuss in more detail below.

Skepticism About Post Context

Because our methodology involved modifying the appearance of people's social media feeds, sometimes our modifications did not make sense in context. Some participants were aware when content appeared that did not fit their mental model of what someone or some group would post. After seeing our false Lettuce meme, P14 said, "We don't typically post news in the group. So this is a sort of odd post... So I would just pass it over." And P2 stated, "[Group Name]... Wait, is this the post what they usually do?... Good post, but not so related to what they do." Both participants scrolled past without investigating the claims or why the account posted it; it was unclear what their assessment was of the content itself. In the Discussion section, we return to this observation of our participants having strong mental models of their own social media feeds.

In some cases, participants were also skeptical about or annoyed by ads on their feeds. For example, P17 mistook a sponsored post for a post by his friend, and then expressed that he did not like "what appears to be a personal post from a person but who's not my friend, [and] it's sponsored. He's not somebody I follow."

Getting Different Perspectives

At least one participant interacted with a misinformation post specifically to learn more about a viewpoint different from their own, rather than (exclusively) to investigate its accuracy: P9 opened the anti-vaccination article from our false tweet to read later because, "In my work and in my life, I encounter a lot of people of the opposite political denomination from me, and so just I want to understand their viewpoint, and I also want to have counter arguments." More generally, we note that a number of participants (P5, P7, P9, P13, P21, and P24) mentioned following differing political viewpoints on their social media feed to gain a broader news perspective.

Misinterpreting Posts

Finally, sometimes participants misinterpreted our false posts, leading to incorrect determinations about their accuracy. For example, P24 saw the Church post appear as if posted by a

LGBTQ group on Facebook, and thus did not actually read the post before quickly clicking "like" on it in order to support the group. P10 glanced at the CA Bill post and said she would normally click to open the article because she identifies as LGBTQ, while not fully understanding the headline. P22 misinterpreted the Billionaires post as being supportive of a politician rather than misquoting and critiquing her.

Reasons for Not Investigating Posts

We now turn to the reasons why participants did *not* further investigate posts—whether or not they were skeptical of them at face value—both based on their comments during the think-aloud portion of the study as well as their more general self-reported behaviors in the interview portion.

Political Burnout

Many participants noted they were too exhausted or saddened by current politics to engage with political news (potential misinformation or not) on social media. For example, P3 stated, "Sometimes, it's like if I'm burnt out, I'm not necessarily going on Twitter to read the news. I just want to see my friends' posts or funny things." Likewise, P2 ignored all posts about politics, including fake ones.

P10 also said, "I feel like a lot of the political posts about so and so has done this, and it's like, is that really true or did someone make that up, or are they exaggerating it? So I think posts like that I kind of question, but I try not to get into political stuff, so I don't ever research it, I don't look into it or anything like that 'cause that's just, I don't know, it's stressful." This finding supports Duggan et al.'s work, which noted that one-third of social media users were worn out by political content on those sites [11].

Not in the Mood, or Uninterested

In some cases, participants were simply not interested in the topic of a particular article—either ever, or at the present moment. Our observations of participants' social media use highlighted the broad range of use cases and content that are combined in a single feed, including news, entertainment, and professional and personal communication. Sometimes participants were just not using social media for the purpose of news, and so investigating potential misinformation did not fit into their task. For example, P20 said, "When somebody [likes] a lot of very political things, I generally don't like to engage with those so much or even read them because I feel like that's not what I want to be using Facebook for." P13 made the same observation about Twitter: "It's a lot of people talking about really politicized issues, so I'm not always in the head space to like really want to dive into some of this stuff."

Would Take Too Long

Participants also sometimes balked at the time and effort it would take to deeply investigate a post, article, or claim. P5 spelled out this calculation: "It wasn't worth me investigating further and then clearing up to them personally [with] how much time or energy it would take from me but then also how important it would be to them." P18, when asked how long they spend on a confusing tweet before moving on, said, "Probably less than 8 seconds."

Hard to Investigate on Mobile

While our study was conducted on a laptop, some participants also discussed using mobile versions of Facebook or Twitter. P15 preferred the desktop versions: "There's a number of things I do with the phone, but I prefer having the laptop experience in general, of having tabs and then I can switch between the tabs more easily than the phone. I don't like how the phone locks you into one thing". On a desktop browser, someone can easily open a post they would like to investigate in a background tab and return to it later, without interrupting their current flow of processing their social media feed.

Overconfidence About Misinformation

One reason that people may sometimes take misinformation at face value is that they incorrectly assume they will be able to recognize it, or that they will not encounter it. For example, Pll mentioned not actively worrying about misinformation online because they believed that it was typically targeted at groups of people they did not belong to: "I tend to associate [fake news] with the [political] right, and I don't follow anything on the right." While prior work does suggest that conservatives shared more misinformation than liberals during the 2016 U.S. election [39], and while P11 was not fooled by any of our false posts, we note that disinformation campaigns have been shown to target left-leaning groups as well (e.g., [4]), and that it is possible that a false sense of security may cause someone to be more susceptible. (This hypothesis should be tested by future research.) As another example, P9 believed the NZ Fox post, despite believing "I guess I don't fall for things with no source documentation or things that aren't true." We discuss other examples of cases where people's stated strategies were contradicted by their behaviors below.

Investigative Strategies

Finally, we turn to the strategies that our participants used—or self-described using outside the context of the study—to investigate potential misinformation posts. That is, once someone has decided that they are unsure about a post, but has not yet decided to dismiss it entirely on those grounds, what do they do to assess its credibility?

Investigating Claims Directly

Participants described several strategies for directly investigating claims in a post. The most straightforward is to click on the article in a post to learn more. For example, when P22 saw the Eggs Facebook post, he was skeptical: "I've never heard anything about Bernie Sanders throwing eggs at black civil rights protesters. So I think I would click on the news story here and see what more it's about." He clicked the article and learned that the post image was miscaptioned. However, clicking through is not always effective: P7 described previously having been fooled by the debunked Pizzagate conspiracy on Twitter: "I click[ed] on that hashtag because [it was] trending of specific region... I saw the number of retweets and the number of [favorites]. It has very, very high numbers of all three elements... I clicked the external link to the website. I read the whole article, and, yeah, I was fooled."

Sometimes participants described previously using a web search or in-person conversation to investigate the claims of a post or article (though no one did this during the study). For

example, P3, P7, P8, P9, P17, P22, and P23 self-reported that they attempted to verify among multiple sources a story they found suspicious, and/or to see if news outlets they trust were reporting on the same story. P23 and her spouse use their smart assistant to check news stories. P8 described asking her friends and family for opinions on news articles.

Another approach—mentioned by P14, P17, and P23—is to use a fact-checking site like Snopes to investigate claims. P14 said about Snopes, "So if somebody posts something and I'm like, 'Um, I don't know about that,' that's the first place I'd go." P7, fooled by Pizzagate as described above, eventually learned it was false when friends directed him to Snopes.

Investigating Article Source

A number of participants discussed using the source website of a posted article to assess its credibility. For example, P24 described the following heuristic: "If it's like 'Al and Bob's website,' I'm not going to click on it....If it's like 'CNN Bob' or 'CNN South' or something, I'm not going to click on it because they put 'CNN' to make it seem factual where it isn't. But if it's like cnn.com, then I'll click on it, but if it's something I've never heard of, then I won't even click on it." Sometimes these lessons are learned the hard way: after P24 (prior to our study) was fooled by a clickbaity headline from a non-reputable site saying a basketball player had been traded, he now only trusts the ESPN site or the social media posts of well-known sports reporters.

P22 explained that if he reads an article (not necessarily from social media) that sounds unbelievable, he double-checks whether the source is a satirical one (such as The Onion); P19 does the same after having once been fooled by an Onion article. Similarly, P25 described using a web search to investigate sources with which he is not familiar.

Investigating Poster

Some participants tried to gain more context about a post by investigating the account that posted it. P3, P5, P7, P9, and P13 hovered over Twitter account icons to gain more context. For example, P3 had to double-check the poster because "they changed their display name and their picture." (We did not notice anyone on Facebook doing the same thing, perhaps because people are more likely to encounter accounts of people they do not personally know well on Twitter versus Facebook.)

Using Comments on Posts

Participants described using the comments or replies on posts as a fact-checking strategy—both to help them assess the credibility of a post themselves, and to proactively help correct others. For example, P18 has used Twitter comments to determine a post's veracity. After hearing about a conspiracy of a town being intentionally set on fire, P18 looked for news on Twitter and learned from "an overwhelming majority of... downvoting" comments on a post that the conspiracy was false.

Others participants mentioned commenting on posts to alert the poster that they posted something incorrect. For example, P14 will "often then post on their post, and be like, 'Um, no."' P22 has gotten themselves "into so many heated arguments on Facebook". However, others reported avoiding engaging in comments, either because of prior bad experiences, because of a desire to avoid conflict, or because they left it to others. For example, P16 explained that they do not reply to anti-vaccination posts because, "I feel like pediatricians and infectious disease people are just doing such a good job with it that I don't have much to add. So I prefer to sit back and concentrate my arguing online to other things." As an example of a bad experience, P13 had to block a friend on Instagram because that person posted a screenshot of P13 trying to debunk the friend's antisemitism and conspiracy theories.

Platform Affordances

As described in the Related Work section, both Facebook and Twitter provides some platform affordances that might aid people in noticing or investigating potential misinformation.

Facebook. We observed *none* of our participants using Facebook's "i" button that shows more information about the source websites (despite investigating the source website being a common strategy, discussed above). When prompted, most of our participants had not previously noticed this button. One of our participants (P22) did mention having seen or heard about a Facebook warning about misinformation (perhaps the "Disputed Article" label from earlier versions of Facebook [23] or a "Related" fact-checking article [21]).

In some cases, platform design may *hinder* participants' ability to accurately assess content's credibility or source. For example, P11 noted a design choice that troubled her on Facebook: when friends "like" Facebook Pages like The New York Times, Facebook may show sponsored articles from that Page associated with the name(s) of the friend(s) who like the Page. P11 explained: "And so, that's always sort of troubled me because it looks like it's associating them liking New York Times with the content of the article. It's kind of disturbing."

Twitter. On Twitter, participants sometimes used or discussed the blue "verified" badge used by some accounts. Confirming prior work [49], we did not observe our participants confusing the badge's meaning as implying credibility of the content. For example, when P7 was asked if the verified status helped him determine an article's veracity, he replied, "Not at all. Because I mean, I don't really know about the verification process of Twitter, but sometimes, I do feel that there are some users where, even famous users, who have very radical political opinions, have verified accounts." While P24 does use the badge to help assess content credibility, he does so by identifying whether the account is an expert (in this case, a known sports journalist) who can be trusted on the topic.

Contradictory Behavior

Finally, we sometimes observed participants act in ways that contradicted their stated strategies—typically incorrectly taking a particular false post at face value. For example, P21 assumed that the NZ Fox tweet from MSNBC was true, despite later explaining that she did not trust MSNBC as a news source: "If they tweet something that interests me that I have never seen anywhere before, I might click through to get more details. I don't trust them. I wouldn't use it for a news source per se...but I might look to see what they're saying so I can go investigate it some place else." P21's explanation for this inconsistency was that the NZ Fox tweet did not contain "criti-

cal information" that necessitated fact-checking (and, perhaps due to confirmation bias, may have been more believable to someone who welcomed the news).

This finding raises an important consideration: there are many different factors that influence how and what people choose to believe, disbelieve, or investigate when they are interacting with their social media feeds. As a result, people's self-reported and well-intentioned fact-checking aspirations can sometimes be trumped by the specifics of the current context.

DISCUSSION

Stepping back, our results provide more ecologically-valid support for previous work on how people determine credibility (as our study was conducted in the context of people's actual personal feeds), while being specific to the mediums of Facebook and Twitter. In this section, we tie our findings into a broader discussion of how people assess the credibility of (mis)information on social media, and we highlight avenues that our results suggest for future work.

Determining Credibility

Motivation

The Elaboration Likelihood Model of persuasion (ELM) [67] and the Heuristic-Systematic Model (HSM) [9] both suggest that people more rigorously evaluate information when their motivation is higher [62]. Further work suggests that dual cognitive processing pathways exist, one "fast, automatic" and the other "slow, deliberative" [15], which researchers have proposed means people use different heuristics for evaluating credibility depending on whether they are motivated enough to spend more time with the information [31]. We echo that "understanding users' motivations to process information using more or less cognitive effort is an important first step toward understanding how often and when specific heuristics may be invoked during credibility evaluation," [61] and offer some empirical examples of different motivations on social media.

Several participants noted that they do not use social media for political news (either at the moment or all the time), and they skipped such content. This lack of interest provides a possible explanation for why exposure to political content on social media has little effect on civic engagement [82]. Our findings support a theory from Tucker et al. that despite the high availability of political news online, people may focus their attention on entertainment news instead [84].

For some participants who were interested in political news, they were either slow and attentive to articles going through their feeds, or they skipped articles because they would take too long to process and instead focused on text or meme posts. Given that medium may affect credibility [89], and that previous work often only focused on article-based misinformation, we recommend future work investigate what cues people use to trust text, image, and meme social media posts. we would also like to see future work studying to what extent people remember (and incorporate into their worldview) misinformation when they do not read a full article (but rather simply scroll past headlines).

Heuristics

Metzger et al. discuss the following credibility heuristics that people may use when processing information with low motivation: reputation, endorsement, consistency, expectancy, and persuasive intent [62]. Indeed, some of our users self-reported using the news source to determine credibility (reputation), relying on trust of the poster (endorsement), searching for multiple other sources corroborating the same headline (consistency), being skeptical when a post seemed out of line from the poster's typical content (expectancy violation), or skipping ads (persuasive intent).

The fact that many participants self-reported using the source (i.e., website) of a news article to evaluate its credibility also echoes findings from Flintham et al. [33]. However, during our study, only P24 actually skipped a post because he did not recognize and trust the article's source. Meanwhile, several participants on Twitter hovered over the icons of posters, rather than look into the article's source. On both platforms, degree of closeness or trust in the poster seemed to be a more salient factor when people decided to pay attention to a post or take it at face value. This observation supports previous work which found social endorsements or trust in the poster to be more important than the article's source [60, 81].

One reason this picture is complicated is that motivation affects strategy. Tandoc et al. [44] found that people relied more on friend endorsement for articles where they had low motivation, but relied more on news organization reputation for articles where they had high motivation. Thus, we emphasize that future work must control for motivation when studying how people interact with (mis)information, and must consider these nuances when designing affordances to support heuristics.

For example, reliance on the endorsement or other heuristics may have played into why almost no participants used or even knew about the Facebook "i" button for learning more about the sources of articles; this interface may work better for more deliberate information processing. Unobtrusive user-facing solutions that aim to describe an article's source may in practice have little effect on users who rely on heuristics.

While the "i" button aims to help people assess article sources, both Facebook and Twitter have blue "verified" checkmarks to help people assess the validity of the accounts of public figures and others (i.e., supporting the endorsement heuristic). These checkmarks are intended to verify authenticity of account ownership, not the content posted by that account, and we found that our participants generally interpreted it this way. While previous work has shown that Twitter's verified badge does not impact credibility perception when participants do not know the poster [87], we note that P24 did use it (reasonably, we believe) as a credibility heuristic for posted content related to the poster's field of expertise.

Understanding and Curation of Own Social Feeds

A recurring theme throughout our observations was that our participants had precise (but not necessarily complete) mental models about their own social media feeds. For example, participants generally expressed an awareness that their feeds are algorithmically curated by Facebook or Twitter (suggesting

broader awareness than reported in prior work from 2015 [14]), and sometimes hypothesized why they were seeing a certain post or ad. Additionally, almost all participants actively identified the ads in their feeds (in contrast to prior work in other contexts suggesting that people perform poorly at identifying ads [2, 40, 90, 91]). And though the fake posts we inserted appeared in the context of participants' own feeds, they often noticed that posts seemed out of place (e.g., noting that a certain person or group does not usually post this type of content), and expressed skepticism at the headlines of inserted posts. Moreover, participants frequently took or discussed taking an active role in the curation of their own feeds—including unfollowing people who posted types or volumes of content they did not like, or frequently using Facebook's "Hide this Ad" option to remove odd and uninteresting posts (including some of those we inserted).

Additionally, despite social media curation raising concern over ideological "echo chambers" [42], researchers have found that social media increases exposure to a variety of political viewpoints [5, 11, 32]. Some of our qualitative results support this finding, as several participants follow and pay attention to news of a different ideological bent in order to gain perspective of "the other side". Twitter and Facebook have provided users with a convenient way of curating multiple perspectives onto their feed. However, further research is needed also on how individuals who *are* embedded in a homogeneous echo chamber [84] interact with misinformation.

There are several possible reasons why our participants displayed so much understanding and agency over their feeds. It could be that people's awareness in general has improved over time (prior work indeed suggests that people improve at identifying ads with more experience [45]), and because online misinformation has received widespread attention. It could also be in part a reflection of our participants' demographics (people who use social media often). However, these findings give us hope: in designing solutions for fake news on social media, we must not assume that all users are ignorant of potential media manipulation and inaccuracy. We are optimistic that education solutions in this area can be impactful (for example, P8 learned from being duped by an article and has adjusted his consumption behavior), as well as user interface solutions that help empower users with additional knowledge and agency. In other words, we should view solutions as a partnership with users, rather than something social media platforms should merely impose on or solve for them.

Avenues For Future Work

Our work suggests possible directions for future designs to better support current ad hoc strategies for evaluating social media post credibility. For example, a common reason that our participants did not further investigate posts was that they did not have or want to take the time and mental energy to dig deeper into something they happened to see while scrolling through their social media feeds (also, on a mobile device such side-investigations may be more challenging). Quickly-updating and algorithmically-curated social media feeds can also make it hard to return to posts later. To better support returning to posts, we propose an "Investigate Later" option that users can select, filing the post away to return to later. User investigative efforts could be offloaded and aided by automated means, e.g., alerting users if posts that they wanted to "investigate later" were subsequently debunked by a reputable fact-checking site.

Finally, we emphasize that the interplay of motivation, demographics, medium, and other effects all play a role in how people interact with misinformation online. Our results presented various user strategies for credibility evaluation, such as fact-checking in comments, crowd-sourced skepticism, and searching for corroborative headlines, and presented lack of motivation for investigating claims in the moment.

Future work should explore interactions and strategies focusing on images, text, and meme-style fake news with a more diverse group of people, false information with different topics and sources, fake news interactions on phones and tablets, as well as investigative strategies during differently-motivated social media use sessions.

LIMITATIONS

Most fundamentally, our study is exploratory qualitative work, so we cannot make generalizations about our findings to the broader U.S. (or any other) population. Our participant demographics had limited variety: one-third are attending a large public university, and all of our participants identified as either politically left-leaning or independent. We also did not control for which types of posts were modified during the study (e.g., poster relationship, sponsored or not, news salience); our results suggest variables that should be controlled in future experiments. Due to technical limitations, neither Twitter nor Facebook posts showed the comments that were originally attached to the fake news post, and some participants did not see all of the posts we intended (or could not complete the feed scrolling portion of the study). Finally, while we encouraged participants to scroll how they normally would through social media, their behavior was observed in a lab setting with a researcher sitting next to them. Despite these limitations, our study expands on prior findings and theories, and presents avenues for future work.

CONCLUSION

Our qualitative study provides a detailed look at how people interact with fake news posts on Twitter and Facebook through both observation and self-reports, using a browser extension to modify posts in participants' social media feeds to look like fake news. Participants often took posts at face value or looked to the poster for context when they were uncertain. Strategies they used to investigate suspicious posts included investigating the source or poster and looking at comments, though many participants did not investigate for a variety of reasons. Our work presents a broad view of social media misinformation consumption and raises important questions for future work.

ACKNOWLEDGMENTS

We are grateful to our study participants. We thank our anonymous reviewers and shepherd for helping improve the paper. This work was supported in part by the National Science Foundation under Award CNS-1651230.

REFERENCES

[1] Alliance for Securing Democracy. 2017. Hamilton 68: Tracking Russian Influence Operations on Twitter. (2017). http://dashboard.securingdemocracy.org/.

[2] Michelle A. Amazeen and Bartosz W. Wojdynski. 2019. Reducing Native Advertising Deception: Revisiting the Antecedents and Consequences of Persuasion Knowledge in Digital News Contexts. *Mass Communication and Society* 22, 2 (2019), 222–247. DOI:http://dx.doi.org/10.1080/15205436.2018.1530792

[3] Athanasios Andreou, Giridhari Venkatadri, Oana Goga, Krishna P. Gummadi, Patrick Loiseau, and Alan Mislove. 2018. Investigating Ad Transparency Mechanisms in Social Media: A Case Study of Facebook's Explanations. In *Network and Distributed System Security Symposium (NDSS)*. DOI: http://dx.doi.org/10.14722/ndss.2018.23204

[4] Ahmer Arif, Leo Graiden Stewart, and Kate Starbird. 2018. Acting the Part: Examining Information Operations Within #BlackLivesMatter Discourse. *Proceedings of the ACM on Human-Computer Interaction (CSCW)* 2, Article 20 (Nov. 2018), 27 pages. DOI:http://dx.doi.org/10.1145/3274289

[5] Pablo Barberá, John Jost, Jonathan Nagler, Joshua Tucker, and Richard Bonneau. 2015. Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber? *Psychological science* 26 (08 2015). DOI:http://dx.doi.org/10.1177/0956797615594620

[6] Leticia Bode and Emily K. Vraga. 2015. In Related News, That Was Wrong: The Correction of Misinformation Through Related Stories Functionality in Social Media. *Journal of Communication* 65, 4 (2015), 619–638. DOI: http://dx.doi.org/10.1111/jcom.12166

[7] David A. Broniatowski, Amelia Jamison, SiHua Qi, Lulwah Alkulaib, Tao Chen, Adrian Benton, Sandra Crouse Quinn, and Mark Dredze. 2018. Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. *American Journal of Public Health* 108 10 (2018), 1378–1384. DOI:http://dx.doi.org/10.2105/AJPH.2018.304567

[8] Mike Caulfield. 2017. Web Literacy for Student Fact-Checkers. (2017). https://webliteracy.pressbooks.com/.

[9] Shelley Chaiken. 1980. Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion. *Journal of Personality and Social Psychology* 39, 5 (1980), 752–766. DOI: http://dx.doi.org/10.1037/0022-3514.39.5.752

[10] Pranav Dixit and Ryan Mac. 2018. How WhatsApp Destroyed A Village. Buzzfeed News. (Sept. 2018). https://www.buzzfeednews.com/article/pranavdixit/whatsapp-destroyed-village-lynchings-rainpada-india.

[11] Maeve Duggan and Aaron Smith. 2016. The Political Environment on Social Media. (2016). https://www.pewinternet.org/2016/10/25/the-political-environment-on-social-media/.

[12] David Emery. 2019a. Did California Democrats Introduce an LGBTQ Bill to "Protect Pedophiles Who Rape Children"? (Feb. 2019). https://www.snopes.com/fact-check/ca-democrats-lgbtq-bill-pedophiles/.

[13] David Emery. 2019b. Does This Photograph Depict an Actual "Abortion Barbie" Doll? (April 2019). https://www.snopes.com/fact-check/abortion-barbie-doll/.

[14] Motahhare Eslami, Aimee Rickman, Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong, Karrie Karahalios, Kevin Hamilton, and Christian Sandvig. 2015. "I Always Assumed That I Wasn't Really That Close to [Her]": Reasoning About Invisible Algorithms in News Feeds. In *33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. DOI: http://dx.doi.org/10.1145/2702123.2702556

[15] Jonathan Evans. 2008. Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual review of psychology* 59 (02 2008), 255–78. DOI: http://dx.doi.org/10.1146/annurev.psych.59.103006.093629

[16] Dan Evon. 2019a. Does U.S. Rep. Ocasio-Cortez Oppose Daylight Saving Time Because It Speeds Up Climate Change? (March 2019). https://www.snopes.com/fact-check/aoc-daylight-saving-time/.

[17] Dan Evon. 2019b. Is This 450-Pound Dog Real? (April 2019). https://www.snopes.com/fact-check/450-pound-dog/.

[18] Dan Evon. 2019c. Is this a Photograph of a Large Storm Over Sydney? (April 2019). https://www.snopes.com/fact-check/storm-over-sydney/.

[19] Dan Evon. 2019d. Is this Crystal Methodist Church Sign in Effing, SC, Real? (March 2019). https://www.snopes.com/fact-check/crystal-methodist-church-sign/.

[20] Facebook. 2017a. New Test to Provide Context About Articles. (Oct. 2017). https://newsroom.fb.com/news/2017/10/news-feed-fyi-new-test-to-provide-context-about-articles/.

[21] Facebook. 2017b. New Test With Related Articles. (April 2017). https://newsroom.fb.com/news/2017/04/news-feed-fyi-new-test-with-related-articles/.

[22] Facebook. 2017c. New Updates to Reduce Clickbait Headlines. (May 2017). https://newsroom.fb.com/news/2017/05/news-feed-fyi-new-updates-to-reduce-clickbait-headlines/.

[23] Facebook. 2017d. Replacing Disputed Flags With Related Articles. (Dec. 2017). https://newsroom.fb.com/news/2017/12/news-feed-fyi-updates-in-our-fight-against-misinformation/.

[24] Facebook. 2018a. Helping Ensure News on Facebook Is From Trusted Sources. (Jan. 2018). https://newsroom.fb.com/news/2018/01/trusted-sources/.

[25] Facebook. 2018b. Helping People Better Assess the Stories They See in News Feed with the Context Button. (2018). https://newsroom.fb.com/news/2018/04/news-feed-fyi-more-context/.

[26] Facebook. 2018c. Making Ads and Pages More Transparent. (April 2018). https://newsroom.fb.com/news/2018/04/transparent-ads-and-pages/.

[27] Facebook. 2019. Combatting Vaccine Misinformation. (March 2019). https://newsroom.fb.com/news/2019/03/combatting-vaccine-misinformation/.

[28] FactCheck.org. 2017. Misinformation Directory. (2017). https://www.factcheck.org/2017/07/websites-post-fake-satirical-stories/.

[29] Factmata. 2019. Trusted News. (2019). https://trusted-news.com/.

[30] Robert M. Faris, Hal Roberts, Bruce Etling, Nikki Bourassa, Ethan Zuckerman, and Yochai Benkler. 2017. *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election*. Technical Report. Berkman Klein Center for Internet & Society Research Paper.

[31] Andrew Flanagin and Miriam Metzger. 2008. Digital Media and Youth: Unparalleled Opportunity and Unprecedented Responsibility. *The MacArthur Foundation Digital Media and Learning Initiative* (2008).

[32] 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. DOI:http://dx.doi.org/10.1093/poq/nfw006

[33] 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. DOI: http://dx.doi.org/10.1145/3173574.3173950

[34] B. J. Fogg, Cathy Soohoo, David R. Danielson, Leslie Marable, Julianne Stanford, and Ellen R. Tauber. 2003. How Do Users Evaluate the Credibility of Web Sites?: A Study with over 2,500 Participants. In *Proceedings of the 2003 Conference on Designing for User Experiences (DUX '03)*. ACM, New York, NY, USA, 1–15. DOI: http://dx.doi.org/10.1145/997078.997097

[35] Adam Fourney, Miklos Racz, Gireeja Ranade, Markus Mobius, and Eric Horvitz. 2017. Geographic and Temporal Trends in Fake News Consumption During the 2016 US Presidential Election. 2071–2074. DOI: http://dx.doi.org/10.1145/3132847.3133147

[36] Dilrukshi Gamage, Humphrey Obuobi, Bill Skeet, Annette Greiner, Amy X. Zhang, and Jenny Fan. 2019. What does it take to design for a user experience (UX) of credibility? (2019). https://misinfocon.com/what-does-it-take-to-design-for-a-user-experience-ux-of-credibility-f07425940808.

[37] Jeff Gary and Ashkan Soltani. 2019. First Things First: Online Advertising Practices and Their Effects on Platform Speech. (2019). https://knightcolumbia.org/content/first-things-first-online-advertising-practices-and-their-effects-on-platform-speech.

[38] Nathaniel Gleicher. 2019. Removing Coordinated Inauthentic Behavior From China. (2019). https://newsroom.fb.com/news/2019/08/removing-cib-china/.

[39] Andrew Guess, Jonathan Nagler, and Joshua Tucker. 2019. Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances* 5 (01 2019), eaau4586. DOI: http://dx.doi.org/10.1126/sciadv.aau4586

[40] David A. Hyman, David J. Franklyn, Calla Yee, and Mohammad Rahmati. 2017. Going Native: Can Consumers Recognize Native Advertising? Does it Matter? 19 Yale J.L. & Tech. 77. (2017).

[41] Poynter Institute. 2019. What is MediaWise? (2019). https://www.poynter.org/mediawise/.

[42] Shanto Iyengar and Kyu Hahn. 2009. Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use. *Journal of Communication* 59 (03 2009), 19–39. DOI:http://dx.doi.org/10.1111/j.1460-2466.2008.01402.x

[43] Caroline Jack. 2017. Lexicon of Lies: Terms for Problematic Information. Data & Society. (Aug. 2017).

[44] Edson C. Tandoc Jr. 2019. Tell Me Who Your Sources Are. *Journalism Practice* 13, 2 (2019), 178–190. DOI: http://dx.doi.org/10.1080/17512786.2017.1423237

[45] A-Reum Jung and Jun Heo. 2019. Ad Disclosure vs. Ad Recognition: How Persuasion Knowledge Influences Native Advertising Evaluation. *Journal of Interactive Advertising* 19, 1 (2019), 1–14. DOI: http://dx.doi.org/10.1080/15252019.2018.1520661

[46] Alex Kasprak. 2018. Did a Harvard Study Prove That"Unvaccinated Children Pose No Ris" to Other Kids? (Aug. 2018). https://www.snopes.com/fact-check/harvard-study-unvaccinated-children/.

[47] Kim Lacapria. 2017. During a Hurricane, Should You Store Important Items in Your Dishwasher? (Sept. 2017). https://www.snopes.com/fact-check/dishwasher-hurricane/.

[48] David Lazer, 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 Thorson, Duncan J. Watts, and Jonathan Zittrain. 2018. The science of fake news. *Science* 359 (2018), 1094–1096. DOI: http://dx.doi.org/10.1126/science.aao2998

[49] Ji Young Lee and S. Shyam Sundar. 2013. To Tweet or to Retweet? That Is the Question for Health Professionals on Twitter. *Health Communication* 28, 5 (2013), 509–524. DOI: http://dx.doi.org/10.1080/10410236.2012.700391 PMID: 22873787.

[50] Stephan Lewandowsky, Ullrich K. H. Ecker, Colleen M. Seifert, Norbert Schwarz, and John Cook. 2012. Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest* 13, 3 (2012), 106–131. DOI: http://dx.doi.org/10.1177/1529100612451018 PMID: 26173286.

[51] C. G. Lord, L. Ross, and M. R. Lepper. 1979. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology* 37 (1979). Issue 11. DOI: http://dx.doi.org/10.1037/0022-3514.37.11.2098

[52] Dan MacGuill. 2019a. Did Lettuce Kill More People in the U.S. in 2018 Than Undocumented Immigrants Did? (Jan. 2019). https://www.snopes.com/fact-check/lettuce-deaths-illegal-immigrants/.

[53] Dan MacGuill. 2019b. Did New Zealand Take Fox News or Sky News Off the Air in Response to Mosque Shooting Coverage? (March 2019). https://www.snopes.com/fact-check/fox-new-zealand-mosque/.

[54] Dan MacGuill. 2019c. Did Rep. Alexandria Ocasio-Cortez Say it Was "Wrong" for Billionaires to Exist? (Feb. 2019). https://www.snopes.com/fact-check/aoc-alexandria-ocasio-cortez-wrong-billionaires/.

[55] Dan MacGuill. 2019d. Was Bernie Sanders Arrested for Throwing Eggs at Civil Rights Protesters? (Feb. 2019). https://www.snopes.com/fact-check/bernie-sanders-throwing-eggs/.

[56] Dan MacGuill. 2019e. Was Canada Under an E. Coli-Related Boil Water Notice in the Spring of 2019? (May 2019). https://www.snopes.com/fact-check/canada-water-ecoli/.

[57] Alice Marwick and Rebecca Lewis. 2017. Media Manipulation and Disinformation Online. Data & Society. (May 2017). https://datasociety.net/pubs/oh/DataAndSociety_MediaManipulationAndDisinformationOnline.pdf.

[58] Lee C. McIntyre. 2018. *Post-Truth*. MIT Press.

[59] Merriam-Webster. 2019. (2019). https://www.merriam-webster.com/dictionary/meme.

[60] Solomon Messing and Sean J. Westwood. 2014. Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation When Selecting News Online. *Communication Research* 41, 8 (2014), 1042–1063. DOI: http://dx.doi.org/10.1177/0093650212466406

[61] Miriam J. Metzger and Andrew J. Flanagin. 2013. Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics* 59 (2013), 210 – 220. DOI: http://dx.doi.org/https://doi.org/10.1016/j.pragma.2013.07.012 Biases and constraints in communication: Argumentation, persuasion and manipulation.

[62] Miriam J. Metzger, Andrew J. Flanagin, and Ryan B. Medders. 2010. Social and Heuristic Approaches to Credibility Evaluation Online. *Journal of Communication* 60, 3 (2010), 413–439. DOI: http://dx.doi.org/10.1111/j.1460-2466.2010.01488.x

[63] NewsGuard. 2019. NewsGuard: Restoring Trust & Accountability. (2019). https://www.newsguardtech.com/.

[64] Brendan Nyhan and Jason Reifler. 2010. When Corrections Fail: The persistence of political misperceptions. *Political Behavior* 32 (June 2010), 303–330. Issue 2. DOI: http://dx.doi.org/10.1007/s11109-010-9112-2

[65] Observatory on Social Media (OSoMe). 2019. BotSlayer. Indiana University. (2019). https://osome.iuni.iu.edu/tools/botslayer/.

[66] Bethania Palma. 2019. Did the Texas Governor Tweet a Fake BBC Page with False Information About a Convicted Rapist? (Feb. 2019). https://www.snopes.com/fact-check/texas-governor-tweet-rapist/.

[67] Richard E. Petty and John T. Cacioppo. 1986. The Elaboration Likelihood Model of Persuasion. Advances in Experimental Social Psychology, Vol. 19. Academic Press, 123 – 205. DOI:http://dx.doi.org/https://doi.org/10.1016/S0065-2601(08)60214-2

[68] Politifact. 2019. Fact-checking U.S. politics. (2019). https://www.politifact.com/.

[69] David P. Redlawsk, Andrew J. W. Civettini, and Karen M. Emmerson. 2010. The Affective Tipping Point: Do Motivated Reasoners Ever âĂIJGet ItâĂİ? *Political Psychology* 31, 4 (2010), 563–593. DOI: http://dx.doi.org/10.1111/j.1467-9221.2010.00772.x

[70] Jon Roozenbeek and Sander van der Linden. 2019. Fake news game confers psychological resistance against online misinformation. *Palgrave Communications* 5, 1 (2019), 65. DOI: http://dx.doi.org/10.1057/s41599-019-0279-9

[71] Yoel Roth and Del Harvey. 2018. How Twitter is fighting spam and malicious automation. (2018). https://blog.twitter.com/en_us/topics/company/2018/how-twitter-is-fighting-spam-and-malicious-automation.html.

[72] Victoria L. Rubin, Yimin Chen, and Niall J. Conroy. 2015. Deception Detection for News: Three Types of Fakes. In *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community (ASIST '15)*. American Society for Information Science, Silver Springs, MD, USA, Article 83, 4 pages. DOI: http://dx.doi.org/10.1002/pra2.2015.145052010083

[73] Ryan Gallagher. 2019. Twitter Helped Chinese Government Promote Disinformation On Repression Of Uighurs. (Aug. 2019). https://theintercept.com/2019/08/19/twitter-ads-china-uighurs/.

[74] Twitter Safety. 2019. Information operations directed at Hong Kong. (2019). https://blog.twitter.com/en_us/topics/company/2019/information_operations_directed_at_Hong_Kong.html.

[75] Chengcheng Shao, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. 2016. Hoaxy: A Platform for Tracking Online Misinformation. In *Proceedings of the 25th International Conference Companion on World Wide Web (WWW '16 Companion)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 745–750. DOI: http://dx.doi.org/10.1145/2872518.2890098

[76] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kaicheng Yang, Alessandro Flammini, and Filippo Menczer. 2018. The spread of low-credibility content by social bots. In *Nature Communications*. DOI: http://dx.doi.org/10.1038/s41467-018-06930-7

[77] Elisa Shearer and Jeffrey Gottfried. 2017. News Use Across Social Media Platforms 2017. (2017). http://https://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017.

[78] Snopes Media Group. 2019. Snopes. (2019). https://www.snopes.com/.

[79] Kate Starbird, Ahmer Arif, Tom Wilson, Katherine Van Koevering, Katya Yefimova, and Daniel Scarnecchia. 2018. Ecosystem or Echo-System? Exploring Content Sharing across Alternative Media Domains. In *International AAAI Conference on Web and Social Media (ICWSM)*.

[80] Shyam Sundar. 2007. The MAIN Model : A Heuristic Approach to Understanding Technology Effects on Credibility.

[81] The Media Insight Project. 2017. "Who Shared It?": How Americans Decide What News to Trust on Social Media. (March 2017). https://www.americanpressinstitute.org/publications/reports/survey-research/trust-social-media/.

[82] Yannis Theocharis and Will Lowe. 2016. Does Facebook increase political participation? Evidence from a field experiment. *Information, Communication & Society* 19, 10 (2016), 1465–1486. DOI: http://dx.doi.org/10.1080/1369118X.2015.1119871

[83] Tony Romm. 2018. "Pro-Beyoncé" vs. "Anti-Beyoncé": 3,500 Facebook ads show the scale of Russian manipulation. (May 2018). https://washingtonpost.com/news/the-switch/wp/2018/05/10/here-are-the-3400-facebook-ads-purchased-by-russias-online-trolls-during-the-2016-election/.

[84] Joshua Tucker, Andrew Guess, Pablo Barbera, 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. *SSRN Electronic Journal* (01 2018). DOI: http://dx.doi.org/10.2139/ssrn.3144139

[85] Twitter. 2017. Our approach to bots and misinformation. (June 2017). https://blog.twitter.com/en_us/topics/company/2017/Our-Approach-Bots-Misinformation.html.

[86] Twitter. 2019. Helping you find reliable public health information on Twitter. (May 2019). https://blog.twitter.com/en_us/topics/company/2019/helping-you-find-reliable-public-health-information-on-twitter.html.

[87] Tavish Vaidya, Daniel Votipka, Michelle L. Mazurek, and Micah Sherr. 2019. Does Being Verified Make You More Credible?: Account Verification's Effect on Tweet Credibility. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 525, 13 pages. DOI: http://dx.doi.org/10.1145/3290605.3300755

[88] Claire Wardle and Hossein Derakhshan. 2017. *Information Disorder: Toward an interdisciplinary framework for research and policymaking*. Technical Report. Council of Europe.

[89] Elizabeth J. Wilson and Daniel L. Sherrell. 1993. Source Effects in Communication and Persuasion Research: A Meta-Analysis of Effect Size. *Journal of the Academy of Marketing Science* 21, 2 (1993), 101–112. DOI: http://dx.doi.org/10.1177/009207039302100202

[90] Bartosz W. Wojdynski. 2016. The Deceptiveness of Sponsored News Articles: How Readers Recognize and Perceive Native Advertising. *American Behavioral Scientist* 60, 12 (2016), 1475–1491. DOI: http://dx.doi.org/10.1177/0002764216660140

[91] Bartosz W. Wojdynski and Nathaniel J. Evans. 2016. Going Native: Effects of Disclosure Position and Language on the Recognition and Evaluation of Online Native Advertising. *Journal of Advertising* 45, 2 (2016), 157–168. DOI: http://dx.doi.org/10.1080/00913367.2015.1115380

[92] Thomas Wood and Ethan Porter. 2018. The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence. *Political Behavior* 41 (01 2018). DOI: http://dx.doi.org/10.1007/s11109-018-9443-y