Stanford Law School
Law and Policy Lab

Fake News & Misinformation Policy Practicum

Fake News and Misinformation:
The roles of the nation's digital newsstands, Facebook, Google, Twitter and Reddit

2017 PRACTICUM RESEARCH TEAM:

Jacob Finkel, JD '19, Steven Jiang, BS '17,
Mufan Luo, PhD '22, Rebecca Mears, JD/MPP '19,
Danaë Metaxa-Kakavouli, PhD '20,
Camille Peeples, JD '18, Brendan Sasso, JD '19,
Arjun Shenoy, JD '19, Vincent Sheu, JD/MS '18,
Nicolás Torres-Echeverry, JSM '17

INSTRUCTOR AND PROJECT LEADS:

SENATOR RUSS FEINGOLD
Edwin A. Heafey, Jr., Visiting Professor of Law

Luciana Herman, Ph.D.
Lecturer in Law
Program Director, Law and Policy Lab

Ashwin Aravind, JD '18
Teaching Assistant

Katie Joseff, MA '19
Research Assistant

POLICY CLIENT:

Hewlett Foundation Madison Initiative

October 2017

Acknowledgements

This report reflects the research and analysis of an inter-disciplinary law and graduate student team enrolled in the Stanford Law School Fake News and Misinformation Policy Lab Practicum (Spring 2017). Under the guidance of instructor Senator Russ Feingold, the Edwin A. Heafey Visiting Professor of Law, the practicum surveyed the roles of four major online platforms in the spread of fake news beginning with the 2016 U.S. election. Assisting Senator Feingold in the practicum were Policy Lab Program Director and Lecturer Luciana Herman, Ph.D., and Teaching Assistant Ashwin Aravind, J.D. '18. Brendan Sasso, J.D. '19, served as the exceptional lead student editor for the report.

The project originated in conversations between Paul Brest, faculty director of the Policy Lab, Larry Kramer, President of the Hewlett Foundation, and Senator Feingold, who discussed issues related to the fake news phenomenon in the immediate wake of the U.S. 2016 state and federal elections. The Hewlett Foundation Madison Initiative sought to track the effects of misinformation on democratic discourse and decision-making and proved a vital policy client in guiding and funding the research of the practicum. The research team extends a special thanks to Kelly Born, Program Officer for the Madison Initiative, who offered extensive guidance to the research teams and detailed input on drafts of this final report. Her expertise proved invaluable in orienting the research around the roles of the platforms and grounded students in complex issues of market incentives. Hewlett President Larry Kramer deftly guided the research team with connections between legal and policy issues that now ground continuing research in the 2017-18 academic year, under the direction of Professor Nathaniel Persily.

The project was enriched by the guidance and editorial input of scholars and experts. We would like to thank Stanford Professors Jeff Hancock (Communications) and Michal Kosinski (Graduate School of Business) for guiding our research on evaluating the underlying algorithms and consumer micro-targeting techniques of the four platforms; Daphne Keller, Director of Intermediary Liability at the Stanford Center for Internet & Society, for her crucial direction on international laws and regulations governing intermediary liability; and to the 2016-17 Knight Journalism Fellows who engaged us in a lively roundtable about the impact of the four platforms' business models on contemporary journalism.

Additionally, we would also like to acknowledge the support of people and departments at Stanford Law School: Sean Kaneshiro, Reference Librarian for the Robert Crown Law Library, who aided students in researching breaking news as well as underlying legal cases and policy issues; Professor Phillip Malone, Director of the Juelsgaard Innovation and Intellectual Property Clinic, who served as faculty support for the online survey tool and provided patient guidance through the IRB process; and to Stephanie Basso who offered essential administrative and organizational support for all the activities of the practicum, including travel by the research team and a special symposium co-hosted by the Handa Center for Human Rights.

We are grateful to the Handa Center for feedback on the project and for co-hosting a symposium featuring the initial research findings (May 2017). Acting Associate Director Jessie Brunner recognized the significance of the research in its early stage and helped to arrange access to Handa Center faculty and scholars. Meredith Vostrejs, Handa Center Program Manager, generously gave her time to help organize the symposium on our findings.

We are also grateful to the Newseum and to the inimitable Esther Wojcicki, in her role overseeing the journalism program at Palo Alto High School. Newseum Chief Operating Officer, Gene Policinski, invited us to observe a Newseum pilot high school news literacy curriculum, "Fighting Fake News: How to Help Your Students Outsmart Trolls and Troublemakers," taking place in Ms. Wojcicki's journalism class. We joined students and Palo Alto High School journalism faculty in providing feedback on the pilot ahead of its formal launch as part of the Newseum Educational Institute.

Finally, the project continues in academic year 2017-18 under the guidance of Nathaniel Persily, the James B. McClathchey Professor of Law, who has assembled an interdisciplinary team of law, communications, public policy, and computer science students to research and extend conclusions from this report. The continuing work of the Fake News and Misinformation practicum will produce a second set of findings and recommendations in summer 2018.

About the Stanford Law School Policy Lab

Engagement in public policy is a core mission of teaching and research at Stanford Law School. The Law and Policy Lab (The Policy Lab) offers students an immersive experience in finding solutions to some of the world's most pressing issues. Under the guidance of seasoned faculty advisers, Law and Policy Lab students counsel real-world clients in an array of areas, including education, intellectual property, public enterprises in developing countries, policing and technology, and energy policy.

Policy labs address policy problems for real clients, using analytic approaches that supplement traditional legal analysis. The clients may be local, state or federal public agencies or officials, or private non-profit entities such as NGOs and foundations. Typically, policy labs assist clients in deciding whether and how qualitative or quantitative empirical evidence can be brought to bear to better understand the nature or magnitude of their particular policy problem, and identify and assess policy options. The methods may include comparative case studies, population surveys, stakeholder interviews, experimental methods, program evaluation or big data science, and a mix of qualitative and quantitative analysis. Faculty and students may apply theoretical perspectives from cognitive and social psychology, decision theory, economics, organizational behavior, political science or other behavioral science disciplines. The resulting deliverables reflect the needs of the client with most resulting in an oral or written policy briefing for key decision-makers.

Directed by former SLS Dean Paul Brest, the Law and Policy Lab reflects the school's belief that systematic examination of societal problems, informed by rigorous data analysis, can generate solutions to society's most challenging public problems. In addition to policy analysis, students hone the communications skills needed to translate their findings into actionable measures for policy leaders and the communities they serve. The projects emphasize teamwork and collaboration, and many are interdisciplinary, giving law students the opportunity to work with faculty and colleagues from across the university with expertise in such fields as technology, environmental engineering, medicine, and international diplomacy, among others.

Table of Contents

Executive Summary

On December 4, 2016, a man walked into a pizza restaurant in Washington, D.C., armed with an assault rifle. He pointed the gun at employees before firing several shots into a locked door. After his arrest, he explained to police that he was investigating claims he had read on the Internet that the pizza restaurant, Comet Ping Pong, was at the center of a massive child sex ring run by Democratic Party officials.[1] After failing to find any abused children at the restaurant, the man admitted that his "intel on this wasn't 100 percent."[2]

The terrified patrons, employees, and owner of Comet Ping Pong aren't the only recent victims of false information. Evidence reveals that false news is distorting politics, sowing confusion, and undermining trust in democratic institutions. An analysis by Buzzfeed News found that during the last few months of the 2016 U.S. presidential campaign, the 20 top-performing fake election stories generated more total engagement on Facebook than the 20 top-performing real stories from mainstream news outlets.[3] A survey released by the Pew Research Center in December 2016 found that 64 percent of Americans believe that fabricated news has caused "a great deal of confusion."[4] Multiple studies have found that trust in the media is at an all-time low.[5] In the year following the 2016 U.S. national elections, a literature is emerging illuminating the presence of fake news during and since the election and its impact on civic discourse. At the time of our research (April 2017 to July 2017), most of these studies focused on the role of the media and how stories were shared across social media, without attention to the role of search engines. In recent months, scholars, journalists, and the Senate Intelligence Committee have begun to track the roles of social media, with particular attention to Facebook and, to a lesser extent, Google and Twitter. This report offers a foray into the roles of social media and the Google search engine in proliferating fake news and misinformation.[6]

This report was produced by ten Stanford University students and overseen by former Senator Russ Feingold as part of a Law and Policy Lab practicum, a course that analyzes a current policy issue on behalf of a client. The students have diverse educational backgrounds in law, communications, and computer science. The report was prepared at the request of the William and Flora Hewlett Foundation's Madison Initiative, which is dedicated to helping "create the conditions in which Congress and its members can deliberate, negotiate, and compromise in ways that work for most Americans." For this

[1] Adam Goldman, *The Comet Ping Pong Gunman Answers Our Reporter's Questions*, N.Y. TIMES (Dec. 7, 2016), https://www.nytimes.com/interactive/2016/12/10/business/media/pizzagate.html.
[2] *Ibid.*
[3] Craig Silverman, *This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook*, BUZZFEED (Nov. 16, 2016), https://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook.
[4] Michael Barthel, Amy Mitchell, and Jesse Holcomb, *Many Americans Believe Fake News Is Sowing Confusion*, PEW RESEARCH CENTER (Dec. 15, 2016), http://www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/.
[5] *See e.g.*, Art Swift, *Americans' Trust in Mass Media Sinks to New Low*, GALLUP (Sept. 14, 2016), http://www.gallup.com/poll/195542/americans-trust-mass-media-sinks-new-low.aspx.
[6] Since completing research for this report in August 2017, Yale

report, Hewlett is particularly interested in examining proposals that could help "re-establish relevant facts as the primary currency for political discourse."[7]

The research focuses on four major online platforms: Facebook, Google, Twitter, and Reddit, in order of their probable effect on proliferating misinformation through policies, algorithms, and user actions. The research team divided into groups to analyze each platform individually, focusing primarily on how users accessed and shared news stories.[8] The groups reviewed recent scholarship, news articles, legal cases, laws, and regulations; they examined the effect of each platform's algorithms and performed user experiments to examine how false information spreads on each platform. The Google research team also analyzed the presence of fake news in Google searches for Congressional candidates in nine states. To gain perspective on journalists' thinking about the impact of fake news on mainstream journalism outlets, the teams met with the Stanford Knight journalism fellows and national news reporters with the *Washington Post* and *New York Times*, among other major news organizations. To better understand the national security implications of fake news, one team interviewed in person several high-level past and present government officials. The teams reviewed corrective measures that the platforms have recently taken to enhance users' awareness of fake news and they outlined an array of options for next steps to address the problem. A second phase of research is continuing in the 2017-18 academic year under the guidance of Nathaniel Persily, the James B. McClatchey Professor of Law at Stanford Law School.

As a starting place for understanding the scope of the problem of "fake news" on the platforms, the research teams identified a few main categories of misinformation:

- **Bogus Click-Bait**: These are entirely fabricated stories designed to gain ad revenue by tricking social media users into viewing and sharing them. The scale of this problem gained more attention in late 2016 when Macedonian teenagers told NBC News that they had made tens of thousands of dollars in just a few months by producing false stories about U.S. politics.[9]
- **Conspiracy Theories:** These are baseless and usually outlandish claims, typically asserting some nefarious secret plot. Unlike intentionally false stories, the promoters of conspiracy theories often believe their own claims.
- **Reckless Reporting:** This category refers to false stories or unsubstantiated rumors that are reported without regard for the truth.
- **Political Propaganda**: These are false stories designed to achieve a particular political outcome. U.S. intelligence agencies agree that Russia undertook an

[7] Kelly Born, *Memo: D/Misinformation and Propaganda*, Madison Initiative (April 2017).
[8] This report does not address the issue of targeted ad campaigns. For further information, see, for example, Manu Raju, Dylan Byers and Dana Bash, *Russian-linked Facebook ads targeted Michigan and Wisconsin*, CNN POLITICS, 10/4/17, amp.cnn.com.
[9]Alexander Smith and Vladimir Banic, *Fake News: How a Partying Macedonian Teen Earns Thousands Publishing Lies*, NBC NEWS (Dec. 9, 2016), http://www.nbcnews.com/news/world/fake-news-how-partying-macedonian-teen-earns-thousands-publishing-lies-n692451.

extensive political propaganda campaign in key states to influence the 2016 election.[10]

- **Hoaxes**: These are fake stories that are typically created as a joke or to make people look foolish for believing them.
- **Misleading News:** This category refers to all news content that is sensational, biased, slanted, unfair, or that otherwise leaves the reader with an inaccurate impression. While this material clearly has a negative effect on public discourse, it is nearly impossible to identify objectively. Not all "advocacy journalism" (reporting with a stated viewpoint) is necessarily bad for public understanding.

Even careful news outlets that subscribe to high journalistic standards sometimes make mistakes. Thus, this research did not examine stories that contained inadvertent errors and corrections. Nor does it address inaccurate statements by political figures. The research focuses only on provably false or misleading news content shared on four major online platforms, Facebook, Google, Twitter and Reddit. At the time of the research, very little information was publicly available about the ways in which the platforms enabled foreign actors and other purveyors of false information to target groups through advertising and simple posts. The next wave of research now focuses on those phenomena with Facebook revealing 3000 targeted ads, garnering 10 million views, paid for by the Russian-controlled Internet Research Agency. Jonathan Adler, Research Director at the Tow Center for Digital Journalism at Columbia University, offers further data showing the far more extended reach of Facebook simple posts, which were shared hundreds of millions of times.[11]

Examination of the problem of "fake news" hinges on how the term has transformed into an attack on mainstream professional journalists. Between his inauguration on November 8 and the completion of this research project on July 3, 2017, President Trump tweeted about the "fake" news media at least 30 times.[12] On February 17, the president called major media outlets "the enemy of the American People."[13] On July 2, he tweeted a video clip of himself wrestling and punching a person whose head was replaced with CNN's logo.[14] It seems inevitable that this sustained attack on the press undermines public trust in journalism and harms the effort to "re-establish relevant facts as the primary currency for political discourse."[15] The president's tweets, however, are not the focus of this report. In recognition of the inflammatory, and abused, connotation of the phrase "fake

---

[10] Patricia Zengerle, *Russian 'propaganda on steroids' aimed at 2016 U.S. election*, REUTERS (Mar. 30, 2017), http://www.reuters.com/article/us-usa-russia-congress-idUSKBN17126T.http://www.reuters.com/article/us-usa-russia-congress-idUSKBN17126T This is one of many articles describing Russian interference in the U.S. 2016 national election.

[11] For data on Facebook posts, see, Jonathan Albright, Itemized Posts and Historical Engagement - 6 Now-Closed FB Pages, Tableau data, https://public.tableau.com/profile/d1gi#!/vizhome/FB4/TotalReachbyPage; for background and context, Craig Timberg, Washington Post, Oct. 5, 2017, https://www.washingtonpost.com/news/the-switch/wp/2017/10/05/russian-propaganda-may-have-been-shared-hundreds-of-millions-of-times-new-research-says/?tid=a_inl&utm_term=.5b83cb77c838.

[12] Donald Trump, (@realDonaldTrump), TWITTER https://twitter.com/search?l=&q=fake%20from%3Arealdonaldtrump%20since%3A2017-01-20%20until%3A2017-07-03&src=typd&lang=en (last visited July 3, 2017).

[13] Donald Trump, (@realDonaldTrump), TWITTER (Feb. 17, 2017), https://twitter.com/realDonaldTrump/status/832708293516632065.

[14] Donald Trump, (@realDonaldTrump), TWITTER (July 2, 2017), https://twitter.com/realDonaldTrump/status/881503147168071680.

[15] K. Born, Memo (April 2017).

news," the report uses such terms as "misinformation" or "false news," which more clearly convey provable irregularities in facts and analysis.

Addressing this problem is a difficult task without any quick or obvious solutions. In many ways, trying to elevate nuanced and accurate information about politics is fighting against human nature. People are psychologically pre-disposed to read and share news that reinforces their existing political beliefs. Sensational headlines get more attention than those that are cautious or complex. Nevertheless, this research is a worthwhile step in understanding the role that major social media platforms play today in disseminating information – and misinformation – to the public. There is broad resistance to the idea of turning Silicon Valley companies into the ultimate arbiters of truth in our society. At the same time, these global platforms are already making countless decisions that influence how consumers respond to information. Should social media algorithms favor cat videos or news content based solely on the number of user clicks? Should conspiracy theories on fringe blogs be presented equivalently with *New York Times* articles in searches on the same subjects? Should uplifting feel-good stories get more algorithmic exposure than do articles about pressing social problems? In a vibrant, deliberative democracy, answers to these questions evoke legitimate public discussion and examination, and reveal the inherent limitations of the platforms' algorithms.

This first phase of inquiry into the phenomenon of false news and misinformation in Election 2016 reached the following findings and outlined these policy options:

Facebook

Findings:

- Users are inclined to believe the news they see on Facebook, whether it is real or not.
- Users' education levels are positively correlated with vigilance in assessing source credibility and identifying false information.
- Users express confidence in their own individual understandings of Facebook's user tools, but a majority of Facebook users did not know that Facebook has added a tool that enables users to report false news, and an even larger group of users have not used the tool.
- Facebook users' knowledge of the new tools is positively associated with believing that Facebook is resolving the fake news problem. The mere presence of the fact-checking tool seems to reassure users that news shared on the platform is valid.
- By implication, this "trust bias" suggests an inherent irony where heightened trust in the platform's actions lowers users' vigilance in checking the sources of news shared on their pages, thereby subverting the value of the new tools in limiting the spread of fake news across the platform.

Options for Facebook:

- More aggressively fact-check news across the platform, flagging it as needed.
- More aggressively market the false news flagging tool to users, but enhance cautionary guidelines.
- Monitor and make public users' behavior in checking sources.

- Increase scholars' access to site data for research purposes. To preserve users' privacy, the data should be limited to news-related content and engagement. Such information will benefit users by enabling third-party researchers to independently assess the reliability of Facebook as a source of news and information.

Google

Findings:

- There has been a long-running "arms race" between search engines and spammers who seek to leverage or monetize the platform's algorithm for personal gain.
- Sites known for promoting false information frequently appeared in the top search results for politicians' names during the 2016 election revealing the effects of an algorithm based, in part, on volume of clicks.

Options for Google:

- Explain search result rankings in plain terms so that average users better understand the limitations of their searches,
- Continue to guard against spammers tricking the algorithm, both through algorithmic monitoring and improvements and through terms of use to defend against abuse,
- Continue to adapt self-regulation tools but also explore possible regulatory frameworks that would help the company limit abuse by spammers.

Twitter

Findings:

- Twitter invites and amplifies adversarial interactions and debates between users.
- Users are more likely to believe content tweeted from an account that is verified with a blue checkmark, sometimes conflating the authenticity of accounts associated with people Twitter deems to be "in the public interest" with accurate information. Users may conflate an officially verified account as a source of "verified" news, though Twitter offers no such assurance.
- Users view tweets with links as more trustworthy but often do not click on the links to check the sources, again revealing a "trust bias" that diminishes users' vigilance and skepticism.
- While Twitter's public-facing platform may encourage users to be accountable for content they post, conversely, it also makes Twitter conversations susceptible to bots and cyborgs, which can highjack conversations to proliferate misinformation.

Options for Twitter:

- Enhance and clarify the verification system.
- Encourage users to clarify or link the source of material they share.
- Pilot a crowd-sourced, curated false news flagging and fact-checking system.
- Even with the understanding that not all bots are negative, implement stricter Captcha gateways to diminish bot activities.

- Develop terms of use that help diminish the proliferation of cyborgs and bots.
- Create more interaction options to supplement "linked response," "retweet," and "favorite" and allow users to disapprove of tweets.
- Share usage data with scholars.

Reddit

Findings:

- Bogus click-bait is less a problem on Reddit than on the other major sites.
- Conspiracy theories are a significant problem on Reddit in comparison to other major sites.
- Individual communities (called "subreddits") set their own rules and expectations.
- Reddit administrators are generally hands-off but changes in algorithmic rankings, including the front page, seem aimed at reducing the prominence of controversial communities.[16]
- Reddit is more transparent with changes to its platform than are the other major sites.[17]

Options for Reddit:

- Use reminders or alerts to encourage users to be skeptical of unverified content.
- Limit the exposure of the particular subreddits commonly known to promote false information.
- Work directly with media organizations to ensure that verified news articles load quickly, especially on mobile devices.

This report is organized into seven sections:

- Section 1: An overview of the legal and regulatory landscape for online platforms and misinformation.
- Sections 2 – 5: Reports on each major online platform: Facebook, Twitter, Google, and Reddit. The order of these reports more or less reflects the research team's general sense of the public influence of these platforms in proliferating misinformation beginning with the 2016 U.S. election through July 2017.
- Section 6: Next steps and priorities in further research.
- Section 7: Analysis of implications for democratic institutions.

This report surveys the roles of four major online platforms in proliferating and, more recently, self-regulating the spread of false news and misinformation as an essential part of U.S. civic culture. It offers insight into the platforms' internal policies, the effect of search engine optimization algorithms in the U.S. 2016 election, and documents how the four major platforms are now beginning to self-regulate the spread of misinformation,

[16] Like the other platforms, Reddit went public with algorithmic changes shortly after the 2016 election. See, for example, https://techcrunch.com/2016/12/06/reddit-overhauls-upvote-algorithm-to-thwart-cheaters-and-show-the-sites-true-scale/ and https://www.theverge.com/2017/2/15/14632390/reddit-front-page-popular-change-new-users.
[17] See Reddit Changelog, https://www.reddit.com/r/changelog/.

including hoaxes, malignant links and propaganda by foreign actors, and amplification of that information through bots. It further investigates users' responses to some of the methods that the companies are using to self-regulate, revealing that fact-checking mechanisms are not a panacea, and in some instances, may exacerbate the spread of misinformation.

While the U.S. regulatory environment continues to protect intermediary platforms from liability in content posted by users, the 2016 election raised public awareness of the ease by which the platforms' algorithms are gamed and manipulated. This report is intended as a further catalyst to the dialogue about how the platforms can effectively counter bad actors in gaming those systems and develop effective internal tools to prevent the spread of false information and promote public access to the factual information necessary to a healthy and vibrant democratic marketplace of ideas.

Section 1. Legal and Regulatory Analysis

I. Introduction

This section explores the legal framework for intermediary liability and the possible options available to contain the spread of misinformation. Defamation lawsuits, for example, could be used to combat some types of false statements that damage a person's reputation. The regulatory powers of the Federal Trade Commission (FTC) and the Federal Communications Commission (FCC) are circumscribed by two main barriers to legal action: The First Amendment and Section 230 of the Communications Decency Act (CDA), which shields intermediary platforms from liability.

Although some legal scholars and policy makers think that the time has come to reexamine the limitations of CDA Sec. 230, legal reforms are just one part of broader public policy and civic engagement in solving the problem of misinformation in our public discourse. This section investigates the legal landscape with this broader vision in mind.

II. Defamation

Some publishers of false information could be liable for civil damages in a defamation lawsuit. The purpose of defamation law is to protect against harm to one's reputation, and it has a long history, dating back to pre-industrial England.[18] Under U.S. common law, to win a defamation case, a plaintiff must show: 1) the defendant made a false statement about the plaintiff; 2) the statement was published to a third party; 3) the defendant was negligent (or worse); and 4) the statement injured the reputation of the plaintiff.[19] Additionally, the statement must be a factual claim, not opinion. There are two kinds of defamation: libel and slander. Generally, libel is written, while slander is spoken.

Defamation law is well-developed and the most obvious tool for attacking misinformation. But the biggest stumbling block is that defamation law is focused on protecting the reputation of individuals, as opposed to preventing the spread of false statements more broadly. Declaring that 2+2=5 is not defamatory, for example, because the equation, though false, is not about any particular individual. Additionally, even if a statement is about an individual, that statement must have harmed the individual's reputation to support a defamation suit. It is not enough to show that the person was annoyed by a lie — the lie would have to expose the person to "hatred, contempt, or ridicule" or subject the person to a "loss of the good will and confidence in which he or she is held by others."[20] Some statements, such as accusing a person of committing a serious

[18] Robert C. Post, *The Social Foundations of Defamation Law: Reputation and the Constitution*, 74 CAL. L. REV. 691 (1986).
[19] Restat 2d of Torts, § 558.
[20] *Romaine v. Kallinger*, 109 N.J. 282, 306 (1988).

crime, are considered "defamation *per se*"— meaning that courts assume that the statement harmed a person's reputation without any further analysis.[21]

Additionally, even if a plaintiff can prove all the elements of defamation, the Constitution may still shield the defendant from liability. In *New York Times v. Sullivan*, the Supreme Court held that the First Amendment requires a plaintiff who is a public figure to prove "actual malice" — that is, that the defendant knowingly lied or acted with reckless disregard to the truth.[22] The Court explained that "erroneous statement is inevitable in free debate, and that it must be protected if the freedoms of expression are to have the 'breathing space' that they 'need . . . to survive.'"[23] Actual malice is a high bar that dooms many defamation suits. It is difficult to find evidence to prove that defendants were knowingly lying or unconcerned with whether their statements were true. A defendant could acknowledge making a careless mistake and still rely on the First Amendment for total protection.

If the plaintiff is a private individual, she must show that the defendant acted with at least negligence.[24] But even in these cases, the plaintiff must show actual malice to get punitive damages.[25] Since the cost of litigation can be so high, many defamation lawsuits only make sense if punitive damages are a possibility. So, for practical purposes, even private individuals often only sue if they can show that the defendant knowingly lied or acted with reckless disregard for the truth.

Despite these significant obstacles, libel suits (or the threat of libel suits) can still be effective in combatting some limited forms of misinformation. Alex Jones, the conspiracy theorist who owns InfoWars.com, publicly apologized and retracted a story in March 2017 claiming that the pizza restaurant Comet Ping Pong was a front for child abuse.[26] In May, he retracted a story claiming that yogurt-maker Chobani was "importing migrant rapists."[27] In both cases, Jones only backed down when threatened with libel lawsuits.[28]

Generally, the remedy in a defamation case is monetary damages. Some states have laws that impose criminal penalties for defamation, but these laws are rarely, if ever,

---

[21] *What is a Defamatory Statement*, DIGITAL MEDIA LAW PROJECT, http://www.dmlp.org/legal-guide/what-defamatory-statement (last visited July 3, 2017).
[22] New York Times v. Sullivan, 376 U.S. 254, 272 (1964).
[23] *Id.* at 274, quoting N. A. A. C. P. v. Button, 371 U.S. 415, 433.
[24] Gertz v. Robert Welch, 418 U.S. 323, 339 (1974).
[25] *Id.*, at 349.
[26] Alex Jones, *A Note to Our Listening, Viewing and Reading Audiences Concerning Pizzagate Coverage*, INFOWARS (March 24, 2017), https://www.infowars.com/a-note-to-our-listening-viewing-and-reading-audiences-concerning-pizzagate-coverage/.
[27] David Montero, *Alex Jones settles Chobani lawsuit and retracts comments about refugees in Twin Falls, Idaho*, L.A. TIMES (May 17, 2017), http://www.latimes.com/nation/la-na-chobani-alex-jones-20170517-story.html.
[28] Paul Farhi, *Conspiracy theorist Alex Jones backs off 'Pizzagate' claims*, WASH. POST (March 24, 2017), https://www.washingtonpost.com/lifestyle/style/conspiracy-theorist-alex-jones-backs-off-pizzagate-claims/2017/03/24/6f0246fe-10cd-11e7-ab07-07d9f521f6b5_story.html.

enforced.[29] Injunctions blocking speech prior to publication are known as "prior restraints," and are extremely rare.[30] Courts consider these injunctions a direct form of censorship that almost always violates the First Amendment. Defamation law is therefore an imperfect tool for preventing the spread of misinformation. It can, however, force producers of misinformation to retract harmful lies and to compensate their victims. The potential for punitive damages can also discourage the reckless publication of misinformation in the first place.

While defamation lawsuits might curb some harmful lies, they also have the potential to chill legitimate speech. A rich plaintiff could use defamation lawsuits as a weapon to punish journalists who publish true information that the plaintiff simply doesn't like. Litigating a defamation case can cost millions of dollars, even if the defendant ultimately prevails.[31] An increase in defamation litigation could drive more news organizations out of business or intimidate others into silence.[32] So there is a risk that more defamation litigation could harm, rather than help, access to true information.

To try to address this potential for abuse, dozens of states have passed "anti-SLAPP" statutes.[33] These statutes, which aim to block "Strategic Lawsuits Against Public Participation," give defendants a mechanism for getting certain lawsuits dismissed early in litigation before they become too costly. The details of the statutes vary from state to state, but they generally cover lawsuits that target written or oral statements about a matter of public interest.[34] The plaintiffs can still overcome these "anti-SLAPP" motions, but they must produce evidence to support their claims. Some states, including California, allow defendants to recover attorney fees if they win an anti-SLAPP motion.[35]

While President Trump has said he wants to "open up" libel laws,[36] any major changes are unlikely. There is no federal defamation law — all lawsuits are based on state law. In theory, Congress could create a federal cause of action for interstate libel using its

---

[29] *Criminal Defamation Laws in North America*, COMMITTEE TO PROTECT JOURNALISTS, https://cpj.org/reports/2016/03/north-america.php (last visited July 3, 2017).

[30] *See* Neb. Press Ass'n v. Stuart, 427 U.S. 539, 559 (1976). "[P]rior restraints on speech and publication are the most serious and the least tolerable infringement on First Amendment rights."

[31] *See e.g.*, Clara Jeffery and Monika Bauerlein, *We Were Sued by a Billionaire Political Donor. We Won. Here's What Happened.*, MOTHER JONES (Oct. 8, 2015), http://www.motherjones.com/media/2015/10/mother-jones-vandersloot-melaleuca-lawsuit/.

[32] Although Hulk Hogan's suit against Gawker was based on a privacy invasion claim, not defamation, it is still a prime example of a civil suit driving a media company into bankruptcy. Notably, the case was bankrolled by a billionaire, Peter Thiel, who disliked Gawker's coverage of him. *See* Tiffany Kary and Steven Church, *Gawker Founder in Bankruptcy After Losing Hulk Hogan Case*, BLOOMBERG (Aug. 1, 2016), https://www.bloomberg.com/news/articles/2016-08-01/gawker-media-founder-nick-denton-files-personal-bankruptcy.

[33] *State Anti-SLAPP Laws*, PUBLIC PARTICIPATION PROJECT, https://anti-slapp.org/your-states-free-speech-protection/ (last visited July 3, 2017).

[34] *Id.*

[35] *Anti-SLAPP Law in California*, DIGITAL MEDIA LAW PROJECT, http://www.dmlp.org/legal-guide/anti-slapp-law-california (last visited July 3, 2017).

[36] Hadas Gold, *Donald Trump: We're going to 'open up' libel laws*, POLITICO (Feb. 26, 2016), http://www.politico.com/blogs/on-media/2016/02/donald-trump-libel-laws-219866.

power under the Commerce Clause,[37] but the First Amendment would still significantly curtail any efforts to expand the power of plaintiffs in defamation suits. A new federal law could preempt state anti-SLAPP statutes, making it more difficult for defendants to defeat defamation lawsuits early in litigation. These kinds of proposals, however, risk stifling legitimate speech without much effect on the spread of misinformation.

III. False Light

Defamation isn't the only civil tort action that could be used to combat misinformation. Courts in many states also recognize a "false light" claim, which is a kind of privacy tort. To win this claim, the plaintiff must show that the defendant cast the plaintiff in a "false light" that 1) would be highly offensive to a reasonable person; and 2) that the defendant acted with reckless disregard for the truth.[38] False light is similar to defamation—the main difference is that defamation is about harm to one's reputation, while false light is about harm to dignity.

IV. Intentional Infliction of Emotional Distress

Another possible claim is "intentional infliction of emotional distress," or "IIED." A defendant is liable under this claim if he engages in "extreme and outrageous conduct [that] intentionally or recklessly causes severe emotional distress."[39] Often, plaintiffs will bring IIED, false light, and defamation claims all at once. For example, the family of Seth Rich, a murdered former Democratic National Committee staffer, appeared to threaten an IIED lawsuit in a cease-and-desist letter to a private investigator who was fueling right-wing conspiracy theories about their son's death.[40] Like defamation claims, IIED and false light are still subject to First Amendment limitations and anti-SLAPP motions.

A potential area for further research is whether non-wealthy plaintiffs can afford to pursue lawsuits over defamation and other similar claims. Although many plaintiff law firms work on a contingency basis (collecting fees as a portion of their client's recovery), there is not much data available on whether this business model provides access to everyone with legitimate defamation claims. Are defamation lawsuits primarily a tool for the rich? Any effort to make defamation lawsuits more common, however, should also consider the potential harm to free speech and journalism.

---

[37] Michael Walsh, *Libel Law Under President Trump*, LAW360 (May 2, 2016), https://www.law360.com/articles/791471/libel-law-under-president-trump.
[38] Restat 2d of Torts, § 652E (2nd 1979).
[39] Restat 2d of Torts, § 46 (2nd 1979).
[40] "Your statements and actions have caused, and continue to cause, the Family severe mental anguish and emotional distress. Your behavior appears to have been deliberate, intentional, outrageous, and in patent disregard of the Agreement and the obvious damage and suffering it would cause the Family." Alex Seitz-Wald, *Slain DNC Staffer's Family Orders Blabbing Detective to 'Cease and Desist'*, NBC NEWS (May 19, 2017), http://www.nbcnews.com/politics/justice-department/slain-dnc-staffer-s-family-orders-blabbing-detective-cease-desist-n762211.

V. FTC Enforcement

Long before the recent attention on "fake news," one federal regulator was already cracking down on the problem. In 2011, the Federal Trade Commission sued 10 operators of "fake news websites" that were selling acai berry weight loss supplements.[41] The companies had allegedly set up websites with titles like "News 6 News Alerts" or "Health News Health Alerts" and included logos of major media organizations. The articles presented fabricated stories of incredible weight loss as if they were being told by legitimate investigative journalists. The FTC accused the companies of engaging in "deceptive" business practices in violation of Section 5 of the Federal Trade Commission Act. In 2015, the FTC released guidelines to businesses on how to produce "native advertising" (paid ads that look like news articles or other content) without deceiving consumers.[42]

Ari Melber, the chief legal affairs correspondent for MSNBC, has suggested that the FTC should launch a broader attack on fake news. In an article for the New Jersey Bar Association in January, Melber argued that the FTC could sue websites that promote false content for engaging in a "deceptive" practice.[43] The FTC has the power to obtain court orders to halt violations of the FTC Act, and, if those orders are violated, to impose civil fines.[44]

The key problem with Melber's argument is that the FTC only has jurisdiction over commercial conduct. The weight loss ads were trying to trick consumers into buying a product, which is different than a general false statement about politics. In an interview with the *Washington Post*, David Vladeck, a former FTC official who is now a professor at Georgetown University Law Center, said he doubts the agency can tackle most forms of misinformation on the Internet. "The FTC's jurisdiction extends only to cases where someone is trying to sell something," he said. "Fake news stories that get circulated or planted or tweeted around are not trying to induce someone to purchase a product; they're trying to induce someone to believe an idea. There are all sorts of First Amendment problems, apart from, I think, the insuperable jurisdiction problems that the FTC would have."[45]

In his article, Melber argued that the FTC and the courts should adopt a more creative view of the agency's powers. Fraud news sites, Melber argued, are pushing a

---

[41] *FTC Seeks to Halt 10 Operators of Fake News Sites from Making Deceptive Claims About Acai Berry Weight Loss Products*, FED. TRADE COMM'N (April 19, 2011), https://www.ftc.gov/news-events/press-releases/2011/04/ftc-seeks-halt-10-operators-fake-news-sites-making-deceptive.

[42] *Native Advertising: A Guide for Businesses*, FED. TRADE COMM'N https://www.ftc.gov/tips-advice/business-center/guidance/native-advertising-guide-businesses.

[43] Ari Melber, *Regulating 'Fraud News'*, NEW JERSEY STATE BAR ASSOCIATION (Jan. 30, 2017), https://tcms.njsba.com/PersonifyEbusiness/Default.aspx?TabID=7512.

[44] *A Brief Overview of the Federal Trade Commission's Investigative and Law Enforcement Authority*, FED. TRADE COMM'N, https://www.ftc.gov/about-ftc/what-we-do/enforcement-authority (last visited July 3, 2017).

[45] Callum Borchers, *How the Federal Trade Commission could (maybe) crack down on fake news*, WASH. POST (Jan. 30, 2017), https://www.washingtonpost.com/news/the-fix/wp/2017/01/30/how-the-federal-trade-commission-could-maybe-crack-down-on-fake-news/.

product: misinformation. Their goal is to trick consumers into spending their time reading a fake news article that will expose them to advertising. "We live in a world where most news consumers never purchase their news directly," Melber wrote. "They consume it online in exchange for viewing ads, or in exchange for providing their personal information (instead of money). An FTC framework for fraud news would treat these readers as 'consumers,' and target the websites for deceptive acts against them."

Even if the courts refuse to adopt this expansive view of the FTC's powers, Congress could rewrite the agency's statute to explicitly give it that authority. Any attempt by an agency to regulate news content would face First Amendment challenges, although the Supreme Court has held that the First Amendment provides lower protection for commercial speech.[46] Regulating "fraud" in news content under a framework of commercial deception in the sale of advertising or the harvesting of personal information (that is later sold) may put the regulation on more defensible legal ground, although there is no guarantee that courts would accept this framing.

There would also be significant political pushback to giving the FTC more power to police misinformation. In February, Mike Masnick, a free speech advocate who edits the TechDirt blog, wrote that letting the FTC define fake news would be giving a political agency a dangerous tool for stifling free speech. "Do we really want to give the FTC — whose commissioners are appointed by the President — the power to take down news for being 'fake'?" he asked. "[A]ll of this puts the onus on the government to fix the fact that some people deal in confirmation bias and believe things that aren't true. Censoring 'fake news' doesn't solve that problem. It just creates yet another tool for censorship." [47]

VI. FCC Regulation

The Federal Communications Commission— which regulates communications over the airwaves and through wires[48]—already has an anti-hoax rule. But the rule, which is extremely narrow, has rarely been enforced.[49] The regulation bars over-the-air television and radio stations from broadcasting false information about a "crime or a catastrophe" if 1) the station knows the information is false; 2) it is foreseeable that the broadcast of the information will cause substantial public harm; and 3) the broadcast of the information does in fact directly cause substantial public harm.[50] The FCC enacted the rule after several on-air incidents, including one in which a radio personality falsely claimed the station was being held hostage and another where the station said that a local trash dump was exploding like a volcano.[51]

---

[46] Cent. Hudson Gas & Elec. Corp. v. Pub. Serv. Comm'n, 447 U.S. 557, 563 (1979).
[47] Mike Masnick, *Bad Idea Or The Worst Idea? Having The FTC Regulate 'Fake News'*, TECHDIRT (Feb. 2, 2017), https://www.techdirt.com/articles/20170201/23481336610/bad-idea-worst-idea-having-ftc-regulate-fake-news.shtml.
[48] 47 U.S. Code § 151.
[49] David Oxenford, *On April Fools Day – Remember the FCC's Hoax Rule!*, BROADCAST LAW BLOG (March 28, 2013), http://www.broadcastlawblog.com/2013/03/articles/on-april-fools-day-remember-the-fccs-hoax-rule/.
[50] *Hoaxes*, FED. COMMC'N COMM'N, https://www.fcc.gov/reports-research/guides/hoaxes (last visited July 3, 2017).
[51] Oxenford, *supra* note 43.

Because it only covers radio and TV broadcasts about crimes or catastrophes, the FCC's anti-hoax rule would not be of much use in fighting the broader problem of misinformation. But it is worth keeping in mind that the FCC is the U.S. agency with the most experience regulating the media industry. Even though many of its policies don't cover the Internet, briefly reviewing the history of FCC media regulation may be helpful in understanding how to improve access to true information today. The FCC, for example, has media ownership rules that cap how many radio and TV stations a single entity can own in a market.[52] The rules ban owners of a radio or TV station from also owning a newspaper in the same market.[53] The purpose of these rules is to combat media consolidation and ensure that the public has access to a variety of viewpoints. Current FCC Chairman Ajit Pai, however, has indicated he wants to roll back the media ownership restrictions, which he considers "antiquated."[54]

As the primary regulator of communications industries, the FCC has closely studied the state of media and journalism. Following up on a 2009 media study by the Knight Commission, a blue-ribbon panel of media, policy, and community leaders, the FCC released its own report in 2011 on access to information in the digital age.[55] By law, the FCC is also required to report to Congress every three years on barriers that may prevent entrepreneurs and small businesses from competing in the media marketplace.[56] To fulfill that statutory requirement (and to follow up on its 2011 study), the FCC in 2013 hired an outside research firm to help produce a "multi-market study of critical information needs."

This study, however, ignited a political firestorm that highlights the risk of the FCC even hinting at regulating news organizations. As part of the study, the FCC planned to launch a pilot program in South Carolina that would have sent voluntary questionnaires to news editors to ask how they decide which stories to cover. Conservative critics were outraged that the FCC was trying to intrude into newsrooms. In an op-ed, Ajit Pai (who was then a commissioner but not yet chairman), wrote that the "government has no place pressuring media organizations into covering certain stories,"[57] and Republicans in

---

[52] *FCC's Review of the Broadcast Ownership Rules*, FED. COMMC'N COMM'N, https://www.fcc.gov/consumers/guides/fccs-review-broadcast-ownership-rules (last visited July 3, 2017).
[53] *Ibid.*
[54] Ted Johnson, *FCC Chairman Ajit Pai Interview: Media Ownership Rules 'Quite Antiquated'*, VARIETY (March 14, 2017), http://variety.com/2017/biz/news/fcc-ajit-pai-media-ownership-1202008630/. As FCC Chairman, Ajit Pai recently relaxed these rules by reviving the "UHF" discount rule that permits broadcast companies to exceed regular limits on market ownership. See "How Trump's FCC aided Sinclair's expansion," *Politico*, 8-6-17, http://www.politico.com/story/2017/08/06/trump-fcc-sinclair-broadcast-expansion-241337 (last visited Aug. 14, 2017).
[55] Steven Waldman and the Working Group on Information Needs of Communities, *The Information Needs of Communities: The changing media landscape in a broadband age*, FED. COMMC'N COMM'N (July 2011), www.fcc.gov/infoneedsreport.
[56] Joe Flint, *FCC cancels newsroom study after backlash from lawmakers*, L.A. TIMES (Feb. 28, 2014), http://www.latimes.com/entertainment/envelope/cotown/la-et-ct-fcc-cin-study-20140228-story.html.
[57] Ajit Pai, *The FCC Wades Into the Newsroom*, WALL STREET JOURNAL (Feb. 10, 2014), https://www.wsj.com/articles/the-fcc-wades-into-the-newsroom-1392078376.

Congress began working on legislation to block the study.[58] The FCC first altered and then abandoned the entire project.[59]

Despite the risk of triggering these kinds of controversies, the FCC does have tools for shaping the media industry, such as its ability to impose conditions on corporate mergers and other deals. Transactions that involve the transfer of FCC licenses must receive approval from the agency. While these merger reviews largely focus on traditional antitrust issues such as market concentration and predicted effects on prices, the FCC's mandate is to block any deal that does not serve the "public interest."[60] So the FCC can use its merger review power to extract binding concessions from companies even if it lacks legal authority to pursue industry-wide regulations on those issues. For example, in order to buy NBC-Universal in 2011, Comcast agreed to offer a minimum number of hours of local news and information programming on NBC and Telemundo stations.[61] The company also agreed not to place competing cable news networks in undesirable channel spots and to increase its Spanish-language programming. With multibillion dollar deals on the line, companies are often willing to make substantial concessions to the FCC.

It is plausible that the FCC could use this merger power to require cable, telecom, and broadcast companies to do more to combat misinformation and invest in quality journalism. These promises could extend to online properties owned by these companies. The commitments, however, are only binding on the companies that are seeking transaction approval and are usually only temporary. Comcast's NBC-Universal commitments, for example, will expire next year. Another point of leverage over TV and radio stations comes when they must periodically renew their broadcast licenses. All broadcast stations are legally required to "serve the public interest as a public trustee."[62] The FCC could use this renewal process to demand more aggressive steps to provide access to truthful news and information.

The FCC's most famous media regulation was the "Fairness Doctrine." The rule, which was first enacted in 1949, required TV and radio stations to 1) "provide coverage of vitally important controversial issues of interest in the community served by the station" and 2) to "afford a reasonable opportunity for the presentation of contrasting viewpoints on such issues."[63] The Reagan administration's FCC abolished the Fairness Doctrine in 1987 and the FCC under President Obama formally eliminated the defunct policy from its code of regulations.

---

[58] Katy Bachmann, *Is the FCC's Latest Study an Attempt to Bring Back the Fairness Doctrine?*, ADWEEK (Dec. 10, 2013), http://www.adweek.com/digital/fccs-latest-study-attempt-bring-back-fairness-doctrine-154422/.
[59] Flint, *supra* note 50.
[60] *Mergers and Acquisitions*, FED. COMMC'N COMM'N, https://www.fcc.gov/proceedings-actions/mergers-and-acquisitions (last visited July 3, 2017).
[61] *FCC Grants Approval of Comcast-NBCU Transaction*, FED. COMMC'N COMM'N (Jan. 18, 2011), https://apps.fcc.gov/edocs_public/attachmatch/DOC-304134A1.pdf.
[62] *License Renewal Applications for Television Broadcast Stations*, FED. COMMC'N COMM'N, https://www.fcc.gov/media/television/broadcast-television-license-renewal (last visited July 3, 2017).
[63] *Report Concerning General Fairness Doctrine Obligations of Broadcast Licensees*, FED. COMMC'N COMM'N, 102 F.C.C.2d 143, 146 (1985).

When the rule was in effect, it allowed anyone who believed a station was failing to honor its obligations to file a complaint with the FCC, which would evaluate the complaints on a case-by-case basis. The Supreme Court upheld the constitutionality of the Fairness Doctrine in 1969, distinguishing broadcasts over scarce public airwaves from other forms of media.[64] Eventually, the FCC adopted two related rules: 1) the "personal attack rule," which required that if a person's integrity was attacked during a program on a controversial issue of public importance, the station had to inform the subject of the attack and provide an opportunity to respond on the air; and 2) the "political editorial rule," which required a station that endorsed a candidate for political office to offer the other candidates for that office an opportunity to respond.[65] Like the Fairness Doctrine itself, these rules applied only to broadcast TV and radio—not cable or satellite.[66] The FCC eliminated the "personal attack" and "political editorial" rules in 2000 after the D.C. Circuit ruled that the commission had failed to adequately justify them.[67] One requirement that is still in effect is the "equal time rule," which requires that if a station offers air time to one candidate, then it must offer all other candidates for that same office "equal opportunities" for time.[68] Although the rule applies to scripted spots, it exempts news programs and talk shows.

The Fairness Doctrine and its related rules have long been deeply controversial, and many critics argue that they stifled press freedom. Recalling his time as a journalist at a radio station owned by the *Houston Chronicle*, Dan Rather said:

> "I became aware of a concern which I previously had barely known existed—the FCC. The journalists at the Chronicle did not worry about it: those at the radio station did. Not only the station manager but the news people as well were very much aware of the Government presence looking over their shoulders. I can recall newsroom conversations about what the FCC implications of broadcasting a particular report would be. Once a newsroom has to stop and consider what a Government agency will think of something he or she wants to put on the air, an invaluable element of freedom has been lost."[69]

While some have blamed the demise of the Fairness Doctrine for the rise of conservative talk radio, other observers argue that a bigger factor was new technology that made nationwide syndication cheaper. Stations no longer had to develop their own local talent. "Rush Limbaugh was around before 1987," said Andrew Schwartzman, a media attorney and legal scholar. "In the 1980s what really happened was national syndication, and it happened in a big way."[70]

---

[64] Red Lion Broadcasting Co. v. Fed. Comm'n Comm'n, 395 U.S. 367 (1969).
[65] Waldman, *supra* note 49, at 278.
[66] *Ibid.*
[67] *Repeal or Modification of the Personal Attack and Political Editorial Rules*, Order, 15 FCC Rcd 20697 (2000).
[68] 47 CFR §73.1941.
[69] Waldman, *supra* note 49, at 278.
[70] Dante Chinni, *Is The Fairness Doctrine Fair Game?*, PEW RESEARCH CENTER (July 18, 2007), http://www.journalism.org/2007/07/18/is-the-fairness-doctrine-fair-game/.

Given the conservative loathing of the Fairness Doctrine and reluctance even among most liberals to police political speech, it seems unlikely that the FCC will revive the policy anytime soon. It is also unclear whether the Supreme Court, which has adopted a more expansive view of free speech in recent years, would follow its own precedent in upholding the regulations, particularly in an environment where the options for delivering news to consumers have expanded so greatly.

If the FCC did decide, however, to more aggressively regulate the media, could its policies cover Internet publications? The FCC enacted its initial media regulations when over-the-air TV and radio were primary ways that the public accessed information. Would it make sense now to apply similar regulations to the dominant form of communication in the 21st Century? A key reason that the Supreme Court has upheld FCC regulations of TV and radio content is that those transmissions rely on a scarce public resource — the airwaves. That rationale would presumably not apply to the Internet. The FCC, however, does have at least some authority to regulate Internet service.

In its 2015 Open Internet Order, the FCC dramatically expanded its authority over broadband Internet access by declaring it a "telecommunications service" under Title II of the Communications Act. This classification, which the FCC has used for more than 80 years to regulate landline phone service, essentially treats broadband as a public utility.[71] The FCC invoked this authority to enact regulations to protect net neutrality—the principle that Internet service providers should not be allowed to control or manipulate the online content that their customers can access. The current FCC, however, is considering a proposal to repeal the Title II classification of Internet service. [72]

The Obama-era FCC was careful to emphasize that it was only regulating the "transmission component of Internet access service" and not "any Internet applications or content."[73] So while the regulations apply to broadband providers like Comcast and Verizon, Web services like Facebook and Google are exempt. In November 2015, the FCC rejected a petition from a consumer advocacy organization to apply Title II privacy regulations to web services, holding that the "Commission has been unequivocal in declaring that it has no intent to regulate" Web services.[74]

---

[71] Both supporters and critics of applying Title II to the Internet have described it as "utility" regulation. *See, e.g.*, Sen. Ed Markey, *To Protect Net Neutrality, Markey Leads Senate Dems in Call to Reclassify Broadband as a Utility* (July 15, 2014) https://www.markey.senate.gov/news/press-releases/to-protect-net-neutrality-markey-leads-senate-dems-in-call-to-reclassify-broadband-as-a-utility; and Ajit Pai, *Why I'm trying to change how the FCC regulates the Internet*, L.A. TIMES (April 26, 2017), http://www.latimes.com/opinion/op-ed/la-oe-pai-fcc-internet-regulation-20170426-story.html.

[72] *In the Matter of Restoring Internet Freedom: Notice of Proposed Rulemaking*, FED. COMMC'N COMM'N (April 27, 2017), https://apps.fcc.gov/edocs_public/attachmatch/DOC-344614A1.pdf.

[73] *In the Matter of Protecting and Promoting the Open Internet: Report and Order on Remand, Declaratory Ruling, and Order*, FED. COMMC'N COMM'N (Feb. 26, 2015), https://apps.fcc.gov/edocs_public/attachmatch/FCC-15-24A1.pdf, ¶ 382.

[74] *In the Matter of Consumer Watchdog Petition for Rulemaking to Require Edge Providers to Honor 'Do Not Track' Requests: Order*, FED. COMMC'N COMM'N (Nov. 6, 2015), https://apps.fcc.gov/edocs_public/attachmatch/DA-15-1266A1.pdf, ¶ 1.

Even if the FCC wanted to expand its Title II authority to police Web content, it is not clear that the statute would support that interpretation. The Communications Act defines a "telecommunications service," in part, as the "offering of telecommunications for a fee directly to the public," while "telecommunications" means the "transmission" of information. It is possible, however, that the FCC could argue that it has "ancillary" authority over Web companies. For decades, the FCC has successfully argued in court that it has jurisdiction over issues related to its core regulatory functions.[75]

Another key regulatory power is Section 706 of the Telecommunications Act, which directs the FCC to "encourage the deployment" of broadband Internet service. The FCC could argue that misinformation is undermining trust in the Internet and discouraging people from adopting broadband service. Regulations combatting online misinformation could therefore "encourage the deployment" of broadband within the FCC's authority under Section 706, which is bolstered by its broad "ancillary" authority.

A federal appeals court has upheld a similar version of this argument. The FCC had first tried to enact net neutrality rules in 2010 without invoking its Title II utility-style powers, relying instead on Section 706. The D.C. Circuit in 2014 ruled that it was reasonable for the FCC to claim that its net neutrality rules would "preserve and facilitate the 'virtuous circle' of innovation that has driven the explosive growth of the Internet."[76] The court ultimately still struck down the rules because they too closely resembled common carriage regulation (which is not allowed for services not classified under Title II), but this expansive view of Section 706 is still a key precedent on FCC authority. There is, of course, no guarantee that a court would extend this logic to the regulation of misinformation, and there would also still be First Amendment barriers to FCC regulation of online content.

VII. First Amendment

The free speech protections of the First Amendment are the most significant barrier to any new law or regulation addressing the spread of misinformation. The Supreme Court has held that content-based restrictions of speech are presumptively unconstitutional.[77] The government can overcome this presumption only if the restriction is narrowly tailored to serve a compelling government interest.[78] In *Texas v. Johnson* in 1989, the Supreme Court explained that "[i]f there is a bedrock principle underlying the First Amendment, it is that the government may not prohibit the expression of an idea simply because society finds the idea itself offensive or disagreeable."[79] The Supreme Court, however, has allowed content restrictions on a "few historic categories of speech"— including obscenity, fraud,

[75] The FCC may regulate issues "reasonably ancillary to the effective performance of the Commission's various responsibilities." United States v. Sw. Cable Co., 392 U.S. 157, 178 (1968).
[76] Verizon v. FCC, 740 F.3d 623, 628 (D.C. Cir. 2014)
[77] *See, e.g.*, Austin v. Michigan Chamber of Commerce, 494 U.S. 652, 655 (1990).
[78] *Id.*
[79] Texas v. Johnson, 491 U.S. 397, 414 (1989).

incitement to violence, and defamation.[80] Commercial speech receives a lower level of "intermediate scrutiny."[81]

A key case for evaluating the constitutionality of any new misinformation measure is *United States v. Alvarez*. In that 2012 decision, the Supreme Court struck down the Stolen Valor Act, which made it a crime to falsely claim to have received a military decoration or honor. The government had argued that false statements have no value and deserve no First Amendment protection. The Supreme Court rejected this argument, warning that it could lead to the creation of an Orwellian "Ministry of Truth."[82] The Court warned that upholding the law "would give government a broad censorial power unprecedented in this Court's cases or in our constitutional tradition."[83] False statements can still fall outside First Amendment protections, but only when they are coupled with other elements as in defamation, fraud, or perjury.

This decision poses a serious challenge for any law targeting misinformation. For example, Ed Chau, a member of the California State Assembly, introduced legislation earlier this year that would make it unlawful to knowingly make a false statement designed to influence an election.[84] Given the fact that the Supreme Court has been especially protective of political speech,[85] it seems unlikely that simply adding the element of attempting to influence an election would allow the law to survive First Amendment scrutiny. After criticism from civil liberties advocates, Chau pulled the bill and revised it to more narrowly target websites that impersonate political candidates or groups with the intent to mislead or deceive.[86]

It is possible that the Supreme Court could one day backtrack on its expansive views of the First Amendment. Sustained academic and political advocacy for a new approach to evaluating free speech rights could eventually shift the Supreme Court's jurisprudence. Given the steady trend of the Court's First Amendment decisions and the difficulty in reversing Supreme Court precedent (and the legitimate concerns with regulating political speech), however, we recommend focusing on proposals that could fit within the existing legal framework.

VIII. Section 230: A Shield for Internet Platforms

Section 230 of the Communications Decency Act (CDA) largely forecloses legal action against online platforms for hosting false content created by others. The law states

[80] United States v. Stevens, 559 U.S. 460, 460 (2010).
[81] Cent. Hudson Gas & Elec. Corp. v. Pub. Serv. Comm'n, 447 U.S. 557, 563 (1979).
[82] United States v. Alvarez, 567 U.S. 709, 723 (2012).
[83] *Id.*
[84] AB-1104 THE CALIFORNIA POLITICAL CYBERFRAUD ABATEMENT ACT (2017-2018), California Legislative Information, http://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB1104.
[85] *See, e.g.*, Snyder v. Phelps, 562 U.S. 443 (2011).
[86] *Ibid.*

that "[n]o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."[87] So a plaintiff could sue a false news site directly for libel but would be barred from bringing a claim against the social media platforms that spread the misinformation. This is a departure from the common law, in which publishers and distributors of defamatory material may be held liable in certain circumstances.[88] Notably, Sec. 230 does not even require websites to take down defamatory material after being notified about it. Supporters of Sec. 230 argue that it has been crucial in enabling free speech and innovation online, and that without it, social media platforms would be forced to engage in massive self-censorship or sued out of existence.[89] Critics, however, argue that while Sec. 230 might have helped the Internet grow in its early days, it is now unnecessarily coddling tech giants who facilitate harmful behavior.[90]

Congress passed the CDA in 1996 by attaching it to the Telecommunications Act, a broad overhaul of FCC regulations. The core purpose of the CDA was to regulate obscenity on the Internet, but the Supreme Court struck down those provisions.[91] Sec. 230, which was authored as an amendment by Reps. Chris Cox (R-Calif.) and Ron Wyden (D-Ore.), remains. Courts quickly interpreted the provision as a broad grant of immunity for Web companies. One of the first cases to examine Sec. 230 was *Zeran v. America Online, Inc.*, in which the Fourth Circuit Court of Appeals held that AOL was not liable for a false advertisement placed on one of its bulletin boards.[92]

The key question in a case involving Sec. 230 is whether the online platform was responsible for the "creation or development" of the objectionable material.[93] A platform that participates in the creation of material transforms into an "information content provider" for purposes of Sec. 230 and loses access to the shield. But this standard is a high bar for plaintiffs to meet. In a particularly broad ruling in 2014, the Sixth Circuit held that a website was immune from liability despite encouraging users to post negative information. The court reasoned that site operators are not liable if they "did not require users to violate the law as a condition of posting, did not compensate for the posting of actionable speech, [or] did not post actionable content themselves."[94] This ruling illustrates

---

[87] 47 U.S.C. § 230.

[88] Amanda Sterling, *Defamation 2.0: Updating the Communications Decency Act of 1996 for the Digital Age*, N.Y.U. J. OF LEGIS. AND PUB. POL'Y (May 25, 2016), http://www.nyujlpp.org/wp-content/uploads/2013/03/Sterling2016nyujllpquorum1-1.pdf.

[89] *Section 230 of the Communications Decency Act*, ELECTRONIC FRONTIER FOUND., https://www.eff.org/issues/cda230 (last visited July 3, 2017).

[90] *See, e.g.*, Andrew Bolson, *The Internet Has Grown Up, Why Hasn't the Law? Reexamining Section 230 of the Communications Decency Act*, INTERNATIONAL ASSOCIATION OF PRIVACY PROFESSIONALS (Aug. 27, 2013), https://iapp.org/news/a/the-internet-has-grown-up-why-hasnt-the-law-reexamining-section-230-of-the/.

[91] Reno v. American Civil Liberties Union, 521 U.S. 844 (1997)

[92] Zeran v. America Online, Inc., 129 F.3d 327 (4th Cir. 1997).

[93] *See, e.g.*, Blumenthal v. Drudge, 992 F. Supp. 44, 50 (D.D.C. 1998).

[94] Jones v. Dirty World Entm't Recordings LLC, 755 F.3d 398, 414 (6th Cir. 2014).

general protections for the platforms from liability when users' share false news and misinformation.

Last year, families of U.S. government contractors killed by ISIS sued Twitter for providing "material support" to terrorists. The families claimed that they were suing Twitter, not as a publisher of any online content, but for providing communications capability to terrorists by allowing them to set up accounts. A federal judge dismissed the case, holding that "no amount of careful pleading can change the fact that, in substance, plaintiffs aim to hold Twitter liable as a publisher or speaker of ISIS's hateful rhetoric, and that such liability is barred by the CDA."[95] The case is now on appeal to the Ninth Circuit. The families of victims of the shootings in San Bernardino, California, and at the Pulse nightclub in Orlando, Florida, have filed similar claims against social media companies.[96]

Some recent cases have chipped away at Sec. 230's broad grant of immunity. In June 2016, a California state court ordered Yelp to take down a critical review of a lawyer. The court held that Sec. 230 did not shield Yelp because the suit was against the reviewer and the order to remove the content did "not impose any liability on Yelp."[97] The case is currently on appeal to the California Supreme Court. Last September, the Ninth Circuit held that the CDA did not bar an aspiring model from suing a modeling industry networking website. The plaintiff, who was raped by two people she met on the site, alleged that the website knew about the risk of rapists and failed to warn her.[98] Also last year, a federal judge held Twitter liable for sending automated text messages to a woman who did not want to receive them (she had recently switched to a phone number that had belonged to a Twitter user who had opted-in to the alerts).[99] Despite these developments around the edges of Sec. 230, social media platforms still have broad protections from liability.

We recommend more research into possible changes to Sec. 230 that would help combat misinformation while not suppressing free speech or the growth of online services. One possible reform would be to require online platforms to take down defamatory content in a timely manner. This process essentially already exists in the United States for copyrighted content (which is exempt from CDA Sec. 230). Under the Digital Millennium Copyright Act (DMCA), online platforms are immune from liability — but only if they respond "expeditiously" to remove infringing material after given notice of its existence.[100] It is a strange feature of U.S. law that websites have an obligation to delete copyrighted videos but can leave up false statements that smear a person's reputation. The United Kingdom's Defamation Act of 2013 creates a system for handling defamatory content that

[95] Fields v. Twitter, Inc., 217 F. Supp. 3d 1116, 1118 (N.D. Cal. 2016).
[96] Matt Hamilton, *Families of San Bernardino attack victims accuse Facebook, Google and Twitter of aiding terrorism in lawsuit*, L.A. TIMES (May 3, 2017), http://www.latimes.com/local/lanow/la-me-ln-san-bernardino-tech-lawsuit-20170503-story.html.
[97] Hassell v. Bird, 247 Cal. App. 4th 1336, 1362 Cal. Rptr. 3d 203 (2016).
[98] Doe v. Internet Brands, Inc., 824 F.3d 846 (9th Cir. 2016).
[99] Nunes v. Twitter, Inc., 194 F.Supp.3d 959 (N.D. Cal. 2016).
[100] 7 U.S. Code § 512.

is similar to the U.S. DMCA "notice-and-takedown" scheme.[101] A proposed law in Germany would give tech companies just 24 hours to delete "obviously criminal content" after notification before facing hefty fines. Germany's Cabinet approved the law but its passage through the country's Parliament is still in doubt.[102]

This approach is not without its drawbacks, however. There is the risk that, to avoid exposing themselves to liability, online platforms could become overly aggressive in deleting content as soon as they receive a complaint.[103] The system could give any user a "heckler's veto" to silence the speech of anyone else, undermining free expression online. Nevertheless, we believe that further research into appropriately tailored reforms to CDA Sec. 230 would be worthwhile. The law made sense in the early days of the Internet, when Congress was especially cautious about not hindering a nascent technology. It is worth exploring further whether the law still strikes the right balance today.

Next Steps: Examine CDA Sec. 230 further in the context of a maturing Internet marketplace.

[101] Sterling, *supra* note 82, at 25.
[102] Joseph Nasr, *Racist post fines on social media firms illegal: German parliament body*, REUTERS (June 14, 2017), http://www.reuters.com/article/us-germany-socialmedia-hatecrime-idUSKBN195227.
[103] *See* Daphne Keller, *Making Google the Censor*, N.Y. TIMES (June 12, 2017), https://www.nytimes.com/2017/06/12/opinion/making-google-the-censor.html.

Section 2. Facebook

I. Introduction and Overview

In the wake of the 2016 U.S. elections, commenters rightly pointed out the disruptive, adverse impact that fake news and misinformation may have had on our democratic discourse. The inability of many Americans to distinguish fact from fiction potentially altered electoral outcomes and subsequently influenced the development of public policy. In the middle of the controversy stands Facebook, which some critics have accused of exacerbating the problem through various features on its popular platform. Given Facebook's prominence as a source of news for the average citizen, this much is clear: Any serious attempt to counteract the creation or effects of fake news and misinformation will necessarily address the company's policies and actions and the sharing habits of its users.

This section examines Facebook users' habits, the effect of the platform's algorithms, its financial incentives, and how other factors contribute to the proliferation of fake news and what Facebook has labeled "false news." The analysis addresses the following questions: How does Facebook as a platform contribute to the dissemination and proliferation of fake news and misinformation? What are Facebook users' habits regarding news consumption, past exposure to fake news, and attitudes toward fake news? How do users react to Facebook's new tools to counter this problem? What other remedies are available to mitigate this problem?

To answer these questions, this section analyzes current relevant case law and it presents a tripartite methodology to investigate the effectiveness of Facebook's recent interventions. Through interviews with officials at Facebook, legal scholars, and policy advocates, this analysis examines fake news as well as what Facebook calls "false news." Company representatives describe false news as narrow news stories premised on false or inaccurate information which, unlike the broader category of "fake news," do not include biased or slanted information. A small-scale user survey, hosted on Amazon Mechanical Turk, investigates Facebook users' interactions with fake news and their exposure to Facebook's recently developed tool for flagging and tagging false news stories on the platform. Finally, a case study of how Facebook users engaged with "Pizzagate" articles posted by three mainstream news sources (NPR, NY Times, and Washington Post) illustrates the genesis and subsequent development of that fake news event. The case study involved both quantitative (e.g., number of likes, comments, shares) and qualitative measures of engagement. The methodology is limited by certain caveats, however, such as a lack of quantifiable data from Facebook regarding user activity and limited granularity with respect to the sentiment analysis in the case study.

Under current law, holding Facebook accountable for the fake news shared on its

platform would be exceedingly difficult, if not entirely impractical. Our analysis focuses on private parties' attempts to hold Facebook liable under defamation and libel laws, because those are the traditional means to address harms born of demonstrably false statements. Section 230 of the Communications Decency Act contains language that provides Facebook and similarly situated content hosts with broad immunity against claims of third-party liability. Courts in multiple jurisdictions have repeatedly held that Facebook's decisions to regulate (or in the context of fake/false news, not regulate) content posted by its users are protected under law. Absent legislative or regulatory action—an unlikely scenario in the current climate—holding Facebook legally liable for the proliferation of fake news on its site is not a viable solution.

Platform Analysis

- Although Facebook has rules to protect against hate speech, it prefers not to be the arbiter of content appearing on its platform; the company's management would prefer that users and third parties evaluate content quality.[104]
- Facebook users' initial exposure to news, whether fake or not, is associated with believing such information; as a result, early exposure to false information is positively associated with believing fake news, even when confronted with fact-checking and later corrections. This suggests early interventions to prevent exposure to fake news will be most effective.
- Our case study of user engagement with Pizzagate stories on Facebook suggests that (1) user sentiments about a fake news story were consistently negative over the lifespan of the story, and (2) that user engagement with analytical stories actually was higher than engagement with initial reports of facts.
- The results from our survey suggest that individuals' education levels are positively associated with user vigilance in assessing source credibility and identifying false information.
- Users express confidence in their own individual understanding of Facebook's user tools, but our survey findings indicate that a majority of users did not know Facebook has a tool available to report false news, and an even larger group of users have not used the tool.
- Facebook users' knowledge of the new tools is positively associated with believing that Facebook is resolving the fake news problem.
- For certain groups, the "trust bias" seems to contravene the intention of the flagging tool. Higher trust in the platform causes some groups to be less likely to check the source of information they share.

[104] Over the past decade, the company has developed hundreds of rules to distinguish between hate speech and legitimate political expression. Facebook has recently expanded monitoring of violent hate speech by expanding its team of monitors from 4500 to 7500. For a fuller description of Facebook's hate speech controls, see Julia Angwin and Hannes Grassegger, "Facebook's Secret Censorship Rules," ProPublica, 6-28-17, https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms.

A limited media analysis presents highlights from a roundtable focus group with eight Knight Journalism Fellows at Stanford University. Their comments revealed that some journalists believe that Facebook's business model forces media sources to design content and headlines to elicit readers' emotional responses. They contended that Facebook's ad revenue business model emphasizes clickbait over quality content and that profit off the spread of fake news will outweigh incentives to effectively deal with the problem. While there is little reason to contest the former point, the latter criticism does not do justice to Facebook's recent actions. Finally, certain Knight Fellows argued that Facebook's tool for flagging false news is ineffective because the viral nature of such news cannot be countered in a meaningful way by a fact-checking process that takes anywhere between 12-48 hours. On face, this criticism has merit.

Near-term Recommendations:

- More aggressively monitor and fact-check news across the platform, flagging it as needed.
- Facebook should market its false news flagging tool more aggressively to users, but enhance cautionary guidelines to counter the trust bias.
- Monitor and make public users' behavior in checking sources.
- Increase scholars' access to site data for research purposes. To preserve users' privacy, the data should be limited to news-related content and engagement. Such information will benefit users by enabling third-party researchers to independently assess the reliability of Facebook as a source of news and information.

Suggested Next Steps for Research:

- Better understand Facebook's business model, including targeted ads, to explore platform-level changes to reduce the spread of fake news on the site.
- Continue to monitor user knowledge of and interaction with Facebook's tools to combat fake news via periodic surveys.
- Align communications across the company about its role in proliferating "false news." Does the Board's approach and assessment align with internal data?
- Engage the company's fact-checking partners in an empirical study of the success of the flagging and fact-checking program. Some studies, including this one, show mixed success and sometimes detrimental effects.
- Explore legislative and regulatory options at the federal and state levels to require social media platforms to disclose data for research purposes. In particular, develop a study of California – where Facebook is headquartered – regulatory options. Such a study would address political feasibility of enhancing or employing such options.

Because the company does not make its internal data public, this study does not determine definitively whether Facebook has effectively addressed the problem of fake news and misinformation on its platform, even if the issue is constrained to Facebook's narrower "false news" conception. Preliminary research, however, reveals grounds for skepticism. While Facebook's proactive steps to resolve the fake news/misinformation problem are encouraging, our study calls into question the effectiveness of these efforts. Additional research and data sharing from Facebook will be necessary to ascertain how helpful these tools are, and whether new platform-based efforts or regulatory action are warranted.

II. Problem Statement

The months leading up to the 2016 U.S. elections brought the issue of fake news and misinformation to the forefront of the national discourse. As millions of Americans were exposed to factual inaccuracies, conspiracy theories, and propaganda masquerading as honest news, commenters rightly began to worry about the adverse short- and long-term effects such information might have on the health of our democracy. If the average citizen couldn't tell the difference between fact and fiction, the quality of public policy would invariably suffer. At the center of the fake news and misinformation controversy stands Facebook, the social media platform vocally criticized for exacerbating the problem through its Trending News algorithm, news feed feature, and ad revenue model. Given Facebook's ascendance as a prominent source for news, any serious attempt to address the creation and spread of fake news and misinformation will require addressing the issue as it exists on the platform. To meet public demand for transparency, such study should not be limited to company findings. Rather, the company should make the data available to scholars' empirical analysis. Transparent scholarly data-sharing, anonymized to protect users' privacy, is fundamental to public trust in the platform.

III. Mission Statement / Goals

The goal of this research is to better understand the nature of the fake news/misinformation problem as it applies to Facebook and its users. This goal could be reached from three perspectives on the issue: (1) users' demographics and habits, (2) the platform's algorithms and financial incentives, and (3) the company's extant solutions. Following this framework, we pursued the following research questions: How does Facebook as a platform contribute to the dissemination and proliferation of fake news and misinformation? What are Facebook users' habits regarding news consumption, past exposure to fake news, and attitudes toward fake news? How do users react to Facebook's new tools? What remedies are available to mitigate this problem? We employed both

qualitative and quantitative research methods to investigate these issues, including interviews, a case study and an online survey.

At an initial research stage, this section takes the first step in examining the fake news problem on Facebook, and builds common ground for further public discussion. At the outset, the memo distinguishes between the broader definition of fake news and Facebook's preferred term for the problem: "false news." Fake news, drawing from both media outlets and research articles, refers to a wide spectrum of misinformation ranging from disinformation and false information to biased/slanted news. Facebook, as a platform and company, has defined "false news" as news articles that are intended to *mislead* readers, involving hoaxes, false information, and misinformation. While Facebook's scoping of the problem may be too constrained, the company's scoping of what it terms "false news" offers a sensible place to begin. This report uses the term "fake news" to denote a more expansive understanding of "false news" as stipulated in the definition outlined on pages 10-11.

IV. Context for the Problem

Globally, as of March 2017, Facebook averages 1.28 billion daily active users,[105] and is the most popular social media platform in the United States.[106] According to recent studies, the majority of Americans use Facebook, and a majority of those users read news sources disseminated through their social network on the platform.[107] Pew Research Center found that 79 percent of Americans use Facebook, and two-thirds of these adults get news from the platform, making it today's news stand.[108] Facebook is popular with people of all age groups, and it has recently become more popular with older adults; 88 percent of 18- to 29-year-old Americans use Facebook, compared to 84 percent of 30- to 49-year-olds, 72 percent of 50- to 64-year-olds, and 62 percent of those ages 65 and older.[109] Women are slightly more likely to use Facebook than men (83 percent of women compared to 75 percent of men).[110] Based on our personal observations, it appears that youth may receive their news from several online platforms while older users typically seem to rely on Facebook along with traditional print and media outlets.

Facebook has become a common part of many Americans' daily routines. 76 percent of American users check the site daily; increasingly, it appears to be a major source

[105] *Company Info (Stats)*, FACEBOOK (Mar. 31, 2017), https://newsroom.fb.com/company-info.
[106] Shannon Greenwood, Andrew Perrin & Maeve Duggan, *Social Media Update 2016*, PEW RESEARCH CENTER (Nov. 11, 2016), http://www.pewinternet.org/2016/11/11/social-media-update-2016.
[107] *Ibid.*
[108] *Ibid.*
[109] *Ibid.*
[110] *Ibid.*

of news for its users.[111] Pew Research Center found that 62 percent of U.S. adults get news on social media and 18 percent do so often.[112] In terms of Facebook's specific reach, 66 percent of Facebook users get news on the site; given Facebook's large user base, this means that 44 percent of the general U.S. population gets news from Facebook.[113] This percentage has increased in recent years, with the percentage of Facebook users who get news from the site increasing from 47 to 66 percent between 2013 and 2016.[114] Facebook users are also more likely to happen upon news while doing other activities online, as compared to Reddit and Twitter users, who are more likely to seek out news on their respective sites.[115]

This study shows that Facebook users are most likely to interact with URLs and news stories that their friends share with them or where they see that their friends have liked or interacted with an article, or because the user follows the particular news source's Facebook page. Facebook has low anonymity, especially compared to Twitter or Reddit, as users' profiles are linked to their personal identity and the items they like, share, or those on which they comment can be traced back to their profile via their friends' news feeds. We assume, and Facebook officials agreed during our interviews with them, that Facebook users (on average) want accurate information to appear on the site, as they do not want to be sharing or interacting with fake content that reflects poorly on them individually. One of Facebook's primary levers is its users' general investment in their personal reputations.

Facebook has come under increasing scrutiny for its role in the spread of fake news and misinformation after the 2016 American election and during recent European elections.[116] Fake news presented itself in many forms on Facebook: click bait, intentionally deceptive stories, foreign propaganda, and slanted or biased news. Many have suggested that the spread of fake news on Facebook, where users often did not read past the headlines or confirm the information, contributed to outcomes in national elections across the world.[117] In the U.S. 2016 election stories about the Pope endorsing Donald

---

111 *Ibid.*
112 Jeffrey Gottfried & Elisa Shearer, *News Use Across Social Media Platforms 2016*, PEW RESEARCH CENTER (May 26, 2016), http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016.
113 *Ibid.*
114 *Ibid.*
115 *Ibid.*
116 Will Oremus, *The Real Problem Behind Fake News*, SLATE (Nov. 15, 2016 11:58 AM), http://www.slate.com/articles/technology/technology/2016/11/the_problem_with_facebook_runs_much_deeper_than_fake_news.html.
117 See, among others: Angie Drobnic Holan, *2016 Lie of the Year: Fake News*, POLITIFACT (Dec. 13, 2016 5:30 PM), http://www.politifact.com/truth-o-meter/article/2016/dec/13/2016-lie-year-fake-news; Mike Isaac, *Facebook, In Cross Hairs After Election, Is Said to Question Its Influence*, N.Y. TIMES (Nov. 12, 2016), https://www.nytimes.com/2016/11/14/technology/facebook-is-said-to-question-its-influence-in-election.html?_r=0; Howard Kurtz, *Fake News and the Election: Why Facebook Is Polluting the Media Environment with Garbage*, FOX NEWS (Nov. 18. 2016), http://www.foxnews.com/politics/2016/11/18/fake-news-and-election-why-facebook-is-polluting-media-environment-with-garbage.html; Max Read, *Donald Trump Won Because of Facebook*, N.Y. MAG. (Nov. 9, 2016 2:37 PM), http://nymag.com/selectall/2016/11/donald-trump-won-because-of-facebook.html; Craig Silverman, *This Analysis Shows How Viral Fake Election Stories Outperformed Real News on Facebook*, BUZZFEED

Trump were shared almost 1 million times on Facebook. Despite claims by Facebook that false news makes up only a small percentage of news on the site, this research study found no corroborating evidence. During the recent French election, Facebook disabled 30,000 accounts that were associated with spreading false news, misinformation or deceptive content. The crackdown indicates a major shift of policy, led by Mark Zuckerberg immediately following the American election.[118]

In light of the public backlash, Facebook has implemented new approaches and tools to combat false news. On February 16, 2017, Mark Zuckerberg addressed the ongoing criticisms regarding filter bubbles and fake news by emphasizing the need for safe, supportive, informed, and civically engaged communities on Facebook, saying:

> Accuracy of information is very important. We know there is misinformation and even outright hoax content on Facebook, and we take this very seriously. We've made progress fighting hoaxes the way we fight spam, but we have more work to do. We are proceeding carefully because there is not always a clear line between hoaxes, satire and opinion. In a free society, it's important that people have the power to share their opinion, even if others think they're wrong. Our approach will focus less on banning misinformation, and more on surfacing additional perspectives and information, including that fact checkers dispute an item's accuracy.
>
> While we have more work to do on information diversity and misinformation, I am even more focused on the impact of sensationalism and polarization, and the idea of building common understanding.[119]

Although Facebook originally pushed back against the suggestion of a fake news problem on the site, it changed its approach following the U.S. election, showing new receptiveness to changing its approach to meet users' concerns. While Facebook makes clear that it does not want to be an arbiter of truth and should not position itself as such, it has made changes to its platform and is beginning to share limited data publicly.[120] For purposes of the tools it has implemented, which we explain below, Facebook is targeting false news rather than the umbrella issue of fake news. As the company's Communications Team explained, the term "fake news" has become overly polarizing and conflated with politics, so they coined the term "false news" to identify what they perceive as a problem

---

NEWS (Nov. 16, 2016 4:15 PM), https://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook?utm_term=.mwzYJNDmG#.yanGDkLE4.

118 Ivana Kottasová, *Facebook targets 30,000 fake accounts in France* (April 21, 2017: 3:23 AM), http://money.cnn.com/2017/04/14/media/facebook-fake-news-france-election/index.html.

119 Mark Zuckerberg, *Building Global Community*, FACEBOOK (Feb. 16, 2017), https://www.facebook.com/notes/mark-zuckerberg/building-global-community/10154544292806634.

120 Interview with Tom Reynolds, Communications Team, Facebook, in Stanford, Cal. (Apr. 25, 2017). In response to the ongoing investigation by the Senate Intelligence Committee into Russian influence on the 2016 election, Facebook produced evidence showing over 3000 Russian-affiliated, targeted ads appeared on its platform in three swing states. See, for example, M. Raju, D. Byers and D. Bash, *Russian-linked Facebook ads targeted Michigan and Wisconsin*, CNN POLITICS, 10/4/17, amp.cnn.com

that can be tackled within the frame of users' First Amendment rights. Facebook seeks to target hoaxes, overtly and intentionally false information and misinformation that fall outside First Amendment protections—what they refer to as the "worst of the worst."[121]

In recent months, Facebook has implemented a cross-functional approach of its own to combat false news.[122] It has implemented tools to deal with the problem in the short-term, including an on-site PSA campaign to help users identify false content when they see it, as well as a reporting device so users can flag issues they recognize as false news. If an article or URL is reported by enough users to cross an internal threshold, the article is sent to a set of independent fact-checking organizations.[123] According to our interviews with Facebook officials, if two of the fact-checking organizations identify the information as false, the URL will be flagged as "debunked" to anyone who sees it, and users will be prompted to consider that independent fact-checkers have identified this information as debunked before they can share it. Finally, the new tool includes a change to the news feed algorithm so that the debunked stories are less likely to show up on users' news feeds.[124] This report describes findings from our online survey of user responses to Facebook's new tools.

As another short-term solution, Facebook has disrupted the financial incentives for users to propagate some types of false news. According to Facebook's communication team, ad farms and clickbait-style content are pervasive, but users do not enjoy seeing them. These ad farms are landing pages with little actual content but full of advertisements, and they are typically built on sensational headlines to draw clicks. Facebook has changed the internal algorithm to limit these pages' ability to show up in news feeds and to cut down on their access to ad revenue.

Facebook has also implemented a long-term solution of working on news literacy through their Facebook Journalism Project (FJP). The FJP goal is to help users understand what qualifies as real news, how to identify credible news sources, and how and when to trust sources they see shared on the site.[125] Facebook is working with other technology companies and journalists to develop a curriculum for news literacy for future distribution.[126]

In analyzing the phenomenon of fake news on Facebook, this report focuses on stories premised on overtly false or inaccurate information, regardless of intention. This analysis of assessing the effectiveness of Facebook's efforts within the framework of "false news" provides a foundation for later consideration of the presence and impact of slanted

---

121 *Ibid.* We explain later how our study's operative definition of fake news differs from Facebook's.
122 Telephone Interview with Facebook Spokespersons (May 23, 2017).
123 Interview with Tom Reynolds, *supra* note 15.
124 *Ibid.*
125 *Ibid.*
[126] Facebook is not the only entity developing news literacy programs. The Newseum is piloting a news literacy curriculum for high schools, which we observed personally at Palo Alto High School in May 2017.

and biased news, which is more difficult to assess with empirical certainty.

Such empirical uncertainty is further compounded by both the platform's policy position on biased news and First Amendment concerns. First, Facebook does not perceive biased news as a problem it should address. Second, attempts by the company to curtail biased news are contrary to its fundamental mission for users, which is to serve as a tool of free expression for individuals with their communities of friends and family.[127] This project focuses primarily on circumstances where biased information in news articles crosses a line to grossly mislead readers about underlying facts.

V. Roadmap

Section 1 of this report offered a brief legal analysis of litigation involving Facebook, and the other major platforms, and Section 230 of the Communications Decency Act, the statute that provides social networks and other content hosts with broad immunity against liability claims (see Section 1: Legal Analysis). This second section presents findings about Facebook as a platform based on (i) stakeholder interviews, (ii) a case study of a well-publicized fake news story, and (iii) a survey of users. Following this internal analysis, this section discusses findings from conversations with journalists about the problems they see with Facebook's business model and new tools, and highlights broad concerns with democratic discourse in light of Facebook's role in the Fourth Estate. The study of Facebook concludes with proposed next steps and future research questions.

VI. Platform Analysis

Facebook is distinct from the other platforms in this project for a number of reasons. Users on Facebook have relatively low anonymity compared to Twitter or Reddit, as posts are associated with users' personal accounts and are distributed to friends and family in their social networks. Readers typically have some relationship to the content they see (with the exception of advertisements), as links on their news feeds will appear because they follow the source or their friends interacted with it in a public way.

Because users select their friends and what to follow, some critics accuse Facebook of creating "filter bubbles" and echo chambers, where users essentially self-select into social networks that echo and reinforce the user's own political or social views. While confirmation bias likely plays a significant role in users' choices about what information to seek out on Facebook and how they perceive what they read, the Facebook algorithm seems to promote content which users are most likely to enjoy—namely content they agree with and will interact with through likes, shares, or clicks. So, while individual users may

[127] This recognizes that the company's financial model is built on ad revenue.

be less psychologically inclined to disregard or disbelieve information that contradicts their beliefs, Facebook seems to benefit from a system—both in terms of revenue and user satisfaction—that promotes content that users want to see on their individual news feeds.

Like other intermediary platforms, Facebook generally prefers not to assess user content appearing on the site; company management would prefer that users themselves and third parties evaluate content quality. Nevertheless, Facebook has implemented a cross-functional initiative to combat false news—a subset of the fake news that might appear on the site. This section presents findings about how users engage with fake news on Facebook and how they have interacted with the new tools.

Finding 1: Exposure to Fake News Is Positively Associated with Believing Fake News

Our study indicates that Facebook users' initial exposure to news, whether fake or real, is associated with believing such information; as a result, early exposure to false information is positively associated with believing fake news, even when confronted with fact-checking and later corrections. This suggests early interventions to prevent exposure to fake news will be most effective in countering its proliferation.[128]

Through a logistic regression, our survey results reveal that the odds[129] of believing fake news are approximately 8.5 times higher for people who read fake news versus those who read real news ($\boldsymbol{\beta = 2.14, \rho < .05}$). The odds of believing real news are approximately 32 times higher for people who read real news versus those who read fake news ($\boldsymbol{\beta = 3.46, \rho < .05}$). That means, exposure to fake (or real) news is positively associated with believing fake (or real news).

These findings suggest that the more fake or real news that an individual consumes, the more likely that reader will believe that particular news is real. This finding supports the familiarity bias of cognitive processing: The more an individual hears particular news, the more familiar it becomes, and the more likely she will believe it is real.[130]

This seemingly intuitive result indicates that reducing the initial exposure of Facebook users to fake news may be more effective than a subsequent correction of misbeliefs and misperceptions. One recent study supports this notion.[131] It found that the correction of misinformation will rarely change people's misbeliefs, and may even backfire. Labeling can be a prominent means of correction. But scholars have argued that

---

[128] Since the completion of our research, a Yale team has published a more extensive study of the exposure effect: See, G. Pennycock, et al., *Prior exposure increases perceived accuracy of fake news*, Aug. 26, 2017, SSRN-id2958246.pdf.
[129] In logistic regression, odds = p(0)/1-p(0), p(0) refers to the probability of the success of some event.
[130] N. Schwarz, L. J. Sanna, I. Skurnik, & C. Yoon (2007). Metacognitive experiences and the intricacies of setting people straight: Implications for debiasing and public information campaigns. *Advances in Experimental Social Psychology*, *39*, 127-161.
[131] N. Mele, D. Lazer, M. Baum, N. Grinberg, L. Friedland, K. Joseph & C. Mattsson, *Combating fake news: An agenda for research and action*, May 2, 2017, **https://shorensteincenter.org/combattingfakenews**.

labeling misinformation as fake may be harmful because nonbelievers may use this label to reinforce their view, while believers may ignore the label and remain steadfast in their misperceptions.[132]

This result suggests that Facebook's debunking tool may not be effective in combating false news, tying into the finding that front-end reduction of exposure to such news might be more effective than back-end corrections. Journalists and legal scholars worry that labeling stories as disputed by independent fact-checkers actually has the opposite effect, increasing user engagement with disputed stories.[133] Given Facebook's alleged history of preventing conservative stories from appearing in the Trending News section of an individual's news feed, conservative groups especially seem reluctant to accept the "disputed" label that could result from use of the tool. In addition, the current version of Facebook's debunking tool does not incorporate real-time tagging; thus, the potentially false news may have already gone viral and the damage may have already been done before Facebook creates a "disputed" label.[134] Further, it appears that some Facebook users will still conflate the actions of third-party fact-checkers with Facebook selectively reviewing content, which our survey suggests detracts from users' perceived credibility of these fact-checking tools on the platform.

Finding 2: Users' Sentiments Are Consistently and Proportionally Negative Over the Lifespan of the Reporting of a Fake News Story, and User Engagement Increases Over Time

Findings from this case study of user engagement with Pizzagate stories on Facebook suggests that (1) user sentiments about a fake news story were consistently negative over the lifespan of the story, and (2) user engagement with analytical stories actually was higher than engagement with initial reports of the facts.

First, this study finds that user sentiments were negative in comments about Facebook posts regarding Pizzagate coverage from different news outlets. We looked at three news outlets, the New York Times, Washington Post, and NPR.[135] We looked at three posts from each outlet reflecting the initial reporting, discussion, and aftermath. For each of these, the sentiments from Facebook users in the comment sections were consistently and proportionally negative across time frames and across news outlets. In an abstract way,

---

132 d. Boyd, Did media literacy backfire? 1/9/2017, http://www.zephoria.org/thoughts/archives/2017/01/09/did-media-literacy-backfire.html.

133 Sam Levin, *Facebook Promised to Tackle Fake News, But the Evidence Shows It's Not Working*, THE GUARDIAN (May 16, 2017 5:00 AM EST), https://www.theguardian.com/technology/2017/may/16/facebook-fake-news-tools-not-working (highlighting a story where traffic skyrocketed after being flagged as disputed).

134 An example of this occurred on the morning of the Las Vegas mass shooting when Google users on the 4chan message board falsely identified the shooter, which was then picked up by Gateway Pundit and broadly amplified on Facebook before monitors could remove the false material. See, for example, Maxwell Tani, "Fake news about the Las Vegas shooting spread wildly on Facebook, Google, and Twitter," *Business Insider*, Oct. 2, 2017, businessinsider.com.

135 *New York Times*, *Washington Post* and NPR are convenient samples of this case study. These three media outlets posted a number of news articles regarding "Pizzagate" on their Facebook pages. Conservative media such as Breitbart and Fox News did not post anything relevant to Pizzagate story, or if they had, it was no longer online at the time of this research (we could not find relevant news stories on their Facebook pages).

it appears that the Facebook users who engaged with these stories were talking about things in the same relative (negative) terms. See Figure 1 for the sentiment analysis charts.

Second, the study finds that user engagement levels did not track our initial hypothesis that users will engage more with short facts and less with long-form analytical articles. We hypothesized this based on conversations with journalists, who suggested that Facebook's model of catering to users' immediate emotional responses to posts tends to favor short and sensational pieces rather than these longer analytical articles.

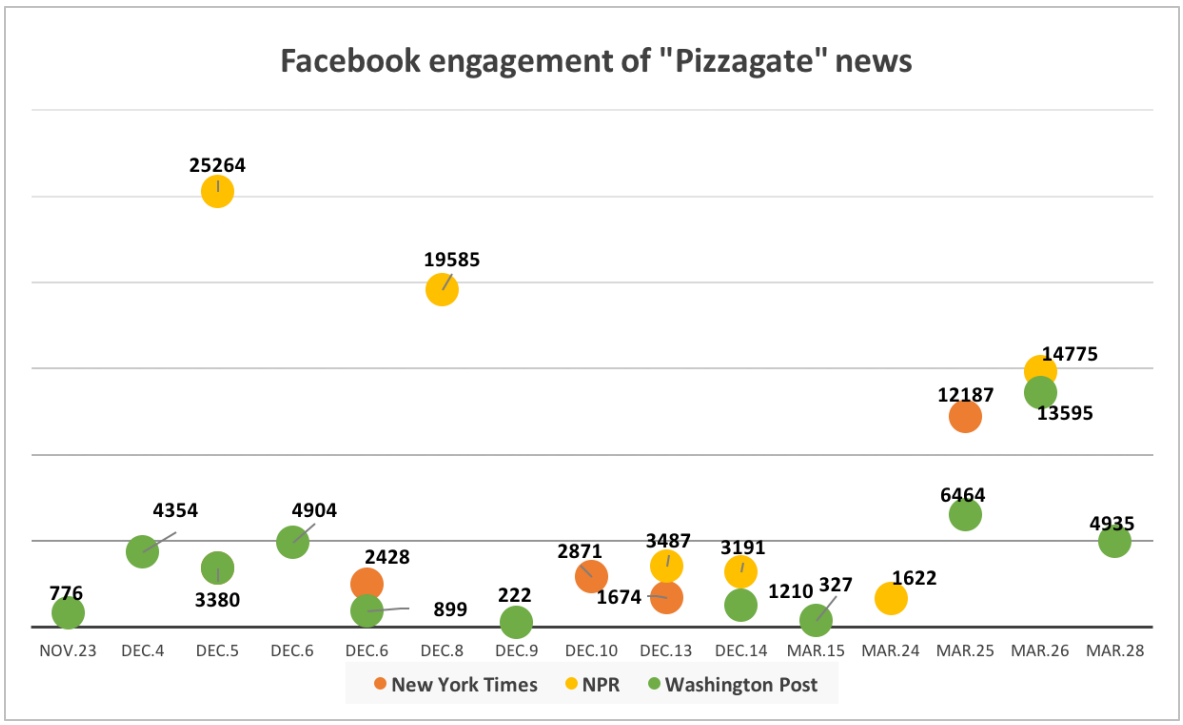
What we found, however, was that for the *New York Times* and *Washington Post*, reader engagement numbers actually increased slightly over time as the outlets posted more analytical articles engaging with how and why the Pizzagate conspiracy developed, leading up to the shooting at Comet Ping Pong. See Figure 2 for Facebook Engagement graphs and Appendix D for brief summaries of the articles we used in the case study. This finding suggests that *Post* and *Times* readers on Facebook may be more interested in these kinds of longer articles. More research, however, on other case studies and other news outlets would be helpful to try to recreate our findings.[136]

We also had an extreme (unexplained) outlier with NPR, which had almost seven times as much engagement for its initial short post about the Pizzagate shooting. We hypothesize that this engagement was so high because NPR may have been one of the first outlets to cover the shooting. NPR has fewer Facebook followers than the *New York Times* or *Washington Post*, so follower numbers cannot account for the difference.

[136] Due to a lack of information on the Breitbart and other conservative media sites, our analysis did not include a study of Facebook readers' engagement with conservative news outlets' coverage of the Pizzagate story. We believe that these sites had removed the stories by the time our study took place.

Figure 1. Sentiment distribution of comments on news report of "Pizzagate" across media outlets.

Figure 2. Facebook engagement of news coverage of "Pizzagate" across media outlets.



Finding 3: Users' Education Level and Habits are Positively Associated with Vigilance Against Misinformation While Reading News

Survey results suggest that education, and the desire to seek more information, are positively associated with user vigilance in assessing source credibility and identifying fake news (see Appendix A methodology for detailed measures). The survey measured individuals' vigilance levels when reading news on Facebook by asking the extent to which they assessed the source's credibility, the writer's intentions, the false aspects of the content, and the general truthfulness of the news story. We also tracked clicking for more information as one predictor of vigilance, hypothesizing that Facebook users who seek additional information while reading news on the site are also more likely to exercise skepticism of fake news. Results suggest that users' education levels and habits of clicking for more information are positively associated with their vigilance level in assessing the quality of the news they read. "Vigilance" in this context refers to a cognitive mechanism for epistemic vigilance against the risk of being misinformed.[137] Facebook users with higher education (***i.e.***, high school degree, some college, bachelor's degree, and graduate degree) are more likely to exercise vigilance against misinformed content compared to participants without a high school degree. Moreover, education is correlated to vigilance in checking news sources. In other words, the higher the education level, the more likely users were to check for background information before sharing a story.

These findings shed light on the effectiveness of the solutions that Facebook has developed, including the Facebook news literacy curriculum. The survey shows a positive association between education level and vigilance against misinformation—a finding that supports Facebook's decision to facilitate media literacy. Notably, implications for Facebook lie in the need to improve media literacy in elementary and middle school education, and possibly all the way through high school. It may be useful to further examine users' habits of news assumptions across different education levels. In addition, although our survey does not assess age as a factor, it does show that media literacy education may be important for adults who have not completed high school. Results suggest that media literacy should be part of general educational interventions geared toward the public online. Moreover, seeking additional information rather than merely reading the news headlines is an effective way to evaluate news trustworthiness. Media literacy is a full package of abilities and skills; thus, educational programs should consider specific solutions (e.g., additional information seeking) to enhance the ability to distinguish fake news.

[137] D. Sperber, F. Clément, C. Heintz, O. Mascaro, H. Mercier, G. Origgi & D. Wilson. (2010). Epistemic vigilance. *Mind & Language*, *25*(4), 359-393.

Finding 4: 70 Percent of Users Were Unaware of Facebook's New Tool for Reporting False News, and 90 Percent Had Not Used the Tool

Although users declare high confidence in their personal knowledge of Facebook's tools, our survey findings indicate that a majority of users did not know Facebook has a tool available to report false news, and an even larger group of users have not used the tool.

According to the survey results, users believe that they have fair amount of knowledge of Facebook's new tool for combating false news (M^{138} = 3.95, *SD* = .98). However, when asked about their knowledge of the specific tool for reporting such news, 70.4% (n = 447) users were unaware of the tool while 18.0% (n = 114) were aware of the tool before participating in the survey. About 90.0% (n = 569) users have not used the tool before taking the survey while only 7.2% (n = 46) users had ever used the tool on Facebook. When users were asked about their knowledge of other recent tools that Facebook has launched to help spot and defend against false news, we found results more consistent with data documenting users' awareness rather than perceived beliefs about the new tool (*M* = 1.81, *SD* = 0.9). Taken together, these findings indicate that Facebook users do not know much about Facebook's new tool for combating false news. Thus, we advise Facebook to put more efforts in marketing – and monitoring the effects of – its new tool.

Finding 5: Knowledge of Facebook's "False News" Solutions Are Positively Associated with Belief that Fake News and Misinformation Will Be Less Prevalent on Facebook

Our findings reveal that Facebook users' knowledge of the new tools is positively associated with believing that Facebook is resolving the fake news problem. According to the survey results, the perceived knowledge of Facebook's tools for curbing "false news" ($\boldsymbol{\beta = .11, \rho < .05}$), and the perceived severity of fake news ($\boldsymbol{\beta = .23, \rho < .05}$) are positively associated with the confidence that the collaboration with fact-checkers will reduce fake news. This means that the more a Facebook user believes that they know of new tools, and the more the user perceives the problem of fake news to be severe, the more likely that s/he will believe that fact-checking sites will resolve the issue in the future. This finding shows that public awareness of the new tools at Facebook can increase public confidence that fake news will become less of a problem. Although Facebook spokespeople were not forthcoming in how the new tool is performing thus far, our survey results indicate that users' knowledge of the reporting tool for false news, and their perceived severity of the problem of fake news, can predict users' confidence overall in the effectiveness of Facebook's interventions.

Significantly, the study shows that user confidence is tied to the knowledge that such tools exist, not to knowledge that the tools are functioning effectively or changing the amount or type of fake news on Facebook. Arguably, Facebook should not be able simply to advertise tools as a means of resolving public backlash. Instead, Facebook should

[138] M refers to Mean; SD refers to Standardized deviation.

implement functional tools that actually reduce fake news on the site.

Subsequent studies have shown "an illusory truth effect" tracking the effects of familiarity and exposure. Pennycock, et al., found that Facebook's warning "Disputed by 3rd Party Fact-Checkers" "did not disrupt the familiarity effect." Their study revealed that, while users were overall more skeptical of all news shared on the site, they were not more skeptical of the specifically tagged as disputed stories. [139] where some users are *more* likely to share false news when they are aware of the flagging tool. Our findings suggest that such sharing is related to user confidence in the platform. For some users, the mere existence of the flagging tool on the platform is enough, in their minds, to ensure the veracity of the news being shared, whether or not it has been fact-checked. Thus, ironically, the presence of the flagging tool, for some groups, may exacerbate the spread of misinformation.

VII. Media Analysis

Our research discovered consistent concerns from journalists about the intersection of their profession and the fake news problem on Facebook, as well as general concerns about how new financial struggles have correlated with a decline in journalism. For this section, we obtained our findings through interviews and discussions with the Knight Fellows, journalists in residence at Stanford University. In addition to a roundtable discussion with the Knight Fellows, this report reflects an in-depth conversation with Frederic Filloux, a reporter and editor from the European Union, who shared his conclusions from his year-long Knight research project focusing on how Facebook's financial model impacts traditional journalism. While we think that some of these concerns are overstated—or that Facebook has shown efforts to address these—they are helpful lenses through which to view the ongoing fake news problem on Facebook.

Issue 1: Financial Incentives for Journalists and Facebook

In our conversation, the Knight Fellows highlighted the perceived problem that Facebook's large share of the market for advertising revenue limits the ability of journalists to compete effectively. The first result, according to the Knight Fellows, is that news sources, especially local news sources, cannot be financially viable online because they cannot compete effectively with Facebook and Google for advertising. They see the platforms as co-opting the kinds of advertising sources that once powered print journalism without feeding any of the revenue back to the actual content providers. They consider this especially problematic for local news, and they tied this issue to the decline of local

[139] See G. Pennycock, T. Cannon, D. Rand, Prior exposure increases perceived accuracy of fake news, August 26, 2017, SSRN-id2958246.pdf. The report argues that "[A]ny benefit arising from the disputed tag is immediately wiped out by the exposure effect. That result, coupled with the persistence of the familiarity effect we observed, suggests that larger solutions are needed that prevent people from ever seeing fake news in the first place, rather than qualifiers aimed at making people discount the fake news that they see" (22).

journalism, citing the increasing movement of journalists to national news outlets in Washington D.C. or New York City. While these causal claims cannot be supported without empirical analysis, the Knight Fellows claim a correlation between the rise of Facebook and the decline of journalism, especially at the local level.

The next perceived problem is that in order for journalism to succeed, journalists must engage with Facebook's model. According to the Knight Fellows, the content that generates the most revenue from placement on Facebook is that which is geared toward heightened emotional responses from Facebook users and user reaction, not in-depth analysis or reasoned discussion. As a result, they say journalists are now being trained (and partnering with strategic companies like RebelMouse) to optimize their article titles and content to match user preferences on Facebook. They try to model their headlines on the success of Buzzfeed's sometimes audaciously worded headlines and they produce short form articles intended to generate clicks rather than longer user engagement. Although sensationalized news content is not a new problem in journalism, the Knight Fellows, particularly Frederic Filloux, perceive today's sensationalized, sometimes misleading, headlines and news content as a primarily a Facebook effect.

The final issue within this topic is how Facebook's own profit model intersects with the fake news problem. The Knight Fellows believe that Facebook gets more money when URLs get more clicks, and articles intended to be inflammatory, sensational, or otherwise emotion-inducing tend to generate the most revenue. They perceive this financial model as opposed to the goal of promoting serious, truthful journalism and information sharing on the platform. We are skeptical of this criticism, however. It appears as though Facebook has taken steps that are in tension with revenue generation—or perhaps the company does not rely on revenue from clicks on these types of URLs in the first place. For example, Facebook has stopped allowing obvious clickbait and ad farms to appear on the Trending Stories section of news feeds, which would appear to cut down on potential revenue. In our conversations with Facebook spokespeople, they indicated that news stories in general, and especially false news (as they define it), is a very small proportion of their total daily traffic, so it was not a serious financial loss to do so. That said, we currently have no means of corroborating such a claim.

Issue 2: Journalists Perceive Facebook Efforts as Inadequate

One Knight Fellow claims that Facebook's crowd-sourced fact-checking solution is a "drop in the bucket" compared to the vast scope of the problem. This cuts down on Facebook's ability to credibly tackle the fake news problem. He claimed that Facebook handles approximately 100 million URLs each day, but that the fact-checking process takes between 12-48 hours in order to debunk fake news. Between concerns that users are not adequately informed of the available tools and the additional concern about the slow nature of fact-checking, these journalists are concerned that Facebook's efforts are too small and

too slow to adequately combat fake news. In response to these concerns, additional information is necessary, including:

- What percentage of Facebook's URL traffic consists of news articles?
- How essential to Facebook's business model is sharing news articles? For example, could the company stop allowing sharing of news articles entirely? Because Facebook's algorithm prioritizes personal posts from users' friends and families over shared news articles, these journalists assume that news is a minor part of Facebook's daily traffic.
- How many users are familiar with these tools and actually use them?
- Would more prominent buttons or a more prominent interface to flag false news be more effective?
- Would additional manpower at the fact-checking organizations increase responsiveness?

When we spoke to Facebook spokespeople about the slow response time, they acknowledged that the speed of responsiveness is a key issue to deal with the virality of false news on social media. They returned to the theme of not wanting Facebook to become the arbiter of truth or content, so they stressed that user reporting and third-party assessment is their preferred system, although they did acknowledge the need to become faster in vetting flagged news articles. Facebook officials also discussed that the new changes to the algorithm ranking for news stories deemphasizes stories that receive clicks but little engagement—such as shares and likes—but that this tool is still early on in its development and will improve to address this problem as well.

Issue 3: Journalists Perceive Lack of Regulation of Platforms or User Data as Problematic

An additional area of concern for these journalists was the lack of regulation regarding user data collected by Facebook that can be sold for profit. For example, Cambridge Analytica, a strategic communications firm, uses Facebook as a source for many of the 3,000-5,000 data points it possesses on nearly 230 million adult Americans.[140] While this kind of data is often used for advertising, it has also been employed since the 2008 presidential election to target political advertising and news stories to potential voters. This provides political entities the opportunity to target potential voters with fake news stories, potentially escalating the threat. Unlike individual health data, which is closely regulated under HIPAA, individual social media data is owned by the platforms and available for sale without user consent. Journalists would like to see more policy discussions about how this data can be better regulated and protected, or how the profits can be split between the platforms and content companies who traditionally rely on

[140] McKenzie Funk, *The Secret Agenda of a Facebook Quiz*, N.Y. TIMES (Nov. 19, 2016), https://www.nytimes.com/2016/11/20/opinion/the-secret-agenda-of-a-facebook-quiz.html.

advertising as their business model.

The Knight Fellows also mentioned antitrust law as a particular area of interest for reducing the influence and total market share for Facebook. They view Facebook's total percentage of the news URL sharing that occurs online as a fundamental challenge to local and small-scale journalism. While we understand the frustration that comes from Facebook's dominance in gaining advertising revenue, we worry that antitrust laws may not be a realistic or helpful solution. Facebook's model as a social network relies on a monopoly of user participation; if most people were not on the same network, Facebook would have less network functionality and appeal (consider, as a comparison, the lack of participation in the much-hyped competitor, Google Plus). Antitrust challenges to social media companies like Facebook would likely destroy the underlying social function of the site, which by definition requires a monopoly of usership to make a social network useful. Antitrust solutions would likely drive people to other new websites, which would likely need to be smaller, and might ultimately become more "siloed" by political or other preference.

VIII. Facebook and Challenges to Democracy

The aftermath of the 2016 presidential election brought increased focus on the role Facebook plays in our democracy. At the time, observers criticized what they perceived to be deficiencies in Facebook's Trending News feature, namely that the algorithm did not differentiate between real and fake news.[141] As a result, they claimed, Facebook had exposed countless unwitting users to false stories that may have influenced the outcome of election. For its part, Facebook has denied that the fake news phenomenon on its platform made any significant difference in electoral outcomes, but at the same time, the company has acknowledged the role its social network plays in shaping our democracy.

In addressing the Facebook community, Mark Zuckerberg has emphasized the company's efforts to create a more informed and more civically engaged citizenry.[142] On the information axis, he noted the greater diversity of viewpoints available on social media v. traditional media. And speaking in relatively abstract terms, he recognized Facebook's obligation to help provide its users with a more complete range of perspectives such that they could "see where their views are on a spectrum and come to a conclusion on what they think is right."[143] Facebook, he posited, stands in a unique position: Because the platform facilitates individuals connecting based on mutual interests beyond politics (e.g., sports,

[141] Nathan Heller, *The Failure of Facebook Democracy*, NEW YORKER (Nov. 18, 2016), http://www.newyorker.com/culture/cultural-comment/the-failure-of-facebook-democracy.
[142] Mark Zuckerburg, *Building Global Community*, FACEBOOK (Feb 16, 2017), https://www.facebook.com/notes/mark-zuckerberg/building-global-community/10154544292806634.
[143] *Id.*

music, culture), it can help create a shared foundation that could make it easier for those same individuals to have a dialogue on issues of political disagreement.

Regarding civic engagement, Zuckerberg emphasized Facebook's role in encouraging voter participation in elections, facilitating dialogue between elected representatives and their constituents, and enabling issue advocates to rally support around their causes. During the 2016 election cycle, the company launched a series of public service announcements that targeted eligible voters across the country and provided them with the information necessary to register to vote. By Zuckerberg's estimates, the campaign helped more than 2 million citizens across the country register and eventually cast ballots.

Nevertheless, there are steps Facebook can take to improve its commitment to informed citizenship while still allowing for the free flow of information on its platform. First, it should do a better job of marketing its immediate-term solution to the false news problem. Although the company touts its tool for flagging and eventually tagging false news, our survey data supports the notion that few users have heard of the system and fewer still have actually used it. In conversations with Facebook officials, we were told that the company sees low use as an indicator that false news is not as much of a problem on the platform as many observers believe.[144] But that kind of reasoning is difficult to credit when our findings show that a large majority of surveyed users can't use the tool even if they wanted to because they don't know that the tool exists.

Second, Facebook should not lose sight of its long-term efforts to educate users on how they can serve as their own information gatekeepers. Facebook's new PSA, which lists tips on how to identify false news, is a start, but we believe a more sustainable solution requires Facebook's continued investment in educational curricula that teach users from a young age how to approach news with a critical eye. Survey data collected through this project support the intuitive notion that education is positively associated with willingness to verify assertions of fact, leading to the hypothesis that such a relationship exists because of higher education's emphasis on skepticism and critical thinking.

IX. Policy Options and Further Research

Recommendations:

- Facebook should market more aggressively to users its false news flagging tool. It is understandable that the company wants to take an approach based on notifying users of false news rather than censoring the information altogether, but if consumers don't know that they have access to a potentially effective tool to combat misinformation, Facebook's user-

[144] Interview with Facebook Spokespersons, *supra* note 17.

centric strategy for solutions will not work.

- Facebook should increase access to its data for scholarly research purposes. Even if such data is limited to news-related content and engagement, users will benefit from the ability of scholars and objective third-party organizations to independently assess the reliability of Facebook as a source of news and information. At the moment, the company's closed and protective nature of its data leaves no way for scholars or third parties to scrutinize company claims regarding the breadth and depth of the fake (or false) news problem on the site.

Suggested Further Research:

- Obtain a deeper sense of Facebook's business model in order to explore what platform-level changes at Facebook could further reduce the spread of fake news on the site. For example, should Facebook temporarily limit news feed visibility for information outlets that it determines consistently false news stories?
- Continue to monitor user habits of consuming news, and knowledge of and interaction with Facebook's tools to combat fake news. Facebook claims that it's too early in the lifecycle for these tools to determine whether they are successful, so it may be helpful for outside researchers to continue to run survey assessments over time as a reference point for future dialogue with the company.
- Reach out directly to members of the Board of Directors at Facebook to see if their approach and assessment of the fake news problem overlaps with what company officials have communicated. Do they see the scope of the problem similarly? How do they see it influencing the reputation of the company as a platform for reliable information?
- Engage the fact-checking organizations with which Facebook partnered, in order to assess how successful the fact-checkers believe this effort has been. Additionally, assuming the effort has been fruitful on the fact-checking side, assess whether additional resources can shrink the turnaround time on flagged articles.
- Explore the pros and cons of possible federal or state-level (in California, particularly) legislative and/or regulatory options, including means of requiring social media platforms to disclose data for research purposes to assess the success of implemented solutions.

X. Conclusion

Ultimately, the question is: Has Facebook effectively addressed the problem of fake news/misinformation on its platform? Despite Facebook's recent willingness to share information about its own internal review, this question is still difficult to answer without transparent data or metrics available internally from Facebook so that scholars and policy analysts may assess the success of its efforts. Facebook seems confident that the new tools are making false news less prevalent on the site and that users interact with false news less frequently. Our evaluation shows, however, that current efforts to solve the problem are not fully understood by users. There is room for the company to publicize its efforts to users and gain public support without intervening directly or proactively in how stories are shared. Research also suggests that the new tool that flags articles as disputed by fact-checkers may actually exacerbate users' negative engagement behaviors.[145] It is not clear if these tools are fast enough to deal with the virality of false news in real time, and we cannot be sure that user reporting of false news is effective when our research suggests that few users are familiar with the tool itself. Finally, the definition of "false news" is a purposely narrow one, which means that concepts such as biased news are not within the purview of Facebook's efforts. While we are encouraged by Facebook's proactive steps to resolve what it considers to be the fake news problem, we are not convinced that these efforts have yet been effective. Additional research and data sharing from Facebook will be necessary to ascertain how helpful these tools are, and whether new platform-based efforts or regulatory action are necessary.

[145] Sam Levin, *Facebook Promised to Tackle Fake News, But the Evidence Shows It's Not Working*, The Guardian (May 16, 2017 5:00 AM EST), https://www.theguardian.com/technology/2017/may/16/facebook-fake-news-tools-not-working.

XI. Appendices

Appendix A: Methodology

Our methodology consists of three sections: interviews, case studies, and an online survey. Our interviews aided us in learning more broadly about Facebook's efforts to combat false news, how those efforts are situated in the broader efforts by the social media platform market, and how those efforts into the legal framework governing social media platforms. Our case study began to identify how Facebook users engage with an example of a fake news story through the major milestones of the story's life online. Our quantitative research helped us understand the public's attitudes and perceptions about fake news and Facebook's tools to inhibit its spread.

Interviews

In order to obtain a greater understanding of the algorithms used and policies implemented at Facebook, we contacted people within Facebook departments and experts outside of the company who could shed light on these matters. For each of our interviewees, we scheduled either an in-person interview or a phone call. We recorded some interviews but primarily took notes without a recording during the majority of interviews. Many of our interviews were off the record, those that are on-the-record include:

- Tom Reynolds, Facebook, *Communications at Facebook*
- Alex Stamos, *Chief Security Officer at Facebook*
- Eugene Volokh, *1st Amendment Scholar at UCLA Law*
- Morgan Weiland, *Junior Scholar at Stanford CIS*
- Sophia Cope, *Staff Attorney on EFF's Civil Liberties Team*

Case Studies:

Our model for a case study involved tracking user engagement on Facebook during each milestone in the lifecycle of a fake news story. First, we selected a prominent fake news story, Pizzagate, and defined the milestones associated with the story—the initial report of the story, origin tracking, and the aftermath of the story. For each of these milestones, we identified and assessed engagement with the corresponding article published by each of the following news outlets: ***The New York Times***, ***Washington Post***, and NPR.

Our engagement metrics included the total number of likes, shares, and comments on each Facebook news story post (see Figure 1 for Facebook engagement of "Pizzagate" news stories). We not only quantified the public engagement with a post, but also examined the qualitative nature of such engagement through the sentiment analysis. We used the NRC Emotion Lexicon, a list of English words and their associations with eight emotions and two sentiment categories, to categorize comments on each post. Through this analysis,

we aimed to gain a better understanding of how Facebook users interact with and respond to fake news as it develops and evolves on the platform (see Figure 2). We sought to answer questions, such as: How does engagement vary with each milestone? Does it differ across the lifecycle depending on the outlet? What are Facebook users' emotions regarding the news coverage of the fake news story?

Online Survey:

We conducted an online survey to examine Facebook users' habits of news consumption, their attitude toward fake news, and their confidence in existing tools to inhibit the dissemination of fake news. Participants were recruited through Amazon Mechanical Turk, and the survey questions were accessed through Qualtrics. Participation in the survey was voluntary and anonymous and limited to adults who describe themselves as "regular" users of at least one of the platforms.

Procedure: In total 861 participants completed the survey on Amazon Mechanical Turk. The survey began with screening questions to identify whether participants have used Facebook, Twitter, Reddit, and Google. If participants haven't used Facebook, they will not see Facebook related questions, and so on for the other platforms. The final sample size for Facebook responses on the Mechanical Turk Survey is 635 survey respondents.

Demographic information: In terms of educational background, 36% of the participants had bachelor's degree (n = 229), followed by 27% (n = 174) with some college degree, 16.4% with a graduate degree (n = 104), 12.3% with an associate's degree (n = 78), 7.4% with a high school degree (n = 47), and 4.7% with less than a high school degree (n = 3). For race and ethnicity, most participants self-identified as White (75.3%, n = 478), followed by 8% Black or African American (n = 52), 7.4% Asian or Pacific Islander (n = 47), 6.0% Hispanic (n = 38), and 1.4% American Indian (n = 9). With regards to political identity, 32.1% of participants self-identified as Independent (n = 204), followed by 41.7% Democrats and 22.7% Republicans (n = 144). (See Figure 3 for detailed information.)

Key measures: The Facebook survey contains four components: users' Facebook habits, past exposure to fake news, attitudes toward fake news, and perceptions of Facebook's new tool to spot false news.

- Facebook intensity: Participants reported their Facebook usage by indicating the extent to which they were emotionally connected to Facebook. This 6-item scale was adapted from Ellison, Steinfield and Lampe (2007). Responses were on a scale ranging from 1 = *Strongly disagree* to 5 = *Strongly agree* (*M* = 3.97, *SD* = 1.14, Cronbach's alpha = .86).
- Facebook habits also examined the frequency of how participants 1) consume news on Facebook (five-point scale from 1 = *Never* to 5 = *Always*; M = 2.64, SD = 1.09), and 2) click for more information while reading news on Facebook (five-point scale from 1 = *Never* to 5 = *Always*; M = 2.81, SD = 1.08).

- Epistemic vigilance: On a 4-item-scale, we measured the extent to which participants tend to avoid the risks of being misinformed on Facebook by measuring the extent to which they assessed the truthfulness of news content, source credibility, intentions of the news writers, and the falsity of the news. Responses were on a scale ranging from 1 = *Strongly disagree* to 5 = *Strongly agree* (*M* = 4.32, *SD* = .97, Cronbach's alpha = .89).
- Sharing truthful news: On a single-item scale, participants self-identified how truthful a news article was to them when they decided to share it on Facebook. The scale ranged from 1 = Definitely false to 5 = Definitely true (M = 4.12, SD = .84).
- Exposure to fake/real news: We measured Facebook users' past exposure to fake and real news by asking them to read 5 top-performing fake and real news headlines, displayed in Facebook's Trending page during March 2017. Participants were asked to identify if they have read these news headlines (1 = No, 2 = Maybe, 3 = Yes), and if they believed the headlines were true or false (1 = False, 2 = Not sure, 3 = True) (See Figure 4.)
- Attitudes toward fake news were measured by the perceived seriousness of the fake news problem and its perceived prevalence: Participants first reported the extent to which they thought fake news was a serious social issue (3-item scale ranging from 1 = Strongly disagree to 5 = Strongly agree; Cronbach's alpha = .75, *M* = 4.17, *SD* = 1.05); second they reported the extent to which they thought news on their own (M_{self} = 3.12, SD_{self} = 1.08) and their friends' Facebook news feed and "Trending Page" was true (M_{friend} = 2.76, SD_{friend} = 1.01).
- Perceptions of new tools: Participants reported the extent to which they knew of some changes on Facebook to address fake news (M = 3.95, SD = .99). They were unaware of the reporting tool (70.4% unaware; 18% aware), they didn't use the reporting tool (90.0% didn't use it; 7.2% did), and they knew other recent efforts to curb the spread of fake news (M = 1.81, SD = .90).
- Confidence in fact-checkers: This scale consisted of four items, such as "I think independent fact-checkers are helpful for weeding out fake news." Participants were asked to indicate the extent to which they agree or disagree with the statements on a scale ranging from 1 = *Strongly disagree* to 5 = *Strongly agree* (*M* = 4.21, *SD* = .88, Cronbach's alpha = .82)

Appendix B: Figures

Figure 3. Descriptive data of education level, vote history, political affiliation, and race/ethnicity.

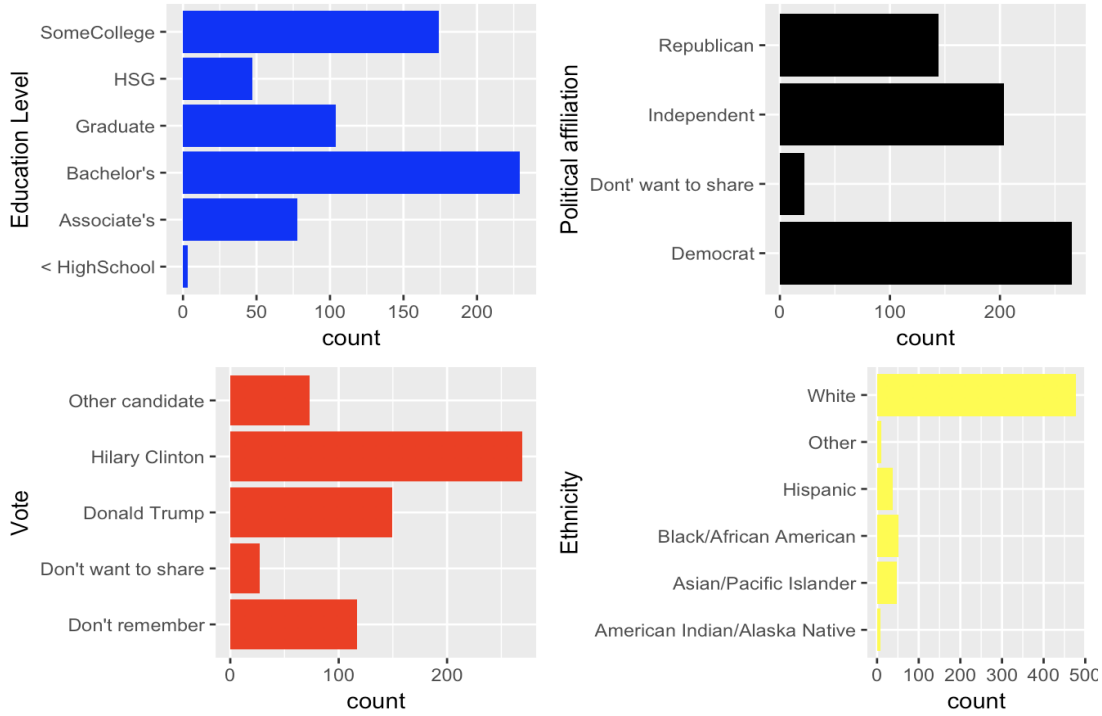
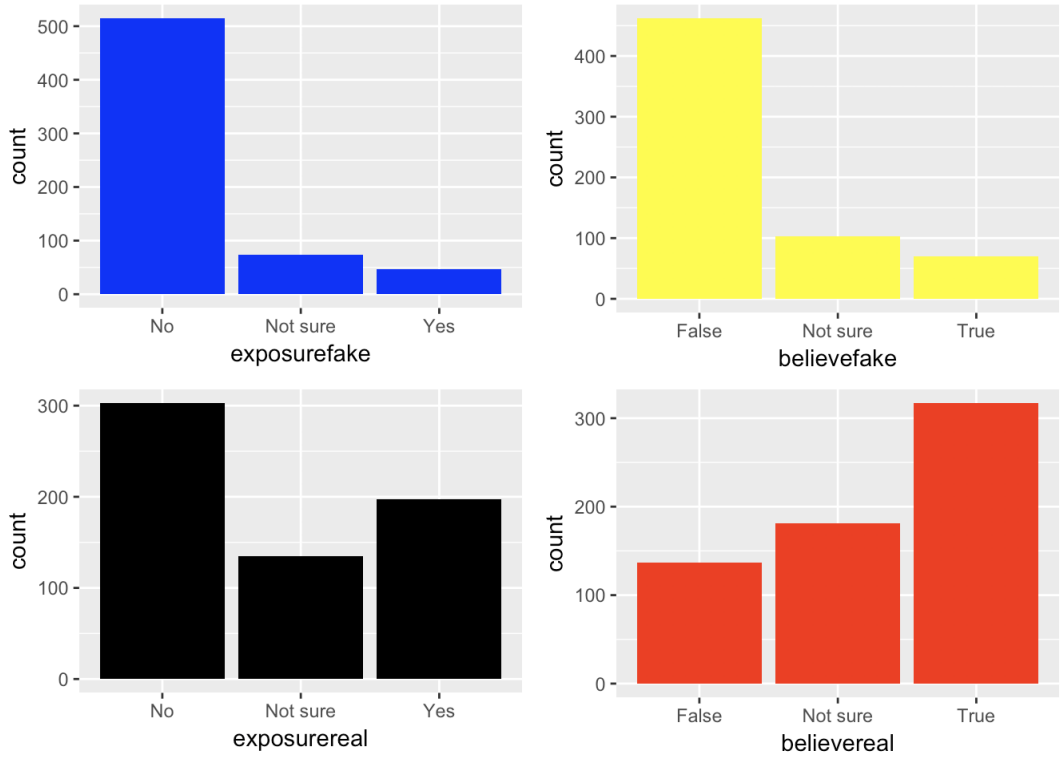


Figure 4. Frequency of exposure to fake/ real news, beliefs of fake/ real news.



Appendix C: Summaries of Case Study Articles

New York Times:

- *Man Motivated by "Pizzagate" Conspiracy Theory Arrested in Washington Gunfire*, N.Y. Times (Dec. 5, 2016): This article provides a factual overview of the shooting, the gunman and his belief in the "Pizzagate" conspiracy theory regarding Comet Ping Pong, and quotes from businesspeople in the neighborhood about the danger of fake information to their businesses and safety.[146]
- *Dissecting the #PizzaGate Conspiracy Theories*, N.Y. Times (Dec. 10, 2016): This article provides examples of the emails from WikiLeaks that online communities examined in order to develop the conspiracy story. It proceeds to give a step-by-step analysis of how the conspiracy theory developed and led up to the shooting at Comet Ping Pong.[147]
- Eli Rosenberg, *Alex Jones Apologizes for Promoting "Pizzagate" Hoax*, N.Y. Times (Mar. 25, 2017): This article discusses aftermath of the Pizzagate conspiracy,

[146] Eric Lipton, *Man Motivated by "Pizzagate" Conspiracy Theory Arrested in Washington Gunfire*, N.Y. TIMES (Dec. 5, 2016), https://www.nytimes.com/2016/12/05/us/pizzagate-comet-ping-pong-edgar-maddison-welch.html?smid=fb-nytimes&smtyp=cur&_r=0.
[147] Gregor Aisch, Jon Huang, & Cecilia Kang, *Dissecting the #PizzaGate Conspiracy Theories*, N.Y. TIMES (Dec. 10, 2016), https://www.nytimes.com/interactive/2016/12/10/business/media/pizzagate.html?smid=fb-nytimes&smtyp=cur.

focusing on Alex Jones, who helped popularize and spread the hoax. It quotes the apology from Alex Jones and discusses the impact of the shooting on the restaurant and its owner. It also discusses how the conspiracy theory has survived on beyond the shooting and discrediting.[148]

Washington Post:

- Faiz Siddiqui and Susan Svriuga, *N.C. Man Told Police He Went to D.C. Pizzeria with Gun to Investigate Conspiracy Theory*, Washington Post (Dec. 6, 2016): This article focuses on the facts surrounding the shooting and the gunman's motivation to investigate the "Pizzagate" conspiracy theory. The article also quotes local business owners regarding the dangers posed by fake news.[149]
- John Woodrow Cox, *'We're Going to Put a Bullet in Your Head': #PizzaGate Threats Terrorize D.C. Shop Owners*, Washington Post (Dec. 6, 2016): This article dives deep into the nature of the threats faced by various businesses located on the same block as Comet Ping Pong.[150]
- Margaret Sullivan, *So Sorry, Alex Jones. Your 'Pizzagate' Apology Doesn't Change a Thing*, Washington Post (Mar. 28, 2017): This article focuses on the authors contention that Alex Jones' apology for spreading the Pizzagate conspiracy was insincere. It notes that Jones had an incentive to issue an apology as a means to avoid a potential lawsuit.[151]

NPR*:*

- Camila Domonoske, *Man Fires Rifle Inside D.C. Pizzeria, Cites Fictitious Conspiracy Theories*, NPR (Dec. 5, 2016): This article briefly describes the factual circumstances surrounding the shooting and tracks the shooting's connection to the emails from John Podesta that were obtained and subsequently disseminated by WikiLeaks.[152]
- Rebecca Hersher, *Webpages Linked to Pizzeria Shooting Go Dark Even as Prosecution Moves Forward*, NPR (Dec. 14, 2016): This article walks through the different sources on the Internet that propagated the Pizzagate conspiracy theory and how many of them have, since the shooting, removed Pizzagate content from

---

148 Eli Rosenberg, *Alex Jones Apologizes for Promoting "Pizzagate" Hoax*, N.Y. TIMES (Mar. 25, 2017), https://www.nytimes.com/2017/03/25/business/alex-jones-pizzagate-apology-comet-ping-pong.html?smid=fb-nytimes&smtyp=cur.

149 Faiz Siddiqui and Susan Svriuga, *N.C. Man Told Police He Went to D.C. Pizzeria with Gun to Investigate Conspiracy Theory*, WASHINGTON POST (Dec. 6, 2016), https://www.washingtonpost.com/news/local/wp/2016/12/04/d-c-police-respond-to-report-of-a-man-with-a-gun-at-comet-ping-pong-restaurant.

150 John Woodrow Cox, *'We're Going to Put a Bullet in Your Head': #PizzaGate Threats Terrorize D.C. Shop Owners*, WASHINGTON POST (Dec. 6, 2016), https://www.washingtonpost.com/local/were-going-to-put-a-bullet-in-your-head-pizzagate-threats-terrorize-dc-shop-owners.

151 Margaret Sullivan, *So Sorry, Alex Jones. Your 'Pizzagate' Apology Doesn't Change a Thing*, WASHINGTON POST (Mar. 28, 2017), https://www.washingtonpost.com/lifestyle/style/so-sorry-alex-jones-your-pizzagate-apology-doesnt-change-a-thing.

152 Camila Domonoske, *Man Fires Rifle Inside D.C. Pizzeria, Cites Fictitious Conspiracy Theories*, NPR (Dec. 5, 2016), http://www.npr.org/sections/thetwo-way/2016/12/05/504404675/man-fires-rifle-inside-d-c-pizzeria-cites-fictitious-conspiracy-theories.

their sites.[153]

- James Doubek, *Conspiracy Theorist Alex Jones Apologizes for Promoting 'Pizzagate'*, NPR (Mar. 26. 2017): This article covers Alex Jones' apology and notes the belief on the part of many that it was motivated primarily for financial purposes. It also notes that many hardcore believers took Jones' apology to be further evidence of a conspiracy.[154]

Appendix D: Survey Questions

F1 In total, about how many total Facebook friends do you have?

F2 In the past week, on average, approximately how many minutes per day did you spend on Facebook?

F3 To what extent do you disagree or agree with the following statements?

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
Facebook is part of my everyday activities. (1)					
I am proud to tell people I'm on Facebook. (2)					
Facebook has become part of my daily routine. (3)					

[153] Rebecca Hersher, *Webpages Linked to Pizzeria Shooting Go Dark Even as Prosecution Moves Forward*, NPR (Dec. 14, 2016), http://www.npr.org/sections/thetwo-way/2016/12/14/505577985/webpages-linked-to-pizzeria-shooting-go-dark-even-as-prosecution-moves-forward.
[154] James Doubek, *Conspiracy Theorist Alex Jones Apologizes for Promoting 'Pizzagate'*, NPR (Mar. 26. 2017), http://www.npr.org/sections/thetwo-way/2017/03/26/521545788/conspiracy-theorist-alex-jones-apologizes-for-promoting-pizzagate.

I feel out of touch when I haven't logged onto Facebook for a while. (4)					
I feel I am part of the Facebook community. (5)					
I would be sorry if Facebook shut down. (6)					

F4 In the past 15 days, on average, how often did you read, watch or listen to news on Facebook in comparison with news websites and apps?

- ❑ Never (1)
- ❑ Sometimes (2)
- ❑ About half the time (3)
- ❑ Most of the time (4)
- ❑ Always (5)

F5 In the past 15 days, on average, how frequently did you click for more information rather than only reading the headlines of news articles.

- ❑ Never (1)
- ❑ Sometimes (2)
- ❑ About half the time (3)
- ❑ Most of the time (4)
- ❑ Always (5)

F6 In this section, you will read several news headlines. Please answer questions after

reading each of them.

Have you read, watched or listened to the following headlines of news stories?

	No (1)	Maybe (2)	Yes (3)
Obama Signs Executive Order Banning The Pledge Of Allegiance In Schools Nationwide (abcnews.com) (1)			
Police Find 19 White Female Bodies In Freezers With “Black Lives Matter” Carved Into Skin (tmzhiphop.com) (2)			
Florida man dies in meth-lab explosion after lighting farts on fire (thevalleyreport.com) (3)			
ISIS Leader Calls for American Muslim Voters to Support Hillary Clinton (worldnewsdailyreport.com) (4)			
Rage AGAINST THE MACHINE To Reunite And Release Anti Donald Trump Album (heaviermetal.com) (5)			
GOP rep: ’Nobody dies because they don’t have access to healthcare’ (hill.com) (6)			
Stephen Colbert’s Diatribe Against Trump to Be Reviewed by FCC (bloomberg.com) (7)			

ABC, CBS, and NBC joined CNN in refusing to air Trump advertisement (timesfreepress.com) (8)			
The Defense Dept is Easing a Trump Tower Apartment to House the Nuclear Football (deathandtaxesmag.com) (9)			
French Candidate Emmanuel Macron Says Campaign Has Been Hacked, Just Before Election (npr.org) (10)			
This Question is To Check That You Are Reading Carefully; Please Select Yes (thankyou) (11)			

F7 Whether each of the following headlines of news stories do you believe is false or real?

	False (1)	Not sure (2)	True (3)
Obama Signs Executive Order Banning The Pledge Of Allegiance In Schools Nationwide (abcnews.com) (1)			
Police Find 19 White Female Bodies In Freezers With “Black Lives Matter” Carved Into Skin (tmzhiphop.com) (2)			
Florida man dies in meth-lab explosion after lighting farts on fire (thevalleyreport.com) (3)			

ISIS Leader Calls for American Muslim Voters to Support Hillary Clinton (worldnewsdailyreport.com) (4)			
Rage AGAINST THE MACHINE To Reunite And Release Anti Donald Trump Album (heaviermetal.com) (5)			
GOP rep: 'Nobody dies because they don't have access to healthcare' (hill.com) (6)			
Stephen Colbert's Diatribe Against Trump to Be Reviewed by FCC (bloomberg.com) (7)			
ABC, CBS, and NBC joined CNN in refusing to air Trump advertisement (timesfreepress.com) (8)			
The Defense Dept is Easing a Trump Tower Apartment to House the Nuclear Football (deathandtaxesmag.com) (9)			
French Candidate Emmanuel Macron Says Campaign Has Been Hacked, Just Before Election (npr.org) (10)			

F8 Fake news, also called misinformation or false stories, can be broadly defined as news stories that are hoaxes or designed to mislead people. In the following questions, you will be asked to answer several questions related to this phenomenon.

To what extent do you agree or disagree with the following statement?

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
The news shared on my news feed is true. (1)					
The news shared on my Facebook friends' news feed is true. (2)					
The news on my Facebook "Trending" page is true. (3)					
The news on my Facebook friends' "Trending" page is true (4)					

F9 To what extent do you agree or disagree with the following statements?

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
Fake news is currently a very serious social issue. (1)					
Donald Trump would NOT have been elected for president WITHOUT fake news. (2)					
I don't want to read fake news on my Facebook newsfeed. (3)					
Fake news will generally impair the society. (4)					

F10 To what extent do you agree or disagree with the following statements?

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
When I read news articles on Facebook, I assess the source credibility. (1)					
When I read news articles on Facebook, I assess the writers' intentions. (2)					
When I read news articles on Facebook, I assess whether or not the content has falsity. (3)					
When I read news articles on Facebook, I gauge the truthfulness of content. (4)					

F11 How frequently did you perform the following sharing-related behaviors on Facebook?

	Never (1)	Sometimes (2)	About half the time (3)	Most of the time (4)	Always (5)
I share fact-based news articles. (1)					
I share comments-based news articles. (2)					
I share news if the content is interesting to me. (3)					
I consider if my friends on Facebook like the news before I share it. (4)					

F12 In general, how truthful to you is the news that you decide to share on Facebook?

- ❑ Definitely false (1)
- ❑ Probably false (2)
- ❑ Neither true nor false (3)
- ❑ Probably true (4)
- ❑ Definitely true (5)

F13 How much do you know about some changes on Facebook to address fake news?

- ❑ A great deal (1)
- ❑ A lot (2)
- ❑ A moderate amount (3)
- ❑ A little (4)
- ❑ None at all (5)

The following screenshot shows a new tool on Facebook for users to report false news. (Disclaimer: the researcher is not affiliated with, endorsed by, or acting on behalf of any of the entities whose names and logos are included in the survey.)

F14 Were you aware of this new tool before participating in the survey?

- ❑ No (1)
- ❑ Maybe (2)
- ❑ Yes (3)

F15 Have you used this tool before participating in the survey?

- ❑ No (1)
- ❑ Maybe (2)
- ❑ Yes (3)

F16 In addition to this new tool, how much do you know about Facebook's other recent efforts to curb the spread of fake news?

- ❑ None at all (1)
- ❑ A little (2)
- ❑ A moderate amount (3)
- ❑ A lot (4)
- ❑ A great deal (5)

F17 Fact-checking is the act of checking factual assertions in non-fictional text in order to determine the veracity and correctness of the factual statements in the text. Given the definition above, to what extent do you agree or disagree with the following statements?

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
I think independent fact-checkers are helpful for weeding out fake news. (1)					
I think the collaboration between Facebook and fact-checking websites can curb the spread of fake					

news on Facebook. (2)					
I think Facebook's new tool of spotting fake news will be effective in weeding out fake news. (3)					
I think there will be less fake news on Facebook in the future. (4)					

D1 What is the highest degree or level of school you have completed?

- ❑ Less than high school (1)
- ❑ High school graduate (2)
- ❑ Some college (3)
- ❑ Associate's degree (4)
- ❑ Bachelor's degree (5)
- ❑ Graduate degree (e.g., MA, MBA, JD, PhD) (6)

D2 Which race/ethnicity best describes you? (Please choose only one)

- ❑ White (1)
- ❑ Black or African American (2)
- ❑ American Indian or Alaska Native (3)
- ❑ Asian or Pacific Islander (4)
- ❑ Hispanic (5)
- ❑ Other (6)

D3 Which of the following best describes your political affiliation?

- ❑ Republican (1)
- ❑ Democrat (2)
- ❑ Independent (3)
- ❑ Don't want to share (4)

D4 Who did you vote for in the last presidential election?

- ❑ Donald Trump (1)
- ❑ Hillary Clinton (2)
- ❑ Other candidate (3)
- ❑ Did not vote (4)
- ❑ Don't want to share (5)

D5 Which of the following brackets best describes your household income level?

- ❑ Less than $25,000 (1)
- ❑ $25,000 - $50,000 (2)
- ❑ $50,000 - $90,000 (3)
- ❑ $90,000 - $150,000 (4)
- ❑ $150,000-$250,000 (5)
- ❑ More than $250,000 (6)
- ❑ Don't want to share (7)

Section 3. Google

I. Introduction

This section of the report reviews the role of Google, and specifically Google Search, in the misinformation landscape. This section tracks the problem of misinformation in search engines from the advent of search engine optimization and spam through the present day, focusing on Google's efforts to curb its role in spreading fake news following the 2016 U.S. elections.

In Part 1, the "arms race" between search engines and spammers exploiting weaknesses in search algorithms informs the analysis of Google's role in proliferating fake and/or biased news in the 2016 elections. Although a full accounting of the impact of fake news and misinformation on the 2016 elections is ongoing, this report tracks search results for senate and presidential candidates in that election. The data set finds that up to 30% of candidates had their search results affected by fake or biased content.

Part 2 summarizes Google's recent efforts in 2017 to curb misleading or offensive content through user reporting and human reviewers, along with the opinions of users and experts who are largely supportive of these changes. We include a discussion of the influence of Internet on journalism broadly, and describe efforts Google has made to invest in initiatives that bolster investigative journalism and news. The analysis of Google concludes with suggestions for policy and research directions, recommending that the company increase data transparency for researchers and find new ways to educate communities to evaluate information and determine truth. To summarize the results of our study, we conclude that transparency, unbiased review, and grassroots education are the next critical steps towards fully benefiting from the ubiquitous and powerful technologies that surround us.

II. Background

Just as the technologies of radio and television once did, the Internet is reshaping democracy. The 2016 elections in different parts of the world proved the importance of social media networks like Facebook and Twitter in influencing voting behavior and political views. Fake news and misinformation proliferated on these platforms, posing new challenges to democracy. While we do not yet know how the spread of fake news and misinformation across these platforms affected political views and outcomes, the events of the recent election – compounded by elections in France and perhaps the U.K. – reveal the importance of the platforms on democratic institutions. Although Google did not attract the attention that social media platforms did as a driver of false information, it is, however, a

crucial actor in this landscape. The Google search engine is a monolithic intermediary between users and content on Internet, providing information that helps to shape ideas, including political perspectives.

This section answers questions regarding misinformation and modern technology in relation to Google Search: How are democratic and political processes in the United States today affected by Google's search engine? What technical solutions, if any, could be implemented at the level of this platform to more positively impact our democracy?[155] This section answers these questions in a six-part analysis. First, it provides an overview of Google, and specifically the relevance of Google Search in today's society. Second, it explains how the search engine's algorithm has evolved over time to address historical manifestations of spam and misinformation. Third, it discusses the role of Google Search in the 2016 U.S. national elections, and the reasons for the recent wave of misinformation. Fourth, it assesses the current state of response in the context of Google's interventions with these recent political challenges. Fifth, it describes the legal framework for online intermediary platforms. Sixth, it provides a brief overview of the relationship between the fake news phenomenon and journalism. And last, it offers public policy recommendations, ranging from the platform level to civil and regulatory lines of action.

This summary draws on information in three forms: (1) two quantitative analyses, one of a data set of Google search results for U.S. congressional candidates in 2016 and the other of survey data from 475 Google Search users collected in May 2017 through Amazon Mechanical Turk; (2) a qualitative analysis of eight semi-structured interviews with key academic and industry figures; and (3) a review of relevant published documents, including academic articles and popular press coverage.[156]

The Role of Google Search in Today's Information Economy

Nearly half of the world's population – 3.5 billion people – accesses the Internet regularly.[157] According to the International Telecommunication Union 2016 Report, "People no longer *go* online, they *are* online…. Internet users read, shop, bank and date online, thanks to a growing number of websites, services and applications that did not exist a decade ago."[158] The quantity of data available on the Internet is similarly astounding. As

[155] The recent reports of the research institute Data and Society, and Harvard and Northeastern universities, highlight the importance of addressing these questions at the platform level. See Alice Marwick and Rebecca Lewis, "Media Manipulation and Disinformation Online" (Data&Society, 2017), https://datasociety.net/output/media-manipulation-and-disinfo-online/ ; see also, David Lazer, et al., "Combating Fake News: An Agenda for Research and Action," May 2, 2017, https://shorensteincenter.org/combating-fake-news-agenda-for-research/; and Robert Faris, et al., *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election, August 2017,* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3019414.

[156] For more information on the quantitative and qualitative methodologies see the Appendix.

[157] International Telecommunication Union, "Measuring the Information Society Report," 2016, http://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2016/MISR2016-w4.pdf.

[158] *Ibid.*, 182.

emblematized by Wikipedia, the number of articles reached 40 million in 2016. In order to make this information accessible in human scale, it must be preprocessed, indexed, and made searchable; this is the role of a search engine. Legal scholar James Grimmelman aptly points out that "The Internet today is usable because of search engines."[159]

Google's search engine receives between 3.5 and 5.5 billion queries per day.[160] This traffic has grown exponentially. When Google launched in 1998, there were 10,000 daily queries; by 2006, the same amount was searched every second, and by 2017 in less than a tenth of a second.[161] Table 1 presents the demographics of search engines from a 2012 survey made by the Pew Research Center.[162] There is no gender difference in the use of search engines but its use is biased towards white, young, more educated, and higher income populations. These demographics are for all search engines. Pew, however, reports that more than 80% of search engine users use Google Search, suggesting that these demographics describe Google users.

Who uses search?

% of online adults in each group who use search engines

	% of each group who ever use search	% of each group who used a search engine yesterday
All online adults	91%	59%
Gender		
Male	90	59
Female	92	60
Race/Ethnicity		
White	93*	63*
African American	89*	44
Hispanic	79	44
Age		
18-29	96	66*
30-49	91	65*
50-64	92	52*
65+	80	38
Education		
Some high school	78	34
High school	88*	45*

[159] James Grimmelmann, "The Google Dilemma," *NYL Sch. L. Rev.* 53 (2008): 941.
[160] "Google Search Statistics - Internet Live Stats," accessed June 10, 2017, http://www.internetlivestats.com/google-search-statistics/; Danny Sullivan, "Google Now Handles at Least 2 Trillion Searches per Year," *Search Engine Land*, May 24, 2016, http://searchengineland.com/google-now-handles-2-999-trillion-searches-per-year-250247.
[161] Sullivan, "Google Now Handles at Least 2 Trillion Searches per Year."
[162] Kristen Purcell, Joanna Brenner, and Lee Rainie, "Search Engine Use 2012" (Pew Research Center, March 9, 2012), http://www.pewinternet.org/2012/03/09/search-engine-use-2012/.

Some college	94*	65*
College graduate	95*	74*
Household income		
< $30,000	84	45
$30,000 - $49,999	93*	54*
$50,000 - $74,999	97*	66*
$75,000+	95*	76*

* Denotes statistically significant difference with other rows in that category

Source: The Pew Research Center's Internet & American Life Project Winter 2012 Tracking Survey, January 20-February 19, 2012. N=2,253 adults age 18 and older, including 901 cell phone interviews. Interviews conducted in English and Spanish. The margin of error is plus or minus 3 percentage points for internet users.

Table 1. Pew Research Center 2012 report on the demographics of search engine users.[163]

In addition to its widespread use, Google has also appropriated the biggest share of this market. Internationally, 89% of searches occurred using Google in mid-2015;[164] in the US, 65% of searches took place on Google.[165] The search engine owes its success to popular satisfaction with its search results. 88% of the respondents in our Practicum survey reported that they are somewhat or extremely satisfied with the information they find using Google Search. Although the Practicum survey is not representative of the U.S population, the percentage may highlight a trend in users' satisfaction. Such satisfaction likely reflects Google's agile algorithm which responds automatically to the creation of new webpages and the creativity of its users; indeed, 15% of the queries Google receives each day have never been searched on Google before.[166]

As a tool that enables individual users access to a vast body of knowledge, Google and other search engines are also critical to defining ideas and shaping the success of businesses, events, and the political landscape. Grimmelmann exemplifies this role with five iconic cases.[167] One shows that there is information we can find only because of search engines, for instance the existence of "mongolian gerbils." A second explains how open source content allows information to flow freely, and this available content enables the search engine to organize and display information.[168] But this also implies that humans can affect the rankings being displayed: for example, a third case describes the way the query "jew" caused search engines, using algorithms based on user behavior, to display

[163] *Ibid.*, 6.
[164] Martin Moore, "Tech Giants and Civic Power" (Centre for the Study of Media, Communication and Power, April 2016), https://www.kcl.ac.uk/sspp/policy-institute/CMCP/Tech-Giants-and-Civic-Power.pdf.
[165] "Google Search Statistics - Internet Live Stats."
[166] Ben Gomes, "Our Latest Quality Improvements for Search," *Official Google Blog*, April 25, 2017, http://blog.google:443/products/search/our-latest-quality-improvements-search/.
[167] Grimmelmann, "The Google Dilemma."
[168] In interviews with our research team, Google officials highlighted this aspect of the web as well.

controversial content. A fourth case presents the importance of search engine rankings to business, recounting how companies have argued that a low ranking would irreparably damage their economic viability. Finally, Grimmelmann's fifth case relates to politics online: a search for "Tiananmen" displayed different results in different countries, based on how the algorithm responded to queries related to those countries' political contexts. Grimmelmann concludes, "Search results matter: to culture, to business, and to society. Every decision about search, and every lawsuit about search, has these inescapably political questions at its heart. That is Google's dilemma, and ours."[169]

Important to an understanding of Google's role in this socio-cultural landscape is an understanding of the market structure in which Google operates. Google receives 89% of its revenues from advertising.[170] The market structure of search engines is two-fold. On one side, search engines provide a service to users in exchange for their user data and who consent to see ads linked to their queries. On the other side, companies pay search engines to place the companies' advertisements on relevant webpages. The first market is closer to a monopoly, the second one to an oligopoly. In the first case, Google has more freedom to fix the prices by extracting more information from users, for example, and then using that information to reap profits that benefit its business model. In the latter, Google must compete with other platforms and media companies for the business of companies seeking advertisement outlets.

Google's Historical Efforts to Combat of Misinformation

The algorithm behind Google's search engine has evolved constantly since its inception. In addition to regular improvements, changes in the algorithm have frequently been triggered in response to spammers attempting to manipulate the system. This process has been termed an "arms race" and has long affected Google and all other search engines. Historical search engines like AltaVista and Lycos were also known to change their algorithms to thwart spammers.[171]

In the early 1990s, first generation search engine algorithms were based on *vector models* of documents, a theoretical formulation from the field of Information Retrieval. Metaxas writes, "the more rare words two documents share, the more similar they are considered to be."[172] This model soon came under attack from spammers, who began a practice that became known as *keyword stuffing*—creating pages with many rare keywords (sometimes

[169] Grimmelmann, "The Google Dilemma," 940-950.
[170] "Google: Distribution of Revenue by Source 2016," *Statista*, 2017, https://www.statista.com/statistics/266471/distribution-of-googles-revenues-by-source/.
[171] Interview with Takis Metaxas, May 1, 2017.
[172] Panagiotis Takis Metaxas, "Web Spam, Social Propaganda and the Evolution of Search Engine Rankings," in *Web Information Systems and Technologies* (International Conference on Web Information Systems and Technologies, Springer, Berlin, Heidelberg, 2009), 176, doi:10.1007/978-3-642-12436-5_13.

hidden by matching the text to the background color to avoid detection by users on the page), so as to get their pages ranked highly for many different, unrelated user queries.[173]

In response, by 1996, search engines had developed second generation algorithms with more sophisticated techniques, for instance relying on connections in the network of pages to determine credibility: popularity—where many pages link to any single given page—was taken as an indicator of quality. In response, spammers created *link farms*, clusters of spam pages all linking to each other to help each other appear popular and, therefore, rise in the rankings.

The third generation of search engine algorithms, introduced with Google's famous PageRank algorithm in 1998, built on the idea of a popularity network but weighted the votes of highly reputable pages (*i.e.*, a page with more links to it than pages linked to by it) more than less reputable pages. Unfortunately, this method is also vulnerable to spam: spammers acquire legitimately high rankings on pages about unrelated topics, and then use that reputation to elevate other pages (which might, for instance, be full of ads providing revenue to the spammer).[174] This tactic was known as forming a *mutual admiration society.*

This process of gaming a search engine to move one's webpages higher in a search engine's ranking is called *Search Engine Optimization* (often abbreviated "SEO"). Notably, this kind of spamming is not uncommon; legitimate companies and individuals often employ SEO experts to help publicize their content online. In this sense, many of us are small-time propagandists without posing much risk to our communities. The danger lies in those who would use these techniques to spread falsehoods and misinformation, and lead others astray in high-risk and high-stakes contexts: for instance, in matters of politics.

One notable example of the entrance of politics into SEO is the "miserable failure" hack. Third wave search engine algorithms, in addition to PageRank, started using *anchor text* (the text that a website matches with a URL link) to learn something about the linked URL. For instance, if a page wrote something like "Many newspapers [link: nytimes.com] have sports sections," Google would learn that the concept "newspapers" was related to the Times' website. Capitalizing on this feature, a group of activists started using the phrase "miserable failure" to link to President George W. Bush's official webpage, causing Google searches for George W. Bush to yield this phrase.[175] This tactic, termed a *Googlebomb*, gained significant publicity and was deployed against other politicians as well. The enormous potential for political impact of web (and particularly web search) technologies, has been recognized as a significant new form of politicking.[176]

The unspoken fuel behind this phenomenon is the implicit trust placed in search engines

[173] Metaxas, "Web Spam, Social Propaganda and the Evolution of Search Engine Rankings."
[174] *Ibid.*
[175] *Ibid.*
[176] Grimmelmann, "The Google Dilemma."

like Google by their users. "Users have come to trust search engines as a means of finding information, and spammers have successfully managed to exploit this trust."[177] More often than not, we do not interpret the results of every Google search we make with skepticism. (After all, how many did you make today? That would be exhausting.) Instead, we have learned a heuristic: that Google's results are almost always helpful and well-ordered. This pattern ingrains an unconscious instinct to trust Google's ranking deeply, and to correlate authority with ranking—meaning that companies ranked highly will see their sales and subscriptions rise, politicians ranked highly may be perceived as more credible, and highly ranked spammers' pages will garner more clicks (and, along with that, ad revenue).[178]

In addition to spammers manipulating Google's search rankings, another source of bias is an accidental byproduct of crowdsourced data that Google's algorithms learn from. Like many algorithms powered by machine learning and fed data generated by (flawed, biased) human users, Google Search is vulnerable to displaying the same biases. For example, in 2016 Google had to manually alter its search suggestions (the suggestions that appear when a user begins typing a search query) to remove autocomplete results that appeared when a user started a query with "are jews" or "are women." Before those changes, Google's algorithm, which used machine learning to autocomplete with common phrases from other users, would suggest ending either phrase with the word "evil."[179] This example, among others, indicates another potential site of misinformation in Google's Search features—those caused by the accidental effect of machine learning trained with biased data. Such algorithmic bias is the topic of a growing field of study, and points to the need for better data collection and transparent research surrounding search engines. While problematic patterns in autocorrect might become obvious to end users who see them directly, subtler patterns (for instance, a consistent slant in search results for certain political topics) might be harder to catch, and nearly impossible to prove as a systematic problem. Without a thorough research framework for collecting and analyzing data, these sorts of issues risk go unexamined.

To address a third possibility for bias, our research further considered both the possibility of Google intentionally influencing its search results or user search suggestions to achieve some political outcome and the company's occasional explicitly political messages to users. Our external examination of the algorithm did not reveal evidence of manipulation of the algorithm for political ends. In fact, our research suggests that Google strives for transparency when it seeks to achieve particular political outcomes. The company has

---

[177] Metaxas, "Web Spam, Social Propaganda and the Evolution of Search Engine Rankings," 171.

[178] Bing Pan et al., "In Google We Trust: Users' Decisions on Rank, Position, and Relevance," *Journal of Computer-Mediated Communication* 12, no. 3 (April 1, 2007): 801–23, doi:10.1111/j.1083-6101.2007.00351.x. This article studies the perception of Google's search algorithm by college students. Using an eye tracking experiment, the authors show that students are strongly biased towards links higher in Google's results (even if those pages were less relevant.)

[179] Carole Cadwalladr, "Google, Democracy and the Truth about Internet Search," *The Guardian*, December 4, 2016, sec. Technology, https://www.theguardian.com/technology/2016/dec/04/google-democracy-truth-internet-search-facebook.

occasionally leveraged its logo, for example, to communicate a particular political message. In January 2012, "Google blacked out its logo in protest against the Stop Online Piracy Act (SOPA) and Protect IP act (PIPA). It also urged people to sign a petition against the bills. On 22nd May 2015 Google's homepage in Ireland told its users they should #VoteYes in the Irish referendum on gay marriage."[180] These occurrences are akin to any company lobbying for its business values, and not a novelty of the technology era. However, given Google's unprecedented availability to billions of people around the world, this form of influence is worth keeping in mind.

III. Misinformation and the 2016 U.S. National Elections

Related Literature

In the months following the 2016 U.S. national elections, a literature is emerging, studying the impact of fake news the election and its outcome. Most of these studies have focused more on social media than on search engines, making this report one of the few studies in the field.

Imperative to the field is a survey conducted by Allcott & Gentzkow (2017) that tracks online news search behaviors. The survey, which is representative of the U.S population, finds that 28.6% of respondents reported receiving political news primarily online, either from social media (13.8%) or websites (14.8%).[181] In comparison, a majority (57.2%) report receiving their political news primarily through television. (The survey did not examine respondents' secondary or tertiary sources of political information.) With regard to search engines, specifically, Allcott & Gentzkow find that between October and December of 2016, 30.6% of visits to pages hosted by the 690 top news sites came from search engines in contrast with 10.1% coming from social media sites. Further, 22.0% of visits to a list of 65 fake news sites came from search engines compared to 41.8% from social media sites.[182] These twin findings suggest that social media sites, rather than search engines, may have been primary drivers in the spread of fake news in 2016.

While there may be less cause for concern regarding the magnitude of the role of the Google search engine in amplifying fake news, some vulnerabilities persist. Some psychologists argue that exposure to different balances of favorable and unfavorable search result rankings can sway voters' opinions on candidates, even when those voters do not visit or fully process the information on the pages shown in the ranking. This effect has

[180] Moore, "Tech Giants and Civic Power," 28.
[181] Hunt Allcott and Matthew Gentzkow, "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives* 31, no. 2 (May 2017): 211–36, doi:10.1257/jep.31.2.211.
[182] *Ibid.*

been termed the Search Engine Manipulation Effect, or "SEME."[183] Further, psychologists have also found that content can be digested unconsciously by users who are unaware of the influence on their beliefs. Such unconscious or biasing effects include peripheral persuasion, confirmation bias, and misinformation correction. We believe that further studies with methodologies able to capture psychological effects may arrive at different conclusions, give a more comprehensive understanding, and find that fake news may, in fact, have a more significant role in political outcomes.

A recent report by Marwick and Lewis (2017) of Data and Society reflects on these psychological factors. The study details media manipulation by various actors including trolls, the "Alt-Right", conspiracy theorists, politicians, and others. The report highlights how distrust and polarization created a feasible environment for the fake news phenomenon and concludes: "we can expect the continuation of some current trends: an increase in misinformation; continued radicalization; and decreased trust in mainstream media."[184] While this report does not address these topics from the internal perspective and management decisions of the platforms, it does examine the role of platforms including Google in perpetuating the spread of misinformation, both directly and indirectly, in terms of the impact the platforms have had on journalism.[185]

Reasons for Misinformation

In our interviews, Google management and experts in the field both were careful to distinguish between two separate causes of fake news and misinformation: financial incentives and political manipulation.

Evidence of the role of financial incentives in proliferating fake news—news that is patently and purposely false—implicated Google. Well-documented news reveal that Macedonian teens developed websites peddling sensationalist, patently false, stories and hoaxes in order to gain money through a Google advertising algorithm that paid them every time their pages were viewed.[186] Indeed, Google AdSense turns a profit via an algorithm that connects companies advertising their products to websites looking to host ads, and delivers a small portion of that profit to the host sites. Google does place some restrictions on which online publishers are allowed to host ads through AdSense, but those restrictions do not bar "fake news" sites.[187] In this way, people producing fake stories, for instance

[183] Robert Epstein and Ronald E. Robertson, "The Search Engine Manipulation Effect (SEME) and Its Possible Impact on the Outcomes of Elections," *Proceedings of the National Academy of Sciences* 112, no. 33 (August 18, 2015): E4512–21, doi:10.1073/pnas.1419828112.
[184] Marwick and Lewis, "Media Manipulation and Disinformation Online," 44.
[185] Marwick and Lewis, "Media Manipulation and Disinformation Online."
[186] Samanth Subramanian, "Inside the Macedonian Fake-News Complex," *WIRED*, February 15, 2017, https://www.wired.com/2017/02/veles-macedonia-fake-news/.
[187] Ginny Marvin, "Google Isn't Actually Tackling 'Fake News' Content on Its Ad Network," *Marketing Land*, February 28, 2017, http://marketingland.com/google-fake-news-ad-network-revenues-207509.

propagating the so-called Pizzagate conspiracy theory, could bring in revenue (with a cut going to Google) by allowing Google to match advertisements to sites. Google has recently pledged to remove from its network any publishers presenting information under "false or unclear pretenses."[188] Notably, this would still allow the Pizzagate pages to continue working with Google AdSense.

The second motivation for producing misinformation is for political gain, which is perhaps especially dangerous for an informed democracy. In such cases, actors, either foreign or domestic, may pursue Google as a vehicle for spreading their biased or incorrect perspectives with the hope of manipulating American voters. For instance, Stanford Professor Larry Diamond believes there is sufficient evidence to show that Russia intentionally manipulated the 2016 U.S. presidential election for its own political gain. In such cases, curbing ad revenue is insufficient to address the problem, and online sources used by citizens to inform themselves will be a primary target. Unfortunately, it is also harder to quantify this subset of the problem, since many websites promote biased content for political gain, and the metric with which to measure that bias can be subjective.

2016 Election URL Search Analysis Results

To investigate the issue of biased news on Google and its role in the 2016 elections, we accessed an unpublished data set collected by researchers at Wellesley College. The data are comprised of top 10 Google Search result URLs (the first page of results) collected twice per week in the 26 weeks prior to the November 2016 U.S. national elections, a total of 141,313 URLs. The results come from Google searches of all congressional candidate names (incumbents and all other candidates running) in six U.S. states (Arkansas, Arizona, Colorado, California, Alaska, and Alabama). Additionally, the data set contained URLs from searches for four presidential candidates: Clinton, Rubio, Sanders, and Trump. Building on the Wellesley data set, we compare these URLs against the 150 URLs in PolitiFact's "guide to fake news websites," which flags sites that have in the past produced disputed content.[189] (We also considered other available lists, such as that curated by Melissa Zimdars,[190] but settled on PolitiFact because it is not only carefully vetted but also relatively short, focusing on popular sites.) That comparison enabled us to check how many, if any, of the top ten results in the data set are from fake or questionable news sites.

	Overall number	Number flagged by PolitiFact's list	Percentage flagged of total

[188] *Ibid.*

[189] Joshua Gillin, "PolitiFact's Guide to Fake News Websites and What They Peddle," *PunditFact*, April 20, 2017, http://www.politifact.com/punditfact/article/2017/apr/20/politifacts-guide-fake-news-websites-and-what-they/.

[190] Melissa Zimdars, "Resource-False-Misleading-Clickbait-Y-and-Satirical-'News'-Sources-1.pdf," November 2016, http://d279m997dpfwgl.cloudfront.net/wp/2016/11/Resource-False-Misleading-Clickbait-y-and-Satirical-%E2%80%9CNews%E2%80%9D-Sources-1.pdf.

All search result URLs	141,313	2,152	1.52%
Unique search result URLs	9,573	283	2.95%
Politicians	356	103	28.9%

Table 2. Findings in the URL analysis show that over 1.5% of all results shown for the politicians in our data set were from disputed sites, and that this amounted to nearly a third of all politicians in the data set having their Google Search results affected by the presence of these sites.

As summarized in Table 2, our analysis shows that over 1.5% of all URLs in our data set belonged to websites disputed by PolitiFact. When filtering for only unique sites, this rises to 2.95%. This indicates that the non-flagged sites were more likely to repeat (appearing consistently from week to week) in our data set than the flagged sites—in other words, flagged sites were more likely to appear and disappear, rather than stay consistently in the top 10, a fact which speaks to the ephemeral, quick-reaction, non-authoritative nature of these sites. Furthermore, we find that when measuring the 356 politicians in our data set, nearly a third had their results affected by these flagged sites in the time period over which data was collected.

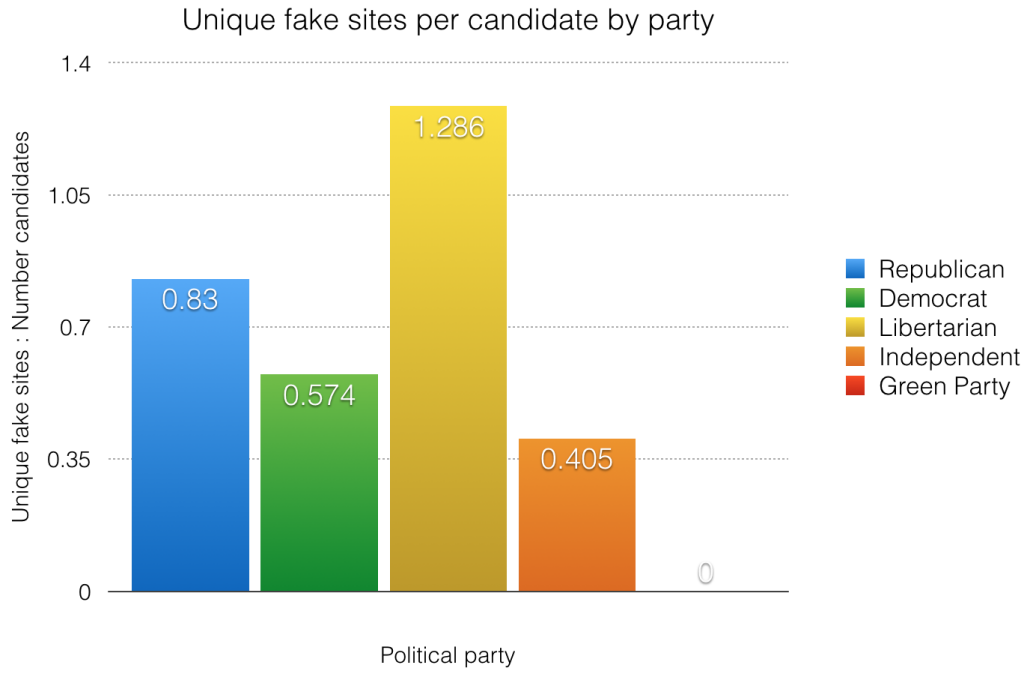


Figure 1. Considering the presence of disputed sites in search results by candidate, we find that Libertarian candidates were much more likely to have their results display URLs from

disputed sites than were other candidates.

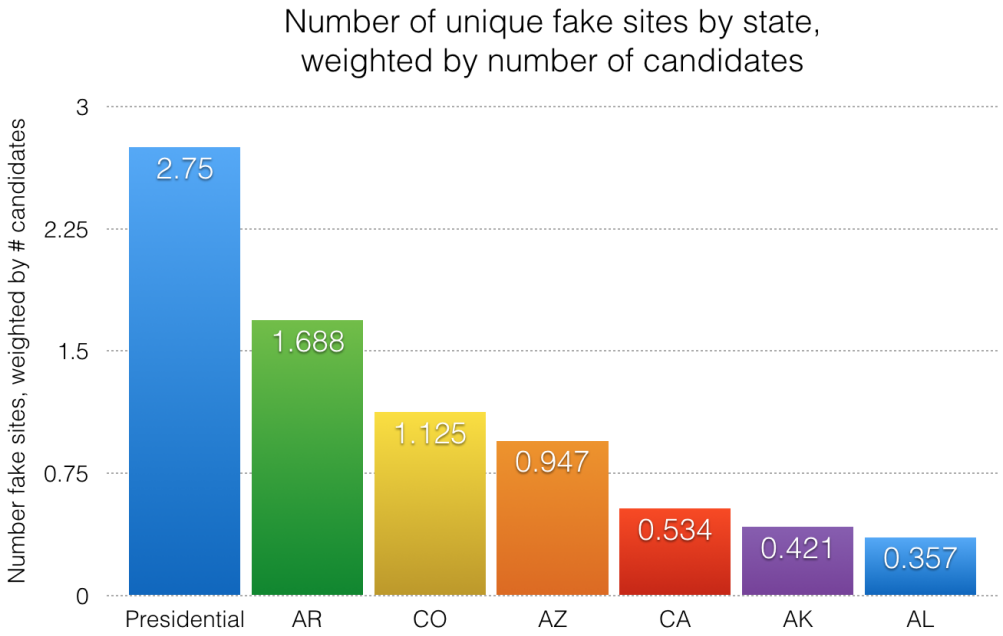


Figure 2. When grouping candidates by position, we find that the candidates in the presidential race were more likely than congressional candidates to see disputed sites in their search results, confirming our hypothesis.

When analyzing by political party, we find that Libertarian candidates were much more likely than others (and Republicans more likely than Democrats) to have their results affected by disputed sites (Figure 1). This may be in part because these less-mainstream candidates inspire coverage from less-mainstream sites whose journalistic standards are sometimes lacking, or because the political positions of those candidates or the journalistic standards of their constituents inspire more editorializing. We also observe that the four presidential candidates were more likely to have their search results affected than were congressional candidates, confirming our hypothesis that the highest-stakes race would attract more spammers and others peddling misinformation or biased content (Figure 2).

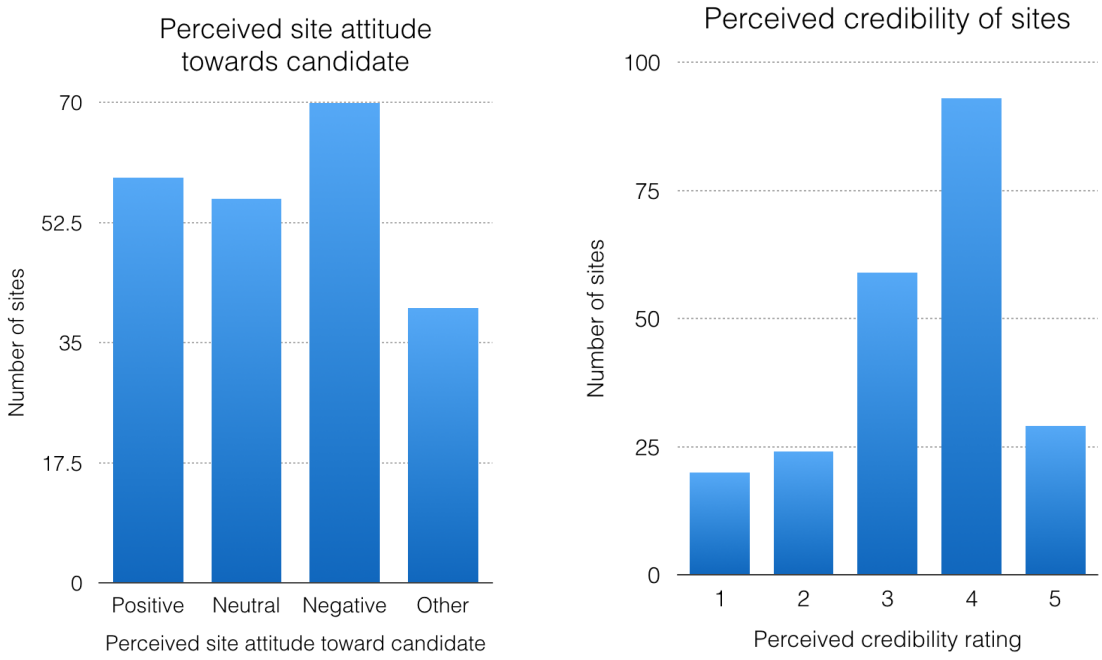


Figure 3. When web users on Mechanical Turk were asked to evaluate the disputed webpages in our data set, they reported that those pages were not overwhelmingly slanted in one ideological position relative to the candidate, nor were the sites perceived as low in credibility.

In order to better understand the URLs in our data set and the perception of those webpages by web users, we posted the URLs on Amazon's Mechanical Turk crowdwork marketplace, along with instructions to read and evaluate each webpage on its perceived attitude towards the candidate in whose search results it appeared, as well as the respondent's perceived credibility of the site. Our results, shown in Figure 3, found that the webpages were not perceived as overwhelmingly positive, negative, or neutral with regards to the candidate and, even more surprisingly, that the sites were perceived as relatively credible, scoring an average of 3.38 out of 5. These results have a couple possible implications. From an optimistic perspective, this could indicate that although sites that have been known to distribute misinformation were flagged in our data collection, the majority of their content is reasonably credible and not overtly biased propaganda. This could point to a tactic of those engaging in misinformation: to sprinkle the misinformation in throughout a larger body of quality content, as was done by early spammers in their mutual admiration societies. On the other hand, more serious issue may be that those pages, while not actually credible, were perceived as such by users, who are unable to effectively evaluate the information with which Google presents them. This is supported by the fact that users did report pages as having a slant either for or against the candidate in the majority of cases, but still found most pages moderately to highly credible.

IV. Current Efforts to Combat Misinformation

To address the problem of misinformation, Google, like other platforms, has included mechanisms to fact-check content and engage users in reporting on low quality search queries and "Featured Snippets" ("which shows a highlight of the information relevant to what you're looking for at the top of your search results"). The company has avoided taking content down unilaterally. Instead, it has tried to improve its algorithm by considering user feedback and hiring quality evaluators. Both Facebook and Google have reacted quickly to public pressure generated by the fake news phenomenon in 2016. They have an incentive to act quickly and avoid social pressure for regulation of their markets; some of our interviewees alluded to this: 2016 was a year that put many eyes on these platforms, and, as Daphne Keller pointed out, "Google is being smart." These platforms know that the government, think tanks, and the public in general are looking at them, and therefore now is an opportune time to push them to address this issue pro-actively.[191]

Google has claimed that the fake news problem is rather minimal, but still worth addressing. As rationale for their recent adjustments, Google claims that only a fraction of a percent (around 0.25%) of user queries "have been returning offensive or clearly misleading content, which is not what people are looking for,"[192] however the company is doing things to address the problem. In April 2017, Google publicly announced three solutions to combat misinformation, as summarized in a recent post by Danny Sullivan.[193] Two solutions include feedback forms with which users can send feedback about search suggestions and "Featured Snippets" answers.[194] A third solution tries to enhance authoritative content by hiring 10,000 "search quality evaluators" to give feedback on search results.

Figures 4 and 5 display the forms with which users can provide feedback to the platform. The form in Figure 4 is for search suggestions, which allows users to report which of the search suggestions seem inappropriate and why. The form in Figure 5 is for Snippets, and includes more predetermined options for reporting as well as a box for comments or suggestions.

[191] Interviews with Jonathan Zittrain and Daphne Keller, May 10, 2017.

[192] Gomes, "Our Latest Quality Improvements for Search."

[193] Danny Sullivan, "Google's 'Project Owl' -- a Three-Pronged Attack on Fake News & Problematic Content," *Search Engine Land*, April 25, 2017, http://searchengineland.com/googles-project-owl-attack-fake-news-273700; Gomes, "Our Latest Quality Improvements for Search."

[194] Gomes, "Our Latest Quality Improvements for Search."

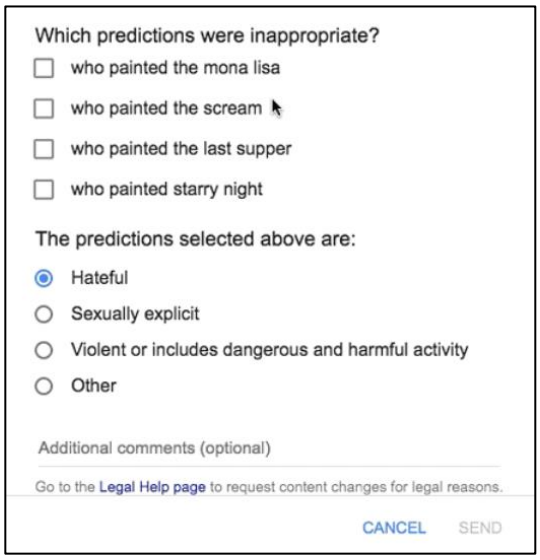


Figure 4. Form for users to give feedback on search suggestions.

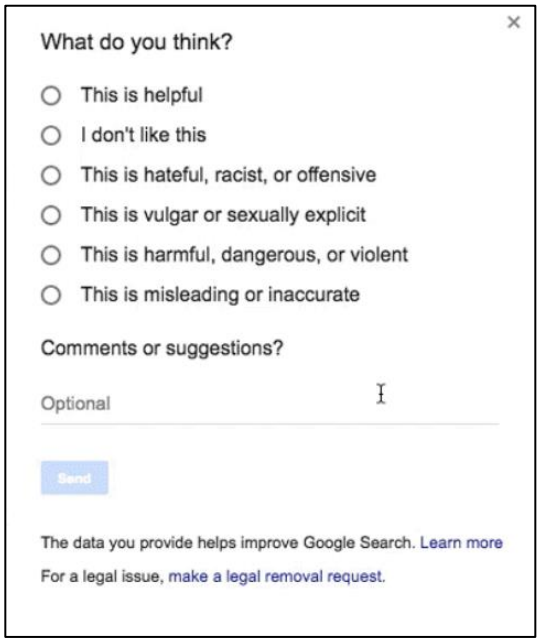


Figure 5. Form for users to give feedback on Snippets.

Notably, at the present time Google does not allow individual users to give feedback on search results. Google attempted to allow user-level feedback of search results in 2008 but discontinued this feature, claiming that it was widely unused.[195] Instead, the company has hired professional evaluators to play this role, a decision that may be influenced by the way Google has scoped the problem and restricted it to a very small proportion of searches. These quality evaluators follow a protocol established by Google to rate search results according to specific and very detailed criteria; the protocol is 157 pages long.[196] Evaluators' provided rankings, weighted into the Search algorithm, combining with the existing infrastructure. This means that in addition to such other features of the algorithm as key words, user historical queries and place, among others, the evaluators' rankings help determine the search results to specific queries.

The guidelines employed by the quality-control evaluators describe upsetting or offensive content as including the following:[197]

- "Content that promotes hate or violence against a group of people based on criteria including (but not limited to) race or ethnicity, religion, gender, nationality or citizenship, disability, age, sexual orientation, or veteran status."
- "Content with racial slurs or extremely offensive terminology."
- "Graphic violence, including animal cruelty or child abuse."
- "Explicit how-to information about harmful activities (e.g., how-tos on human trafficking or violent assault)."
- "Other types of content which users in your locale would find extremely upsetting or offensive."

Researchers cannot yet assess the impact of these efforts, but no doubt Google is carefully collecting data in order to do so. However, the new features do not seem to have been widely publicized. In our survey of 467 Google users, conducted in May 2017, 72% of respondents did not know Google had made changes to curb fake news, and less than 4% reported having been informed by Google's own communications (see Figure 6). When informed of the details of these features, however, 63% of those surveyed expressed support for the changes.

[195] "SearchWiki: Make Search Your Own," *Official Google Blog*, November 20, 2008, https://googleblog.blogspot.com/2008/11/searchwiki-make-search-your-own.html.
[196] Google, "General Guidelines," May 11, 2017, https://static.googleusercontent.com/media/www.google.com/en//insidesearch/howsearchworks/assets/searchqualityevaluatorguidelines.pdf.
[197] Danny Sullivan, "Google Launches New Effort to Flag Upsetting or Offensive Content in Search," *Search Engine Land*, March 14, 2017, http://searchengineland.com/google-flag-upsetting-offensive-content-271119.

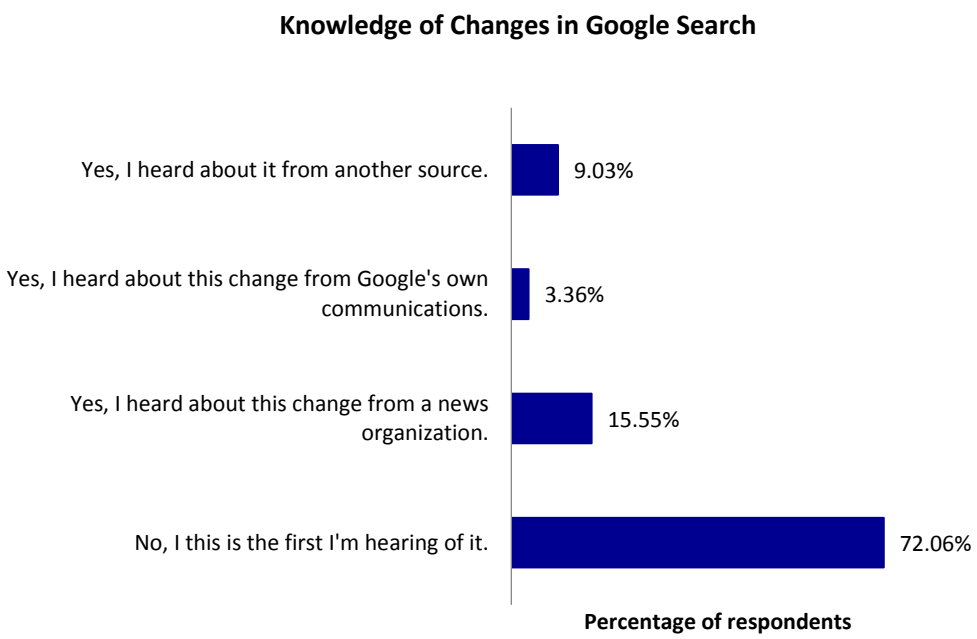


Figure 6. When informed about Google's changes to Search intended to curb fake news and hate speech, users reported being unaware of the changes; only a small minority reported being informed by Google itself of these changes.

Our interviewees, experts in this area, agreed with users in their approval for Google's changes. However, they also pointed out the need for complementary efforts.[198] These experts focused on the need to protect the First Amendment and avoid making Google the "arbiter of truth," instead suggesting that additional efforts be pursued both internally at the platform level, as collaborations between platforms and external groups, and purely externally.

Regarding such collaborative solutions, two of our interviewees, Takis Metaxas and Jonathan Zittrain, suggested that librarians could play a central role in the fact checking mechanisms. Zittrain additionally suggested that schools could be included, involving students in a multifaceted collaboration with platforms by engaging them in evaluating information on these platforms. The value of a critical civic education and critical thinking skills came up consistently in our interviews, and should not be dismissed. However, such solutions would likely be most effective in the longer than the short term.

Regarding complementary solutions outside the platforms, Robert Epstein and Dennis Allison referred us to their own efforts to create a civic organization to conduct research and advise on regulation, for instance by collecting and analyzing data from across the country in a manner modeled after the Nielsen Ratings system. Their proposed organization is called the Sunlight Society. Zittrain echoed this point, saying we need "a model of research that is collaborative and requires sharing of data early and often" between researchers and the platforms. He noted that in

[198] Metaxas, Keller, Sullivan.

the ideal scenario this effort might have some cooperation from platforms to aid researchers in collecting data and other information, though Epstein and Allison suggest, instead, that this data could be passively collected from a random sample of volunteers in order to decrease dependency on the platform.

Another relevant point that several respondents stressed in interviews was that opportunities for platforms to deal with the issue of misinformation, though important, are relatively limited. While many aspects of the issue are directly related to platforms, the underlying issues—distrust in journalism, political polarization, civic education, international interference in domestic elections, among others—are much bigger issues that do not fall cleanly under the purview of companies like Google.

V. The Impact of Google Search on Journalism

It is impossible to reflect on the phenomenon of fake news without reflecting on journalism. The fake news crisis has been predicated on distrust in traditional media sources. People are seeing news from traditional sources less frequently and, increasingly, question whether traditional news sources provide reliable information.[199] This was not a main focus on our research, as Madison Initiative has already scoped philanthropic interventions in this area.[200] Yet the subject of the current state of journalism came up repeatedly throughout our research, leading us to conclude that the subject is essential to this analysis.

Current John S. Knight Journalism Fellows at Stanford rehearsed with us the crisis in the advertising business model that has traditionally supported print and television journalism. In "The Platform Press: How Silicon Valley Reengineered Journalism," Emily Bell and Taylor Owen of Columbia University, agree, arguing: "The 'fake news' revelations of the 2016 election have forced social platforms to take greater responsibility for publishing decisions. However, this is a distraction from the larger issue that the structure and the economics of social platforms incentivize the spread of low-quality content over high-quality material. Journalism with high civic value—journalism that investigates power, or reaches underserved and local communities—is discriminated against by a system that favors scale and shareability."[201]

The Knight Fellows see the very function of social networking sites as promoting the diffusion of low-quality emotional content over higher quality content. Platforms benefit from so-called clickbait articles that entice users to interact and share those materials on the platform, which results in higher revenue advertising. In other words, short, controversial content engages more

[199] The DataSociety report highlights distrust as one of the major factors to why the media is vulnerable. See Marwick and Lewis, "Media Manipulation and Disinformation Online."
[200] Internal memo from the Madison Initiative.
[201] Emily Bell and Owen Taylor, "The Platform Press: How Silicon Valley Reengineered Journalism," *Tow Center for Digital Journalism Blog*, March 29, 2017, http://towcenter.org/research/the-platform-press-how-silicon-valley-reengineered-journalism/.

users than do longer, well-researched, analytical or investigative articles. The Knight journalists point out that the platform companies have little business incentive to diminish emotional content. Moreover, without sufficient demand, quality journalism will not maintain the necessary funding to survive.[202]

On the other hand, representatives from Google strongly refute this account. For one thing, Google claims that emotional content has always been an issue for journalism. A high-level Google official, for example, referred to the new market ecosystem as yet another in a string of adjustments to which journalism will adapt, as it has in the past. In fact, the official argues, the vast amount of data available provides opportunities for journalism to grow in new directions.[203] He pointed to The Trust Project, a Santa Clara University venture in collaboration with Google News, as an example of this emerging transformation. The Trust Project "crafts tangible digital strategies to fulfill journalism's basic pledge: to serve society with a truthful, intelligent and comprehensive account of ideas and event," seeking to "bake the evidence of trustworthy reporting—accuracy, transparency and inclusion—plainly into news practices, tools and platforms."[204]

This rings true in comments received from Google users in response to survey questions. As a final question, the survey asked respondents to share thoughts on the topic of trust in the media. Several comments underscore the decline in trust in the media. One respondent lamented: "BBC, CNN, NY Times - FOX news may be considered traditional, but they are the most egregious offenders when it comes to misleading news, they are absolutely not trustworthy!" Another criticized the integrity of mainstream media, musing, "Yeah, hmm, CNN, MSNC, CNBC, New York Times, Washington Post - these are FAKE NEWS media outlets. Better yet, I should say they are public relations agencies for the Democratic Party and the political left in general." Further, we received comments suggesting that news outlets are not effectively communicating their norms and standards. As one respondent put it: "Many liberal sites are publishing content from unverified anonymous sources as if they ARE news, thus creating fake news because it is more important to them to harm Trump than tell the truth." Another expressed concern about a liberal feedback loop: "Fact checking in and of itself is problematic because liberal sources [fact-]check liberal sources."

While it is not yet possible to definitively conclude whether platforms like Google are helping or hindering journalism, we see an opportunity for collaboration. Companies like Google are not only thriving in the current ecosystem, but also helping to create it. If they can be more fully engaged as strong allies with journalism, they can help to advance high quality news and promote the journalistic integrity necessary to a vibrant democracy.

---

202 This impact over the quality of journalism is seen as part of the changes in technology and the expansion of internet that negatively affected journalism by the Knight Fellows. The other major technological change has been the disappearance of local journalism. They refer to the places with no news coverage as "news deserts" where there are no daily local news outlets. News deserts leave large sectors of the U.S. with no local news and create a distance between news sources and people that is a perfect setting to cause distrust. For a graphic description of "news deserts," by *Columbia Journalism Review*, see: "America's Growing News Deserts," *Columbia Journalism Review*, accessed June 10, 2017, https://www.cjr.org/local_news/american-news-deserts-donuts-local.php.

[203] Interview with Google officials, May 12, 2017.

[204] Taken from "The Trust Project," http://thetrustproject.org/.

VI. Policy Options

The 2016 election serves as a case study for escalating concern with the proliferation of misinformation. This study focusing on the role of the platforms in that proliferation reveals opportunities for partnerships with journalism and civic institutions to enhance civic engagement through online forums. We present four forums where such opportunity may be most impactful in strengthening democratic institutions and governance: The platforms, civil society organizations, journalism, and government.

1) The Platform: Increase Transparency and Collaboration

Google should engage in research partnerships with universities and think tanks, sharing data collected from Google Search that pertains to democratic institutions. A first stage of this project should address two questions: (1) what data can and should Google share? And (2) with whom can these different data be shared with optimal effect on strengthening democratic institutions?

Fundamental to considering what data to share is the balance between the potential benefits of transparency with privacy and spamming. Privacy law limits the type of data than can be shared, barring, for instance, personally identifiable information from becoming public. In addition to this concern, Google is restrained in what it can make public about the workings of its algorithms, as spammers will take advantage of any openness to game the system more effectively for purposes ranging from ads and other forms of revenue to political manipulation.

A second stage of the project might encourage Google to share information more widely among research scholars and the public. Protocols could be set in place dictating which data is shared with the general public, which with academics, and, potentially, which with third-party auditors to analyze impartially and confidentially. These considerations could lead not only to greater insight into Google's effect on its users, but could also inform users themselves in innovative ways, much like the Google Trends tool does.

2) Civilian Oversight: Civil Society Organizations

In addition to lobbying for Google's explicit collaboration in these efforts, this report encourages further independent research collection and analysis. Existing groups, such as Stanford's Center for Internet and Society and Harvard's Berkman-Klein Center for Internet and Society, for example, have a long history of industry-parallel work. It is also worth considering novel proposals such as Robert Epstein's idea for a Nielsen-like system for collecting data from volunteers across the country to examine the workings of Google search independently of the company, or Takis Metaxas's and Jonathan Zittrain's suggestions to develop a corpus of fact-checked knowledge run by impartial, trusted individuals such as librarians, in which the discourse and debate around every check is documented and publicly visible.

3) Public Accountability

There is widespread agreement in the United States that the First Amendment should be upheld and that platforms should not act as arbiters of truth, deciding what people can and cannot post online. This forestalls efforts that might seek to force platforms to perform such functions. There may, however, be an opportunity to regulate the types of information that these companies must make public or allow to be audited by third parties. The amount of power that platforms gain in collecting data on their users should result in some level of responsibility and accountability in the public interest. Projects that enhance collaboration between journalists and Google and other platforms, with special attention to the value of local and public news outlets, may help bolster journalism and foster public trust in media.

4) Further Research

To better understand Google's impact on democratic institutions, this study recommends four further lines of research:

1. Research on specific elections tracking Google Search results with political outcomes. This could be done in different states or countries and at different political election levels, for instance, elections of local mayors or city councils, or state legislators, or U.S. representatives. Midterm elections could be a good time to execute these studies, allowing ample time to gather information before the elections and to track search results on specific issues or candidates. This line of research could also examine how people who use Google may change their political behavior in response to different kinds of interventions. The studies do not need to be restricted to Google and in fact could be more insightful if they include other platforms.
2. Studies on the psychological and civic impacts of Google search results. This line of study could address the following questions: What subconscious effects do Google results have on citizens' political views? Do citizens gain a more diverse view of politics by using Google? If so, under what conditions? What is the impact on deliberative democracy of new features that engage users in assessing the quality of content on the platform? What platform features produce more civic engagement in political debates of high quality and which ones of low quality? Which features cause or increase political polarization?
3. Further investigation of the legal and regulatory framework for search engines. This line of study could address such questions as: What is the regulatory framework for digital footprints? How is that market regulated? What alternative frameworks exist, as for example in other countries? How might these various regulatory models attach to intermediary liability in the spread of misinformation?
4. Surveys to provide not only demographics but also descriptive statistics on how different users engage with content on the platform, with an eye to misinformation.

VII. Conclusion

As evidenced by the explosion of interest in fake news in late 2016, the issues of misinformation, propaganda, and bias, and their propagation through online tools, are paramount. In our research, we examined a long "arms race" history of how search engines deal with spammers' and other bad actors' intentional efforts to mislead. Google's deep attention to these problems may partially explain why Google Search was not as hard hit by fake news gaming its algorithm as were social networking sites (SNS), and, thus, why it continues to maintain a high level of trust among users.

We also identified the misinformation risk on Google search, which can be subtler (e.g. a biased collection of search results) than on SNS, and influence users unconsciously. This search result risk for unconscious, peripheral persuasion warrants serious attention, since it can only be uncovered by systematic, aggregated analysis that considers the experiences of large swaths of users over time. Using the 2016 election as a case study, our research leads us to conclude that a precise estimate of the effect of fake news is out of reach without such broad data collection and analysis, but that misinformation did spread on Google in political web search results at that time.

We find that Google is taking steps to mitigate the issue of spreading fake news. Our data shows that although users are largely unaware of Google's efforts, once they are informed, they strongly support the company's efforts. Here we identify a tension: systematic, rigorous, third-party analysis is necessary to ensure protection of free speech, but Google has widespread consumer trust and may be less inclined to engage in changes that risk public misunderstanding or bad publicity. We encourage the company to develop infrastructure for rigorous review of search results across users and over time. Creating transparency and review out of a highly personalized, complex, black box algorithm is the next step towards fully understanding the ubiquitous technological environment in which we are all steeped, and its tremendous potential impact on our national civic beliefs and culture.

VIII. Bibliography

Allcott, Hunt, and Matthew Gentzkow. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives* 31, no. 2 (May 2017): 211–36. doi:10.1257/jep.31.2.211.

"America's Growing News Deserts." *Columbia Journalism Review*. Accessed June 10, 2017. https://www.cjr.org/local_news/american-news-deserts-donuts-local.php.

Bell, Emily, and Owen Taylor. "The Platform Press: How Silicon Valley Reengineered Journalism." *Tow Center for Digital Journalism Blog*, March 29, 2017. http://towcenter.org/research/the-platform-press-how-silicon-valley-reengineered-journalism/.

Cadwalladr, Carole. "Google, Democracy and the Truth about Internet Search." *The Guardian*, December 4, 2016, sec. Technology. https://www.theguardian.com/technology/2016/dec/04/google-democracy-truth-internet-search-facebook.

Epstein, Robert, and Ronald E. Robertson. "The Search Engine Manipulation Effect (SEME) and Its Possible Impact on the Outcomes of Elections." *Proceedings of the National Academy of Sciences* 112, no. 33 (August 18, 2015): E4512–21. doi:10.1073/pnas.1419828112.

Gillin, Joshua. "PolitiFact's Guide to Fake News Websites and What They Peddle." *PunditFact*, April 20, 2017. http://www.politifact.com/punditfact/article/2017/apr/20/politifacts-guide-fake-news-websites-and-what-they/.

Gomes, Ben. "Our Latest Quality Improvements for Search." *Official Google Blog*, April 25, 2017. http://blog.google:443/products/search/our-latest-quality-improvements-search/.

Google. "General Guidelines," May 11, 2017. https://static.googleusercontent.com/media/www.google.com/en//insidesearch/howsearchworks/assets/searchqualityevaluatorguidelines.pdf.

"Google: Distribution of Revenue by Source 2016." *Statista*, 2017. https://www.statista.com/statistics/266471/distribution-of-googles-revenues-by-source/.

"Google Search Statistics - Internet Live Stats." Accessed June 10, 2017. http://www.internetlivestats.com/google-search-statistics/.

Grimmelmann, James. "Speech Engines." *Minn. L. Rev.* 98 (2013): 868.

———. "The Google Dilemma." *NYL Sch. L. Rev.* 53 (2008): 939.

International Telecommunication Union. "Measuring the Information Society Report," 2016. http://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2016/MISR2016-w4.pdf.

Keller, Daphne. "Making Google the Censor." *The New York Times*, June 12, 2017, sec. Opinion. https://www.nytimes.com/2017/06/12/opinion/making-google-the-censor.html.

Lazer, David, Matthew Baum, Nir Grinberg, Friedland, Kenneth, Hobbs, and Mattsson. "Combating Fake News: An Agenda for Research and Action," May 2, 2017. https://shorensteincenter.org/combating-fake-news-agenda-for-research/.

Marvin, Ginny. "Google Isn't Actually Tackling 'Fake News' Content on Its Ad Network." *Marketing Land*, February 28, 2017. http://marketingland.com/google-fake-news-ad-network-revenues-207509.

Marwick, Alice, and Rebecca Lewis. "Media Manipulation and Disinformation Online." Data&Society, 2017. https://datasociety.net/output/media-manipulation-and-disinfo-online/.

Metaxas, Panagiotis Takis. "Web Spam, Social Propaganda and the Evolution of Search Engine Rankings." In *Web Information Systems and Technologies*, 170–82. Springer, Berlin, Heidelberg, 2009. doi:10.1007/978-3-642-12436-5_13.

Moore, Martin. "Tech Giants and Civic Power." Centre for the Study of Media, Communication and Power, April 2016. https://www.kcl.ac.uk/sspp/policy-institute/CMCP/Tech-Giants-and-Civic-Power.pdf.

Pan, Bing, Helene Hembrooke, Thorsten Joachims, Lori Lorigo, Geri Gay, and Laura Granka. "In Google We Trust: Users' Decisions on Rank, Position, and Relevance." *Journal of Computer-Mediated Communication* 12, no. 3 (April 1, 2007): 801–23. doi:10.1111/j.1083-6101.2007.00351.x.

Purcell, Kristen, Joanna Brenner, and Lee Rainie. "Search Engine Use 2012." Pew Research Center, March 9, 2012. http://www.pewinternet.org/2012/03/09/search-engine-use-2012/.

"SearchWiki: Make Search Your Own." *Official Google Blog*, November 20, 2008. https://googleblog.blogspot.com/2008/11/searchwiki-make-search-your-own.html.

Subramanian, Samanth. "Inside the Macedonian Fake-News Complex." *WIRED*, February 15, 2017. https://www.wired.com/2017/02/veles-macedonia-fake-news/.

Sullivan, Danny. "Google Launches New Effort to Flag Upsetting or Offensive Content in Search." *Search Engine Land*, March 14, 2017. http://searchengineland.com/google-flag-upsetting-offensive-content-271119.

———. "Google Now Handles at Least 2 Trillion Searches per Year." *Search Engine Land*, May 24, 2016. http://searchengineland.com/google-now-handles-2-999-trillion-searches-per-year-250247.

———. "Google's 'Project Owl' — a Three-Pronged Attack on Fake News & Problematic Content." *Search Engine Land*, April 25, 2017. http://searchengineland.com/googles-project-owl-attack-fake-news-273700.

"The Trust Project." Accessed June 10, 2017. http://thetrustproject.org/.

Volokh, Eugene, and Donald M. Falk. "Google: First Amendment Protection for Search Engine Search Results." *JL Econ. & Pol'y* 8 (2011): 883.

Zimdars, Melissa. "Resource-False-Misleading-Clickbait-Y-and-Satirical-'News'-Sources-1.pdf," November 2016. http://d279m997dpfwgl.cloudfront.net/wp/2016/11/Resource-False-Misleading-Clickbait-y-and-Satirical-%E2%80%9CNews%E2%80%9D-Sources-1.pdf.

IX. Appendices

Appendix 1: Methodology

Quantitative

We have access to data collected twice per week in the 26 weeks prior to the November 2016 U.S. national elections. The data include the top ten URLs returned when searching the names of all Senate and House representatives, as well as both of the presidential candidates. We also have access to lists of fake news sites from various sources, including Professor Melissa Zimdars and PolitiFact. As a first analysis step, we are using available lists of verified fake news URLs to check how many, if any, of the top ten results for these representatives are from sites suspected of spreading misinformation. The precise methodology here is to use computational methods to scan our data for any URLs matching the list of fake news sites.

This analysis, however, cannot allow us to definitively conclude that misinformation or fake news appeared with certainty in these search results, as some organizations publishing fake news may also publish legitimate information. In other words, the appearance of an article from www.breitbart.com in our data set does not guarantee that fake news was being spread, as some might argue that not all of Breitbart's articles are illegitimate. In order to examine this issue more closely, we have asked human annotators to annotate a subset of the data on dimensions including whether the content on that specific page perceived as credible, or is favorable or unfavorable to the candidate. These annotators are online crowd workers hired from Amazon Mechanical Turk and paid the equivalent of minimum wage to evaluate one URL at a time.

Qualitative

We interviewed experts in search engine manipulation, particularly with regard to its political implications, as well as scholars at the forefront of law, policy, and democracy studies:

- Dennis Allison, Lecturer in the Computer Systems Laboratory, Stanford University. Sunlight Society scholar studying the presence of search bias in algorithms
- Larry Diamond, Senior Fellow, Freeman Spogli Institute for International Studies and the Hoover Institution, Stanford University.
  Scholar of democratic institutions.
- Robert Epstein, American psychologist studying the search engine manipulation effect.
- High-level officials at Google who are at the forefront of the debates about Google's roles in news and technology. These officials requested anonymity in the public version of this report.
- Daphne Keller, Director of Intermediary Liability, Stanford University Center for Internet and Society; former Associate General Counsel for Intermediary Liability and Free Speech issues at Google.
- Takis Metaxas, Professor of Computer Science, Wellesley College.

Scholar studying crowdsourcing and social networks in the context of political events and news literacy.

- Danny Sullivan, Search Engine Land and Third Door Media
  Analyst, journalist, and expert on search engines studying changes in Google's search engine and algorithmic interventions for fake news and misinformation.
- Jonathan Zittrain, George Bemis Professor of International Law, Harvard Law School; Faculty Director, Berkman-Klein Center for Internet and Society
  Scholar in digital property and content and the roles of intermediary platforms.

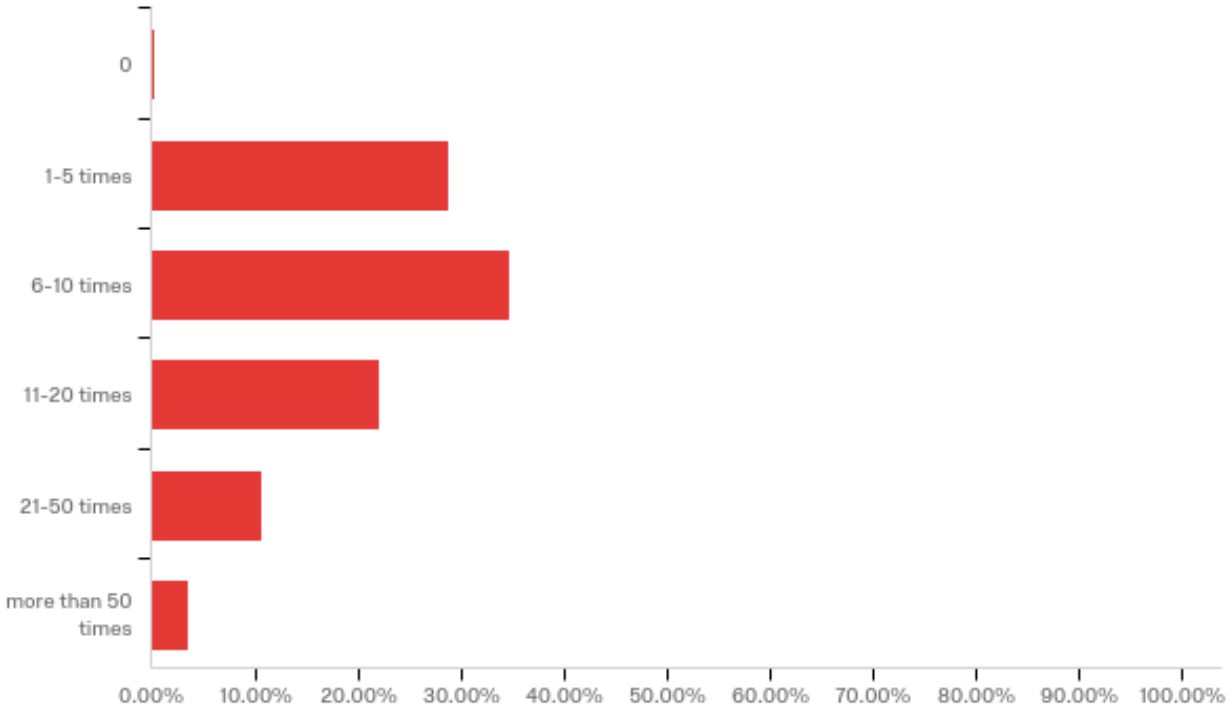
Appendix 2: Amazon Mechanical Turk Report: Google Survey

NOTE: DUE TO SMALL SAMPLE SIZE, THESE RESULTS ARE NOT REPRESENTATIVE OF THE U.S POPULATION.

Q2 - I have read and understand the above consent form, I certify that I am 18 years old or older and, by clicking the submit button to enter the survey, I indicate my willingness to voluntarily take part in the study.

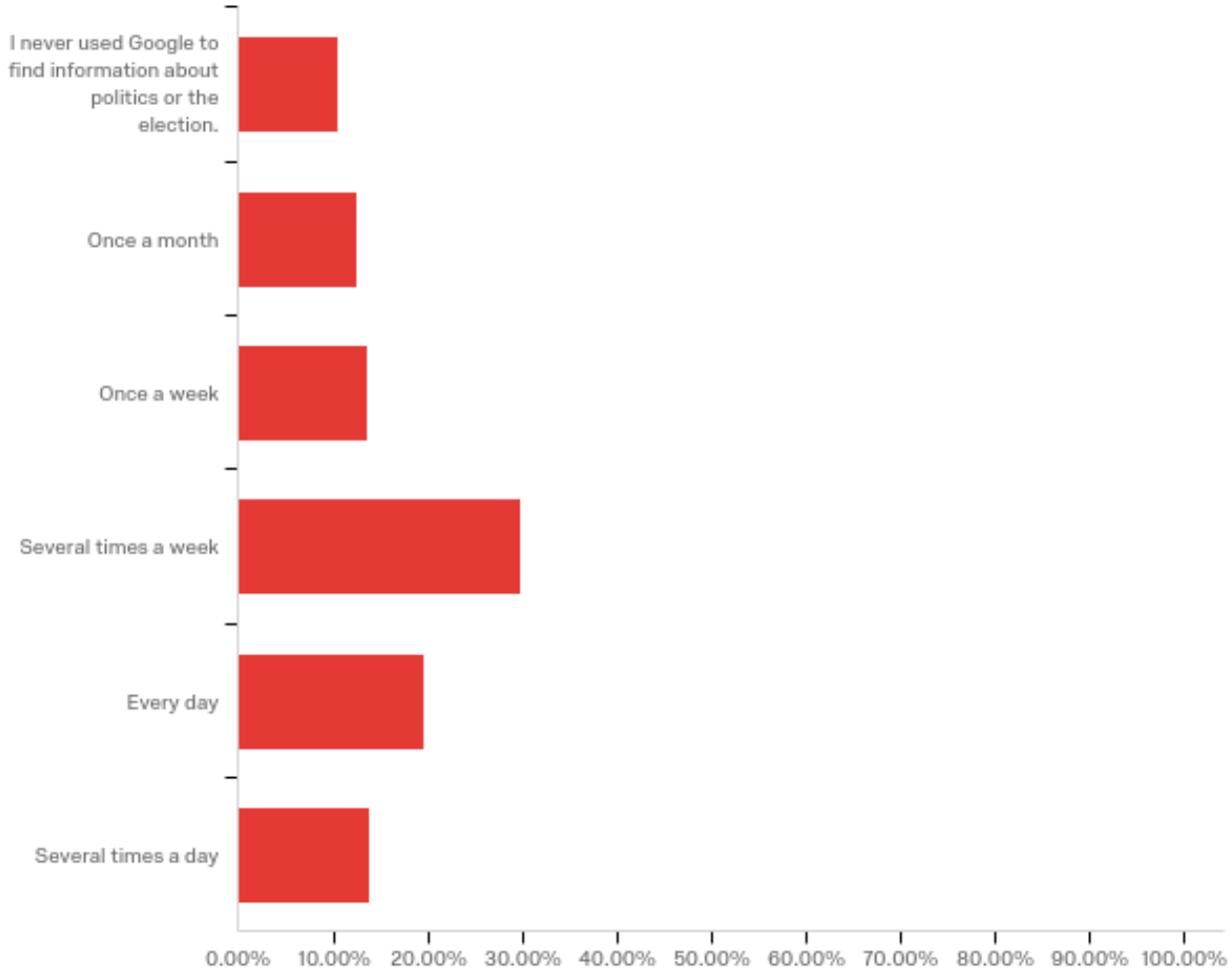
#	I have read and understand the above consent form, I certify that I am 18 y...	Percentage
1	Yes, I want to participate.	100.00%
2	No, I don't want to participate.	0.00%
	Total	475

G1 - How many times per day, on average, do you use Google to search for information?



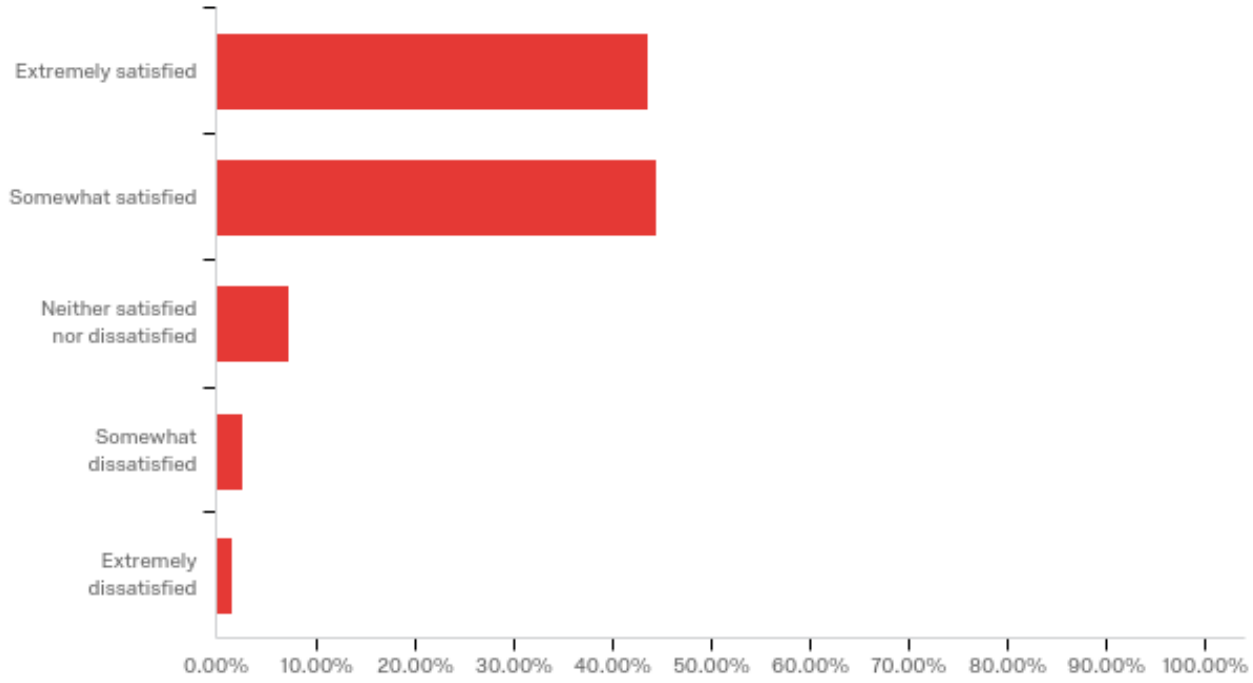
#	How many times per day, on average, do you use Google to search for informa...	Percentage
1	0	0.21%
2	1-5 times	28.78%
3	6-10 times	34.66%
4	11-20 times	22.06%
5	21-50 times	10.71%
6	more than 50 times	3.57%
	Total	476

G2 - Around the time of the recent 2016 presidential election, approximately how often did you use Google to find information about politics or the election?



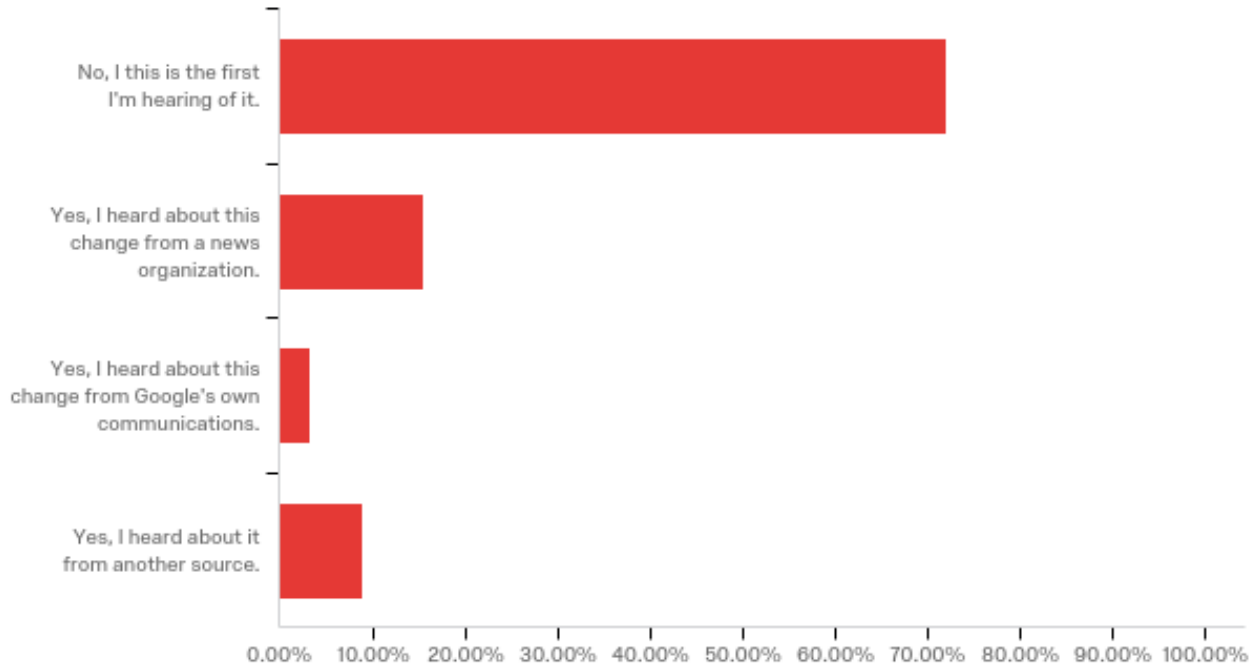
#	Around the time of the recent 2016 presidential election, approximately how...	Percentage
1	I never used Google to find information about politics or the election.	10.50%
2	Once a month	12.61%
3	Once a week	13.66%
4	Several times a week	29.83%
5	Every day	19.54%
6	Several times a day	13.87%
	Total	476

G3 - In general, how satisfied are you with the information you find using Google search?



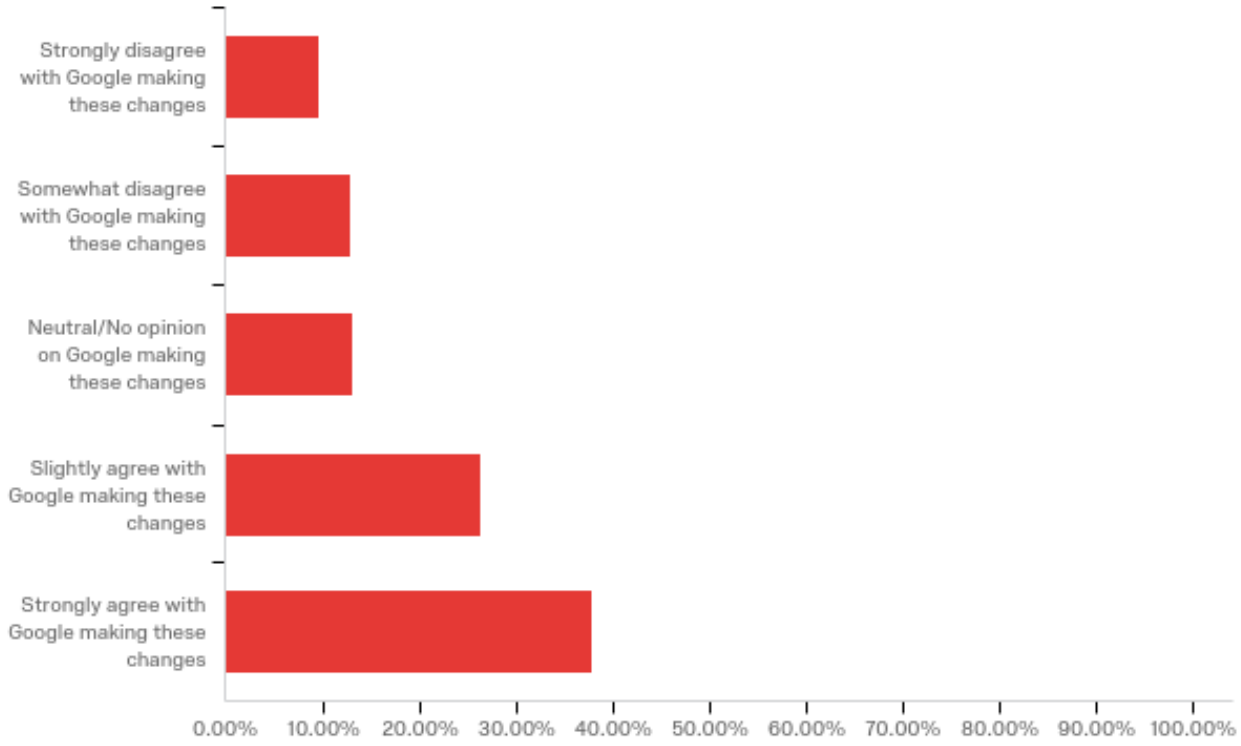
#	In general, how satisfied are you with the information you find using Googl...	Percentage
2	Somewhat satisfied	44.54%
4	Somewhat dissatisfied	2.73%
3	Neither satisfied nor dissatisfied	7.35%
1	Extremely satisfied	43.70%
5	Extremely dissatisfied	1.68%
	Total	476

G4 - Google recently announced a couple changes to Google Search intended to curb fake news and hate speech. Have you heard about this before?



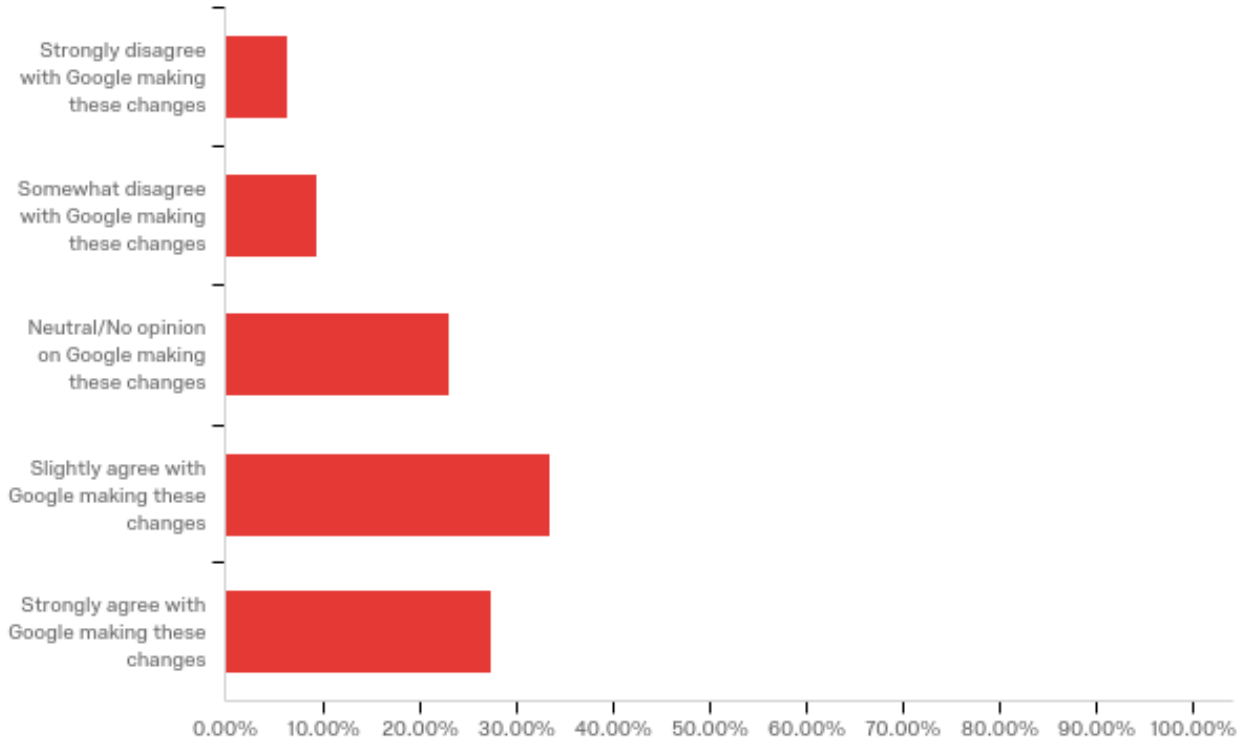
#	Google recently announced a couple changes to Google Search intended to cur...	Percentage
1	No, I this is the first I'm hearing of it.	72.06%
2	Yes, I heard about this change from a news organization.	15.55%
3	Yes, I heard about this change from Google's own communications.	3.36%
4	Yes, I heard about it from another source.	9.03%
	Total	476

G5 - One of Google's proposed changes is to the search algorithm, which determines what results users see, and in what order. The changes include more aggressive demoting or hiding of any information that Google considers "blatantly misleading, low-quality, offensive, or downright false information" from search results. What is your opinion on these changes?



#	One of Google's proposed changes is to the search algorithm, which determin...	Percentage
1	Strongly disagree with Google making these changes	9.66%
2	Somewhat disagree with Google making these changes	12.82%
3	Neutral/No opinion on Google making these changes	13.24%
4	Slightly agree with Google making these changes	26.47%
5	Strongly agree with Google making these changes	37.82%
	Total	476

G7 - What is your opinion on these changes to the autocomplete feature?

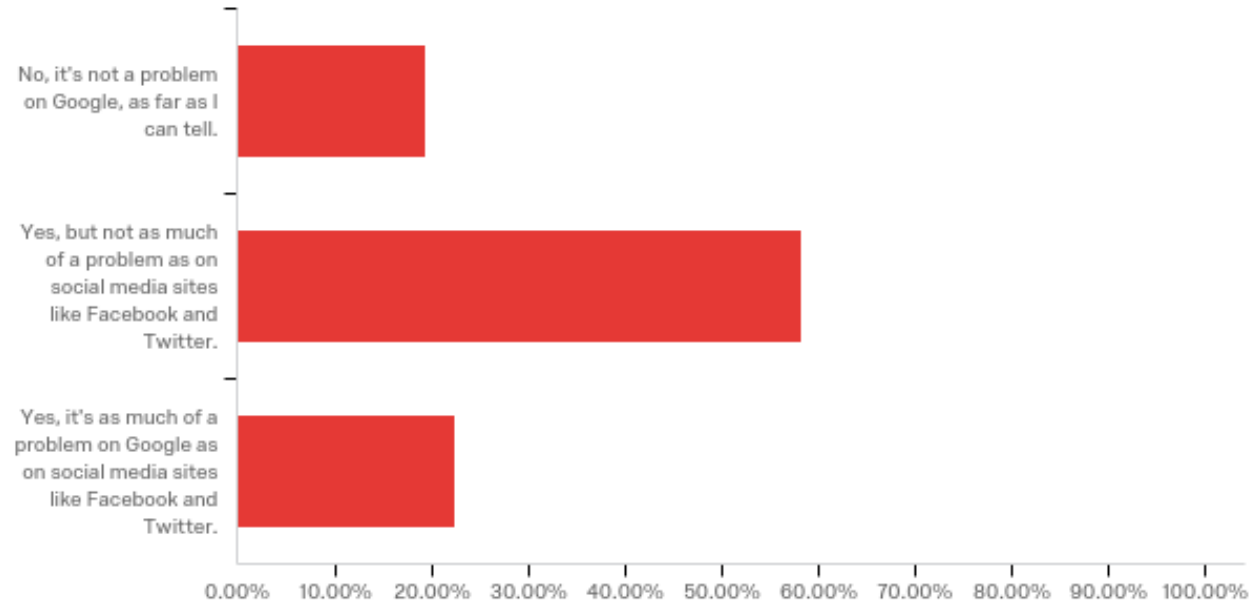


#	What is your opinion on these changes to the autocomplete feature?	Percentage
1	Strongly disagree with Google making these changes	6.30%
2	Somewhat disagree with Google making these changes	9.45%
3	Neutral/No opinion on Google making these changes	23.11%
4	Slightly agree with Google making these changes	33.61%
5	Strongly agree with Google making these changes	27.52%
	Total	476

G8 - In the TWO MONTHS leading up to the recent 2016 presidential election, some Google users reported seeing fake news ("blatantly misleading, low-quality, offensive, or downright false information") in their search results. This question is to check that you are reading carefully. Please ignore the question and answer the word "maybe" as your answer. Thank you for reading carefully. How many times do you remember seeing fake news in Google search results during that time?

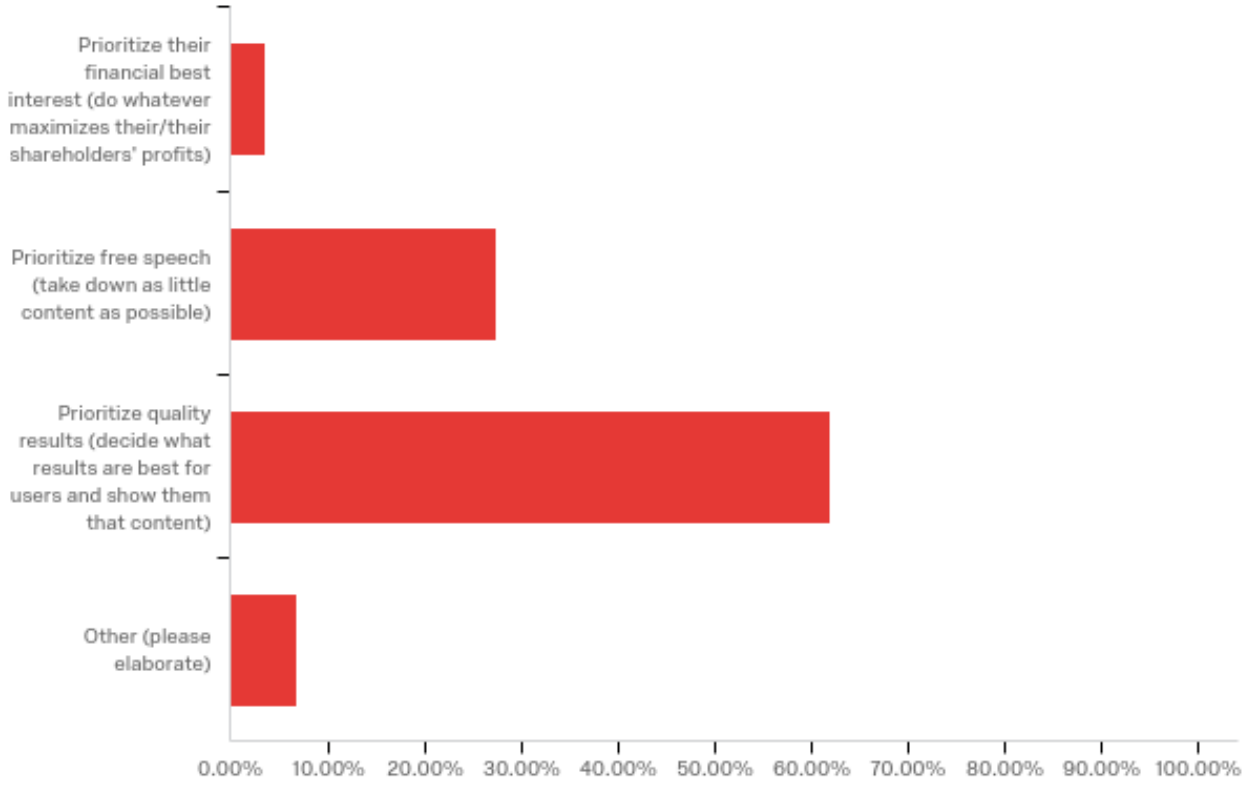
Only those who replied "maybe" are presented in this appendix.

G9 - In your opinion, is "fake news" ("blatantly misleading, low-quality, offensive, or downright false information") a problem on Google, relative to other sites like Facebook and Twitter?



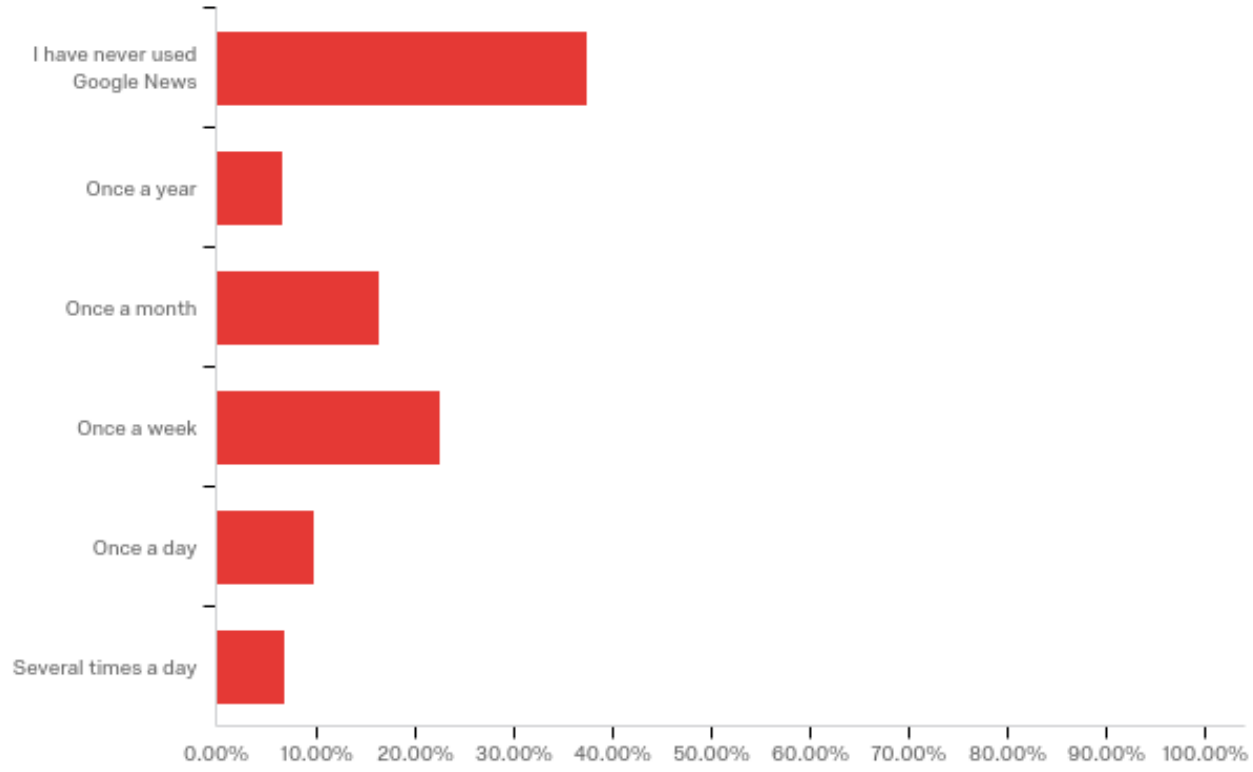
#	In your opinion, is "fake news" ("blatantly misleading, low-quality, offens...	Percentage
1	No, it's not a problem on Google, as far as I can tell.	19.33%
2	Yes, but not as much of a problem as on social media sites like Facebook and Twitter.	58.19%
3	Yes, it's as much of a problem on Google as on social media sites like Facebook and Twitter.	22.48%
	Total	476

G10 - Which of the following do you think should be Google's TOP priority when responding to fake news?



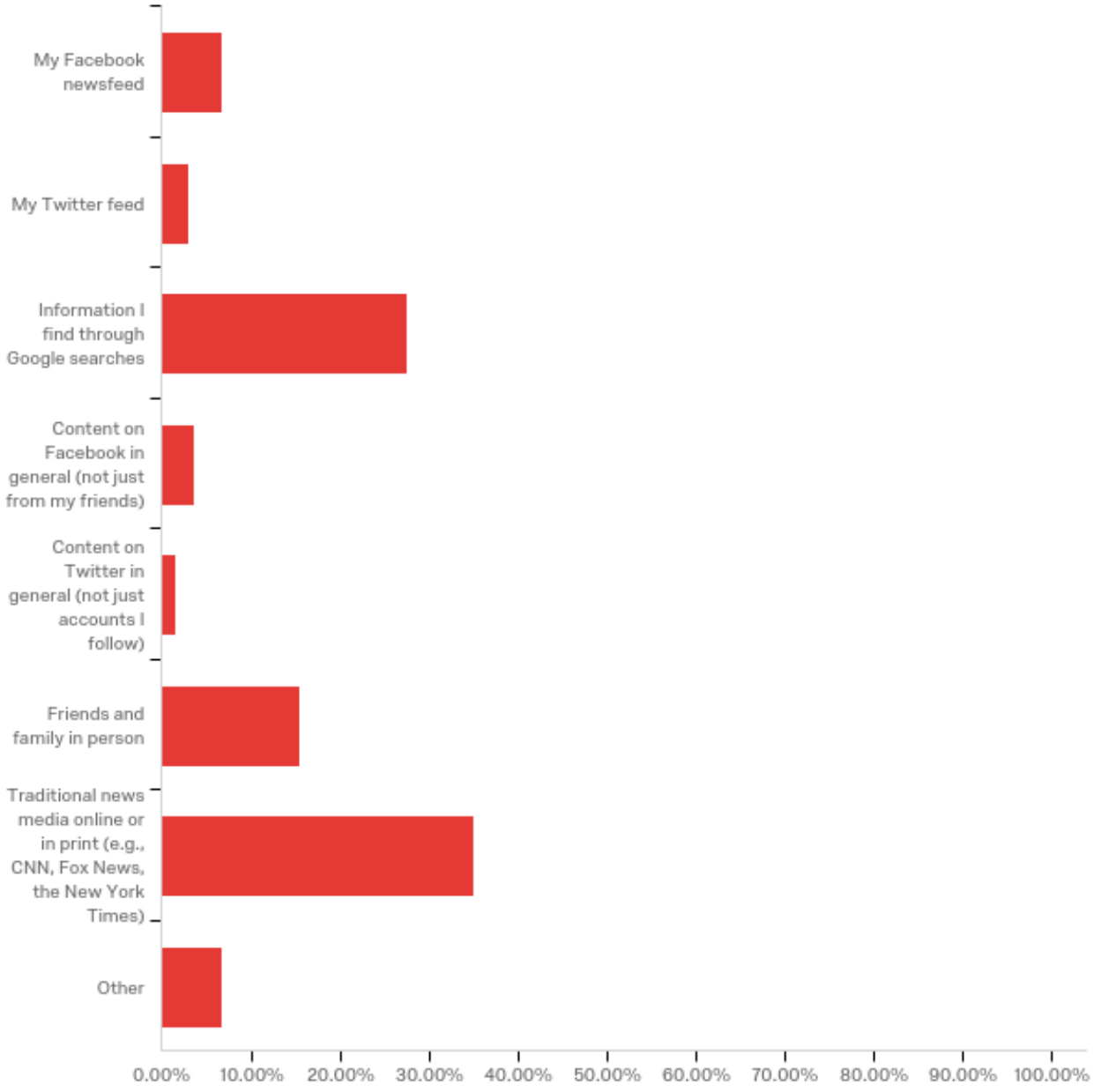
#	Which of the following do you think should be Google's TOP priority when re...	Percentage
1	Prioritize their financial best interest (do whatever maximizes their/their shareholders' profits)	3.57%
2	Prioritize free speech (take down as little content as possible)	27.52%
3	Prioritize quality results (decide what results are best for users and show them that content)	61.97%
4	Other (please elaborate)	6.93%
	Total	476

G12 - Do you ever view your results using Google News? If so, approximately how often?



#	Do you ever view your results using Google News? If so, approximately how o...	Percentage
7	I have never used Google News	37.39%
8	Once a year	6.72%
9	Once a month	16.39%
10	Once a week	22.69%
11	Once a day	9.87%
12	Several times a day	6.93%
	Total	476

G13 - Which of the following do you consider to be trustworthy sources of information about news and politics? Please select all that apply.



#	Which of the following do you consider to be trustworthy sources of informa...	Percentage
1	My Facebook newsfeed	6.78%
2	My Twitter feed	3.00%

3	Information I find through Google searches	27.49%
4	Content on Facebook in general (not just from my friends)	3.68%
5	Content on Twitter in general (not just accounts I follow)	1.65%
6	Friends and family in person	15.49%
7	Traditional news media online or in print (e.g., CNN, Fox News, the New York Times)	35.14%
8	Other	6.78%
	Total	1033

Section 4. Twitter

I. Introduction and Overview

Twitter offers an unprecedented opportunity for users to interact and connect online across the globe. Yet it simultaneously presents a series of unique challenges, complicated by its economic sustainability, that make its influence on the misinformation problem particularly pertinent and challenging. This analysis identifies three key aspects that make Twitter unique in the spread of misinformation:

1. The layout of Twitter's user interface and its 140 character limit render it especially susceptible to false or incomplete information;[205]
2. Twitter is resistant to modifying its platform or algorithms, which other platforms have done, to stem the spread of misinformation, citing its commitment to First Amendment principles; and
3. Twitter has not fully implemented key tools for verification, thus making it difficult for users to identify credible sources and build relationships of trust on the platform.

In seeking to resolve these principal problems, this memo, rather than offering a blanket definition of the content that Twitter defines as misinformation,[206] instead focuses on how misinformation spreads on the platform, and possible interventions. This Section 4 of the full report uses the terms "unverified news" or "misinformation" to describe content that is impermissible on Twitter because it violates the terms of service or falls outside of legal protections for free speech and intermediary platforms.

This analysis of the spread of misinformation on Twitter uses three methods: case studies, user experiments, and research interviews. Our case studies analyze responses across the social network to certain inciting events on Twitter, revealing the platform's dynamics of information sharing. Our user experiments revealed possible improvements to Twitter that can foster trust and enhance online discourse. Our research interviews with key journalists and news editors, high level company decision makers, legislative staff and policymakers, and research scholars expanded the scope of our research to address other – sometimes competing – perspectives on the impact of Twitter in addressing the phenomenon of misinformation.

From a legal perspective, Twitter gains significant protection from the provisions of the Communication Decency Act, which limit the liability of online sites for the content they host and

[205] The company is now testing a new limit of 280 characters.

[206] Although this section uses the term "fake news", the research team recognizes that:

- The term "fake news" has become highly politicized and partisan, which can derail genuine attempts at constructive discourse on the subject.
- The term is overused and hard to define, which could cause skeptical audiences to dismiss the substance of our conclusions.
- Popular use of the term "fake news" implies that the phenomenon of misinformation is unique to our era. Our research shows that misinformation is not unique to, or uniquely virulent on, any individual news media. Misinformation rapidly spread after the introduction of the newspaper as a means of mass communication in the nineteenth century; after the introduction of radio and television, McCarthyism's brand of "fake news" and defamation ran rampant in 1950's America. We emphasize here that we are not proceeding as if misinformation is a phenomenon that occurs only on social media or the internet, or one that demands a particular exception to existing Free Speech Doctrine or societal norms regarding free speech. Rather, we believe that approaching the problem of "misinformation" as simply the latest iteration of a centuries-long trend enables us more accurately to understand the challenges of fighting misinformation, while still respecting America's fundamental, unmistakable protections of Free Speech.

promote freedom of speech. Even tragic cases fall outside Twitter's sphere of liability, such as those where individuals use Twitter to facilitate heinous acts (including ISIS attacks).

Platform Analysis

General Trends

1. **Twitter invites and publicizes adversarial interactions and debates between users.** Users tend to use Twitter as a platform for engaging in and viewing debates between different users.
    a. As a result, misinformation tends to abate more quickly than on Facebook or Reddit because Twitter's user experience ("UX") is not as insulated against competing opinions and ideas. Other aspects of the user interface ("UI") described below, however, counter this finding.
2. **Facebook is more susceptible to spreading misinformation than Twitter**. Facebook users interact less with those with whom they disagree. Rather than inviting debate, Facebook is a platform that users primarily use to share, view, and otherwise consume content that is oriented around friend groups with compatible world views. Yet it is not the case that Twitter reduces the spread of misinformation.

User Interface

1. Even though Twitter's UI can obfuscate the authenticity and veracity of content, **some of Twitter's UI features succeed in facilitating accountability and trust.** Most importantly, tweets are public, so that users can be held accountable by the 'Twittersphere' at large for incorrect content.
2. **Public signifiers of authenticity play a pivotal role in the trust that users accord to content.** On Twitter, the key signifier is the **blue verified badge.** This badge option is currently open only to users whom Twitter judges to be "in the public interest," even when that "interest" is potentially negative.
3. **Users rely on the presence of citations to judge the veracity of content on Twitter.** In particular our research reveals that:
    a. Users regard content that provides citations as more trustworthy than content that does not provide citations.
    b. Users greatly differ on the credibility they ascribe to different sources. This often reflects partisan leanings.
    c. This is not a new phenomenon. The inherent nature of Twitter's platform, however, has obfuscated many of the ways users are able to discern content credibility. For instance, it is difficult to discover the original source of different news content due to the nature of "retweeting"—social sharing—on Twitter.
4. **Twitter reinforces information amplification by speeding up the spread of information sharing.** This leads to the immediate amplification of certain storylines, with fewer opportunities for verification. Twitter's UI additionally exposes users to fewer signals that could enable them to make accurate split-second judgments on content veracity. This counter-balances the finding that Twitter's adversarial and debate-oriented UI diminishes the spread of misinformation.

User Incentives

1. **Twitter positively reinforces incentives to sensationalize content.** The speed of sharing encourages users to gravitate towards attention-grabbing content, incentivizing other actors to model their content on these user preferences.
2. **Fake accounts that join Twitter conversations can hijack those threads, spreading massive influxes of content with relative ease.** While Twitter's public-facing platform helps hold users accountable for content they post, conversely, it also makes Twitter conversations susceptible to bots and cyborgs.
    - Cyborgs and bots manipulate the signals that users look for when discerning content credibility, a problem enhanced by the UI, which strips out origination links and features that help users verify content.
    - Although this needs study and verification, a preponderance of Twitter's "bots" may serve legitimate purposes.

Actionable Recommendations for Twitter

1. **Permit all users to apply for the blue verified badge for their accounts.** Verification may enhance basic rules of conduct, including individual tweeters' responsibility for the content they share.
2. **Create a 'false news' option that enables users to report false content.** To prevent overuse and contamination by partisan motive, there are several options worth exploring for implementing a due process, explored in the "Recommendations" section.
3. **Implement stricter Captcha gateways for user registration and when suspicious bot activity is detected.**
4. **Create more options for reacting to tweets, beyond 'favorite' and 'retweet'.**
5. **Prominently display the original source of content that is repeatedly retweeted and shared between different conversations.**
6. **Refine automatic detection tools for spam and other suspicious account activity.**

Media Analysis

General Trends

1. **Journalists actively optimize their news content for maximum social media traction.** Headlines are written in consultation with data analytics that help predict which keywords will facilitate greatest social sharing, and journalists are now being trained in social media strategy.
2. **Journalists use Twitter to find and publicize breaking news and Facebook to post longer, more in-depth pieces for social news sharing.** While individual journalists use Twitter to find breaking news and then "get the word out" for their own breaking stories, newsrooms tend to emphasize tailoring content for Facebook, for the purposes of advertising revenues and increasing user attention and clickthrough rates.

Increased Competition for Advertising Revenues

1. While the Trump administration has proved a boon to news outlets' bottom lines, **news outlets are struggling to compete with online platforms for advertising revenues**. Local news outlets are particularly struggling, forcing an unprecedented wave of local news closures and consolidation.
2. **News outlets view Facebook and Google as the primary and most aggressive players in siphoning away advertising revenues from news organizations.** Twitter and Reddit occupy very minor market positions in the advertising market.
3. **News organizations accept this new reality of competition for revenues and are trying to adapt to new online audiences and patterns of sharing.** Some journalists believe that superior journalistic quality will ultimately help them triumph in the market and draw revenues.
4. **Large news aggregators pose the most significant threat to traditional journalism's profit margins.** They enjoy three advantages in the marketplace:
    - **a.** They have control over access to end users. This allows them to set the terms when creating contracts with news providers that they host.
    - **b.** They enjoy the benefits of network effects on two-sided platforms to a greater extent than do traditional news organizations. News providers rush to join the aggregators that have access to the greatest number of users, and users prefer to join the aggregators that host the greatest number of news providers.
    - c. Their business model is centered on news distribution. They have the resources to tailor their content to users that traditional news organizations do not. This has tended to scale much more rapidly than does traditional news in gaining and retaining users.
5. **News organizations that traffic in misinformation are better equipped to out-compete traditional journalism for both revenue and audiences, especially on social media platforms like Twitter.** These organizations, ranging from clickbait sites to sophisticated peddlers of false news, have lower costs that reflect the absence of business overhead, including reporters; they also have higher profits that benefit from sensationalized content designed to attract user attention and social media virulence, and hence reap more advertising dollars. These organizations have proven adept at 'gaming' social media platforms for maximum exposure to end users.

Bolstering Local News and Investigative Reporting

1. One prominent suggestion among stakeholders is to **encourage investigative journalism, especially at the local level.** According to stakeholders connected with local news reporting, U.S. audiences are interested in investigative reporting about their local communities, and some in the news industry believe that harder-hitting reporting on the local level will foster deliberative, serious dialogue, re-

engage citizens in their local communities, and drive up subscriptions and advertiser interest in local news, shoring up bottom lines.

2. **This is an area where nonprofit and civil society organizations might help support research to aid local news organizations in developing a more competitive business model.** Investigative journalism's primary constraint is its higher built-in costs. With the shift in advertising revenues from content providers to the platforms hosting that content, news organizations are less able to afford such costs.

Alternatively, there may be a role for civil society organizations to help leverage online platforms to support citizen-based local investigative journalism.

Developing Platform Transparency for Improved Journalism

1. **Twitter should share its relevant internal data with outside researchers who study network dynamics and user behavior.** If the company does not open its data independently, then investigate the possible use of regulatory levers.
2. **Work to develop public-private partnerships on combating misinformation on social media.**
3. **Encourage news organizations to be more public in disclosing details of their editing and review process.**
4. **Invest in grant funding to local news organizations, emphasizing local investigative journalism.**
5. **Provide consultation to and organize training opportunities for smaller-scale news organizations on social media strategy.** Many local news organizations may simply not have the requisite tools to develop effective online presences for modern audiences. Alternatively, **work to persuade Twitter to work directly with journalists to help them spread news content on Twitter.** This could be developed into a platform, such as 'Twitter for Journalists', aimed at journalists using the platform.

Future Research Areas

1. **Conduct larger-scale user experiments on Twitter to document how users interact with fake news.**
2. **Conduct an industry analysis of online advertising business models, including monetizing fake news stories and revenue shifts in journalism**.
3. **Conduct an industry analysis of the news industry, with an emphasis on local news organizations.**
4. **Conduct quantitative analyses of misinformation campaigns on Twitter and other major platforms, including the scope and impact of Russian and foreign state-sponsored campaigns, and other sources engaged in proliferating misinformation**.
5. **Conduct a comprehensive quantitative study of the likely impact of various forms of 'false news' on user behavior and perceptions.**

Next Steps

1. **Facilitate collaboration between social media researchers, journalism researchers, and national security experts to better explore solutions for misinformation**. A model of this collaboration could potentially be Jigsaw.
2. **Analyze the impact on news consumers of local investigative journalism that effectively leverages social media platforms**, including citizen journalism.
3. **Promote civic engagement by encouraging lawmakers, policymakers, and government entities directly to engage their communities through Twitter.**
4. **Encourage Twitter to develop user features that facilitate users' access to accurate information on policies and politics.**
5. **Gain expertise on the technical aspects of 'fake news'** through continued academic partnerships.

Recent political events and the current political climate have generated interest among social media platforms and many civil society organizations to develop effective interventions. Through pressure and persuasion, Twitter may be encouraged to develop interventions that slow or forestall the spread of misinformation and foster public access to accurate information.

II. Demographics

Focusing on the spread of misinformation, this memo describes the characteristics of Twitter's platform structure, and the ways that users behave and respond to content posted on it, with attention to distinctions between Twitter and other online platforms. A brief summary of the demographics of Twitter users frames analysis of the characteristics unique to the platform. Understanding Twitter's user demographics is critical to understanding Twitter's current capacity to address misinformation and the potential for future policies to address misinformation on the Twitter platform. For instance, approximately 80% of Twitter users are located outside of the United States.[207] This suggests that regulatory structures of other countries will inherently impact the way Twitter functions, and that U.S. regulation may not be the most effective at implementing change in the Twitter platform. Additionally, users on Twitter skew young, wealthy, and tend to be college-educated. This suggests that any misinformation spread on Twitter is unlikely to be due to a lack of education or resources, and is more likely the result of users' over-reliance on the Internet for information (due to the youthful nature of users) and other psychological factors such as confirmation bias.

24% of online adults (21% of all Americans) use Twitter

% of online adults who use Twitter

All online adults	24%
Men	24
Women	25
18-29	36
30-49	23
50-64	21
65+	10
High school degree or less	20
Some college	25
College+	29
Less than $30K/year	23
$30K-$49,999	18
$50K-$74,999	28
$75,000+	30
Urban	26
Suburban	24
Rural	24

Note: Race/ethnicity breaks not shown due to sample size.
Source: Survey conducted March 7-April 4, 2016.
"Social Media Update 2016"
PEW RESEARCH CENTER

Figure 1: Demographics of Twitter

III. Misinformation Issues Inherent to Twitter's Platform

Twitter's platform structure is optimized for brevity, spontaneity, and easy access for users to online conversations. This structure makes it possible for bad actors to produce misleading content that can be shared with large online audiences. This analysis of Twitter's

[207] Interview with high-level official at Twitter, Apr. 28, 2017.

platform highlights three specific issues that arise due to how Twitter, as a platform, restructures the way by which information is shared and valued:

Increased Speed

Scholars have distinguished this newest generation of "misinformation" from previous shifts in how society consumes content by noting that Twitter and other online platforms have overwhelmingly reduced the temporal barriers to the spread of information. There are three specific impacts of this trend on modern information gathering and knowledge:

A. Decreased Opportunities for Verification

Content is now able to reach vast audiences on Twitter before third-party fact-checking organizations and other third-party observers are able to evaluate its veracity or tendency to mislead.

B. News Organizations and Individuals Have Actively Gravitated Towards Sensationalist Content to Gain Audiences

Because information consumption is now predicated more than ever on the speed with which it reaches consumers, many media houses and producers of content have consciously gravitated towards content that can spread faster throughout online user bases.

C. Decreased Importance of Speaker Reputation in User Trust of Content

People tend to evaluate truth based on their trust in the speaker. The speed of information transmission, however, has reduced users' abilities to discern the credibility of speakers. Especially when an online conversation is emergent, and different users are entering the conversation and spreading information quickly, users are less likely to base their reposting behavior on the perceived credibility of the content producer.

"Retweet"/"Favorite" System as a Means of Information Amplification

Users often report that one of the primary characteristics they use to judge a given tweet's truth value or credibility is its number of retweets/favorites. This is a proxy for the tweet's actual credibility. There are two issues inherent in this structure:

A. Positive Reinforcement of Information Amplification

Users rely on the number of retweets/favorites as a criterion for judging the credibility of particular tweets. That is then directly correlated with users' likelihood of retweeting or

otherwise sharing that content. This means that online conversations may rapidly amplify certain storylines that are able to initially gain traction among a critical mass of users.

B. Advantages for Attention-Grabbing or Sensational Content

In some circumstances, users have relied on the number of retweets/favorites as a more authoritative indicator of a tweet's veracity than the tweet's substantive characteristics or truth claims. Sensational content that attracts large numbers of retweets due to its inherent novelty thus possesses a structural advantage in this reconfigured marketplace.

Bots

The preponderance of bots and cyborgs on Twitter is a well-documented phenomenon.[208] We note two additional means, apart from intentional misinformation, by which bad actors can use bots to maliciously influence online opinion:

A. Malicious Trolling

There are documented incidents where organizations have assembled automated armies of Twitter bots to bombastically harass certain tweeters, in an effort to intimidate or silence them. Journalists are frequently the target of these concerted attacks.

B. Digital Insurgencies

This is a characteristic of state actors, who can scrape the characteristics of existing users, replicate those characteristics in new, engineered accounts, and then use that veneer of authenticity to infiltrate online conversations and distort opinions.

Finally, Twitter's administration tends towards nonintervention and their policies do not align with those of Facebook. Twitter's culture is rooted in relatively libertarian conceptions of free speech, which has implications for cross-platform solutions. Twitter will not agree to changes that its administration perceives as restricting free speech even when some other platforms may be adopting such changes to help manage the spread of false information.

[208]Twitter asserts that it takes down nearly 3.2 million suspicious accounts weekly. See, Twitter Blog, Sept. 28, 2017, https://blog.twitter.com/official/en_us/topics/company/2017/Update-Russian-Interference-in-2016--Election-Bots-and-Misinformation.html; see also, "Nearly 48 Million Twitter Accounts Could be Bots," CNBC, https://www.cnbc.com/2017/03/10/nearly-48-million-twitter-accounts-could-be-bots-says-study.htm. Nick Pickles, Senior Public Policy Manager for Twitter, emphasized that the company is improving monitoring bots and cyborgs (human users, usually operating with pseudonyms, who deploy bots to accelerate and prolong their posts) and taking down those that violate its terms of service. (Pickles, "Digital Platforms and Democratic Responsibility," Global Digital Policy Incubator Symposium, Stanford CDDRL, October 6, 2017.)

IV. Research Study Questions

This research addresses four questions, which are framed through case studies of the French Election and the Pope endorsing Trump, a Mechanical Turk survey of users, and interviews with scholars, industry experts, policymakers, Twitter officials, and journalists.

- What types of misinformation predominate on Twitter?
- How is this information transmitted and amplified on Twitter?
- Do the modes of disseminating misinformation differ from modes for regular news on Twitter?
- How can we clarify and resolve the problem as it exists today on Twitter?

V. Platform Analysis

Quantitative – Amazon Mechanical Turk Survey

The Mechanical Turk survey provided information through a fine-grain user experiment on how certain elements of Twitter's user interface affect user trust in content. These factors may be determinative in how users assess content and accord plausibility and credibility to different news items.

Survey Set-Up

The four research teams collaborated to develop an Amazon Turk Survey with four parts, each targeted to users of the particular platforms. The survey queried users about their demographics and their use of the platforms. Those who indicated that they use Twitter were then asked about their habits on the platform (for an in-depth look at the survey, see **Appendix A**).

The questions designed for this section probe what makes misinformation appear truthful on Twitter. The questions reflect the hypothesis that the way in which information is presented on Twitter plays a major role in whether individuals believe the information. The study focuses on the following six categories:

1. Confidence of the original Tweeter
2. Influential retweeters
3. Capitalization/spelling/grammar
4. Source of information
5. Citation
6. Verified Symbol

To assess these categories, we developed a series of sample tweets (see example in ***Figure 2*)**. After reading the particular tweet, the user is then asked to respond to the following questions:

On a scale of 1 to 5 how would you rate the truthfulness of this post?

1) True

2) Possibly True

3) Unsure

4) Possibly False

5) False

On a scale of 1 to 5 how likely would you be to hit like on this tweet?

1) I would like it

2) I probably would like it.

3) Unsure

4) I probably would not like it.

5) I would not like it.

On a scale of 1 to 5 how likely would you be to retweet this tweet?

1) I would retweet it

2) I probably would retweet it.

3) Unsure

4) I probably would not retweet it.

5) I would not retweet it.

The only difference between the two tweets is that the first has an abundance of favorites and retweets while the second does not. The hypothesis is that the tweet with more favorites and retweets will be considered more truthful than a tweet that has fewer retweets and favorites.

Figure 2: Sample survey tweets

Respondents see only one of the tweets

Survey Results

Over the course of a few weeks, 202 people took the Amazon Mechanical Turk survey (see Appendix B). The study revealed a strong correlation between the amount of trust individuals have in a given tweet if that tweet is accompanied by the Twitter verification symbol. This suggests that users perceive the blue verification symbol not just as a symbol authenticating a particular tweeter's identity—and often indicating celebrity status because, in Twitter's terms, the tweeter is "of public interest"—but as a marker indicating that the verified tweeter is providing truthful information. This perception among users likely attaches to the phenomenon of personal credibility:

Figure 3. The blue verification badge currently serves to authenticate prominent users' accounts

When a person's reputation is on the line, they are typically more likely to want to share information that reflects well on their character. For instance, last year, Stephen Colbert came under criticism for a racist tweet from a Twitter account (@ColbertReport). People assumed that the account was managed by Colbert himself, and held him accountable for the tweet urging the network to #CancelColbert. Stephen Colbert promptly informed the public that he was not the manager of that

account and subsequently set up a separate Twitter account (@StephenAtHome) which is he, indeed, manages himself.[209] Because this verification symbol seems to heighten followers' expectation of truthfulness, accuracy, and, in the case of @ColbertReport, ethics, it also serves to encourage followers to hold verified individuals accountable for their tweets. Our survey reveals that users associate the blue verification symbol with truthfulness and trustworthiness.

This is the key distinction. While the express purpose of the verification badge is to authenticate the *account* as authentic, users tend to rely on the badge as a proxy for the *content*'s accuracy. In reality, however, tweeters "of public interest" may still tweet content of verifiably no truth value despite the blue verification badge. This does not necessarily mean this behavior is problematic - one of the cornerstones of rhetoric is that an audience judges speech by the perceived authenticity of the speaker. Rather, the implication is twofold: first, because individuals who are not "in the public interest" do not enjoy access to the badge, only a select few individuals gain the benefits of endorsed authenticity that the badge accords. Secondly, because users have few other signals on Twitter's user interface with which to judge tweets' credibility, they may tend to over-rely on this signal, which is a flawed signifier of the content's credibility.

There was also a strong correlation between the amount of trust individuals had in a given tweet and the use of a URL in the tweet, suggesting that, at some level, people still associate citations with facts and are more skeptical of tweets that do not provide citations. Similarly, a lack of typos in a tweet had a strong correlation with trustworthiness, and use of the phrase "here are the facts" before a tweet had a weak correlation with trustworthiness. These results suggest that people use "facts," correctly spelled language, and citations as a shortcut for fact-checking information.

It is important to note, however, that the presence of a URL itself is an incomplete predictor of a tweet's truth value. This is due to the limited amount of time that people spend judging the veracity of a tweet. This result suggests that users generally assume that the inclusion of any URL likely indicates the "truthfulness" of a tweet. While shortcuts can help users quickly assess different tweets' likelihood of truthfulness, they serve, at best, as a flawed proxy for thorough fact

[209] Megan Gibson, *Colbert Tweet Draws Accusations of Racism and #CancelColbert*, TIME (Mar. 28, 2014), http://time.com/41453/stephen-colbert-report-tweet/.

checking. The implication is that there should be additional UI features that better enable users to easily evaluate tweets for their credibility.

Finally, we found no correlation between the number of shares for a given tweet and the amount of trust that individuals have in the tweet. This result appears to run counter to the phenomenon of groupthink theorized and documented in the literature of psychological research.[210] Our results may be explained by the unique communicative circumstances of Twitter: (1) users may be aware of the retweet and favorite numbers' flaws as a measure of tweet credibility; (2) users are more attuned to these numbers' flaws than the potential flaws of other signals, such as verification symbols as indicators of trust and truth; (3) there may be a key distinction between online and offline behavior when forming opinions; and (4) there may be a key distinction between how information is *presented*, which is the mode our experiment tested and which focuses on how users learn about information, in contrast to information *content*, which focuses on how users analyze and form opinions about the content of that information.

Figure 3: Amount of trust in a tweet

Amount of Trust in a Given Tweet

Factors measured	Strong correlation	Weak correlation	No correlation
	Verified Symbol	Use of 'facts' in the tweet	News Source
	Citation (URL)		Number of Shares
	Lack of Typos		

[210] For an example, see: S. E. Asch (1951). Effects of group pressure upon the modification and distortion of judgment. In H. Guetzkow (ed.) *Groups, leadership and men*. Pittsburgh, PA: Carnegie Press; Asch (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological monographs: General and applied, 70(9)*, 1-70.

Qualitative - Case Studies

We examined the role of two case studies in the spread of misinformation on Twitter: 1) a false story about the Pope endorsing Trump and 2) the spread of #MacronLeaks across Twitter in real time.

The Pope endorses Trump

Our first case study concerns the viral news story in October 2016 that the Pope endorsed Trump in the lead up to the United States presidential election.

Figure 4: A tweet alleging that the Pope endorsed President Trump

Our study of the life on Twitter for the story of the Pope endorsing Donald Trump revealed that misinformation dies quickly on Twitter. Firstly, due to the character limit on tweets and the focus on hashtags as a method for tagging, sharing, and spreading information, URLs are not as common on Twitter as on Facebook. Indeed, the URL linking to the fake story of the Pope endorsing Trump was shared only 784 times on Twitter compared to almost one million times on Facebook. This presents a basic structural barrier: it is more difficult to link to external misinformation sources on Twitter than on Facebook. This has a key implication: Misinformation requires the careful cultivation of a narrative that enhances the plausibility of the news and reduces the doubts or skepticism of the audience. Such cultivation often requires thorough development and imagery, features that are often best facilitated on external websites that are not subject to

character limitations. With tweets limited to 140 characters, Twitter's platform is not amenable to such thorough narrative development, which can help disrupt opportunities for cultivated stories to spread. In contrast, while both platforms accommodate external links, Facebook has built-in functionality that allows individual users to write long, detailed narratives, often accompanied by multimedia, which users then share across friend groups. As Twitter moves to 280-character tweets, the platform may become more like Facebook with stronger cultivation of stories grounded in false news and misinformation.

Figure 5: Number of shares for the Pope endorses Trump

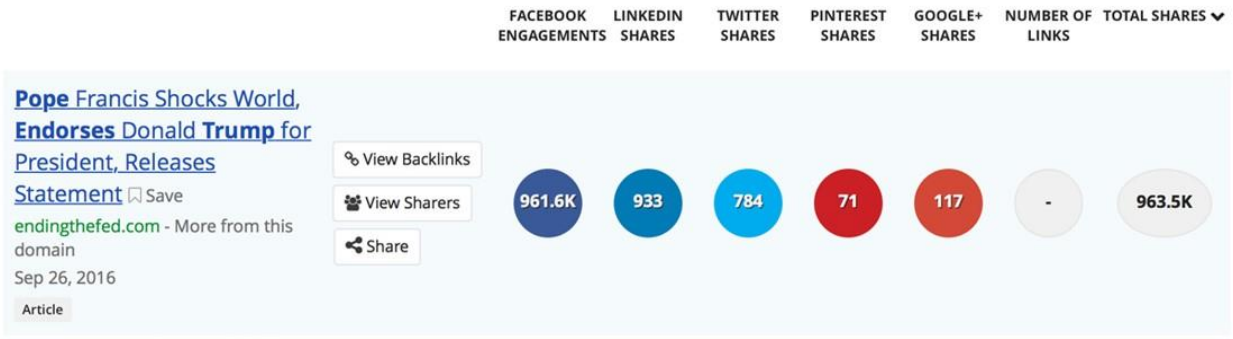
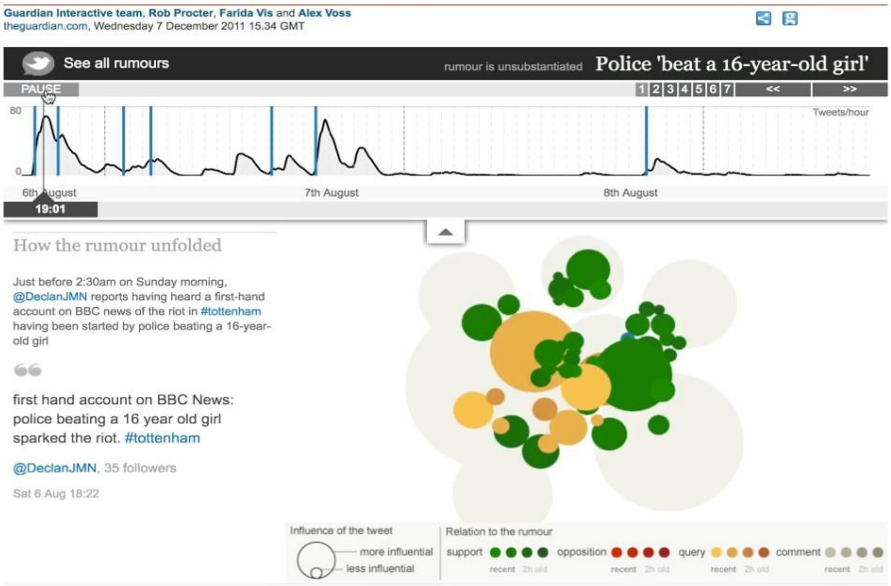


Figure 6: Screenshot of an experiment by The Guardian showing the spread of misinformation on Twitter



Second, Twitter is a very public platform. User tweets are public, meaning anyone can see and comment on your tweets. So, if you post something false on Twitter and it gains prominence (*i.e.*, a substantial number of retweets) someone will likely point out your mistake by commenting on your tweet. When others look at your tweet, they will often check these comments, helping prevent misinformation from spreading too broadly. In the screenshot (Figure 5), notice how a false story of police beating a sixteen-year-old girl dies out in about 24 hours.

#MacronLeaks

By following the spread of the hashtag #MacronLeaks in real time, we discovered two key findings: 1) who tweets matters (*i.e.*, a tweet by a verified, reputable Twitter account gets more traffic); and 2) bots successfully amplify regular tweets to increase their prominence and spread. Two days before the final round of the French presidential election, the hashtag #MacronLeaks went viral. This hashtag referred to a nine-gigabyte leak of presidential candidate Emmanuel Macron's campaign emails on the anonymous document site Pastebin.com. Users who tweeted about these leaked emails alleged everything from Macron being involved in illegal international money transfers,[211] to Macron's team engaging in questionable election practices,[212] to Macron being gay and asking his assistant to buy him drugs.[213]

The hashtag #MacronLeaks reached 47,000 retweets in three-and-a-half hours. The first tweet can be traced back to Twitter user Jack Posobiec, an ardent Trump supporter and currently the Washington Bureau Chief of *The Rebel Media*. "Posobiec's first tweet about the leaked Macron "campaign documents" occurred at 14:49 EST (18:49 UTC). This was then retweeted fifteen times in the first minute and 87 times in five minutes, suggesting the use of automated bots to amplify the signal."[214]

[211] @JackPosobiec, Twitter (May 5, 2017, 1:00 PM), https://twitter.com/JackPosobiec/status/860584992702910464 (Possible evidence of Macron international money transfer, per /pol/ #MacronLeaks.").
[212] @Messsmer, Twitter (May 5, 2017, 1:41 PM), https://twitter.com/Messsmer/status/860595225483456518 ("Nous allons nous séparer de très nombreuses personnnes le 7 mai (Facile en cas de victorie).").
[213] @HollandeOust, Twitter (May 6, 2017, 4:42 PM), https://twitter.com/HollandeOust/status/861003300040298497 ("Macron sur mailing list gay, collaborateur lui achetant sa drogue, etc. Les mails parlent! #MaconLeaks.").
[214] Atlantic Council Digital Forensic Research Lab, *Hashtag Campaign: #MacronLeaks*, Medium (May 5, 2017), https://medium.com/dfrlab/hashtag-campaign-macronleaks-4a3fb870c4e8.

The largest surge in retweets came when WikiLeaks tweeted about the hashtag. In fact, WikiLeaks amounted to almost 20% of all coverage of the leaks.[215] Each tweet by WikiLeaks garnered at least 1,000 retweets and often many more.[216]

Interview Findings

The case studies and quantitative survey experiment provided key research results relating to user behavior, psychology, and UI/UX. Those results, however, are incomplete without a thorough understanding of the broader implications of misinformation on society. To accomplish this purpose, we interviewed stakeholders in four categories: 1) Twitter officials; 2) journalists; 3) law and policy makers; and 4) scholars and experts. Here, we discuss briefly the main findings of our interviews. For a more in-depth look at findings from each interview, see Appendix C. Stakeholders representing the four categories are, as follows:

1) **Twitter**
    - High-level officials overseeing Twitter's internal assessment of dis- and misinformation. These officials requested anonymity for the public version of this report.
2) **Journalists**[217]
    - The Twitter team interviewed senior journalists and editors for national news organizations to gain their perspective on the structural factors contributing to the rapid ascent of 'fake news' and the decline of traditional journalism.
3) **Law and Policy Makers**[218]
    - We met with high-level law and policy makers in Congress and at the FTC. These interviews provided insight into threat of misinformation in the government context as well as potential legislative and regulatory concerns and interventions.

---

[215] *Ibid.*
[216] @WikiLeaks, Twitter (May 5, 2017, 4:31 PM), https://twitter.com/wikileaks/status/860638153471918081.
[217] Journalists include Philip Taubman, former DC Bureau Chief of *The New York Times*, Scott Shane, a reporter in the DC Bureau of the *New York Times*, Craig Gilbert, a reporter in the DC Bureau of the *Milwaukee Journal Sentinel*, Carl Hulse, with the DC Bureau of the *New York Times*, and the Stanford University Knight Journalism Fellows 2016-17.
[218] Government officials included current high-level congressional and former agency staff who requested anonymity for the public version of this report.

4) **Scholars and Experts**[219]
    - These interviews provided rigorous analytical perspectives, which complemented stakeholder interviews.

Twitter Officials

Key takeaways from high-level officials at Twitter highlight how the company defines and polices content that violate its terms of service. Twitter currently defines removable content as that which meets one of the following three categories:

- It is demonstrably untrue and poses a risk of imminent harm or danger to someone **OR**
- It engages in election manipulation (i.e. a tweet which claims that you can text your vote for president and have it counted) **OR**
- It is likely to cause civic unrest

If company monitors determine that the content falls within one or more of the three categories, then Twitter takes action using a system of bounces (*i.e.*, taking the tweet off of public view and turning it into a "read-only" format), strikes (three strikes, and you're suspended), and suspensions.

If the intent to commit harm is clear (*e.g.*, if there are multiple posts of this type and the user refuses to change or take down the tweet when warned) then Twitter will bounce the tweet and give the user one strike. If the intent is not clear (e.g. maybe the individual was not trying to incite civic unrest, but was trying to instead make a tasteless joke), then Twitter bounces the tweet and does not give the user a strike. If this bounced tweet is reposted, then the user will be given a strike. Once a user has three strikes, their account is suspended. If the tweet is overtly violent (*e.g.*, a tweet which says "grab a gun and shoot @realTwitteruser, because they deserve to die"), Twitter does not bother with bouncing the tweet, Twitter will suspend the account immediately.

Twitter officials also emphasized the differences between Facebook and Twitter. Facebook allows users to create silos of information; Facebook users can tailor their preferences so that only their friends see their posts and so that they can only see their friends' posts. Twitter on the other

[219] This category included Larry Diamond, Professor, Stanford University; and Gene Policinski, Chief Operating Officer, Newseum Institute and First Amendment Center.

hand is less prone to silos as your tweets are public and can be seen and commented on by anyone in the world. Twitter's UX allows for a "cross-pollination" of ideas and information which helps prevent and reduce the spread of misinformation.

Finally, Twitter officials suggested that bots themselves are not the problem, as people are unlikely to trust a bot that posts something (since usually these accounts have zero followers and are not reputable, and people like to repost things that other reputable people have posted). The problem is with bots that amplify a human tweet that may contain misinformation, thus making a bad tweet appear more popular than it would have been with exclusively human interaction.

Journalists

All the journalists we spoke to expressed concern with how social media is changing our engagement with information. In their view, social media platforms devalue well-researched, objective reporting by placing it on the same playing field as an anonymous user with an axe to grind. Social media is also fundamentally changing the way that news organizations report stories. Philip Taubman, former DC bureau chief of *The New York Times*, explained:

> If you went to the Post newsroom now you'd see a scoreboard with data on the number of clicks on stories. They would tell you—Marty Barron would say that's not influencing our decisions. I don't see how over time that can't affect what you're writing about and how you value reporters. The old metric for reporters, the rawest metric was number of frontpage bylines—that was an indication you were a great asset to the paper. The question now is going to be—part of the conversation at an annual evaluation—they have a list that shows how your stories are faring [on social media]. They're saying we're not judging based on that, but I don't see how it's avoidable over time. It's not a good trend.

Law and Policy Makers

The Chiefs of Staff we spoke with all had similar views about misinformation. If their Senator/Representative were the target of a misinformation attack, they would not expect Twitter to act. Instead, they said that their strategy would be to try to get ahead of the story and offset it through counter-narratives on social media, in press conferences, and in speeches; their goal would be to publicize the falsity of the story. They also described local newspapers as helpful allies in attacking misinformation because newspapers care about getting the facts right and constituents typically regard local papers as trustworthy news sources.

While U.S. lawmakers typically seem to abide in the rectitude of CDA Section 230, their counterparts in Germany had been considering (and recently dismissed[220]) a law that would have included fines for platforms that fail to remove false news. When asked whether they could see the possibility of the U.S. developing legislation that mirrors Germany's law, which would have allowed social media platforms to be fined if they failed to take fake news off of their platforms, each responded, without hesitation, that such legislation was unthinkable in the United States due to the nature of our free market politics and the strength of First Amendment protections. Thus, it does not appear that the current political climate is ready for regulatory or legislative reform that would institute liability for online platforms.

On the other hand, the congressional officials we spoke with were quick to draw a line between misinformation perpetuated by individuals and misinformation used by foreign nations as a form of attack. Several officials suggested that Congress and the United States government have a duty to step in if a foreign nation like Russia were involved in a cyber-attack involving misinformation. No interviewees were able to offer specifics beyond this statement due to the sensitivity of the information.

Scholars and Experts

Gene Policinski leads the First Amendment section of the Newseum. He discussed the potential constitutional issues with broadening algorithms to police against misinformation. He also acknowledged, nevertheless, the need to do so in certain instances given how misinformation has become "weaponized." "I expect government to protect me from military threats. If information has become weaponized to the point at which it's a threat to democracy, then I can see why. But then the government is making value judgments. The military will err on the side of safety. That's an appropriate response."

Jon Leibowitz spoke with us for internal purposes only. He noted that the FTC has been taking action against organizations for false advertising for quite some time. With the advent of the internet, false advertisements have become more sophisticated, appearing on social media and at the bottom of certain webpages. The FTC has noticed that false advertisements are being

[220] "Facebook Law' that could have fined social media firms up to £45 MILLION for failing to remove offensive or racist content is rejected - because it infringes on freedom of expression," *Daily Mail*, 14 June 2017.

promoted by the same third-party vendors. Unfortunately, the FTC can only go after the actual advertiser for false claims and has no jurisdiction over third parties. Jon Leibowitz suggested that if the FTC were given aiding and abetting jurisdiction then they could prosecute those third-party promoters of false advertising and really make a dent in the issue.

These interviews collectively revealed important insights into the current challenges, as viewed by different sectors, and defined the scope of possible interventions and recommendations.

VI. Options and Recommendations

Twitter

1. **Permit all users to apply for the blue verified badge for their accounts. Verification may help enhance basic rules of conduct, including individual tweeters' responsibility for content they share.** As revealed through our user experiment (see Appendix A), users rely on the blue verification badge as a key means of judging the trustworthiness not only of the individual account but the content shared. Currently, the blue verification badge is available to authenticate only the accounts of users whom Twitter perceives to be "of public interest," including some accounts that promote hate speech.[221] Overall, expanding access to verified status could help encourage responsibility for shared content by tweeters. The drawback of this suggestion is the problem of users conflating trust in the authenticity of the account sharing information with the content shared. Twitter should emphasize that verification of the authenticity of an account does not verify the accuracy or quality of the information shared.
2. **Experiment with a 'false news' flagging mechanism that enables users to report false content.** To avoid users over-reporting either through genuine vigilance or partisan motives, Twitter could implement certain rules of due process. The drawback to such rules would be the loss of real-time information in that the vetting process would likely result in a delay of at least 24 hours to determine accuracy.
    - Experiment with a crowd-sourced voting system, similar to Facebook, whereby Twitter users can upvote or downvote a tweet for its truth content. Tweets that pass

[221] See Twitter Help Center, "About verified accounts," https://support.twitter.com/articles/119135. See, for example, Jim Dalrymple II, "Twitter Grapples with 'Verified' White Supremacists," *Buzzfeed News*, Aug. 19, 2017. https://www.buzzfeed.com/jimdalrympleii/twitter-has-a-verified-conundrum?utm_term=.jcoW1yJGz#.alAaq7G69

a certain threshold number, both by raw count and by percentage, of downvotes, would be sent to fact-checkers for review.

- Place the burden on the user reporting the content to offer evidence for the given content's falsity. This would mirror current trial procedure in defamation suits. An opportunity for defense may be offered thereafter to the user who posted the content. A relevant authority at Twitter could arbitrate.
- A variety of subtle UI features could be implemented to flag content that is deemed 'false' under this option. These include, but are not limited to:
    - Slightly altered background color of 'false' tweets to signal the likely falsity of the content.
    - The addition of subtle text in muted, neutral colors that is not prominent to the tweet. The user would likely see this text after having already viewed the content.
    - The addition of prominent text to the tweet clearly visible before the user views the original content.

Currently, Twitter offers no option to report content that is either demonstrably false or misinformation. The options are currently limited to reporting harassment, inappropriate content, or other forms of legally impermissible content. However, the inclusion of an option to flag false information and misinformation can encourage users to be more vigilant about the veracity of content that they post and view.

As valuable as the flagging mechanism might be in diminishing the proliferation of misinformation, it could also lead users to over-report "false" content, running the risk of flooding Twitter with false positive reports. Thus, implementation would need to be explicit about review procedures. Reports of misinformation would need to be reviewed with great caution and judiciousness. While it is unlikely that Twitter has the financial capacity to hire many fact checkers, as Facebook has done, implementation of this option would allow Twitter to maintain a commitment to reducing the incidence of misinformation on its platform, keep its long-standing commitment to free speech, and conserve its financial and administrative resources. Such crowd-sourced reporting of misinformation could:

- Allow Twitter monitors to evaluate content on a case-by-case basis, rather than expend significant engineering cost to develop artificial intelligence that could automatically detect misinformation but result in unintended consequences.
- Reduce the incidence of misinformation. Further research could examine methods to disincentivize users from over-reporting. For instance, Twitter could penalize users who inaccurately over-report false positives by temporarily suspending their user privileges.
- Help Twitter respond to possible German legislation that would require the company to remove hate speech, incitements to violence, and other possibly incendiary content. This option would enable the company to share the burden with users (and the German state) to diagnose such issues. Users could file reports whenever they see such content on Twitter. By expanding the universe of observers, crowd-sourced flagging of misinformation could help the company contain its costs for detecting broad classes of such content without sacrificing Twitter's inherent free speech principles or running counter to different, less rigid, regulatory regimes in other countries.[222] The downside of such a method is that users might over-report and also not be adept at diagnosing such issues.
- Encourage users to be more active about discerning the credibility of their sources and fact-checking.

3. **Implement stricter Captcha gateways during registration process and when suspicious activity is detected.** To enhance trust and enforce accountability, Twitter should implement Captcha gateways both during the account registration process and when suspicious account activity is detected. This intervention would also help to manage the millions of bots that manipulate Twitter's platform to shape online conversations.
    - This tool aligns with the terms of service for online services that are liable to fraud or spam. There is reason to believe that this standard would not be unduly burdensome, as it has successfully been implemented in a variety of other services.
    - This tool would be noninvasive and would not impinge on Twitter's commitment to free speech on its platform. Although this intervention could help to manage the

[222] For background on the proposed law, see Amar Toor, "Germany wants to fine Facebook over hate speech, raising fears of censorship," *The Verge*, June 23, 2017. Accessed June 24, 2017.

influence of fake bots on Twitter, the intent is content-neutral and aims to preserve free speech.

- While not impermeable, the tool could diminish foreign state actors' ability to manipulate online conversations or amplify false information.
- Twitter already has the capability to automatically detect suspicious account activity in a variety of categories. Simply adding a Captcha gate when this activity is detected would be relatively simple to implement and would not impose significant engineering costs to develop or maintain.

4. **Create more options for users to react to Tweets, beyond 'favorite' and 'retweet'.** This is similar to the change that Facebook implemented that enables users to react in multiple ways to Facebook posts. This would have three benefits on Twitter:
    - This would facilitate more dynamic, multifaceted exchanges. Users may engage more with content if they are given more options for how they can interact with it.
    - This shifts the norms for retweeting away from the current implication of personal endorsement. Users may choose to disagree with content or express negative or other reactions to content while still 'retweeting', thereby encouraging other users to engage critically as well.
    - If prominently displayed, the varied reactions to breaking news can encourage users to be more discerning of the nature of the news and evaluate its substance before sharing.
5. **Prominently display the original source of retweeted content.** This is recommended for two reasons:
    - Users tend rightly to judge content by its source. Twitter's current retweeting mechanism occludes the origin of a post after more than one retweet. Developing a means of retaining access to the original post would help users judge the quality of the original content before deciding to retweet.
    - Original provenance allows original sources to be credited for producing content that is widely shared and read. Original credit can also potentially allow revenues to flow directly to the original source.
6. **Refine automatic detection tools for spam and other suspicious account activity.** This is a long-term improvement and one that Twitter likely already has engineers working on.

Improvements to the code infrastructure would, nevertheless, help resolve the emerging conflict between Twitter's commitment to freedom of speech and public calls for greater oversight for misinformation.

Philanthropic Institutions

1. **Encourage Twitter to share relevant internal data with academic researchers to better understand network dynamics and user behavior.** Twitter is already more transparent with its user data than are other online platforms. Twitter's streaming API allows relatively voluminous data collection (up to 1% of tweets in real-time can be collected), but access to data remains an impediment for researchers. Twitter need not open source its data, however. Rather, by sharing its data selectively under nondisclosure agreements, Twitter could benefit from analysis that yields improvements to the platform and a more contextual understanding of how the platform fits within the social media landscape.
2. **Work to develop public-private partnerships on combating misinformation on social media.** Help establish secure lines of communication between designated Twitter employees and U.S. government officials, including the FTC. One respondent suggested the need for a "point person" or government agency that could work directly with online platforms to help mediate the spread of "misinformation." This respondent pointed out that some European countries have designated government agencies or officials to oversee illegal or inappropriate content on online platforms. Although U.S. law protects platforms from intermediary liability, the U.S. government maintains an interest in preventing state-sponsored misinformation campaigns from leveraging these platforms.
3. **Encourage news organizations to be more public in disclosing details of their editing and review process.** Public trust in news institutions is increasingly divided by partisan affiliation. Transparency in the editing and review process could enhance media literacy and, in turn, help to pressure news organizations to improve their reporting practices.
4. **Develop initiatives that help smaller, independent and local news organizations retool their business models to support investigative journalism.** Stakeholders, representing local news organizations, emphasized the difficulty for those organizations to fund deep investigative reporting. Philanthropic institutions could help grow trusted, scholarly online resources, such as the *Journalists' Resource* (journalistsresource.org), that provide journalists with access to reliable, well-vetted datasets and summaries to ground

investigative reporting. This is one of the few online organizations to support journalists with guidelines for investigative reporting in such complex areas as tax reform and campaign finance, among others. With some expansion, it could provide a possible forum where independent and local news organizations could gain perspective on issues relevant to local concerns.

Stakeholders representing independent and local news organizations further pointed out the challenge of competing for advertising revenues in the online marketplace. Many news organizations may simply not have the requisite tools to develop effective online presences. Philanthropists could fund initiatives that enable smaller news organizations to explore how to reinvent their business models to support investigative journalism. A partnership between a philanthropic foundation or think tank and *Journalists' Resource*, for example could develop a forum where smaller, independent and local news organizations could access information, and possible training, to retool their business models to compete in the online marketplace.

5. **Persuade Twitter to more effectively train and work with journalists to help them spread their content on Twitter.** This could be developed into a platform, such as 'Twitter for Journalists', aimed at journalists using the platform. Twitter clearly has the capacity to develop a new initiative or platform aimed specifically at journalism. Twitter has already developed an extensive platform - 'Twitter Ads' - tailored to advertisers. 'Twitter for Journalists' can be a similarly comprehensive initiative.

Long-Term Research Projects

1. **Conduct or invest in funding for large-scale user experiments on Twitter.** Features could include devoting resources to developing samples representative of different geographic areas or dynamics, pulling sample sizes of at least 1,000, and creating more detailed mockups of entire Twitter threads that employ certain key modifications in UI.
    - This would be a successor the user experiment that was conducted for the purposes of this project (see Appendix A). This practicum was resource constrained in terms of implementation of its experimental design.

- A comprehensive user experiment of this nature could be implemented through a partnership with an academic research institution.

2. **Conduct an analysis of the online advertising industry and its impact on journalism**. Probe the ways in which the platforms strip news outlets of advertising dollars.
3. **Conduct a quantitative analysis of the likely scope and impact of Russian, or otherwise state-sponsored, misinformation campaigns on Twitter and/or other platforms**. This would examine the impact on the behavior and sentiment of users exposed to such misinformation. This would draw on social network analysis with attention to national security concerns. Findings could be persuasive when making a case to social media companies for self-regulation.
4. **Most ambitiously, conduct a comprehensive quantitative study of the likely impact of 'false news' on user behavior and perceptions.** Effective research proxies could include an estimate on the number of voters in the 2016 U.S. presidential election who switched voting preferences based on exposure to false news or misinformation. This could be captured through analyses—both experimental and aggregated—of shifts in online user behavior when exposed to misinformation.

Near-Term Options

1. **Facilitate collaboration between social media researchers, journalism researchers, and national security experts to better explore the scope of the solution space for state-sponsored misinformation**. A model of this collaboration could potentially be Jigsaw.
2. **Conduct an industry analysis of the online advertising industry**. This would be crucial for understanding the evolving financial motives underlying the ways journalists approach social media.
3. **Explore potential alternative models of local investigative journalism that leverage social media platforms**, including citizen journalism, with attention to the democratizing potential of Twitter and other platforms.
4. **Encourage and guide elected representatives and public agencies in more active Twitter presences.** While the tendency for some of today's politicians to assume disruptive, misinformative online presences is now well known, the overall benefit to deliberative democracy nevertheless justifies this recommendation. Elected representatives

should not be encouraged to tweet missives contributing to the coarsening of public discourse. Rather, they should be encouraged to leverage Twitter's democratizing potential, through more frequent Twitter town halls and development of constituent services delivered over social media.[223]

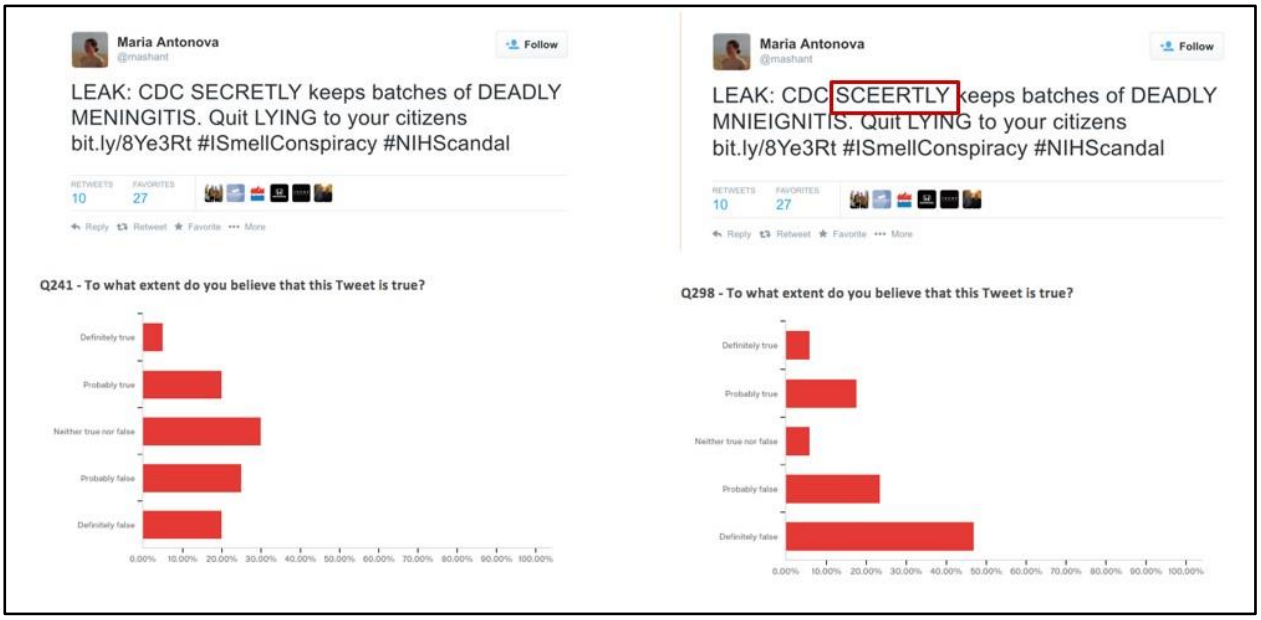
5. **Encourage Twitter to develop user features that specifically deliver key information on policies and politics.** This could build on the "Civic Engagement" initiative of Google in recent elections. Twitter could aggregate lists of reliable, "verified" officials in each country where there is a strong user base to build a "Policy Center" list that users could follow for information about pending legislation, committee information, live video feeds of floor debates, and debates on legislation.
6. **Fund an in-house research fellow to help lead further "fake news" research initiatives.** Such a fellow could continue to work with outside research organizations to guide projects that leverage complementary skills sets, including computer science, law and policy, journalism and communications, psychology, product design, and user interface design. The fellow could oversee a series of projects designed to offer a 360-degree perspective on the problem of fake news and misinformation in the context of today's journalism and the role of online intermediary platforms.

[223] While this has traditionally been the domain of social media strategists and communications directors to public officials, public officials can directly use social media platforms to better engage with their constituents. Politicians build relationships and trust with their constituents by physically meeting them and listening to their concerns. Relationships of trust between public officials and citizens can help disrupt the spread of false stories. With more active presences on social media, public officials may be able to connect with wider audiences and, thereby, build trust but, fundamentally, the antidote to misinformation is integrity and credibility.

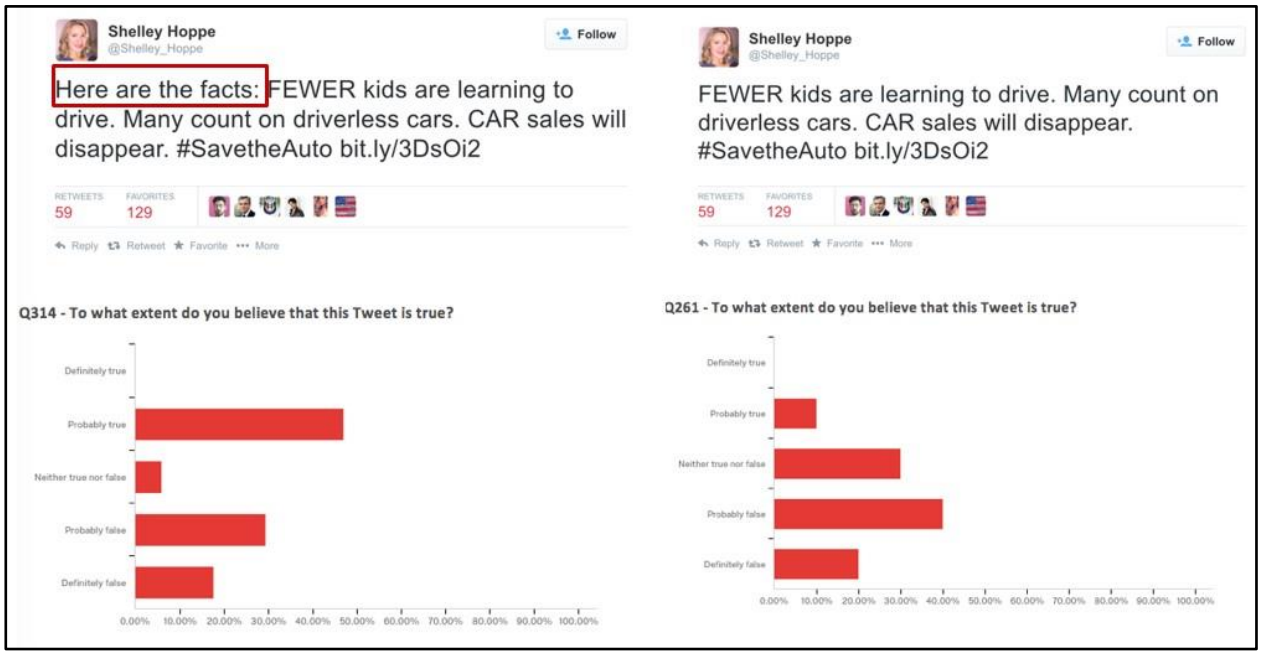
Appendix A - Full Survey Results

Verified Badge

Citation (URL)

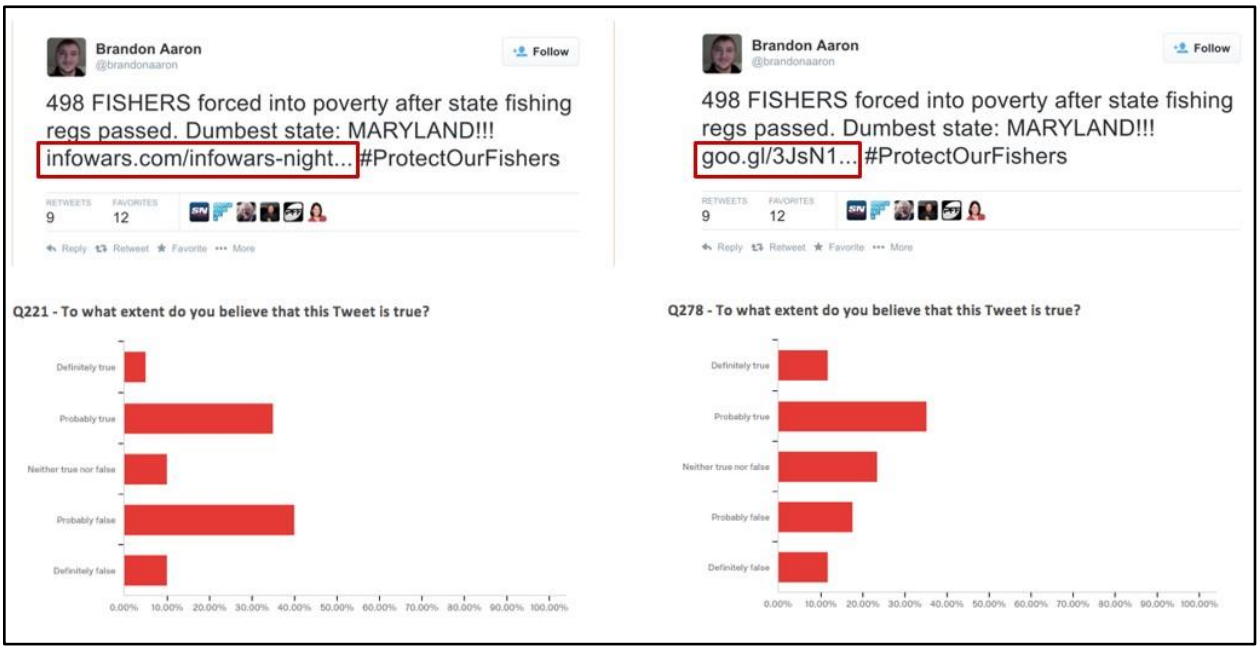


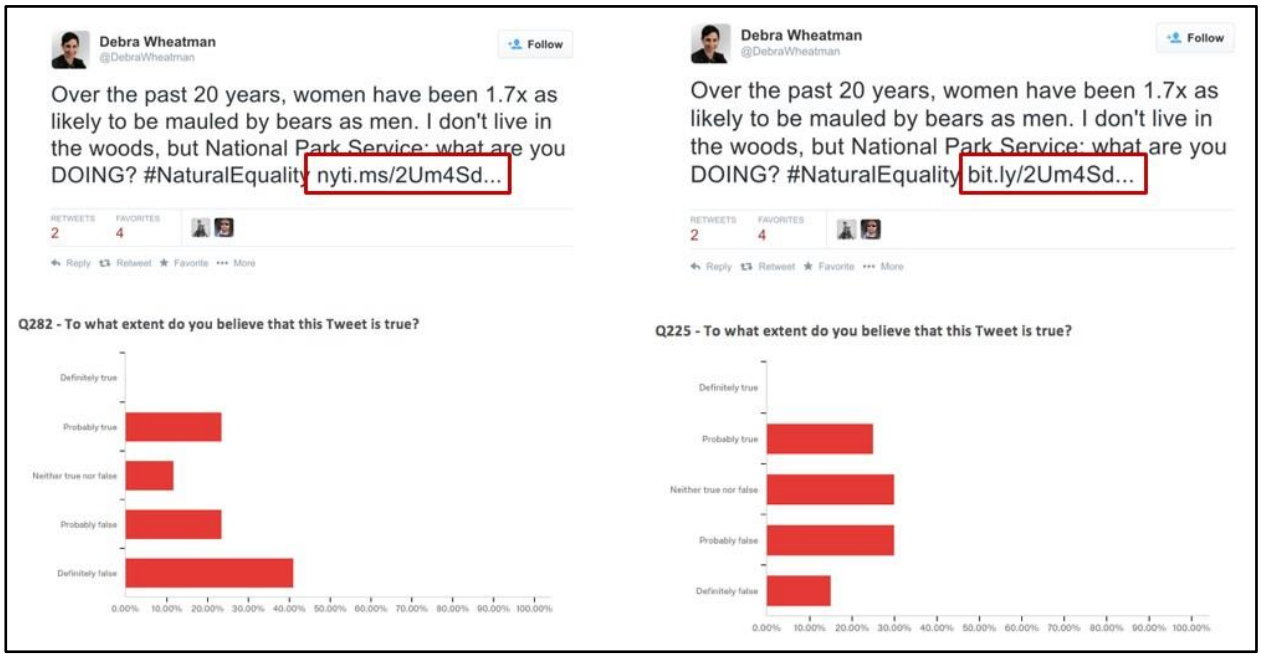
Typos



Use of the Term "Facts" in the Tweet

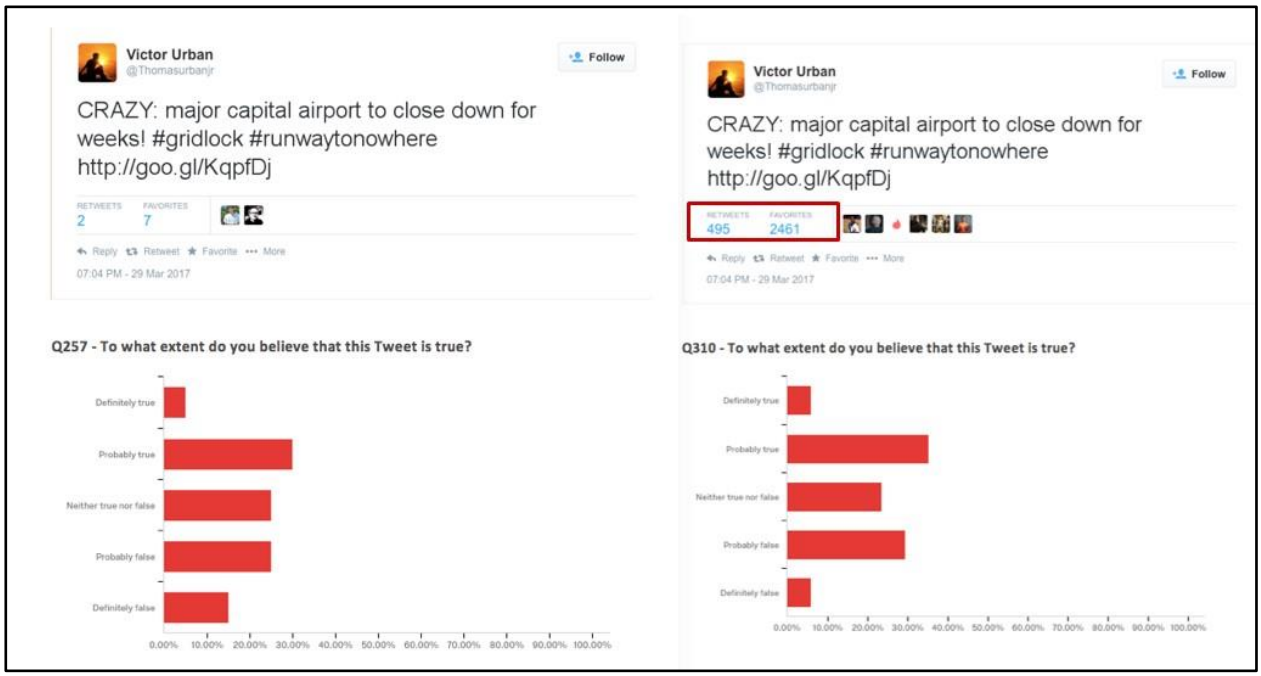
Type of Source (Infowars vs. Google)





Type of Source (New York Times vs. Bit.ly)

Number of Shares



Section 5. Reddit

I. Introduction and Overview

Reddit is the fourth most visited site in the United States, with over 250 million unique users per month. It is also a major source of information for its massive user base, as 70 percent of Reddit users get news from the site. Several aspects of the site, however, make it a possible target for the spread of disinformation, misinformation, and other forms of fake news:

1. Relatively light oversight in how unique user accounts can be created;
2. A content submission system typically open to anyone with a user account;
3. A trade-secret algorithm (with upvotes/downvotes) that gauges popularity by roughly promoting news and posts that have had a high number of upvotes assigned to them;
4. An organization system that subdivides the site into so-called "subreddits"— opt-in communities organized around specific topics thereby enhancing the effect of siloed communities;
5. An arbitrary classification of "default" subreddits that are automatically presented to naive visitors to the site;
6. A community that, at times, has been shown to be susceptible to creating, popularizing, and perpetuating its own "fake news," with real-world impact.

As a result, it is possible for parties to create large numbers of unique user accounts, submit "fake news" content, quickly popularize it using those created user accounts, target specific communities of users susceptible to "fake news" efforts, and rely on the efforts of the communities and users themselves to perpetuate and spread "fake news."

Our research examined how misinformation spreads on Reddit and sought to develop measures that could limit this spread. To accomplish this goal, we reviewed available academic research, studied current moderation policies on the site, interviewed a high-level company spokesperson and a key volunteer moderator, and studied case studies of misinformation that spread on the site.

Findings

- **Fake breaking news stories and bogus clickbait are less of a problem on Reddit than on other platforms.** Both the Reddit community and the volunteer moderators that control topic-based communities on Reddit tend to delete obviously false content.
- **Conspiracy theories are prominent in certain subreddit communities.** These conspiracy theories can gain momentum on Reddit as users may act as amateur detectives, believing they are uncovering some major crime or scandal. Fringe and even mainstream websites often then promote the work found on Reddit, and those articles then end up back on Reddit, continuing the cycle.

- **Reddit administrators tend to be hands off, though in recent months they have taken steps to separate fringe communities from the larger community as a whole.** Reddit takes action if personal information is posted, but otherwise leaves moderation policies to volunteer users, as many Reddit users are hostile towards the company's administrators and would resist more aggressive moderation policies. Where Reddit has made tweaks is in its voting algorithm—changing how content appears on the front page—and, occasionally, an increased heavy-handedness in dealing with uncooperative moderators.
- **Moderators vary in their moderation, but nudges and/or reminders in certain subreddits can improve the quality of the community.** Civil discourse, commenting, voting, and submitting may all be affected by specific moderation measures, which may also slightly reduce the spread of false information.

Recommendations

- **Reddit should nudge users to be skeptical of unreliable content**: These reminders could be in the form of top comments or other alerts placed prominently on the page. These reminders would be placed on the main log-in page for the overall site.
- **Reddit should limit the exposure of "problem" subreddits**: The platform should continue taking steps that limit the prominence of communities that regularly host false content.
- **Reddit should work with media organizations to ensure news articles load quickly**: Although we don't have access to Reddit's internal data, it seems likely that users are more inclined to engage with content that loads quickly. Cutting load times on mainstream news articles could give that content an advantage over unreliable fringe websites. This advantage could result in real news consistently getting more exposure and engagement than false news.

Other options include banning these controversial communities, although we do not recommend that step at this time. Reddit has ambitions to grow its user-base and substantially increase its advertising revenue to rival that of Facebook.[224] It is already experimenting with new advertising features to try to accomplish this goal. We therefore believe that the most likely source of leverage over Reddit is through its relationships with advertisers. Reddit will take the problem of conspiracy theories and misinformation seriously if it believes it is in danger of losing advertisers.

Future research could focus on whether it would be effective to ban major subreddits or top users who share false information. More data could also reveal what kinds of reminders or

[224] George Slefo, *Reddit Intros New Ad Offering, 'Grows Up' and Says It Can Be as Big as Facebook*, AdAge (July 26, 2016), http://adage.com/article/digital/reddit-let-marketers-sponsor-users-organic-posts/305115/

alerts would be the most effective in warning users about false information and where they should be placed on the page.

II. Background

Site Overview

Reddit is a content-aggregation, rating, and discussion website serving its content to users worldwide. Founded in 2005 by Steve Huffman and Alexis Ohanian, its current headquarters are in San Francisco; as of 2015, it had approximately 100 employees.[225] Although it operates independently, Reddit was acquired by Condé Nast Publications in 2006 and Condé Nast's parent company, Advance Publications, is its largest shareholder today. In its latest round of fundraising in October 2014, Reddit was valued at $500 million. In 2015, Reddit saw 82.54 billion pageviews, 73.15 million submissions, 725.85 million comments, and 6.89 billion upvotes on 88,700 active subreddits.[226]

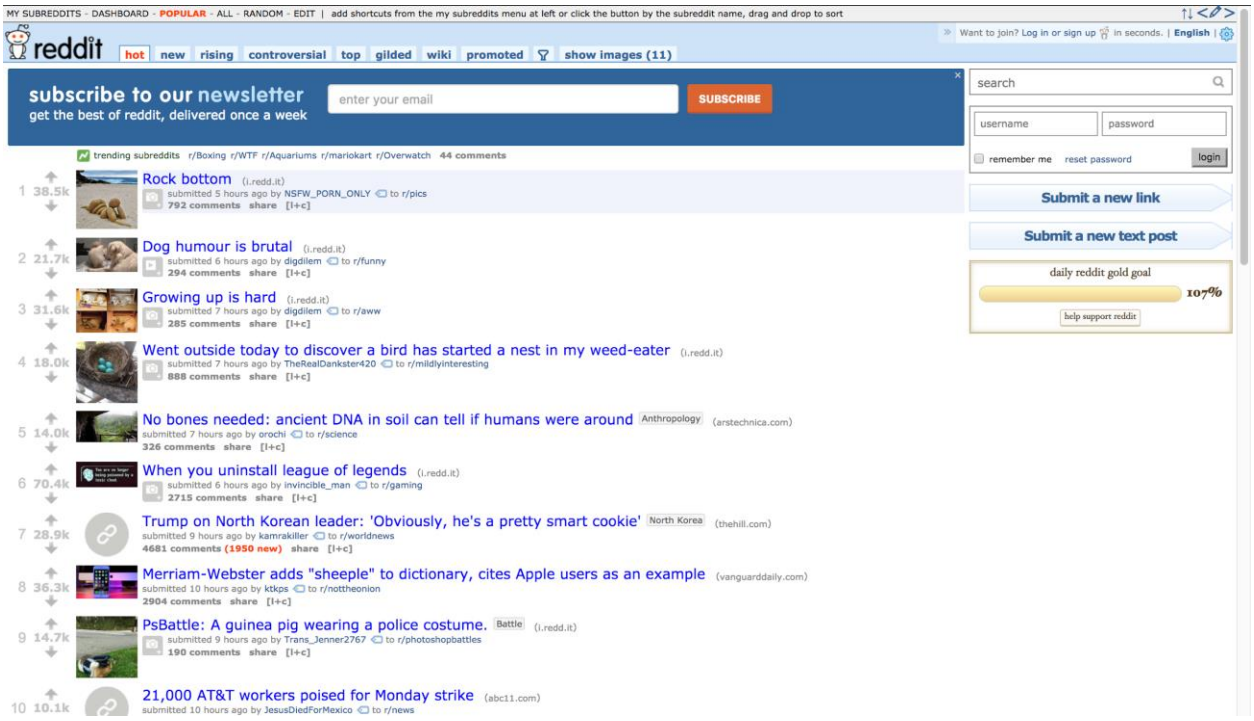
Reddit is organized into subreddits (stylized /r/<subreddit name>), which are opt-in communities organized around particular topics (**Figure A.2**)**.** Subreddits range from just a few subscribers, to tens of millions (e.g. /r/AskReddit, /r/worldnews); subreddit topics range, but can broadly be classified as external content (e.g. /r/news, /r/gaming, /r/science, /r/baseball, /r/videos, etc.) or self-posts (e.g. /r/bestof, /r/OutOfTheLoop, etc.). Each subreddit's page is a collection of user-submitted entries to that particular subreddit.

Registered users, known as Redditors, can upvote or downvote content; Reddit's proprietary algorithm then presents content in rank-order format, with a preference for newer content that has a high total vote count, with a high upvote:downvote ratio. Individual posts have their own pages, on which users can submit comments (**Figure A.3**). Comments themselves can be upvoted and downvoted, and can be sorted by several metrics (e.g. "new", "best", etc.). Reddit's front-page represents a diverse aggregation of the top content (as measured by Reddit's algorithm) taken from many of the most popular subreddits (**Figure A.1**).

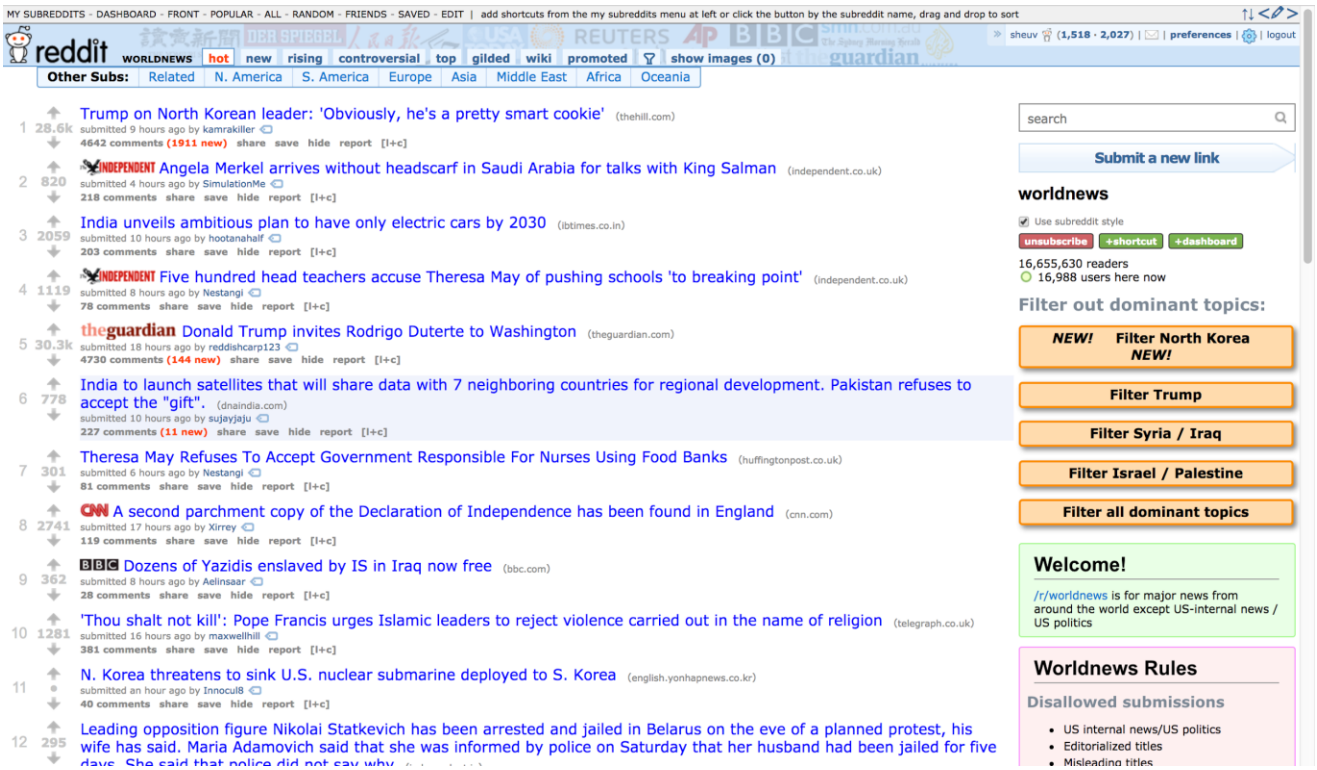
[225] Jessi Hempel, *Inside Reddit's Plan to Recover From Its Epic Meltdown*, WIRED (Oct. 6, 2015), https://www.wired.com/2015/10/reddit-survived-meltdown-can-fix/.
[226] *Reddit in 2015*, REDDIT (Dec. 31, 2015), https://redditblog.com/2015/12/31/reddit-in-2015/.

Figure A: Reddit site layout.

(1) Reddit front page. Diverse, aggregated content from the most popular subreddits, rank-ordered by Reddit's algorithm.



(2) Subreddit (/r/worldnews). Content organized around a particular topic. Users can subscribe and unsubscribe to individual subreddits to curate their Reddit experience.


MY SUBREDDITS - DASHBOARD - FRONT - POPULAR - ALL - RANDOM - FRIENDS - SAVED - EDIT | add shortcuts from the my subreddits menu at left or click the button by the subreddit name, drag and drop to sort
reddit
WORLDNEWS hot new rising controversial top gilded wiki promoted show images (0)
sheuv (1,518 · 2,027) | preferences | logout
Other Subs: Related N. America S. America Europe Asia Middle East Africa Oceania
1 28.6k Trump on North Korean leader: 'Obviously, he's a pretty smart cookie' (thehill.com)
submitted 9 hours ago by kamrakiller
4642 comments (1911 new) share save hide report [l+c]
2 820 INDEPENDENT Angela Merkel arrives without headscarf in Saudi Arabia for talks with King Salman (independent.co.uk)
submitted 4 hours ago by SimulationMe
218 comments share save hide report [l+c]
3 2059 India unveils ambitious plan to have only electric cars by 2030 (ibtimes.co.in)
submitted 10 hours ago by hootanahalf
203 comments share save hide report [l+c]
4 1119 INDEPENDENT Five hundred head teachers accuse Theresa May of pushing schools 'to breaking point' (independent.co.uk)
submitted 8 hours ago by Nestangi
78 comments share save hide report [l+c]
5 30.3k theguardian Donald Trump invites Rodrigo Duterte to Washington (theguardian.com)
submitted 18 hours ago by reddishcarp123
4730 comments (144 new) share save hide report [l+c]
6 778 India to launch satellites that will share data with 7 neighboring countries for regional development. Pakistan refuses to accept the "gift". (dnaindia.com)
submitted 10 hours ago by sujayjaju
227 comments (11 new) share save hide report [l+c]
7 301 Theresa May Refuses To Accept Government Responsible For Nurses Using Food Banks (huffingtonpost.co.uk)
submitted 6 hours ago by Nestangi
81 comments share save hide report [l+c]
8 2741 CNN A second parchment copy of the Declaration of Independence has been found in England (cnn.com)
submitted 17 hours ago by Xirrey
119 comments share save hide report [l+c]
9 362 BBC Dozens of Yazidis enslaved by IS in Iraq now free (bbc.com)
submitted 8 hours ago by Aelinsaar
28 comments share save hide report [l+c]
10 1281 'Thou shalt not kill': Pope Francis urges Islamic leaders to reject violence carried out in the name of religion (telegraph.co.uk)
submitted 16 hours ago by maxwellhill
381 comments share save hide report [l+c]
11 • N. Korea threatens to sink U.S. nuclear submarine deployed to S. Korea (english.yonhapnews.co.kr)
submitted an hour ago by Innocul8
40 comments share save hide report [l+c]
12 295 Leading opposition figure Nikolai Statkevich has been arrested and jailed in Belarus on the eve of a planned protest, his wife has said. Maria Adamovich said that she was informed by police on Saturday that her husband had been jailed for five days. She said that police did not say why
search
Submit a new link
worldnews
Use subreddit style
unsubscribe +shortcut +dashboard
16,655,630 readers
16,988 users here now
Filter out dominant topics:
NEW! Filter North Korea NEW!
Filter Trump
Filter Syria / Iraq
Filter Israel / Palestine
Filter all dominant topics
Welcome!
/r/worldnews is for major news from around the world except US-internal news / US politics
Worldnews Rules
Disallowed submissions
US internal news/US politics
Editorialized titles
Misleading titles

(3) Posts. Discussion (commenting and voting) on individual posts to the subreddit. The items boxed in red are examples of subreddit policy reminders enacted by the subreddit's moderators.

Redditors have extensive control over their experiences on the site. Anyone can create an account with nothing more than a chosen username/password—an email is not required, and account creation is free. Within their account, any Redditor can create any subreddit (public or private); subreddit creators become moderators, but can also appoint any other Redditor as a moderator. Moderators, who are unpaid and unaffiliated with the company, curate content on an individual subreddit—they can delete posts, block users, create filters, and enact moderation policies that affect how others interact with an individual subreddit. Redditors can subscribe/unsubscribe to all public subreddits; their front-page aggregates content from only their subscribed subreddits.

Redditor accounts also present a history of the account's interaction with the site. A Redditor's post and comment history is publicly viewable; accounts also aggregate "post karma" and "comment karma" (representing the net upvotes on a user's posts and comments anywhere on the site). Redditors can also gift each other (or themselves) "Reddit gold"—a premium membership subscription to the site that includes additional features.[227]

[227] Additional features generally improve the user experience in ways that do not affect how content is curated (e.g. ability to hide ads). A full list of Reddit Gold features can be found at https://www.reddit.com/gold/about/.

While Reddit has a few sources of income, it has not been a highly profitable company. Because it is not a publicly traded company, Reddit financial data is hard to come by. What is known: advertisements and Reddit gold—$3.99 a month, or $29.99 a year, respectively—represent Reddit's main revenue streams. As of 2012, the site was not profitable;[228] in 2014, the site brought in $8.3 million in advertising revenue.[229] Unlike other media sites that have engaged in advertising campaigns by placing audio or video ads in front of viewers, Reddit's ads in comparison remain low-key, and the company seems reluctant to change that policy in the near future.

III. Problem/Goal Statements

Several aspects of Reddit make it a possible target for the spread of disinformation, misinformation, and other forms of fake news:

1. Light oversight in how unique user accounts can be created and ease of creating new accounts;
2. A content submission system typically open to anyone with a user account;
3. A trade-secret algorithm (with upvotes/downvotes) that calculates popularity by roughly promoting news and posts that have had a high number of upvotes assigned to them;
4. An organization system that subdivides the site into "subreddits"— opt-in communities organized around specific topics; and,
5. A community that has, at times, been shown to be susceptible to creating, popularizing, and perpetuating its own misinformation, with real-world impact.

As a result, it is possible for parties to create large numbers of unique user accounts, submit "false," "fake," or "misinformed" content, quickly popularize it using those created user accounts, target specific communities of users who are susceptible to misinformation efforts, and rely on the efforts of the communities and users themselves to perpetuate and spread misinformation.

This research report aims to determine the demographic behaviors of Reddit's users, the extent to which misinformation has been a problem on Reddit, the effect of moderation policies and mechanisms on the spread of content, and the feasibility of possible future directions.

IV. Methodology

Our primary evidence consists of reviewing Reddit itself, and interviews with interested and knowledgeable parties at Reddit. We organized our studies into two focuses: content moderation, and content propagation. By studying the different ways in which content propagations through Reddit, we looked to identify patterns in propagation. Analyzing this data provided insight into possible ways of tailoring recommendations to better address

[228] yishan, Comment on *Now is the Time... to Invest in Gold*, Reddit (Nov. 8, 2012), https://www.reddit.com/r/blog/comments/12v8y3/now_is_the_time_to_invest_in_gold/c6yfbuh/.
[229] yishan, *Decimating Our Ads Revenue*, Reddit (Feb. 28, 2014), https://www.reddit.com/r/blog/comments/1z73wr/decimating_our_ads_revenue/.

misinformation propagation on Reddit. In looking at Reddit's past policies and changes with regards to content moderation, we looked to answer the question "What has Reddit done to manage content in the past?" hoping to gain some idea of the efficacy of different measures a site can implement. Any successful measures could be further recommended to other sites on the Internet. The moderation policies and measures we studied included moderator-level policies (subreddit content policies and moderation notifications regarding commenting and voting), administrator-level policies (Reddit's development of its voting algorithm, content policies, and history of enforcement), and user reactions to moderation at both levels.

V. Platform Analysis

Demographics

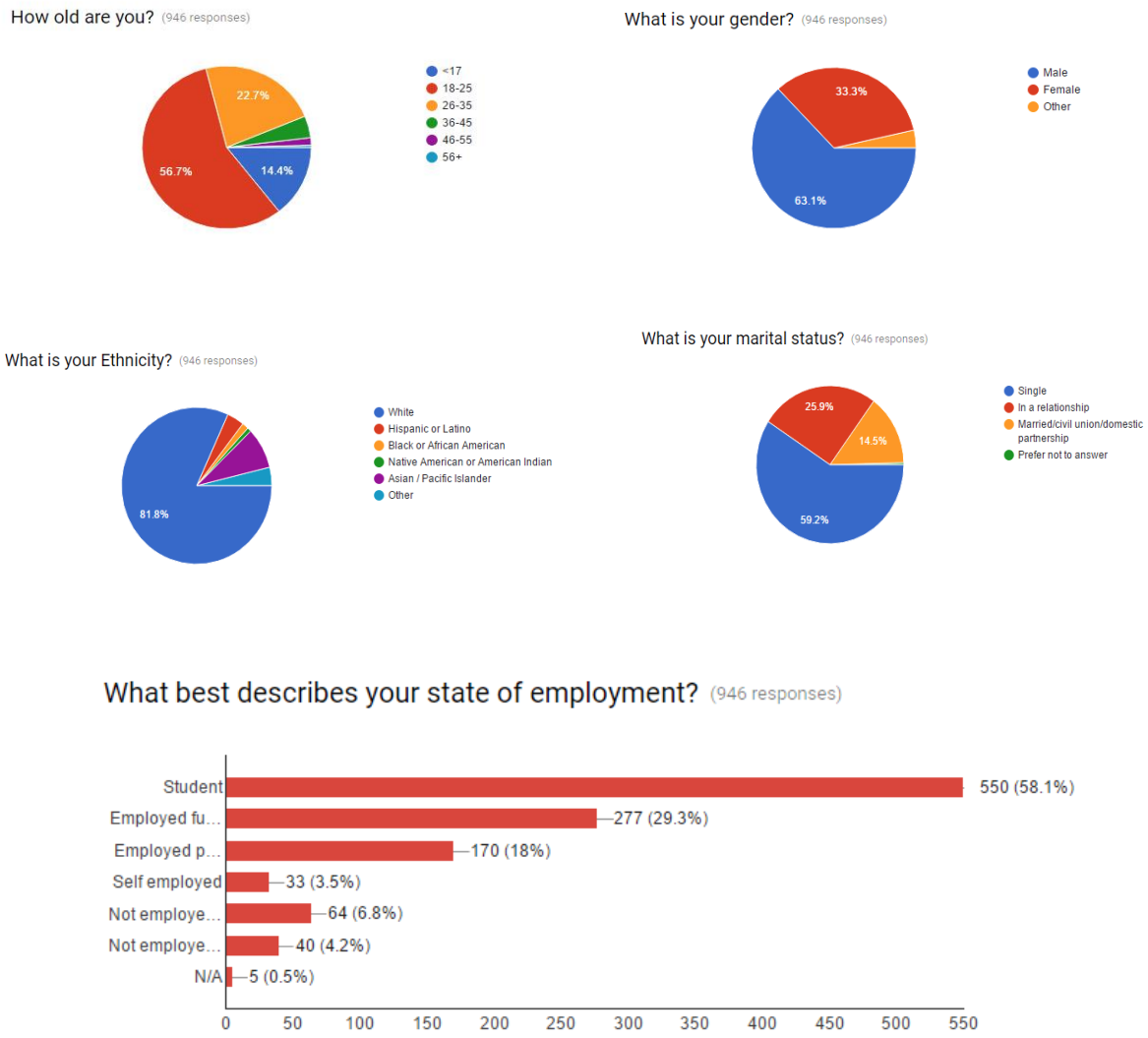
With 250 million unique users each month, Reddit's userbase encompasses a diverse range of backgrounds; however, some patterns emerge when studying its demographics. According to the 2016 user survey conducted on /r/samplesize (n = 946), the average user is a young, white, male, single, American student (**Figure B**).[230] SurveyMonkey Intelligence's statistics on app data for Reddit's official mobile app confirm many of these conclusions; according to app data, 59% of Reddit users are male, 45% are between ages 18 and 29 (an additional 40% between ages 30 and 49), and 46% of users have a college degree or higher (an additional 40% have a high school degree) (**Figure C**).[231] 70 percent of Reddit users said they get news from the site, but post only occasionally.[232]

Based on the proportions of users who have spent 2, 3, and 4 years on the site, Reddit appears to have gained a steady stream of users the last 4 years.

[230] HurricaneXriks, *[OC]The Results of the Reddit Demographics Survey 2016. All data collected from users on /r/Samplesize*, Reddit (Oct. 11, 2016), https://www.reddit.com/r/dataisbeautiful/comments/5700sj/octhe_results_of_the_reddit_demographics_survey/.
[231] Abhinav Agrawal, *The user demographics of Reddit: The Official App*, Medium (Dec. 6, 2016), https://medium.com/@sm_app_intel/the-user-demographics-of-reddit-the-official-app-7e2e18b1e0e1. SurveyMonkey has since shut down its SurveyMonkey Intelligence service.
[232] Michael Barthel, et al., *Seven-in-Ten Reddit Users Get News on the Site*, Pew Research Center (May 26, 2016), http://www.journalism.org/2016/02/25/seven-in-ten-reddit-users-get-news-on-the-site/.

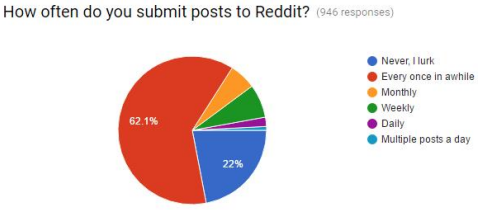
Figure B: Results from the Reddit Demographics Survey 2016 via /r/samplesize.[233]

(1) Reddit demographics.



[233] HurricaneXriks, *[OC]The Results of the Reddit Demographics Survey 2016. All data collected from users on /r/Samplesize*, Reddit (Oct. 11, 2016), https://www.reddit.com/r/dataisbeautiful/comments/5700sj/octhe_results_of_the_reddit_demographics_survey/.

(2) Reddit usage. The majority of Redditors have been on the site for at most 4 years; most users have either never submitted posts, or submit "every once in a while."

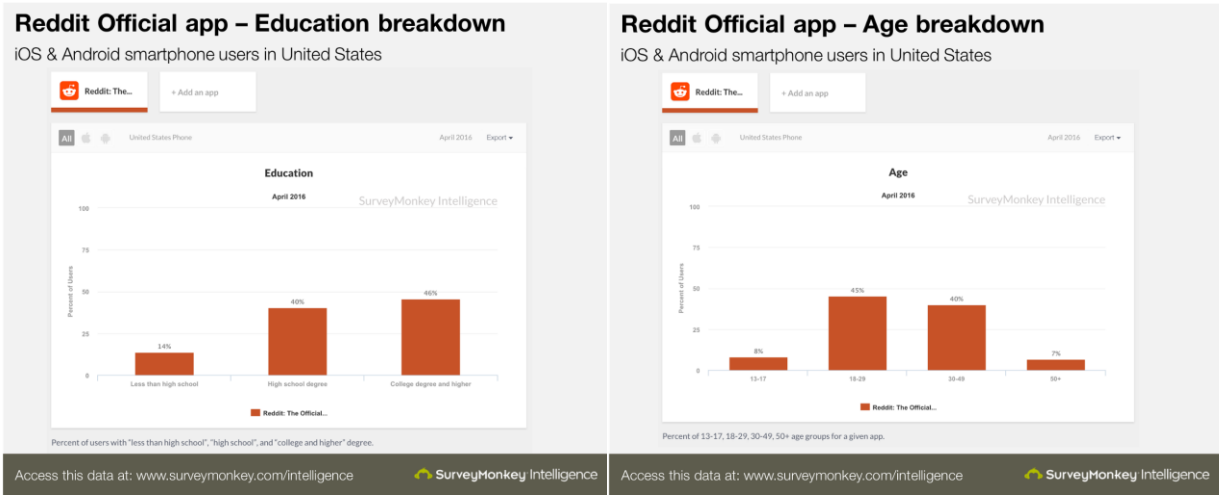


(3) Worldwide Redditor distribution. The legend provides a key for the world map. A clear majority of Redditors are American, though Reddit believes it can reach the "billion-user mark by expanding its platform globally—something it has put little effort into doing until now."[234]

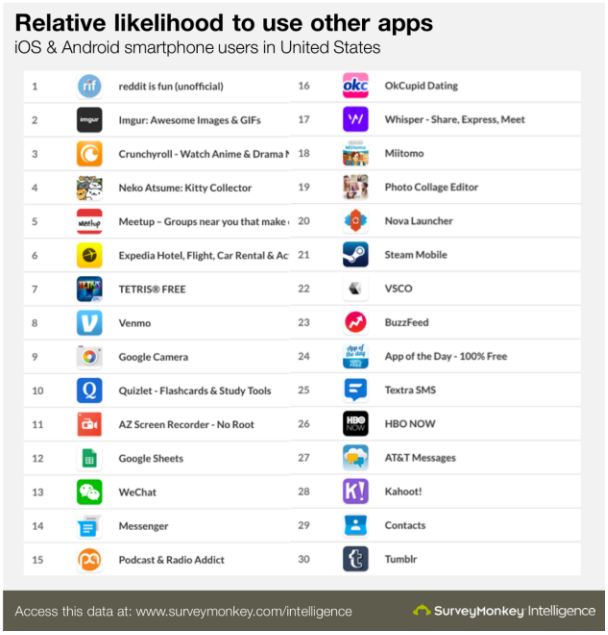
[234] George Slefo, *Reddit Intros New Ad Offering, 'Grows Up' and Says It Can Be as Big as Facebook*, Ad Age (July 26, 2016), http://adage.com/article/digital/reddit-let-marketers-sponsor-users-organic-posts/305115/.

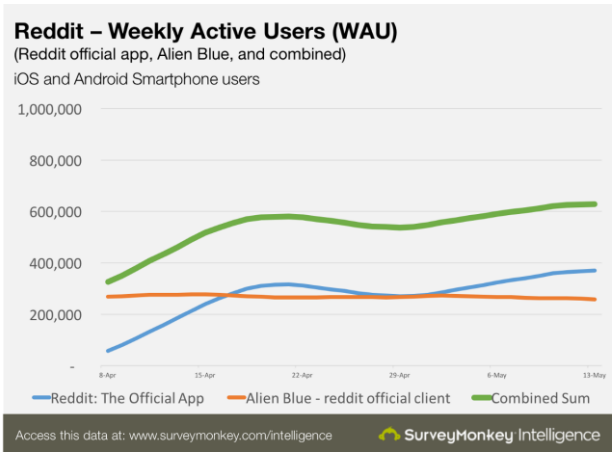
Top 10 Countries	
USA	520
Canada	75
UK	60
Australia	31
Germany	19
The Netherlands	18
New Zealand	11
France	8
Sweden	8
Austria	6

Figure C: SurveyMonkey Intelligence app data for Reddit's official app.[235]



[235] Abhinav Agrawal, *The user demographics of Reddit: The Official App*, MEDIUM (Dec. 6, 2016), https://medium.com/@sm_app_intel/the-user-demographics-of-reddit-the-official-app-7e2e18b1e0e1. SurveyMonkey has since shut down its SurveyMonkey Intelligence service.





VI. Content Propagation Findings

1. Redditors rely on "mainstream media" for breaking news stories, but editorialize discussion.

Despite Reddit's design (opt-in communities organized around topic or ideology—a characteristic that can promote polarization within individual echo-chambers), subreddits (regardless of ideology) tend to rely on mainstream media reporting for breaking news content. Because breaking news reliably leads to competing stories being posted across subreddits within minutes, the stories that are the most upvoted and commented on can provide valuable insight into what users rely on for their news. Accordingly, we tracked a breaking story as news broke and propagated across Reddit: James Comey's dismissal from the FBI.

On May 9, 2017, James Comey was fired from his position as Director of the FBI. News quickly spread to Reddit; within minutes, competing stories had been posted in virtually all major news subreddits. Some Redditors on subreddits (e.g., /r/politics, generally considered centrist/left-leaning) created a self-posted "megathread" aggregating the dozens of news articles appearing in a variety of news sources; Redditors kept this megathread updated virtually to-the-second as more information flowed in.[236] On another of the most popular news subreddits,

[236] PoliticsModeratorBot, *Megathread: FBI Director Comey Fired*, REDDIT (May 9, 2017), https://www.reddit.com/r/politics/comments/6a8jsf/megathread_fbi_director_comey_fired/.

/r/news, a Redditor posted the ABC News article on the situation; over a hundred thousand upvotes and over 24,000 comments followed.[237]

That ABC News article was also reposted to /r/conspiracy and /r/askthe_donald, two right-wing/alt-right subreddits. On /r/the_donald (one of the largest and most-influential right-wing, pro-Donald Trump subreddits), users also looked to "mainstream" media sources over right-wing news sources for their news. /r/the_donald's most upvoted and commented post (17000+ upvotes, 2800+ comments) came from Bradd Jaffy (NBC senior editor); a reporter other /r/the_donald users had called a "fake news [reporter]"[238] and a "whiny littel[sic] b****."[239] A BBC article (8442 upvotes, 298 comments),[240] and AP article (5000+ upvotes, 42 comments)[241] followed closely. In contrast, the /r/the_donald post citing a *Russia Today* piece garnered 343 upvotes and 3 comments,[242] while the corresponding Breitbart article received only 19 upvotes and 2 comments. Notably, that AP article—which was also cited by the moderate/conservative users of /r/conservative—also represented the most-highly upvoted/commented article on both /r/conspiracy and /r/conservative.[243]

However, while subreddits may have near-universally relied on "mainstream" media for their breaking news reporting, they differed greatly in their editorialization and commenting. For example, the AP article, submitted as "President Trump Fired FBI Director Comey" to /r/conservative, instead was given the title "COMEY FIRED!!! REEEEEEEEEEEEEEEEEE!!!!!!!" on /r/the_donald.[244] Commenting (sorted by "best" and/or "top," to bring the most agreed upon viewpoints to the top) took very different turns as well. Predictably, the best/top comments on right-wing/alt-right subreddit tended to be supportive of the President and critical of the mainstream media, the Democratic party, and liberals as a whole. Fringe subreddits sometimes took on a more analytical approach (e.g. /r/law),[245] while varying levels of incredulity and analysis, often critical of the President, dominated on centrist/left-leaning subreddits.[246] In the days that followed Mr. Comey's firing, posted content and

---

[237] JewishDoggy, *James Comey Terminated as Director of FBI*, Reddit (May 9, 2017), https://www.reddit.com/r/news/comments/6a8ji6/james_comey_terminated_as_director_of_fbi/.

[238] khaki54, *Digg.com and Fake News senior editor Bradd Jaffy think Trump can't tell the difference between Kim Jong dynasty members. Forgets that Kim Jong-un wasn't born on the day his father died, and forgets he had previous roles in NK. Shameful.*, Reddit (Apr. 18, 2017), https://www.reddit.com/r/The_Donald/comments/6646yn/diggcom_and_fake_news_senior_editor_bradd_jaffy/.

[239] six5_SMK, *I looked up "whiny littel bitch" and it turns out that this NBC reporter is the literal definition: Bradd Jaffy on Twitter*, Reddit (Mar. 10, 2017), https://www.reddit.com/r/The_Donald/comments/5ynndu/i_looked_up_whiny_littel_bitch_and_it_turns_out/.

[240] freetvs, *FBI Director Comey Sacked by President Trump*, Reddit (May 9, 2017), https://www.reddit.com/r/The_Donald/comments/6a8k0h/fbi_director_comey_sacked_by_president_trump/.

[241] besevens, *COMEY FIRED!!! REEEEEEEEEEEEEEEEEE!!!!!!!*, Reddit (May 9, 2017), https://www.reddit.com/r/The_Donald/comments/6a8juo/comey_fired_reeeeeeeeeeeeeeeeee/.

[242] Boozeman78, Comey Fired, Reddit (May 9, 2017), https://www.reddit.com/r/The_Donald/comments/6a8jp3/comey_fired/.

[243] COMEY FIRED!!! REEEEEEEEEEEEEEEEEE!!!!!!! - Duplicates, Reddit (May 9, 2017), https://www.reddit.com/r/The_Donald/duplicates/6a8juo/comey_fired_reeeeeeeeeeeeeeeeee/.

[244] *Ibid.*

[245] mumbling2myself, Trump fires FBI director James Comey, Reddit (May 9, 2017), https://www.reddit.com/r/law/comments/6a8smx/trump_fires_fbi_director_james_comey/.

[246] The authors recommend that readers read the comment sections of each of the subreddits, linked in the previous 10 footnotes, for a complete picture of how commenting tends to reflect each of the viewpoints of the individual subreddits.

discussion drifted back toward ideological lines.[247] This pattern—the initial spreading of breaking news via mainstream news sources and editorialized comment threads, leading to ideologically-based conclusions and discussion points—reflects the predominant propagation pattern of most important breaking news we reviewed.

2. Redditors tend to spot bogus clickbait from external sources and downvote such content, preventing it from reaching a larger portion of the userbase.

Redditors will also tend to spot news that is verifiably false and downvote or remove it, even if that news would otherwise support their viewpoint. When the satirical news site WTOE 5 (since deleted) released an article with the headline "Pope Francis Shocks World, Endorses Donald Trump for President, Releases Statement,"[248] the article picked up over 960,000 engagements on Facebook.[249] In contrast, when posted to /r/the_donald, the post received numerous downvotes and only 4 comments—quickly ferreting out that the article was fake and from a satirical/fake news website.[250] A similar article from EndingTheFed's was posted on /r/thedonald; it was also quickly downvoted to 0, with the only user comment stating: "Fake news, the pope many times advised against Trump."[251] In another noteworthy example, a fake Denver Guardian (a newspaper which does not exist) story entitled "FBI Agent Suspected in Hillary Email Leaks Found Dead in Apartment Murder-Suicide" was posted to /r/the_donald but only garnered 19 net upvotes and 2 comments;[252] on /r/conspiracy, the most highly upvoted post also quickly called out the Denver Guardian article as "fake," and chastised other users to "vet stuff before you post it."[253]

While it is possible that other bogus clickbait articles did gain traction on the site and have since been deleted, our review of top subreddits found that, to the extent that bogus clickbait does get posted, the subreddit moderators usually remove it quickly if the userbase doesn't address it themselves, making bogus clickbait less of a problem on Reddit than on other platforms.

---

[247] *Cf.* Expert_novice, JAMES COMEY MAINSTREAM TIMELINE!, REDDIT (May 10, 2017), https://www.reddit.com/r/The_Donald/comments/6aenro/james_comey_mainstream_timeline/.
[248] The WTOE 5 news site has been taken down. An Internet Archive version of the article exists and is available here: https://web.archive.org/web/20161115024211/http://wtoe5news.com/us-election/pope-francis-shocks-world-endorses-donald-trump-for-president-releases-statement/.
[249] Hannah Ritchie, *Read all about it: The biggest fake news stories of 2016*, CNBC (Dec. 30, 2016), http://www.cnbc.com/2016/12/30/read-all-about-it-the-biggest-fake-news-stories-of-2016.html.
[250] TinkerMech, *Da frickin' Pope*, REDDIT (July 12, 2016), https://www.reddit.com/r/The_Donald/comments/4shq82/da_frickin_pope/.
[251] Ps1515, *Pope endorses Trump*, REDDIT (Oct. 13, 2016), https://www.reddit.com/r/thedonald/comments/57c4x3/pope_endorses_trump/.
[252] ris4republican, *FBI Anon, Thank You For Your Service and Sacrifice, May President Trump Give You the Presidential Medal of Freedom*, REDDIT (Nov. 9, 2016), https://www.reddit.com/r/The_Donald/comments/5byb2i/fbi_anon_thank_you_for_your_service_and_sacrifice/.
[253] wiseprogressivethink, *FBI Agent Suspected in Hillary Email Leaks Found Dead in Apparent Murder-Suicide*, REDDIT (Nov. 6, 2016), https://www.reddit.com/r/conspiracy/comments/5be1yg/fbi_agent_suspected_in_hillary_email_leaks_found/.

We sent private messages to the moderators of four major news subreddits to get their insights on how they handle misinformation in their communities We received a response from Isentrope, a moderator of /r/worldnews, which has more than 17 million subscribers.[254] "In terms of actual fake news following the Macedonian troll factory model, reddit generally seems to do a good job of getting rid of it," Isentrope said.[255] "Users here and many other parts of the site tend to be very critical of non-reputable sources, although on /r/worldnews, we do see tabloid journalism of dubious quality. Of course, that also depends on whether the source is something that the users want to believe — an issue users across the political spectrum deal with is confirmation bias."

Each subreddit has its own rules, and /r/worldnews is especially restrictive. "On /r/worldnews, we will remove any post where there is affirmative evidence that it is false," u/Isentrope explained. "We also take a very aggressive approach to taking down spam sites which are set up to essentially farm clicks and ad revenue. This takes down some of the worst offenders. Finally, since the sub is straight news only, we do not permit any articles that are analytical/opinionated in nature, eliminating posts where facts may be misconstrued."[256]

3. In contrast to their effectiveness against external misinformation, some subreddits are a major source of conspiracy theories.

Reddit's bigger problem is that some subreddits are a major source of conspiracy theories. Unlike bogus clickbait, conspiracy theories usually have at least some tenuous connection to real facts, and are therefore more difficult to conclusively debunk. The authors of conspiracy theories put much more work into making them believable (and in many cases, probably do in fact believe they are true).

Reddit has often spawned user-generated conspiracy theories. In these cases, Reddit users act as amateur detectives, combing through documents and images to string together a theory. They often trumpet each other's supposed findings and egg each other on to look for more clues to corroborate their claims. Often, but not always, these user-generated conspiracy theories push a particular political narrative. Once these theories gain traction on Reddit, they can get covered by fringe or even mainstream news outlets, further spreading the claims. The news coverage often then finds its way back to Reddit, continuing the cycle. Sometimes the theories start on fringe websites before gaining attention on Reddit and then spreading elsewhere. Reddit administrators will delete threads and even entire subreddits that post personally identifiable information (called "doxing"), but the company has no general policy against spreading baseless conspiracy theories.

[254] *World News*, Reddit, https://www.reddit.com/r/worldnews/ (last visited June 11, 2017).
[255] Direct Reddit message from /u/isentrope, May 19, 2017. Isentrope has been a Reddit user for 6 years and moderates 29 subreddits, including /r/worldnews and /r/politics.
[256] *Id.*

We have examined two case-studies to briefly illustrate how this kind of misinformation spreads on Reddit:

- Boston Marathon Bombing:

While the incident was not strictly speaking a "conspiracy theory," Reddit users played a central role in spreading misinformation about the Boston Marathon Bombing in April 2013. In the chaos following the attack, thousands of users turned to Reddit for news and updates. While many users shared news articles and official statements, others took it upon themselves to try to identify the perpetrators of the attack. Some people tuned into Boston police scanners, sharing whatever they were able to hear, while others scoured photographs from the scene, trying to identify anyone carrying a bag before the blast.[257]

Much of the activity took place on massive threads on /r/news, and users created a whole new subreddit, /r/findbostonbombers, just for the manhunt. Many Twitter users also spread misinformation about the attacks, and there was often a fluid exchange of theories between the two platforms. Based on information they heard on the police scanner, some Reddit users claimed the suspect was "Mike Mulugeta." Although "Mulugeta" had been mentioned on the scanner, the person was unrelated to the bombing and not named "Mike." The officer was just spelling the last name, "M as in Mike."[258]

It's difficult now to reconstruct exactly how various theories began, but at some point, Twitter and Reddit users began to speculate that one of the bombers was Sunil Tripathi, a Brown University student who had been missing for a month. Reddit users flagged a tweet from someone who claimed to have gone to high school with Tripathi who thought she recognized him in surveillance photographs.[259] Tripathi quickly became a top suspect on /r/findbostonbombers, and one comment claimed to hear his name mentioned on the police scanner.[260] Based on the social media rumors, journalists called Tripathi's family all night, leaving dozens of voicemails with his family.[261] Reddit users congratulated each other for solving the case.[262] Police eventually confirmed to media organizations that Tripathi was not a suspect, and his body was ultimately found later that month. He had committed suicide the day he went missing, a full month before the attack.[263]

On April 22, Erik Martin, Reddit's general manager, wrote a blog post apologizing for the misinformation that had spread on the platform. "[T]hough started with noble intentions, some of the activity on reddit fueled online witch hunts and dangerous speculation which spiraled into very negative consequences for innocent parties," Martin wrote. "The reddit staff and the millions of people on reddit around the world deeply regret that this happened. We have

---

[257] Alexis C. Madrigal, *#BostonBombing: The Anatomy of a Misinformation Disaster*, The Atlantic (April 19, 2013), https://www.theatlantic.com/technology/archive/2013/04/-bostonbombing-the-anatomy-of-a-misinformation-disaster/275155/.
[258] *Id.*
[259] Chava Gourarie, *A closer look at the man wrongfully accused of being the Boston bomber*, Columbia Journali□m Review (Oct. 9, 2015), https://www.cjr.org/analysis/sunil_tripathi_was_already_dead.php.
[260] *Id.*
[261] *Id.*
[262] *Id.*
[263] *Id.*

apologized privately to the family of missing college student Sunil Tripathi, as have various users and moderators. We want to take this opportunity to apologize publicly for the pain they have had to endure."[264] The company ultimately shut down the /r/findbostonbombers subreddit.

The incident was a major black eye for Reddit, following a wave of negative publicity it received in 2011 over the /r/jailbait subreddit that featured pictures of scantily-clad young girls (Reddit eventually shutdown /r/jailbait after a public uproar). In an interview with the *New York Times* shortly after the Boston attack, Martin said that the company would be more vigilant in enforcing its rule against posting personal information, but he didn't commit to any changes in company policy.[265]

- Pizzagate:

One of the most widely believed conspiracy theories of 2016-2017 has been "Pizzagate" — the bizarre claim that top Democratic officials ran a child sex abuse ring out of a Washington, D.C. pizzeria. The theory appeared to get its start on Oct. 30, 2016, when a white supremacist Twitter user, citing a supposed police source, claimed that there were emails on Anthony Weiner's laptop exposing a "pedophila ring" [sic] and showing that "Hillary Clinton is at the center."[266] Right-wing and conspiracy-theorist websites quickly picked up the claim, citing the tweet or their own alleged sources. At least two fake news sites run by teenagers in Macedonia also aggregated and amplified the claims.[267] SubjectPolitics.com, a separate right-wing site that regularly promotes false and misleading news, posted a story with the headline, "IT'S OVER: NYPD Just Raided Hillary's Property! What They Found Will RUIN HER LIFE," accompanied by an unrelated picture of police officers carrying bags of evidence.[268] The fabricated story quickly gathered 107,000 shares on Facebooks, and inspired more copycats on other sites.[269]

Soon, users on Reddit, Twitter, and 4Chan (an external discussion board) began combing through hacked Democratic emails (posted online by Wikileaks) trying to find corroborating evidence of a child sex ring. They focused on the hacked email account of John Podesta, the chairman of Clinton's presidential campaign. The amateur sleuths believed there were code words in the emails for child abuse. "Cheese pizza," for example, was supposedly code for "child pornography," while "pizza" was for "girl" and "pasta" was for "boy."[270] Innocent-seeming emails about dinner were thus transformed into nefarious evidence of pedophilia. The theory was especially popular on subreddits including /r/the_donald, /r/conspiracy, and /r/HillaryForPrison.

---

[264] *Reflections on the Recent Boston Crisis*, Reddit (April 22, 2013), https://redditblog.com/2013/04/22/reflections-on-the-recent-boston-crisis/.
[265] Leslie Kaufman, *Bombings Trip Up Reddit in Its Turn in Spotlight,* N.Y. Time□ (April 28, 2013), http://www.nytimes.com/2013/04/29/business/media/bombings-trip-up-reddit-in-its-turn-in-spotlight.html.
[266] Craig Silverman, *How The Bizarre Conspiracy Theory Behind "Pizzagate" Was Spread*, Buzz□eed (Nov. 4, 2016), https://www.buzzfeed.com/craigsilverman/fever-swamp-election.
[267] *Ibid.*
[268] *Id.*
[269] *Id.*
[270] Gregor Aisch, Jon Huang & Cecilia Kang, *Dissecting the #PizzaGate Conspiracy Theories*, N.Y. Time□ (Dec. 10, 2016), https://www.nytimes.com/interactive/2016/12/10/business/media/pizzagate.html.

Users created a separate subreddit, /r/pizzagate, to pour over the supposed evidence. Reddit banned /r/pizzagate on November 24, 2016 for posting personal information, but the other subreddits continued.[271]

The food-based lexicon led Reddit users to focus on Comet Ping Pong, a pizza parlor in Washington, D.C. owned by James Alefantis, who has connections with various Democratic operatives, including his ex-boyfriend, David Brock, the founder of Media Matters for America.[272] Reddit users scrutinized Alefantis's Instagram account and any information they could find out about him or his restaurant. They concluded that a picture of an empty walk-in refrigerator was evidence of a "kill room," and that Comet's basement led to secret underground tunnels (Comet doesn't have a basement). Users believed that two computer-generated police sketches in a 2007 British child abduction case looked like John Podesta and his brother, Tony Podesta (the sketches were in fact of one suspect, not two, and who was reportedly much younger than either Podesta brother at the time).[273] Soon, Alefantis began receiving hundreds of death threats.[274]

Pizzagate escalated into real-world violence on December 4, 2016, when Edgar Maddison Welch, a 28 year-old man, walked into Comet with an assault rifle, pointed it at one of the employees, and fired multiple shots into a locked door.[275] After his arrest, he told police he was investigating the conspiracy theory and wanted to rescue the children he believed were being held captive.[276] Welch pled guilty in March to weapons and assault charges.[277] Many Reddit users, however, dismissed the shooting as a "false flag" — a government cover-up or distraction aimed at discrediting the conspiracy theorists.[278]

One of the most shocking aspects of Pizzagate is how many people believe in it. A YouGov/ Economist poll released on December 27, 2016, found that 46 percent of Trump voters and 17 percent of Clinton voters believed that it was "probably" or "definitely true" that "leaked emails from the Clinton campaign talked about pedophilia and human trafficking – 'Pizzagate.'"[279] The wording of the question clearly matters — a Public Policy Polling survey

---

[271] Abby Ohlheiser, *Fearing yet another witch hunt, Reddit bans 'Pizzagate,'* Wa□h. Po□t (Nov. 24, 2016), https://www.washingtonpost.com/news/the-intersect/wp/2016/11/23/fearing-yet-another-witch-hunt-reddit-bans-pizzagate/.
[272] Glenn Thrush, *David Brock blasts Brooklyn, 'animals' in press*, Politico (Dec. 13, 2016), http://www.politico.com/story/2016/12/david-brock-trump-clinton-media-232562.
[273] Gregor Aisch, Jon Huang & Cecilia Kang, *Dissecting the #PizzaGate Conspiracy Theories*, N.Y. Time□ (Dec. 10, 2016), https://www.nytimes.com/interactive/2016/12/10/business/media/pizzagate.html.
[274] *Ibid.*
[275] Crimesider Staff, *Man pleads guilty in "pizzagate" shooting in D.C.*, CBS New□/A□□ociated Pre□□ (Mar. 24, 2017), http://www.cbsnews.com/news/pizzagate-comet-ping-pong-shooting-washington-dc-guilty-plea-edgar-welch/.
[276] Faiz Siddiqui and Susan Svrluga, *N.C. man told police he went to D.C. pizzeria with gun to investigate conspiracy theory*, Wa□h Po□t (Dec. 5, 2016), https://www.washingtonpost.com/news/local/wp/2016/12/04/d-c-police-respond-to-report-of-a-man-with-a-gun-at-comet-ping-pong-restaurant/.
[277] Laura Jarrett, 'Pizzagate' shooting suspect pleads guilty, CNN (Mar. 24, 2017), http://www.cnn.com/2017/03/24/politics/pizzagate-suspect-pleads-guilty/index.html.
[278] BransonOnTheInternet, Pizzagate: Calling it now Reddit will attempt to shut down any discussion regarding Pizzagate after today's events. We have to remain vigilant and give a voice to the voiceless. Don't be scared into silence, Reddit (Dec. 5, 2016), https://www.reddit.com/r/conspiracy/comments/5gjih4/pizzagate_calling_it_now_reddit_will_attempt_to/.
[279] Kathy Frankovic, *Belief in conspiracies largely depends on political identity*, YouGov (Dec. 27, 2016), https://today.yougov.com/news/2016/12/27/belief-conspiracies-largely-depends-political-iden/.

released on Dec. 9, 2016, found that 14 percent of Trump supporters thought Hillary Clinton was connected to a child sex ring run out of a Washington, D.C. pizzeria, while 32 percent weren't sure.[280]

4. Reddit's model encourages groupthink, leading to isolated pockets of groupthink and echo-chambers that propagate other forms of misinformation.

While conspiracy theories appear to be the most significant form of misinformation on Reddit, it's not the only version of the problem. The subreddit system encourages communities to group into like-minded bubbles. The Reddit algorithm favors the most upvoted content, and the most upvoted content in many subreddits is the content that appeals to the biases of that community in the most sensational way possible. Even if a false article gets corrected, that correction rarely gets nearly as much attention as the original falsehood.

So, while pro-Trump and "alt-right" subreddits have promoted the most outlandish conspiracy theories over the past year, liberal subreddits have hardly been immune from spreading misinformation. Anti-Trump subreddits like /r/MarchAgainstTrump, /r/EnoughTrumpSpam, /r/TrumpforPrison, and /r/Impeach_Trump have copied /r/the_donald's style of posting images and memes without any source material (as opposed to actual news articles). This content is often misleading or outright false. Those subreddits also have hosted unverified and bizarre claims, such as articles from Patribotics.blog, a website run by Louise Mensch. On May 20, 2017, for example, Mensch reported that the Marshal of the Supreme Court had notified President Trump that impeachment proceedings had begun against him.[281] (This is, needless to say, not how the impeachment process works.) That particular article appeared on both /r/MarchAgainstTrump and /r/EnoughTrumpSpam, although neither article received a large amount of attention.

VII. Content Moderation

Reddit's content policy bans content if it:

- Is illegal;
- Is involuntary pornography;
- Encourages or incites violence;
- Threatens, harasses, or bullies or encourages others to do the same;
- Is personal and confidential information;

[280] *Trump Remains Unpopular; Voters Prefer Obama on SCOTUS Pick*, Public Policy Polling (Dec. 9, 2016), http://www.publicpolicypolling.com/pdf/2015/PPP_Release_National_120916.pdf.
[281] Louise Mensch & Claude Taylor, *EXCLUSIVE: Judiciary Committee Considering Articles of Impeachment*, PATRIBOTIC☐ (May 20, 2017), https://patribotics.blog/2017/05/20/exclusive-judiciary-committee-considering-articles-of-impeachment/.

- Impersonates someone in a misleading or deceptive manner; or,
- Is spam.[282]

The site's administrators enforce the policy with warnings, temporary account suspensions, removal of content, and even permanent bans of entire subreddits.[283] If a subreddit's content is "extremely offensive or upsetting to the average redditor," the company can "quarantine" it, which requires that users have a verified email address and explicitly opt-in to access the community.[284] All ads are also removed from the quarantined subreddit, but the penalty is rarely imposed.

Reddit's content policy doesn't prohibit false information or conspiracy theories, and the company has largely shifted the responsibility to the volunteer subreddit moderators to police problems beyond the company's explicit rules. "There's all kinds of different content on the site, from discussions of creepy story-telling to UFOs to political matters," said a Reddit spokesperson who asked not to be named in this report. "Reddit is a platform and therefore allows users to post content that they're interested in. As long as it doesn't break our specific rules, it's allowed on the platform."[285] The spokesperson explained that moderators are ultimately "the ones who determine the content on their communities."

Reddit banned /r/pizzagate for posting personal information about one month before the restaurant shooting, but that step enraged the conspiracy theory's supporters on other parts of the platform.[286] Much of the anger was directed at Reddit CEO Steve Huffman, with users accusing him of censorship and covering up a child sex ring. During this backlash, Huffman edited some comments attacking him by replacing references to his username ("spez") with the usernames of moderators of /r/the_donald. Huffman quickly acknowledged that he was responsible and apologized. "It's been a long week here trying to unwind the /r/pizzagate stuff. As much as we try to maintain a good relationship with you all, it does get old getting called a pedophile constantly," he wrote in a comment in /r/the_donald. "As the CEO, I shouldn't play such games, and it's all fixed now. Our community team is pretty pissed at me, so I most assuredly won't do this again."[287]

In a later interview on the podcast *Reply All*, Huffman explained that he thought that he was "extending [r/the_donald users] an olive branch, which was like, 'Hey, I'm a real person, like, I can play this game, you know, we have something in common in that we can toy with people on the internet.'"[288] But /r/the_donald's moderators and users didn't appreciate Huffman's attempt at humor, and they are still widely suspicious and hostile toward Huffman and other

[282] *Reddit Content Policy*, Reddit, https://www.reddit.com/help/contentpolicy/ (last visited June 15, 2017)
[283] *Ibid.*
[284] *Quarantined Subreddits*, Reddit, https://reddit.zendesk.com/hc/en-us/articles/205701245 (last visited June 15, 2017).
[285] Telephone interview with Reddit spokesperson, May 24, 2017.
[286] OhSnapYouGotServed, *Announcement: In regards to today's shutdown of /r/pizzagate*, Reddit (Nov. 23, 2016), https://www.reddit.com/r/The_Donald/comments/5ee6gs/announcement_in_regards_to_todays_shutdown_of/dabqpof/.
[287] spez, Comment to *The Admins are suffering from low energy - have resorted to editing YOUR posts. Sad!*, Reddit (Nov. 23, 2016), https://www.reddit.com/r/The_Donald/comments/5ekdy9/the_admins_are_suffering_from_low_energy_have/dad5sf1/.
[288] *#83 Voyage Into Pizzagate*, Reply All, Gimlet Media, (Dec. 8, 2016), https://gimletmedia.com/episode/83-voyage-into-pizzagate/.

Reddit administrators. For example, in May 2017, /r/the_donald briefly went private in protest of Reddit removing three moderators for content policy violations.[289]

This saga highlights the complicated and often strained relationship that Reddit has with some of the unsavory portions of its user-base. While it doesn't want to antagonize a large segment of heavy Reddit users, it also doesn't want to let controversial content spook advertisers. In a 2016 interview with *AdAge*, Huffman and Ohanian said they have ambitious plans to grow their company's userbase and advertising revenue to rival that of Facebook.[290] For example, Reddit has already rolled out a feature to allow advertisers to boost the exposure of user-created posts that are favorable to the advertiser's brand.[291]

Sensitive to the desires of advertisers, Reddit has created a "no ads list" (separate from the quarantine list) that ensures major brands don't have their ads placed next to controversial content. Following the shooting at Comet Ping Pong, Reddit added /r/conspiracy to its "no ads list," although it didn't quarantine or ban the community.[292] Reddit also gives advertisers tools to avoid certain keywords, like "Donald Trump."[293] While these measures might reassure advertisers, they don't address the core problem of misinformation on Reddit. The false information is still just as available to users — it just doesn't have ads next to it.

The company has taken some steps recently that limit the exposure of controversial subreddits. Throughout much of 2016, content from /r/the_donald frequently dominated the front-page of Reddit, irritating users who weren't Trump supporters or didn't care about politics.[294] In June, Huffman announced that Reddit was tweaking its algorithm to "promote more diversity in the feed, which will help provide more variety of viewpoints and prevent vote manipulation."[295] Those changes appeared to reduce (but hardly eliminate) the popularity of Trump content on the front page.

On February 15, 2017, Reddit announced that all logged-out users who visit the main homepage will see a version of the platform called "popular." This mode excludes pornographic content, as well as "a handful of subreddits that users consistently filter out."[296] In response to a user question, a Reddit administrator explained that "large and dedicated [subreddits] to specific games are heavily filtered, as well as specific sports, and narrowly focused politically related subreddits, etc."[297] The mode now filters out /r/the_donald, as well as several large anti-Trump

---

[289] stopscopiesme, *The_donald has gone private in protest of their clash with the admin*s, Reddit (May 20, 2017), https://www.reddit.com/r/SubredditDrama/comments/6c7utq/the_donald_has_gone_private_in_protest_of_their/.
[290] George Slefo, *Reddit Intros New Ad Offering, 'Grows Up' and Says It Can Be as Big as Facebook*, AdAge (July 26, 2016), http://adage.com/article/digital/reddit-let-marketers-sponsor-users-organic-posts/305115/.
[291] *Ibid.*
[292] George Slefo, *Amid 'PizzaGate,' Reddit Moves to Prevent Ads From Appearing on Conspiracy-Driven Topics*, AdAge (Dec. 6, 2016), http://adage.com/article/digital/reddit-taking-preventive-steps-pizzagate/307040/.
[293] *Ibid.*
[294] Benjy Sarlin, *Donald Trump's Reddit Fan Club Faces Crackdown, Infighting*, NBC New☐ (July 1, 2016), http://www.nbcnews.com/politics/2016-election/donald-trump-s-reddit-fan-club-faces-crackdown-civil-war-n601441.
[295] spez, *Let's talk about Orlando*, Reddit (June 13, 2016), https://www.reddit.com/r/announcements/comments/4ny59k/lets_talk_about_orlando/
[296]simbawulf, *Introducing /r/popular*, Reddit (Feb. 16, 2017), https://www.reddit.com/r/announcements/comments/5u9pl5/introducing_rpopular/ddsczx/.
[297] *Ibid.*

subreddits, but users can still visit those pages or subscribe to them.[298] Previously, logged-out users only saw select "default" subreddits, but many users chose to visit /r/all, which contained everything on the site. The new "popular" version provides more diverse content than the "default" mode and may lead fewer users to go to /r/all. Limiting the exposure of these controversial subreddits could slow the spread of misinformation to the public, although dedicated users can still seek out the content.

Reddit is also creating new features to encourage journalists to interact with the site. In May 2017, the *Washington Post* became the first news organization to create a public page on Reddit to share its stories and engage with users.[299] The *Washington Post*'s page isn't nearly as popular as large news subreddits, but Reddit is hopeful that more news organizations will follow, allowing redditors to interact with journalists.[300] The feature could ultimately create a kind of verification system for news organizations, helping users to know that certain pages will contain reliable information.

Quantitative Studies: External

One potentially promising method for combating misinformation on Reddit is to use "nudges" to prompt users to be more skeptical of unreliable news content. In their book *Nudge: Improving Decisions About Health, Wealth, and Happiness*, Richard Thaler, a University of Chicago economist, and Cass Sunstein, a Harvard Law School professor, argue that institutions can alter human behavior by reframing choices.[301] This approach, they argue, can lead to more positive behavior without constraining individual liberty.

Inspired by the nudge theory, J. Nathan Matias, an MIT researcher, conducted an experiment with the help of the moderators of /r/worldnews, one of the largest news subreddits.[302] The hypothesis was that "sticky" comments (automated messages that appear at the top of a discussion thread) could cause users to downvote unreliable content or post their own comments with refuting evidence. This change in behavior could limit the spread of misinformation on Reddit, Matias theorized.

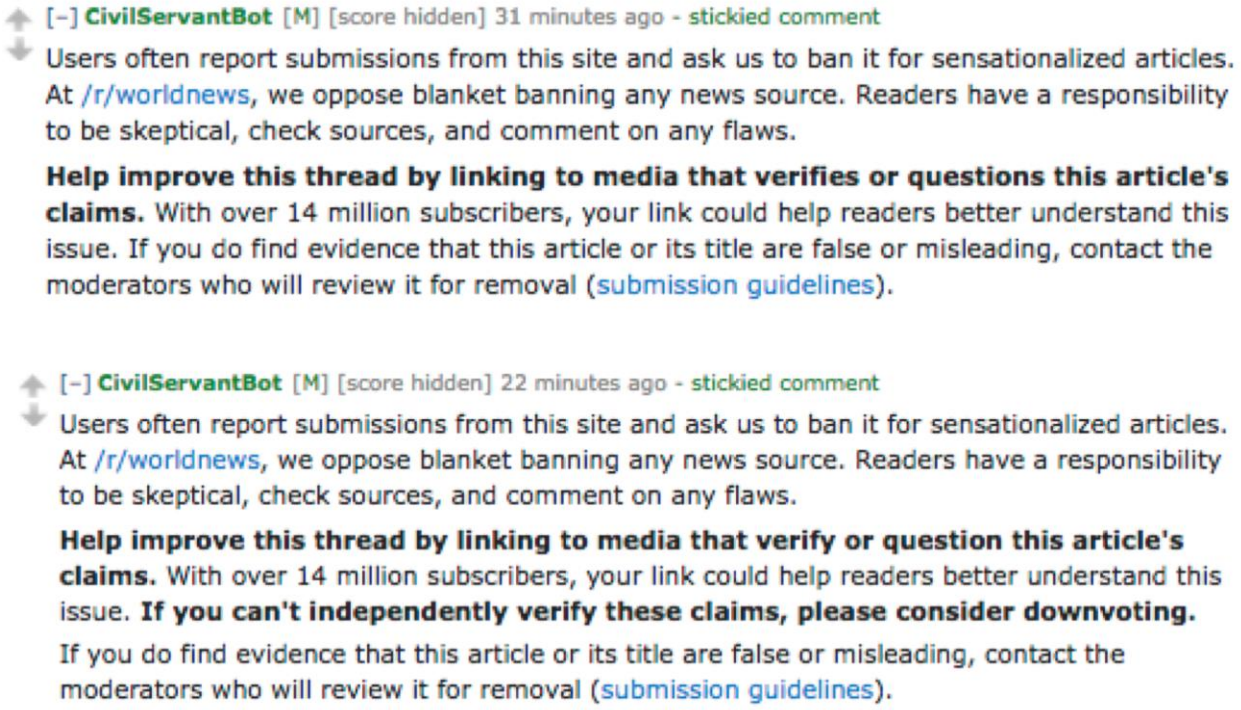
First, Matias and the subreddit moderators compiled a list of eight new sources "that frequently receive complaints" about "sensationalized headlines and unreliable information." Because /r/worldnews focuses on international stories, there was just one U.S. outlet—the *New York Post*—while the other sources were British and Australian tabloids. From Nov. 27, 2016 to

---

[298] *Ibid.*
[299] WashPostPR, *The Washington Post is the first national news publisher to debut a Reddit profile page*, WA□H. PO□T (May 17, 2017), https://www.washingtonpost.com/pr/wp/2017/05/17/the-washington-post-is-the-first-national-news-publisher-to-debut-a-reddit-profile-page.
[300] Interview with Reddit spokesperson, May 24, 2017.
[301] RICHARD H. THALER & CA□□ R. SUN□TEIN, NUDGE: IMPROVING DECI□ION□ ABOUT HEALTH, WEALTH, AND HAPPINE□□, (Penguin Books rev. & exp. ed. 2009).
[302] J. Nathan Matias, *Persuading Algorithms with an AI Nudge*, MEDIUM (Feb. 1, 2017), https://medium.com/mit-media-lab/persuading-algorithms-with-an-ai-nudge-25c92293df1d.

Jan. 20, 2017, each new link from one of those eight sources was randomly assigned (a) no sticky comment, (b) a sticky comment encouraging skepticism, or (c) a sticky comment encouraging skepticism and voting. In total, Matias studied 840 posts from those eight sources.

Figure D: Examples of the sticky comments assigned to content. Note the additional difference between (1) and (2), asking users to downvote sensationalized articles.



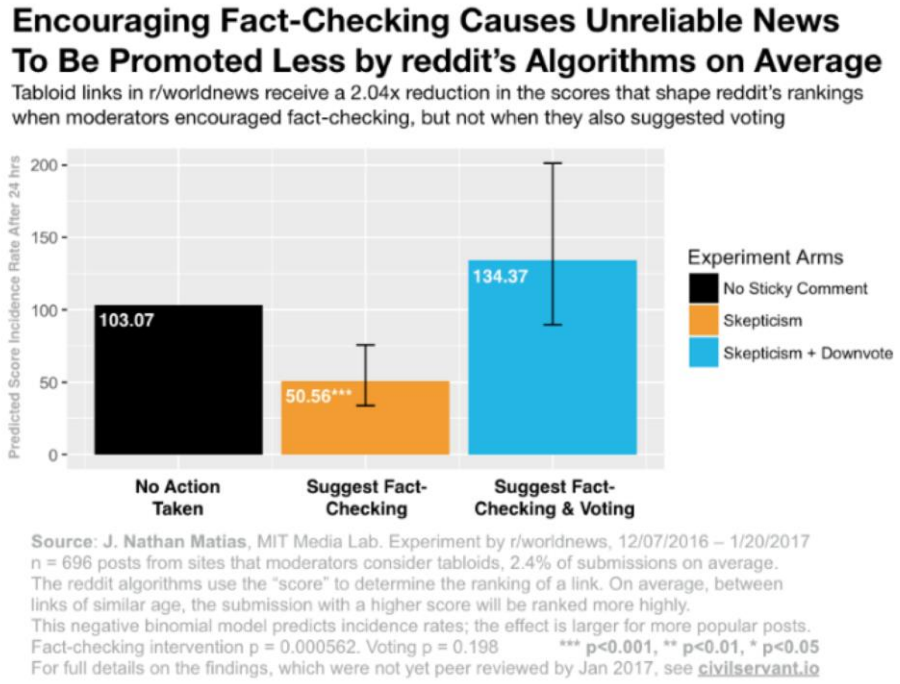
He found that both messages caused users to post more comments citing outside material. On average, the sticky comment encouraging fact-checking resulted in a 201 percent increase in comments with links, while the sticky-comment encouraging fact-checking and down-voting resulted in a 203 percent increase in comments with links. This is a promising result. It suggests that a prominently placed reminder to users to be skeptical of content can cause users to think more critically and to look for refuting or supporting evidence. Other users who visit the discussion page will then see the original source, along with links to outside evidence on the topic.

Matias also found that the sticky comment encouraging fact-checking decreased the overall Reddit score of the tabloid material. After 24 hours, submissions with that sticky comment received on average 49.1 percent of the score of submissions with no comment. The Reddit score is important, not only because it reflects how users are reacting to the content, but because submissions with lower scores get seen by fewer people. Decreasing the Reddit scores of misleading and false submissions means they receive less exposure on the site. The Reddit

score reflects total upvotes and downvotes, but the company keeps the precise details of the algorithm confidential.

Interestingly, however, the sticky-comment encouraging downvoting in addition to fact-checking appeared to backfire with users. Matias found no statistically significant change in Reddit score on submissions with that sticky-comment compared to submissions with no-sticky comment at all. This odd finding might be a result of Reddit's contrarian user-base that doesn't appreciate being told what to do. "It's possible that some people disliked the idea of moderators encouraging downvoting and decided to do the opposite," Matias wrote in a blog post about his study.[303]

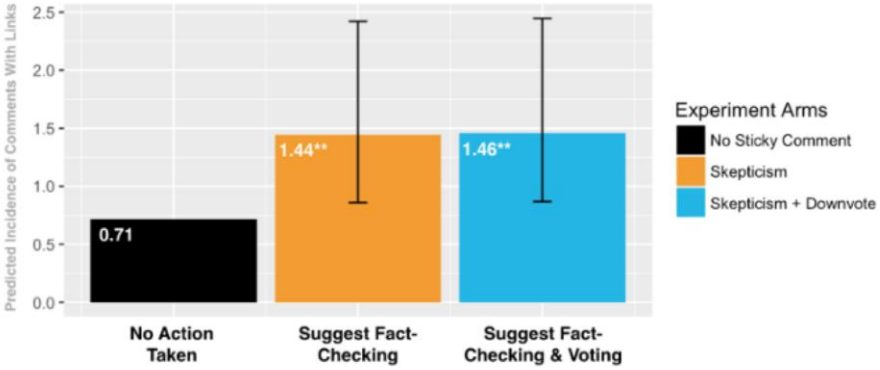
Figure E: Redditor Response to Fact-Checking Encouragement.



[303] *Ibid.*

Encouraging Fact-Checking Causes Unreliable News To Receive 2x More Comments with Links on Average

Tabloid links in r/worldnews receive a 2.01 to 2.03x increase in the number of comments including links to further evidence when moderators use sticky comments to encourage fact-checking.



Source: **J. Nathan Matias**, MIT Media Lab. Experiment by r/worldnews, 11/27/2016 – 1/20/2017
n = 840 posts from sites that moderators consider tabloids, 2.4% of submissions on average.
This negative binomial model predicts incidence rates; the effect is larger for more popular posts.
Fact-checking: p = 0.0083. Fact-checking + Voting: p = 0.0073 ***** p<0.001, ** p<0.01, * p<0.05**
For full details on the findings, which were not yet peer reviewed by Jan 2017, see civilservant.io

Quantitative Studies: Internal

5. Moderation Policies, in a Vacuum, Have an Effect on Redditor Behavior.

Our own studies appear to confirm the impression that moderation policies have an effect on Redditors' behavior. Using an Amazon Mechanical Turk survey, we asked 169 users about their Reddit experience—specifically, whether they had (a) noticed, and (b) had their behavior changed by various moderator-enacted measures regarding content or commenting standards: (1) sidebar information (information on the right-hand side of a subreddit, where a subreddit's general information typically is presented), (2) sticky posts (which remain at the top of any page, regardless of how they rank under the Reddit voting algorithm), (3) notifications when upvoting/downvoting (that instantly pop up when a user mouseovers the upvote/downvote buttons as a real-time reminder), and (4) lists of verified or approved high-quality sources of content (**Figure F**).

Figure F: Redditor Response to Moderation Measures

Moderation Measure	% Who Noticed	% Who Changed Behavior
Sidebar Information	31.19%	14.68%
Sticky Posts	35.19%	10.19%
Upvote/Downvote Notifications	25.00%	9.78%
Verified/Approved High-Quality Content Source List	20.93%	10.47%
None of the Above	11.32%	33.96%

The results are striking—although no one moderation measure was noticed by over 35.19% of Redditors surveyed, in aggregate, 88.68% of Redditors noticed at least one moderation measure, and 66.04% of Redditors surveyed had changed their behavior by at least one of the moderation measures. Some measures (e.g. sidebar information, which users probably read to establish behavioral and content expectations when entering a community) were more

effective than others (e.g. upvote/downvote notifications, which only pop up when a user mouseovers the upvote/downvote buttons, and thus probably only remind the user of a community policy after s/he has already made a decision to upvote/downvote something). While no single measure was unilaterally effective, and while there may be a smaller group of Redditors no moderation measure will affect (e.g. some subset of the 11.32% who did not notice a moderation measure, or of the 33.96% who did not change their behavior), in concert with each other, each of these moderation measures can play an important role in helping moderators moderate content and community.

6. Implemented Moderation Policies are Promising.

We also completed a study of moderation policies across 116 subreddits, consisting of the superset of default subreddits, several news-related subreddits, and a collection of other popular and relevant subreddits.[304] For each subreddit, we determined the following characteristics:

- # of subscribers
- Default status (yes/no)
- Popularity (One of the Top 100 most-subscribed subreddits?)
- General source of content (outside content vs. self-generated)
- News-related (yes/no)
- Content-moderation, via:
    - Sidebar post
    - Sticky post at the top of the subreddit
    - Sticky comment at the top of each comment page
    - Pop-up reminding users of upvote/downvote policies
    - Pop-up reminding users of commenting policies
- The general sense of content/source of content on news-related subreddits

We then asked users for feedback on subreddits they trusted and distrusted for content. User responses tended to be partisan (e.g. no users trusted both /r/politics, known to be centrist-left, and /r/the_donald, known to be far-right/alt-right), but /r/worldnews and /r/news emerged as the most commonly trusted subreddits for news (showing up in 43 and 27 out of 169 responses, respectively). On the other end of the spectrum, 51 respondents mistrusted /r/the_donald than any other.

While the sample size is too small to draw meaningful conclusions, we would like to note that /r/worldnews and /r/news have implemented more moderation policies than /r/the_donald (specifically, /r/worldnews has sidebar information, sticky comments, and voting pop-ups based

[304] The full study is unwieldy and does not fit well as a data table in this report. We have made it available for viewing here: https://docs.google.com/spreadsheets/d/1_ld_spU27dWxFEDEGY9a5U31jRXNfjVniJX6QGwZcsI/edit#gid=0.

on quantitative studies they have researched or read, while /r/news has sidebar information, upvote/downvote pop-up reminders, and commenting reminders enabled as well; in comparison, /r/the_donald only has sidebar information). We suspect that the more active moderation from moderators creates an environment that Redditors have found more conducive to trustworthy sources of news, but recommend further study (e.g. an A-B test to random samples of Redditors on a particular subreddit, testing the effect of moderation measures either alone or in concert with one another) to confirm our intuition.

VIII. Conclusions/Further Study

Recommendations for Reddit

1. Implement Reminders About Misinformation

Based on the Mathias study, we recommend that Reddit administrators work with subreddit moderators to develop reminders encouraging users to be skeptical of content and to provide refuting evidence of misinformation. The company should experiment with different messages to discover the most effective way to word these reminders and should share that data with subreddit moderators. According to Reddit's spokesperson, company administrators already share resources and tools with the moderators to improve their communities.[305] More widespread use of these reminders could result in more down-votes for false content (limiting its exposure) and more outside links in the comments, providing users access to true information.

These reminders could be attached to sites that regularly get flagged for false content (like in the Mathias study) or could get automatically attached to all content as a generalized reminder. Further study is needed to determine the most effective model for the reminders. While some subreddit moderators will likely be interested in this kind of tool, others will be hostile to it. Reddit could override the moderators' wishes and attach the reminders to all subreddits, but the company may be reluctant to engage in this kind of interference in subreddits. The Mathias study focused on "stickied" comments, but there are variety of places on a Reddit page that could host this kind of reminder.

Additionally, Reddit could follow the Facebook model and attach a warning to content that fact-checkers flag as false. This warning could appear next to the headline, so that users see it even if they don't click through to the comment section. But the company should be careful not to use these kinds of warnings if they only give the content more attention. Additionally, Reddit's formula heavily prioritizes new content, so it might not be possible for independent fact-checkers to review the content quickly before it goes viral and dies off on Reddit (with the damage already done).

[305] Telephone interview with Reddit spokesperson (May 24, 2017).

Some subreddit communities, which view the mainstream media and Reddit administrators with suspicion, would likely react with hostility to any mandated reminders about misinformation. So, while reminders might be helpful on subreddits like /r/worldnews, those have never been the communities that have been spreading the bulk of misinformation on Reddit.

2. Limit Exposure of Problem Subreddits

We recommend that Reddit expand its efforts to limit the casual user's exposure to misinformation-spreading subreddits that they haven't sought out. Changes to Reddit's algorithm in 2016 appeared to significantly reduce the /r/the_donald's dominance of the front-page. We don't have access to data on the new "popular" mode, but it does filter out the subreddits on both the right and the left that are the most responsible for pushing false information. We encourage Reddit to make further changes to its algorithm so that subreddits that push false information appear less prominently in /r/all. Reddit could make broader use of its "quarantine" feature to require users to affirmatively opt-in to access problematic subreddits. While some people will always find dark corners of the internet to share conspiracy theories, Reddit can at least prevent that false information from gaining larger audiences through its popular homepage. And because many users don't want to be exposed to that content in the first place, it seems it is in Reddit's interest to contain the reach of the controversial subreddits.

3. Ensure real news loads quickly; possibly slow load times for unvetted or demonstrably false news

Although it was not a focus of our research, we believe that users on Reddit (and other social media sites) are more likely to engage with content that loads quickly. This is especially important on mobile devices, which tend to load more slowly overall. Giving reliable news content an edge of a few seconds in load times over unreliable content could cause it to become more popular and spread more broadly. It might also encourage more users to read the details of fully loaded stories instead of just scanning the headlines of partially loaded stories. Images with quotes and other low-effort (and unreliable) content tend to perform well on Reddit and other platforms. Decreasing the load times on news articles and presenting them in a way that is easy to access could help close that gap.

Facebook already has a similar program called "Instant Articles," in which Facebook hosts the news content on its own site and shares ad revenue with participating publishers.[306] We recommend more research to see if Reddit or other social media platforms could adapt a version of this program aimed at addressing the problem of misinformation. For example, sites that regularly post reckless and false information could be excluded from the program. This recommendation would require collaboration with willing news publishers to succeed.

[306] Instant Articles, FACEBOOK, https://instantarticles.fb.com/.

4. Consider Other Options

There are more aggressive options for tackling false information that Reddit appears unlikely to take. For example, the company could change its content policy to ban baseless conspiracy theories and debunked claims. But like many platforms, Reddit is reluctant to become an arbiter of truth. Because of the small size of its staff, it is also ill-equipped to engage in large-scale review of the content on its site.

It's unclear what the consequences would be if Reddit banned a handful of the largest subreddits that promote false content. Users may quickly create new subreddits, creating a whack-a-mole problem, where Reddit is constantly required to assess the veracity of content. Users might also just turn to more extreme and isolated communities. After the banning of /r/pizzagate and other subreddits, for example, some Reddit users began using Voat — essentially a Reddit clone with an even more lax attitude toward extreme content.

In the *Reply All* interview, Huffman explained that he believes that banning /r/the_donald would only exacerbate political divisions and drive the Trump supporters to other online forums. "When we take a step back and look at the entire country, and we saw this play out through the last election, there's about half the country that feels alienated and unheard. And they elected a candidate, despite many flaws, that they felt like was giving them a voice," Huffman said. "And the last thing we want to do is take away that voice. We will do we the best we can to provide them that voice, and do our best to prevent the toxic minority of users from undermining it."[307]

There are many other steps that Reddit could also consider, such as altering its algorithm to explicitly favor news content from established organizations or demoting less reliable content. By more subtly adjusting the content that users view, Reddit might be able to avoid the backlash it would face by abruptly banning subreddits. Another option could be displaying related articles next to the content that users click on. Reddit could create an algorithm that would only show mainstream news content in the related section, so that any false news story would likely appear with real news stories debunking it directly next to it. By making the feature seem like it's just "related articles" (as opposed to Reddit fact-checking stories), it might dampen user backlash.

[307] *#83 Voyage Into Pizzagate*, Reply All, Gimlet Media, (Dec. 8, 2016), https://gimletmedia.com/episode/83-voyage-into-pizzagate/.

Figure G: Recommendations Summary

Recommendation	Feasibility	Efficacy	Likely User Response
Work with moderators to implement moderation measures that educate and remind users about false content and misinformation	Will require support from either Reddit or subreddit moderators	Could cause users to check outside information and share with others	Likely welcomed by the general userbase
Limit exposure of problematic subreddits	Will require more aggressive action from Reddit	Could reduce the general userbase's exposure to misinformation	Likely welcomed by the general userbase, but fringe and/or targeted elements may rebel—loudly
Decrease load times of news articles	Will require collaboration between Reddit and media organizations	Could increase engagement on true stories	Likely welcomed by user-base
Consider banning subreddits and users that propagate misinformation	Will require persuading Reddit to change fundamental company strategy	Dependent on specific measures considered and implemented	Dependent on specific measures considered and implemented

IX. Next Steps / Areas for Further Research

One area for potential further research is examining how users respond to encountering links and additional evidence in the comments section of a Reddit post. We are assuming that more links are better because they presumably provide outside evidence that might refute false information in the original post. It is possible, however, that users continue to hold false beliefs even when they encounter contradictory information in the comment section. It is also possible that the links themselves contain false information. Matias acknowledged that this was a potential shortcoming of his study, explaining that he looked "at outcomes within discussions rather than individual accounts, so we can't know if individual people were convinced to be

more skeptical, or if the sticky comments caused already-skeptical people to investigate and share."[308]

If future researchers are interested in the effectiveness of sticky comments, another source of data could be the civility reminder that is now attached to all articles in the /r/politics subreddit. The comment reminds users to "be courteous to others," to "attack ideas, not users," and that "personal insults, [...] hate speech, and other incivility violations can result in a permanent ban." Examining whether this reminder has affected the civility of debate on /r/politics could inform similar reminders about false information.

More data on Reddit user behavior would also be helpful. How many users scan headlines without clicking, how many go to the comments, and how many read the actual article? Reddit may be reluctant to share this data, but it could help guide strategies for combating misinformation. For example, the sticky comments will have only limited effectiveness if most Reddit users never read the comments. Would a warning next to the headline prove more effective? How would users react to that kind of false information warning system? Data from Facebook's false news fact-checking system might be helpful in informing strategies for combating misinformation on Reddit. Additionally, it would be helpful to know whether users trust major media outlets and whether some kind of verification symbol would help users differentiate reliable and unreliable sources. Reddit already indicates the source of content in small gray print next to the link, but is that something users even notice?

More research could also inform whether it would be effective to ban problem subreddits altogether. How might users react? Would it be equally effective to ban particular users who post excessive false content or will they just create new accounts? How much false content is spread by only a handful of accounts?

X. Conclusion

As the self-proclaimed "front page of the internet," Reddit plays a central role in the modern media ecosystem. Millions of people get their news every day from Reddit. It is therefore crucial to understand how Reddit spreads information and misinformation.

There is no silver bullet for solving the problem of fake news on Reddit or other platforms. Our research, however, indicates that reminders could encourage users to be more skeptical of content and to share links to refute false information. These links could better inform other Reddit users, limiting the spread of the misinformation. We also encourage Reddit to reduce the exposure of subreddits that regularly propagate false information and to ensure that real news loads quickly. Because of Reddit's reliance on advertising (and its ambition to expand its revenue), the most effective way to push Reddit to take action on misinformation is likely through pressure on its advertisers.

[308] J. Nathan Matias, *Persuading Algorithms with an AI Nudge*, MEDIUM (Feb. 1, 2017) https://medium.com/mit-media-lab/persuading-algorithms-with-an-ai-nudge-25c92293df1d.

Section 6. Democratic Implications of Misinformation

I. Introduction

The online platforms Facebook, Twitter, Reddit, and Google represent one part of a digital revolution sweeping through nations at many different stages of political development. These platforms unlock the potential for frank and unmoderated discourse that runs counter to the elaborate political theater which sometimes typified persuasive communication in earlier eras. Yet we must also recognize that these platforms have set out to change the way we receive information not necessarily for the benefit of the world, but through a business model that profits from the new status quo they have constructed.[309]

The changes that have taken place are seen, for instance, in the contrast between spontaneous tweets and postings of live events made possible by the Internet and the somber scene of Buddhist monk, Thich Quang Duc, burning himself alive in a 1963 protest against the Vietnam War. The latter was carefully scripted with fellow Vietnamese monks carrying signs written in English for American television audiences.[310] Where the Quang Ducspectacle was publicized to reporters in advance, today bystanders are able to tweet and live stream such grisly events live. The democratized marketplace of ideas has amplified exponentially in the age of the Internet.

Indeed, speaking from the distance of two millennia, the Greco-Roman political philosopher Polybius warned "[W]hen a new generation arises," after many years of democratic rule, it often results in developments which "tempt[] and corrupt[] the people in every possible way."[311] These online sites may present a danger to the democratic state that is in the spirit of what Polybius foresaw where democracy is consumed by unrestrained popular force.[312]

In the traditional media environment, the value of information was inherently graded by political and economic forces, with seemingly reputable sources such as *The New York Times* or *TIME Magazine* holding front row status at newsstands for easy viewing and purchase, while more fringe, specialized, or less popular sources would be relegated to the back of the newsstand, there

---

[309] N. Persily, *Can Democracy Survive the Internet?* JOURNAL OF DEMOCRACY, April 2017, 74.

[310] J. M. Bessette, AMERICAN GOVERNMENT AND POLITICS: DELIBERATION, DEMOCRACY AND CITIZENSHIP. Belmont: Wadsworth. 2014, 554.

[311] Polybius, HISTORIES, VI.II "On the Forms of States." Translation by W.R. Paton (Loeb Classical Library, 1922). "But when a new generation arises and the democracy falls into the hands of the grandchildren of its founders, they have become so accustomed to freedom and equality that they no longer value them, and begin to aim at pre-eminence; and it is chiefly those of ample fortune who fall into this error. So, when they begin to lust for power and cannot attain it through themselves or their own good qualities, they ruin their estates, tempting and corrupting the people in every possible way. And hence when by their foolish thirst for reputation they have created among the masses an appetite for gifts and the habit of receiving them, democracy in its turn is abolished and changes into a rule of force and violence. For the people, having grown accustomed to feed at the expense of others and to depend for their livelihood on the property of others, as soon as they find a leader who is enterprising but is excluded from the houses of office by his penury, institute the rule of violence; and now uniting their forces massacre, banish, and plunder, until they degenerate again into perfect savages and find once more a master and monarch."

[312] *Ibid.* Polybius' view reflected a refinement of the classical Greek idea of *anacyclosis*, or the gradual cyclical development of the state, from primitive monarchy to kingship, to tyranny, leading to broader aristocracy and then oligarchy, until finally democracy emerges, only eventually to be overtaken by ochlocracy, or mob rule, until a strong autocrat rises again and the cycle begins anew. Cicero, among other ancient thinkers, suggested the formation that America's Founding Fathers eventually adopted in seeking to arrest this slide away from democracy, using a mixed form of government containing checks and balances (*see* Cicero, *de Re Publica*, I.XXIX).

for anyone who sought them out, but not forced into the view of passersby. In the virtual newsstand of the 21st Century, this gradation system has largely collapsed – no longer bounded by economic necessities, such as printing and distributing newspapers, fringe media is able to intrude into the daily awareness of millions through online platforms such as Facebook, Twitter, and Reddit.

Further, the responsibility of the Fourth Estate, once, in the United States, held firmly by the major news networks, and emblematized by such august figures as Walter Cronkite (or, more enigmatically, Johnny Carson), have been eroded by the tide of content available online. Users are not individually incentivized to take responsibility for the content they are receiving, nor are there great incentives to verify the validity of the content they share. The wave of information to which a user of a major social networking site may be exposed dwarfs the quantity previously provided by nightly news or other media. Although users today retain the capacity to curate their information – and indeed sites such as Facebook expend great effort in tailoring content to individual users – both users and the platforms themselves may understandably hold less attachment to the journalistic ethics of traditional media that previously helped to ensure the authenticity of their news. Accurate news reporting today is being usurped by the inherently nebulous system of provisioning information through social networks.

Perhaps compounding the proliferation of unvetted news and misinformation are the protections for intermediary liability guaranteed by the Communications Decency Act ("CDA §230"). CDA §230 exempts online social media platforms from significant legal responsibility for the content they host. While the EU is developing regulations aimed at holding platforms accountable directly, no legislative action appears likely in the United States.[313] The result is no single actor is vested with responsibility for ensuring that the news content offered to the American people and to users around the world is accurate and truthful,[314] and not an effort at misinformation or malignant persuasion.

While the legal issues implicated in this problem extend beyond this project's current scope, the four research teams, each focused on a major platform, have compiled individual findings and possible solutions to address the issues of both curation and responsibility, with primary recognition that no blanket solution be offered or imposed across the platforms. Each of the four major platforms surveyed in our research operates differently — as do Instagram, SnapChat, and other emerging social networks — and solutions should be carefully adapted to the user needs and business models of each individual platform. Fundamental across the platforms, however, are the underlying, sometimes competing, values of free speech and verified content. That complex matrix challenges the otherwise self-interested structures of social networks to take on the mantel of a role they have not yet assumed: cornerstones of a vibrant democracy.

[313] See, *e.g.*, the European Union's new General Data Protection Regulation (replacing an existing directive on the topic) that will come into effect in 2018. In a more draconian context, consider also authoritative regimes' use of individual responsibility to encourage compliance: new regulations in China may require individuals posting on social media sites to obtain a license from the government. http://www.globaltimes.cn/content/1046281.shtml, retrieved May 31, 2017.

[314] Twitter's user base, for instance, is 80% international.

II. Universal Platform Evaluation

This study identifies four factors across which to evaluate the platforms. The first two attributes are user-facing while the second two are platform-facing (*e.g.*, 'anonymity' refers to users, while 'transparency' refers to the platform).

Anonymity: the control that a user has in revealing his or her identity on the platform. A "Very High" score reflects higher ability to remain anonymous by default.

Control: the influence that a user has over the types and quality of content that he or she sees on the platform. A "Very High" score reflects greater ability to control this experience.

Complexity: the difficulty a user may have in understanding the platform as a whole. A "Very High" score reflects lesser ability for a user to understand how the platform operates.

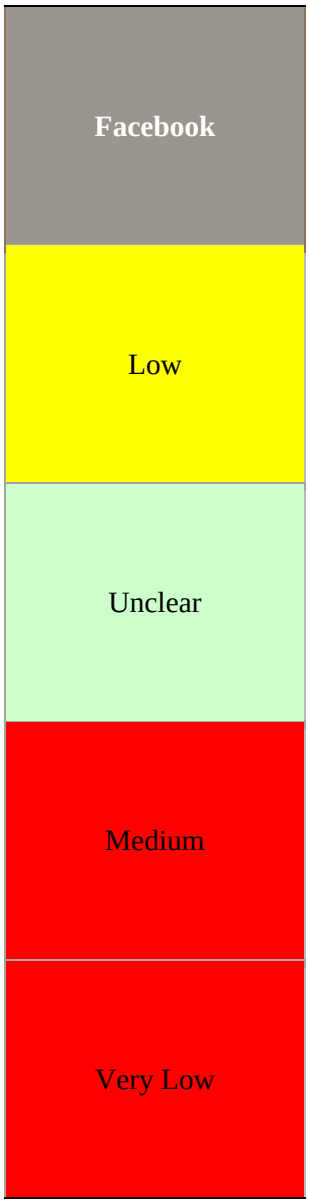
Transparency: the capacity for external entities to examine the workings of the platform and to gain data on how information is being transmitted, etc. A "Very High" score reflects greater ability to obtain this type of information publicly.

(fig. 1)

	Facebook	Twitter	Reddit	Google
Anonymity (lower is better)	Low	Medium	Very High	Medium
Control (higher is better)	Unclear	High	Very High	Low
Complexity (lower is better)	Medium	Low	High	High
Transparency (higher is better)	Very Low	High	Low	Very Low

Key: **Red** is an undesired result, **Blue** is desired, while **Green** is neutral. **Orange** and **Yellow** are intermediate gradations.

Overall, Twitter scores best, broadly doing well across these different categories, with Reddit in last place. Google, however, is different from the social networking platforms and remains difficult to quantify in these terms.

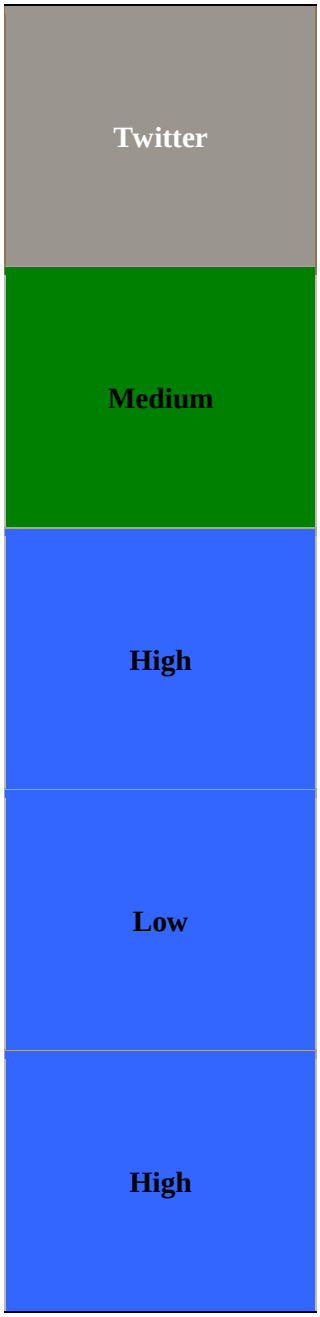


III. Facebook

With 1.2 billion users, Facebook is the largest social networking site, providing users with the ability to amplify their messages across their personal online community of "friends." Yet, as valuable as the platform is to the democratic principle of expansive dialogue, its algorithm is opaque to users; the platform does not provide users with information about how it provides them with the content they see. Facebook remains obstinate in declaring itself content-neutral in its decision-making, yet its complex black box algorithms prevent users, researchers, and policy makers from understanding how Facebook is sorting and managing inevitable content decisions. As Stanford Assistant Professor of Organizational Behavior Michal Kosinski has demonstrated, Facebook users provide a massive amount of data about themselves – Kosinski's psychological models show how even more information may be inferred[315] – enabling Facebook to tailor content according to users' interests. Tailored content has the potential to revolutionize not only advertising (for commercial and political purposes), but also to provide Facebook with substantial new avenues for revenue growth.

However, Facebook does provide one sterling advantage from a democratic, informational, standpoint: Users share information about themselves to enhance their personal profiles aimed at communicating a virtual representation of their lives. This buy-in among users presents the opportunity to encourage users to be more responsible with their use of Facebook. The company is leveraging user responsibility as a core component of its recent misinformation awareness campaign. Findings suggest the value of encouraging Facebook to provide sensitive data to researchers, perhaps with an NDA, allowing access to the information previously available only internally on the platform. The provision of this data will be valuable for researchers seeking to understand Facebook's pivotal role in 21st century communications, and to learn more about how users take advantage of the site's features and how the company's processes respond to such interaction.

[315] See, *e.g.*, M. Kosinski, S. Matz, S. Gosling, V. Popov, D. Stillwell, *Facebook as a Social Science Research Tool: Opportunities, Challenges, Ethical Considerations and Practical Guidelines*, AMERICAN PSYCHOLOGIST, 2015.



IV. Twitter

On Twitter, by contrast, messages are often shared under the veil of anonymity, paired with the spontaneous 140-character communications that typify the platform. In addition, the millions of cyborgs and bots that pervade the medium are, to some extent, given equal footing with human users in terms of ability to engage audiences by posting and retweeting content.[316] Nonetheless, Twitter scores well on the control that it provides users in crafting the content experience they receive. It is also a self-evidently simple framework that enables users to self-observe the propagation of their actions and to understand what influence their tweets may be having on others. Finally, the very structure of the platform permits researchers to access the sharing and retweeting data. However, more direct collaboration between Twitter and researchers would potentially enhance democratic principles; although the company is more transparent with its data and algorithm than is Facebook, for example, a direct invitation from Twitter to use its data would be invaluable to understanding how users interact with the platform more fully.

The ways in which Twitter could further promote democratic engagement begin with implementing a more widely adopted verification framework that gives many users the capacity to be recognized as real individuals and not bots or cyborgs operating under pseudonyms, simultaneously encouraging more responsible tweeting behavior that does not continue to propagate false or misleading content. In addition, users could be encouraged to report misleading content by offering a more readily apparent method for reporting such misinformation when it is tweeted out. The company will need to be vigilant about checking such sources to prevent users' abuse and false reporting. Further, Twitter's financial situation could be negatively impacted by changes that weaken its strong position in providing user control.[317] Twitter should be encouraged to maintain these opportunities for users, while also developing its capacity as a breaking news hub, used by journalists and everyday users reporting fast-moving events. It is in Twitter's own interests to ensure that such information remains generally reliable, providing an economic rationale for maintaining informational clarity on the site.

[316] Bots represented a substantial portion of political conversation on Twitter during the 2016 campaign. *See* https://theconversation.com/how-twitter-bots-affected-the-us-presidential- campaign-68406?utm_content=buffer8bb03&utm_medium=social&utm_source=twitter. com&utm_campaign=buffer; http://journals.uic.edu/ojs/index.php/fm/article/ view/7090/5653. Nick Pickles, Senior Public Policy Manager for Twitter, emphasized that the company is improving monitoring of both bots and cyborgs (human users, usually operating with pseudonyms, who deploy bots to accelerate and prolong their posts) and taking down those that violate its terms of service. (Pickles, "Digital Platforms and Democratic Responsibility," Global Digital Policy Incubator Symposium, Stanford CDDRL, October 6, 2017.)

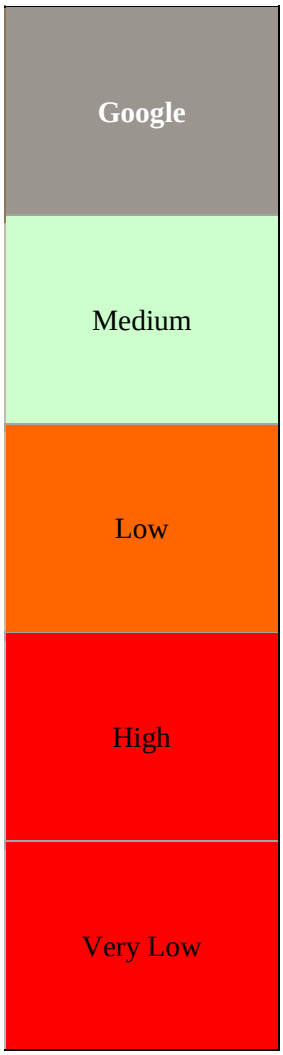
[317] In the final quarter of last year, Twitter lost $167 million. Twitter, Twitter Announces Fourth Quarter and Fiscal Year 2016 Results (2017), http://files.shareholder.com/downloads/AMDA-2F526X/4280851616x0x927284/1A6D9055-6176-4A45-A707-9A90417914F7/TWTR_Q4_16_Earnings_Press_Release.pdf.

V. Reddit

For Reddit, the sheer volume of content on the site may overwhelm casual users, while the ease of anonymously contributing content to the site dulls a sense of collective responsibility for the nature of the content. At the same time, Reddit features a unique up/down voting system that allows users themselves to take control of the site in a truly influential way. This system is not only valuable for encouraging user responsibility (although whether it has thus far done so remains unclear) but also as an example of user control for the other platforms. This serves to prevent a single user from gaining inordinate power over the editorial direction of the site, but allows a group of users to do so. While the engagement that this system fosters may not always produce responsibility – which is not, it must be acknowledged, certain to accompany user engagement – it is a necessary predicate and a valuable step in the right direction.

Improvements to the site involve moderator training and user outreach as well as platform and algorithmic enhancements:

- The site's administrators should work with subreddit moderators to develop helpful tips and reminders to users about how to evaluate the truthfulness of news.
- These reminders could appear at the top of comment sections and could encourage users to down-vote false news and to provide evidence debunking false news in the comments section. The company could test the effectiveness of different ways to phrase these reminders.
- If implemented, the company should also ensure that its algorithm doesn't artificially inflate false news based on the number of fact-checking comments.
- Further, the company could work with subreddit moderators to be more aggressive in deleting debunked news items, and Reddit administrators could develop policies to more effectively depress the popularity of subreddits that regularly traffic in discredited conspiracy theories. Because such actions can be abused and, for some subreddits, weaponized, the company would likely favor a facially neutral way of tweaking its algorithm to help achieve this result.
- Finally, Reddit administrators or subreddit moderators could develop labels for major news outlets and opinion pieces to help readers quickly identify the type of content being shared and its reliability.



VI. Google

Google presents a different, and potentially imposing, misinformation threat in comparison to social networking sites. As a search engine, the metrics of anonymity selected to categorize social media sites are less useful as tools to evaluate Google's control of information. Moreover, the black box phenomenon characteristic of Facebook is also true of Google, which permits very limited access to its internal data for researchers, let alone the general public. Yet, to understand how the site prioritizes information, researchers need access to the underlying algorithms that determine what content is displayed in response to a search. Team Google suggests that Google enhance the ability of users to report content they find questionable or worrisome. These tools should be made more obvious and more easily accessible to users. Simultaneously, Google should deepen its collaboration with independent fact checking organizations, integrating the verification efforts of those groups into search results directly. Such reforms would help to democratize Google's results by enhancing transparency and the role of users in flagging questionable content. By more fully integrating users into monitoring search results, Google elevates the role of users, making them more responsible not only for the content they consume but for the content that others see. A heightened user experience gives users the tools to make Google's' search engine more accurate in its results than it is today.

Section 7. Other Topics for Research

I. Trust in the Media

A key element of the fake news problem is that trust in established media organizations is at an all-time low. Consumers are much more likely to turn to unreliable outlets such as InfoWars if they think established organizations cannot be trusted. A September 2016 Gallup study found that only 32 percent of Americans trust the media to "report the news, fully, accurately, and fairly."[318] This figure fell 8 percent since the previous year, but it doesn't appear to just be the result of election year politics — Gallup has found a steady decline in trust in the media since at least 1997.[319] A separate study from the Media Insight Project indicates that Americans might despise "the media" generally, while trusting their own favorite outlets. For example, the study found that only 24 percent of Americans believe the news media in general is "moral," but that number jumps to 53 percent when the respondents were asked about the news media they use most often.[320] President Trump's repeated attacks on the media as "fake news" and the "enemy of the American people" seem likely to further erode the standing of journalists.

So, what can be done to improve trust in the media? Is the issue hopelessly partisan and unfixable? People are almost certainly psychologically predisposed to believe news that is favorable to their own preferred political positions. We believe though that more research into ways to improve trust in journalism would be worthwhile.

In particular, there may be practices that news organizations could adopt to improve their perceived trustworthiness. These practices could include:

- Post links whenever possible to primary documents such as video clips, official statements, letters, bills, or court documents. People are much more likely to believe a news report if they can see the underlying evidence for themselves, and the primary documents could provide additional context and detail for interested readers. If these documents aren't already available online, the journalists should post them themselves. Tech companies could invest in creating tools to help journalists post these documents more easily, which is especially important given the time constraints that journalists face. News organizations should link to other organization's work that informs their reporting, but journalists should, whenever feasible, verify reporting for themselves.
- Issue corrections promptly and prominently. Some mistakes are inevitable, but news organizations should acknowledge when they make them. This policy should extend to social media such as Twitter; journalists could, for example, delete an incorrect tweet (to prevent it from spreading), then post a screenshot of the incorrect tweet with a correction. News organizations often update online articles—especially breaking news stories—to provide additional detail and context, but they should be transparent about what changes they have made.

---

[318] Art Swift, *Americans' Trust in Mass Media Sinks to New Low*, GALLUP (Sept. 14, 2016), http://www.gallup.com/poll/195542/americans-trust-mass-media-sinks-new-low.aspx.
[319] *Id.*
[320] '*My' media versus 'the' media: Trust in news depends on which news media you mean*, AMERICAN PRESS INSTITUTE (May 14, 2017), https://www.americanpressinstitute.org/publications/reports/survey-research/my-media-vs-the-media/.

- Use anonymous sources only when necessary to obtain meaningful information. Journalists should consider a source's motives and should provide as much information as possible about a source's position so the public can evaluate his or her reliability. Journalists should verify claims with multiple sources except in rare situations when a single source is especially reliable. Representatives from government agencies, companies, and other organizations could help journalists by only demanding anonymity when it is actually necessary.
- Publish the organization's journalistic standards for the public to view. While many news organizations have internal policies on issues such as what stories to publish, how to issue corrections, and how to handle conflicts of interest, these policies are rarely available in an easily accessible location. News organizations should compile these policies into a single webpage that is accessible from an obvious link on articles and the homepage. Organizations that don't yet have complete internal guidelines could consult industry resources, such as the Society of Professional Journalists' Code of Ethics.[321]
- Distinguish opinion pieces from factual reporting. While there is nothing inherently bad about op-eds and editorials, news organizations should be careful to clearly and prominently label opinion pieces. A tiny note at the bottom of an article about an op-ed writer's affiliation is insufficient. Confusing an opinion piece with a straight-news article might cause a reader to disregard that outlet's factual reporting as biased. News organizations could collaborate through a professional association to establish a standardized format for labeling opinion pieces. TV and radio producers could develop similar policies to reduce confusion about opinion shows.

II. Digital Advertising

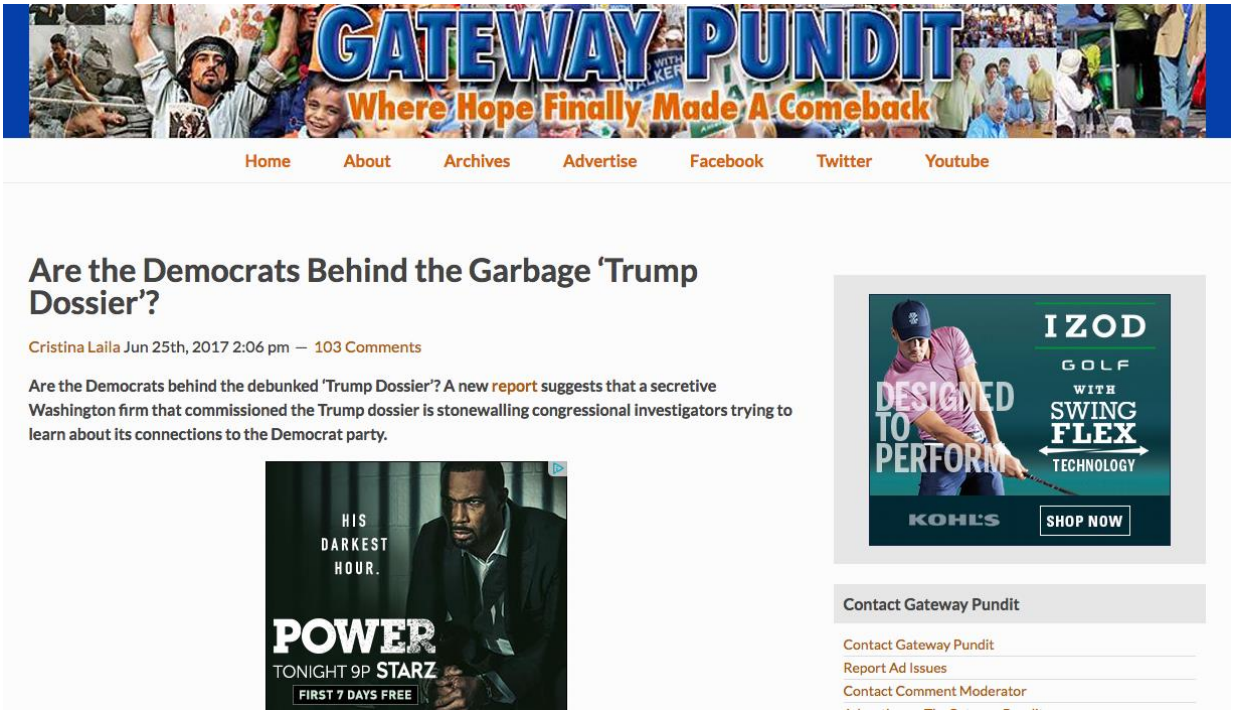
Many false news websites would likely cease to exist without money from advertising. Most websites get advertising revenue by placing a snippet of code on a page that connects to an advertising network. These networks, which often have partnerships with major brands, then display ads to the users. The networks frequently place small tracking "cookies" on users' devices that allow the networks to target ads based on users' browsing histories. The ad networks pass on a portion of the revenue to the sites themselves.

In recent weeks, under pressure from the Senate Intelligence Committee, Facebook, Twitter, and Google have begun to research the role that their own advertising platforms played in supporting false news sites. Facebook has found 3000 targeted ads, reaching 10 million users and typically designed to stoke social tensions, linked to Russian sources. According to Jonathan Albright, research director of the Tow Center for Digital Journalism at Columbia University, the posts were shared hundreds of millions of times.[322] Twitter has released information about 200

[321] *SPJ Code of Ethics*, SOCIETY OF PROFESSIONAL JOURNALISTS (last visited on July 3, 2017) http://www.spj.org/ethicscode.asp.
[322] Quoted in Dwoskin, Entous, Timberg, "Google Uncovers Russian-Bought Ads on Gmail, YouTube, and Other Platforms," *Washington Post*, October 9, 2017, https://www.washingtonpost.com/news/the-switch/wp/2017/10/09/google-uncovers-russian-bought-ads-on-youtube-gmail-and-other-platforms/?utm_term=.d8717b75ae9f. Albright's data on Facebook has since been scrubbed. See, Albright, "Itemized Posts and Historical Engagement – 6 Now-Closed FB Pages," Tableau, Oct. 5, 2017, https://public.tableau.com/profile/d1gi#!/vizhome/FB4/TotalReachbyPage; and Natasha Bertrand, "Facebook Data Scrubbed, *Business Insider*, Oct. 12, 2017, http://www.businessinsider.com/facebook-russia-data-fake-accounts-2017-10?r=UK&IR=T.

accounts associated with those same Russian sources.[323] Google has similarly revealed ads purchased by Russian agents seeking to spread disinformation across the political spectrum via Google's various products: YouTube, Google search, Gmail, and the DoubleClick ad network.[324]

The biggest ad networks are run by Google and Facebook, but there are many smaller competitors that are hardly household names like AdMaven, Content.ad, PulsePoint, Revcontent, Adblade, Taboola, and Outbrain. The potential to make thousands of dollars through these ad networks is clearly the main motivation for many false news sites. A Macedonian teen told NBC News in November 2016 that he made at least $60,000 through Google's AdSense network in just six months fabricating news articles.[325] More established websites that spread misinformation such as InfoWars and Gateway Pundit also rely on ad networks for revenue.



Advertisements for Kohl's and Starz appear on Gateway Pundit, a controversial right-wing website.

[323] For an overview of Facebook's and Twitter's recent public release of information related to Russian-linked ads and congressional and FEC responses, see Tony Romm and Kurt Wagner, "Silicon Valley's Russian ads Problem Explained," *Recode*, October 9, 2017, https://www.recode.net/2017/10/6/16419388/facebook-google-twitter-russia-ads-2016-election-president-donald-trump.

[324] See, for example, Dwoskin, Entous, Timberg, "Google Uncovers Russian-Bought Ads on Gmail, YouTube, and Other Platforms," *Washington Post*, October 9, 2017, https://www.washingtonpost.com/news/the-switch/wp/2017/10/09/google-uncovers-russian-bought-ads-on-youtube-gmail-and-other-platforms/?utm_term=.d8717b75ae9f; and Daisuke Wakabayashi, "Google Finds Accounts Connected to Russia Bought Election Ads," *New York Times*, October 9, 2017, https://www.nytimes.com/2017/10/09/technology/google-russian-ads.html

[325] Alexander Smith and Vladimir Banic, *Fake News: How a Partying Macedonian Teen Earns Thousands Publishing Lies*, NBC NEWS (Dec. 9, 2016), http://www.nbcnews.com/news/world/fake-news-how-partying-macedonian-teen-earns-thousands-publishing-lies-n692451.

In November 2016, Google and Facebook both updated their policies to try to reduce the presence of their ads on fake news sites.[326] Following these changes, Google banned 200 websites from AdSense in the fourth quarter of 2016.[327] In a January 2017 speech, Randall Rothernberg, the president and CEO of the Interactive Advertising Bureau (IAB), which represents large and small ad networks, urged his member companies to do more to cut off ties with fake sites. "If you do not seek to address fake news and the systems, processes, technologies, transactions, and relationships that allow it to flourish, then you are consciously abdicating responsibility for its outcome—the depletion of the truth and trust that undergird democratic capitalism," he said.

It is unclear, however, whether these steps are making a real dent in the problem. Google's new policy, for example, primarily targets sites that misrepresent who they are, while hyper-partisan false news is still allowed in the network.[328] Other ad networks are apparently not effectively cracking down on misrepresentation; Taboola was showing ads in late November 2016 on USAToday.com, as well as USATodaycom.com, a fake news site based in Tblisi, Georgia.[329]

Major brands are often uncomfortable with their ads appearing next to objectionable content and may put pressure on their ad network partners to do a better job cleaning up the problem.[330] Other spam advertisers, however, probably do not care how they reach users so long as they are reaching them.

Although it is illegal for foreign nationals to purchase political ads, some of the ads that circulated (and likely continue to circulate) on Facebook, Twitter, and Google were not political per se. Rather they spanned the political spectrum sowing social unrest, sometimes framed through fake stories and videos, on both the right and the left. Further research on both domestic and foreign sources can show how advertising networks promote fake news and sow social unrest. By releasing the primary sources of the ads purchased on the three platforms, the companies would enable researchers to better understand the strengths and weaknesses of the character of the ads. Such research could probe possible interventions, including whether social and/or regulatory pressure on major brands, ad networks, or advertising trade associations such as IAB is effective in choking off revenue to sites that fabricate news? Ad networks supporting controversial content is not a new problem. The movie and music industries have been working for years to try to get Google and other ad networks to cut off revenue to illegal streaming sites.

---

[326] Nick Wingfield, Mike Isaac, and Katie Benner, *Google and Facebook Take Aim at Fake News Sites*, N.Y. TIMES (Nov. 14, 2016), https://www.nytimes.com/2016/11/15/technology/google-will-ban-websites-that-host-fake-news-from-using-its-ad-service.html

[327] Tess Townsend, *Google has banned 200 publishers since it passed a new policy against fake news*, RECODE (Jan. 25, 2017), https://www.recode.net/2017/1/25/14375750/google-adsense-advertisers-publishers-fake-news.

[328] Ginny Marvin, *Google isn't actually tackling 'fake news' content on its ad network*, MARKETING LAND (Feb. 28, 2017), http://marketingland.com/google-fake-news-ad-network-revenues-207509.

[329] Lucia Moses, *'The underbelly of the internet': How content ad networks fund fake news*, Digiday (Nov. 28, 2016), https://digiday.com/media/underbelly-internet-fake-news-gets-funded/.

[330] Sapna Maheshwari, *Ads Show Up on Breitbart and Brands Blame Technology*, N.Y. Times (Dec. 2, 2016), https://www.nytimes.com/2016/12/02/business/media/breitbart-vanguard-ads-follow-users-target-marketing.html.

What lessons could be learned from these efforts? These and other questions are relevant to ongoing research on this project during the coming year.

III. Conclusion and Next Steps for Research

In sum, the role of the platforms as the virtual newsstands of the 21st century is still taking shape and, to maintain a well-informed public, changes are needed in how the major platforms deliver information to millions daily. The recommendations in this report focus on platform self-regulation and do not approach the major legal changes entailed in altering CDA §230; instead, these suggestions focus on feasible steps that can be taken, with targeted enhancements for each platform that could immediately improve users' responsibility for content. Looking ahead, it remains important to recall the necessarily profit-oriented objectives of the technology companies involved. While Facebook, Google, and Twitter have all modified their policies and algorithms in response to negative publicity about their roles in proliferating false news and misinformation, they have surely done so in ways that enhance their business interests, not for altruism.[331] Anticipating the areas where those business interests align with the needs of a healthy democracy will be perhaps *the* key next step in the years ahead.

As the Stanford Law and Policy project moves into its next phase of research, it will examine the extent to which online disinformation and propaganda is directly affecting the performance of U.S. democracy. A subset of that question is the degree to which users' political beliefs and their ability to recognize such information has shifted in recent months with heightened awareness of the issue. Finally, are citizens polarizing because of exposure to biased information, or are polarized citizens more likely to seek out biased information?

With more information coming to light about the role of the platforms, many questions have opened, including:[332]

- Who are the primary producers of disinformation in the U.S. and elsewhere (government officials, political parties, media companies, individuals on the internet, etc.)?
- Which producers of disinformation/propaganda are most effective at impacting electoral outcomes, and how concentrated are they?
- Which purveyors are more likely to propagate polarizing disinformation/propaganda? What strategies and tactics (e.g., bot networks) are most prevalent/successful in distributing disinformation?
- Which platforms are more conducive to spreading of falsehoods? How significant are echo chamber/filter bubble challenges for each platform? iii. What cues do people use to assess the credibility of an online news story (e.g., brand, byline, expert citations)? What factors of disinformation/propaganda content determine its virality? What are the dominant narratives across different audience segments and platforms driving polarization?

[331] N. Persily, *Can Democracy Survive the Internet?* JOURNAL OF DEMOCRACY, April 2017, 73-74.

[332] These and other questions are some of the research topics underway in the 2017-18 "Fake News and Misinformation" Law and Policy Lab Practicum led by Nate Persily, https://law.stanford.edu/education/only-at-sls/law-policy-lab/practicums-2017-2018/fake-news-and-misinformation/. These questions and others emerged in discussions at the conclusion of this phase of the research project with credit to Kelly Born, Program Officer for the Hewlett Foundation Madison Initiative.

- What are the economic incentives of each platform?
- What is the role of the platforms in personalizing content for political micro-targeting?
- How effective are different interventions? What should the platforms do differently to elevate "quality" content and reduce polarization and/or partisan incivility?
- What are the legal and regulatory options for platform responsibility in spreading disinformation and propaganda?
- Do the platforms have monopoly power in any markets? If so, to what degree are they using such power anti-competitively?
- Should traditional views of free speech be updated to address current technologies and their applications? How does the regulation of disinformation compare to regulation of other types of prohibited speech (e.g., terrorist recruitment/incitement, hate speech, obscenity, intellectual property, bullying, etc.)?
- What are legal obstacles to intermediary liability for self-regulation of fake news? Should takedown procedures similar to those for alleged copyright infringement be extended to other types of content?
- Could FTC enforcement involving "deceptive acts or practices" be extended to include false news and disinformation?
- How could FEC regulations requiring disclosures of paid online elections communications be better enforced?
- How could platforms be incentivized to self-regulate?

These questions seed the research landscape for the digital era. The responsibility for preventing misinformation from overwhelming the facts – be that information digital, in print, or in spoken forums – rests with each producer and consumer of news, shared with the platforms themselves. Yet blanket or legislative solutions need not be the necessary answer to this problem. As the foundation for an ongoing exploration of possible solutions to the proliferation of fake news and misinformation, this research reviews the advantages and drawbacks of each platform. The goal is eventually to cultivate an individualized approach to strengthening the platforms, pruning their approach where necessary, to deliver an important, lively, and accurate civil discourse as the basis of healthy democracy. Ultimately, as Benjamin Franklin pithily observed in Philadelphia in the summer of 1787, in response to a woman who inquired what form of the government the new nation would take: "[a] republic, Madame, if you can keep it."[333]

[333] W. Isaacson, BENJAMIN FRANKLIN: AN AMERICAN LIFE. New York: Simon & Shuster, 2003, 459.