

October 1, 2017



Is Social Media a Threat to Democracy?



Contributors:

Anamitra Deb, *Director*, Omidyar Network
Stacy Donohue, *Investment Partner*, Omidyar Network
Tom Glaisyer, *Program Director*, Democracy Fund

Our Purpose

In this paper, we examine six key issues and implications presented by social media participation and manipulation, and we cite examples of what we and others are doing—or could do—to possibly mitigate their negative impact.

We were inspired to draft this paper because the current instability in democracies—caused in part by increasing political polarization, the trust deficit vis-à-vis institutions and elites, and the spread of mis- and disinformation across the globe—is causing potentially irrevocable harm to our fundamental rights. Our objective in sharing this initial analysis and the questions we have is to begin a productive dialogue among diverse stakeholders on this important topic.

Our contributors have a particular interest in the intersection between technology and democracy as representatives of two philanthropies that are part of The Omidyar Group—[Democracy Fund](#) and [Omidyar Network](#). Both organizations devote significant resources to exploring the impact of technology and digital communication on our global society.

A Call to Action

We look forward to continued sense-making and exploring possibilities to partner with others as we search for answers. Please email the authors at inquiries@omidyargroup.com if you'd like to discuss how we might work together.

About The Omidyar Group

The Omidyar Group is a diverse collection of companies, organizations, and initiatives that strive to catalyze social impact by empowering individuals to create and sustain healthy societies. www.omidyargroup.com.

About The Democracy Fund

The Democracy Fund is a bipartisan foundation established by Pierre Omidyar to help ensure our political system is able to withstand new challenges and deliver on its promise to the American people. Since 2011, Democracy Fund has invested more than \$60 million in support of a healthy democracy, including modern elections, effective governance, and a vibrant public square. www.democracyfund.org.

About Omidyar Network

Omidyar Network is a philanthropic investment firm established by Pierre Omidyar and his wife Pam. It is dedicated to harnessing the power of markets to create opportunity for people to improve their lives. They invest in and help scale innovative organizations to catalyze economic and social change. To date, Omidyar Network has committed more than \$1 billion to for-profit companies and nonprofit organizations across multiple initiatives, including Education, Emerging Tech, Financial Inclusion, Governance & Citizen Engagement, and Property Rights. www.omidyar.com.

Executive Summary

It is becoming increasingly apparent that fundamental principles underlying democracy—trust, informed dialogue, a shared sense of reality, mutual consent, and participation—are being put to the test by certain features and attributes of social media. As technology companies increasingly achieve financial success by monetizing public attention, it is worth examining some of the key issues and unintended consequences arising as a result.¹

¹ Given the focus of this paper, we have not addressed any of the negative individual, psychological, ramifications of social media, such as its effects on depression, addiction, coercion, exhibitionism, self-censorship, and other “chilling effects.”

EXECUTIVE SUMMARY

6 KEY ISSUES:

Echo chambers, polarization, and hyper-partisanship

Social media platform design, combined with the proliferation of partisan media in traditional channels, has exacerbated political divisions and polarization. Additionally, some social media algorithms reinforce divisions and create echo chambers that perpetuate increasingly extreme or biased views over time.

Spread of false and/or misleading information

Today, social media acts as an accelerant, and an at-scale content platform and distribution channel, for both viral “dis”-information (the deliberate creation and sharing of information known to be false) and “mis”-information (the inadvertent sharing of false information). These two types of content—sometimes mistakenly conflated into the term “fake news”—are created and disseminated by both state and private actors, in many cases using bots. Each type poses distinct threats for public dialogue by flooding the public square with multiple, competing realities and exacerbating the lack of agreement about what constitutes truth, facts, and evidence.

Conversion of popularity into legitimacy

The algorithms behind social media platforms convert popularity into legitimacy, overwhelming the public square with multiple, conflicting assertions. In addition, some social media platforms assume user intentionality (e.g. in search queries) and conflate this with interest, via features such as auto-fill search terms. These design mechanisms impute or impose certain ways of thinking, while also further blurring the lines between specialists and laypeople, or between verified and unverified assertions, thus contributing to the already reduced trust in traditional gatekeepers.

Manipulation by “populist” leaders, governments, and fringe actors

“Populist” leaders use these platforms, often aided by trolls, “hackers for hire” and bots, on open networks such as Twitter and YouTube. Sometimes they are seeking to communicate directly with their electorate. In using such platforms, they subvert established protocol, shut down dissent, marginalize minority voices, project soft power across borders, normalize hateful views, showcase false momentum for their views, or create the impression of tacit approval of their appeals to extremism. And they are not the only actors attempting to use these platforms to manipulate political opinion—such activity is now acknowledged by governments of democratic countries (like the UK), as well.²

Personal data capture and targeted messaging/advertising

Social media platforms have become a preferred channel for advertising spend. Not only does this monetization model drive businesses reliant on the capture and manipulation of huge swathes of user data and attention, it also widens the gap between the interests of publishers and journalists and erodes traditional news organizations’ revenues. The resulting financial strain has left news organizations financially depleted and has reduced their ability to produce quality news and hold the powerful to account. In addition, advanced methods for capturing personal data have led to sophisticated psychographic analysis, behavioral profiling, and micro-targeting of individuals to influence their actions via so-called “dark ads.”

Disruption of the public square

Some social media platforms have user policies and technical features that enable unintended consequences, like hate speech, terrorist appeals, and racial and sexual harassment, thus encouraging uncivil debate. This can lead members of frequently targeted groups—such as women and minorities—to self-censor or opt out of participating in public discourse. Currently, there are few options for redress. At the same time, platforms are faced with complex legal and operational challenges with respect to determining how they will manage speech, a task made all the more difficult since norms vary widely by geographic and cultural context.

² Samantha Bradshaw & Philip N. Howard, *Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation*. Samuel Woolley and Philip N. Howard, Eds. Working Paper 2017.12. Oxford, UK: Project on Computational Propaganda. <http://comprop.oii.ox.ac.uk/>

INTRODUCTION: Framing the Problem

It is becoming increasingly apparent that fundamental principles underlying democracy—trust, informed dialogue, shared sense of reality, mutual consent, and participation—are being put to the test by certain features and attributes of social media; they have disrupted our public square. As these platforms achieve financial success by monetizing public attention, it is worth examining some of the key issues and unintended consequences arising as a result.

By any meaningful measure, the category we refer to as “social media” today encompasses massive scale. These platforms are pervasive, and fully and effectively integrated into the public discourse and lives of individuals. By 2017, for example, Facebook-owned platforms already reach [86% of Internet users aged 16 to 64 in 33 countries](#) and effectively act as the gateway to the Internet, if not the Internet itself.³ In a sense, Facebook is becoming the world’s largest news source; 44% of people across 26 countries surveyed say they use it for news. Similarly, [Pew’s analysis](#) shows that in the U.S., two-thirds of Facebook users (66%) get news on the site, nearly six out of ten Twitter users (59%) get news on Twitter, and, highest, 70% of Reddit users get news on that platform. [Similar trends exist for 18-to 24-year-olds](#) and users in emerging economies such as the Philippines and Myanmar.⁴

Facebook has boasted about [the amount of a user’s daily attention it captures](#), and as we learn more about the new “[attention economy](#),” this issue is mission-critical. Google has already commoditized search, email, storage and arguably, information—in many instances, disrupting economies and longstanding arrangements. The advantage social media platforms such as Facebook and Google have in monetizing attention accrues from their unprecedented and large-scale collection and analysis of personal data. Less sanguine is the use of behavioral and psychographic profiling, which can be harvested to deliver personalized content and advertising—much of which is unregulated and invisible to all but the recipient.⁵

It could be argued that this exceptionally clear line to monetization dominates their operating logic, their community architecture, and their decision-making. The business model is simple: capture attention, then monetize it through advertising. However, the effects of social media on public discourse, civility, and fact-based debate—given the massive scale—necessitate increased attention to its impact on democracy.

The early optimism about social media’s potential for democratizing access to information, and giving voice to those who were traditionally marginalized or censored, is eroding. Indeed, as social media platforms have grown, they have been accused of:

- Exacerbating the polarization of civil society via echo chambers and filter bubbles
- Rapidly spreading mis- and disinformation and amplifying the populist and illiberal wave across the globe
- Creating competing realities driven by their algorithms’ intertwining of popularity and legitimacy
- Being vulnerable to political capture and voter manipulation through enabling malevolent actors to spread disinformation and covertly influence public opinion
- Capturing unprecedented amounts of data that can be used to manipulate user behavior
- Facilitating hate speech, public humiliation, and the targeted marginalization of disadvantaged or minority voices

This paper seeks to explore these issues and identify questions that need to be considered if we are to ensure our media structures serve democratic ends.

³ A 2016 survey by Reuters Institute of 50,000 users in 26 countries found that 51% use social media as a source of news each week.

⁴ More than a quarter of 18-to 24-year-olds (28%) say social media is their main source of news—more than television (24%) for the first time. More so in Europe than in the U.S., one argument goes that these platforms have evolved to effectively become media

firms, that their strategies and decisions may now be interpreted as editorial decisions, and they should be regulated accordingly.

⁵ As noted above, these companies also deliver valuable services which we now take for granted—free search, free connection and community, free email and messaging, free storage, etc.—in return for such data collection.

ISSUES: Focusing on Six Key Risks

Though there are a number of different ways to analyze the risks that social media poses for democracy, here we focus on six key issues at the core of the discussion.

Issue One: Echo chambers, polarization, and hyper-partisanship

The “attention economy” is predicated on understanding and targeting individual users with a variety of customized content. It is well documented that an individual’s search results, newsfeed, and advertising offers are dependent on, and generated by, their digital footprint as manipulated by the ever-evolving algorithms of social media platforms. Of these, advertising is the most important, since it underwrites the business model of the social media giants. The prioritization of user preferences results in a feedback loop where the feeding of news, search results, and social network updates that align with user attitudes and interests exacerbates and reinforces user preferences—and on platforms such as Facebook, this tends to promote self-segregation into like-minded groups. As articulated by Ethan Zuckerman in an insightful column comparing Facebook’s [coverage of Ferguson with the ALS ice-bucket challenge](#), Facebook’s preference for creative “viral cascades over surfacing novel content” within a social network can lead to significant isolation from the truth: “If few users in your circle of friends are sharing stories about Ferguson—which is a distinct possibility if you are white and most of your friends are white—Facebook’s algorithm may see this as a story unlikely to “go viral” and tamp it down, rather than amplifying it, as it has with the ALS stories.”⁶ That said, not all social media platforms are designed for such self-segregation, as noted in a Digiday article, [“Facebook is for ice buckets, Twitter is for Ferguson.”](#)

Few users consume purely partisan media, but social media platform design and the proliferation of partisan media in traditional channels have exacerbated

partisanship and identity polarization by creating “echo chambers” where views get reinforced and become entrenched—and more extreme—over time. Some part of this is design. But another important part is due to the way people decide whether news and information online can be trusted. [A recent American Press Institute \(API\) study](#) shows that when American readers see news on social media platforms, it’s not the source of the news that matters as much as whom in their network shares the link.⁷ Attempts to make visible to users the echo chambers in which they operate have been frequent, but we have found little data on whether such information changes minds.

Such trends are likely to continue or accelerate. Worldwide, social media users have multiple platform choices, so they can read single articles from different sources based on what is popular in their network. This increased personalization means [users are more likely to see \(and believe\) what their peers share](#) than what news publishers curate, making them less likely to encounter multi-faceted or counter-attitudinal views. In addition, partisan content delivered deliberately and consistently to echo chambers (or “ideological silos”) [shifts belief and political behavior](#), and the results of partisanship on [social capital are noteworthy](#). More partisans today hold extreme (or “shifted”) political views, and there is less willingness to engage across the aisle. People prefer their social networks to be “like-minded” politically, and for the very first time, [fewer people in America](#) want to see their children married to someone across the political aisle than to someone from another race.

⁶ In a study of social network demographics by Daniel Cox and Robert Jones of the Public Religion Research Institute, data revealed that the average white American has one black friend, and that 75% of white Americans have entirely white social networks.

⁷ American Press Institute, “Who shared it? How Americans decide what news to trust on social media,” Published March 20, 2017, <https://www.americanpressinstitute.org/publications/reports/survey-research/trust-social-media/>

ISSUES: Focusing on Six Key Risks

Issue Two: Proliferation of several types of misinformation and disinformation

The use of false information to change public opinion is at least as old as the newspaper. Ben Franklin used the tactic to generate opposition to Native Americans during the American Revolution. Today, social media acts as an accelerant, and an at-scale content platform and distribution channel, for what is now widely referred to as “fake news.” This much-maligned term actually comprises [several types](#) of “dis”-information (the deliberate creation and sharing of information known to be false) and “mis”-information (the inadvertent sharing of false information) (see Figure 1 below).

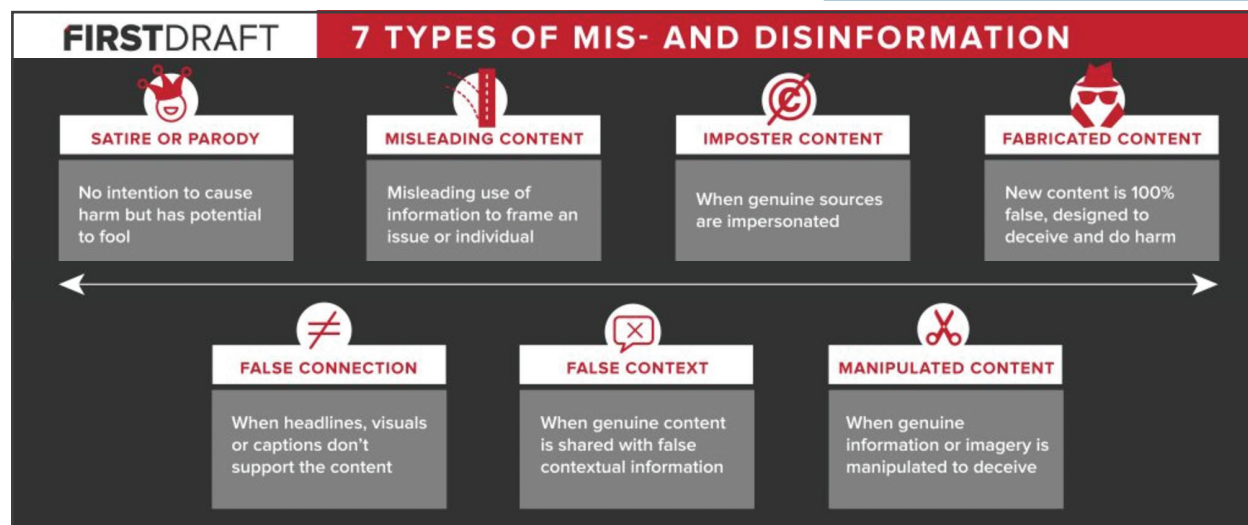
Some of these types of content present a higher risk to democratic discourse—e.g. it’s clear that ‘fabricated content’ is qualitatively different in intention and potential impact than satire or parody, or, for that matter, false context. These types of information are created and disseminated by a variety of actors—including state and private actors who often use bots — and often to different ends.^{8,9}

Increasingly, such information carries consequences in the “offline” or “real” world. [PizzaGate](#) is one example of the dangerous amplification of conspiratorial

speculation: 4Chan and Reddit pranksterism spread across social media platforms, was picked up by mainstream news, and, ultimately, led to a holdup of a restaurant in Washington, D.C.

Even more troubling are recent killings in South Sudan,¹⁰ which some have cited as an extreme example of how fake news and hate speech promoted by social media influencers can be used to incite mass violence against minority groups. Oxford University’s Internet Institute exposed patterns of automated fake news production recently during their real-time investigation of political communication during the Brexit referendum in the UK. [A subsequent study](#) of what Twitter users in Michigan were sharing in the week before voting found that the volume of junk news was much greater than that of professionally produced news. Moreover, the level of professionally produced news content shared over Twitter hit an all-time low in the 24 hours before the election.¹¹

The low barriers to creation and distribution of online content have facilitated massive growth in “news publishers” whose revenue models maximize attention

Figure 1: 7 Types of Mis- and DisinformationSource: <https://firstdraftnews.com/fake-news-complicated/>

⁸ We speak to the danger inherent in the active manipulation and political capture of information on social media below as part of Issue Three.

⁹ It is worth underscoring that the motivations for mis- and disinformation can be monetary and/or ideological. Such motivational distinctions might be important to the deployment of viable tactics for tackling the problem.

¹⁰ Jason Patinkin, *How to Use Facebook and Fake News to Get People to Murder Each Other*, BuzzFeed, Updated January 16, 2017, [https://www.buzzfeed.com/jasonpatinkin/how-to-get-people-to-murder-each-other-through-fake-news-and?](https://www.buzzfeed.com/jasonpatinkin/how-to-get-people-to-murder-each-other-through-fake-news-and?hpid=hp_hp-top-table-main-social-media%3Ahow-to-get-people-to-murder-each-other-through-fake-news-and%3Ahomepage%2Fstory&hpt=hp-top-table-main-social-media%3Ahow-to-get-people-to-murder-each-other-through-fake-news-and%3Ahomepage%2Fstory)

¹¹ John Markoff, *Automated Pro-Trump Bots Overwhelmed Pro-Clinton Messages, Researchers Say*, *The New York Times*, Published November 17, 2016, <https://www.nytimes.com/2016/11/18/technology/automated-pro-trump-bots-overwhelmed-pro-clinton-messages-researchers-say.html>

ISSUES: Focusing on Six Key Risks

Issue Two: Proliferation of several types of misinformation and disinformation

and engagement with low regard for quality control or traditional journalistic ethics. [A review of YouTube recommendations](#) on the eve of the U.S. presidential election showed that “more than 80% of recommended videos were favorable to Trump, whether the initial query was ‘Trump’ or ‘Clinton.’ A large proportion of these recommendations were fake news.”

Egregious instances of this phenomenon also include [the Macedonian teenagers who mastered fake news to generate pocket money](#): stories in [The Guardian](#) and on [BuzzFeed](#) revealed that the Macedonian town of Veles (pop. 55,000) was the registered home of at least 100 pro-Trump websites, many of them filled with sensationalist, utterly fake news (the imminent criminal indictment of Hillary Clinton was a popular theme; another was the Pope’s approval of Trump.) Automated advertising engines, like Google’s AdSense, rewarded the sites’ ample traffic handsomely.¹² Digital publications—including [fake think tanks](#)—that generate provocative and partisan content to resonate with users’ beliefs and biases are enabled by falling costs and easy access to [advanced video and voice technologies](#). Nor is this a phenomenon restricted to one side of the political aisle. As recent articles by the [The Atlantic](#), [Buzzfeed](#) and [BBC](#) document, at least in the U.S. and the UK in 2017, there has been a significant spike in the proliferation of disinformation campaigns and websites catering to the “left.”¹³

Social network platforms have huge incentives to accommodate the creation and distribution of content and feed the “attention economy.” And, unlike regulated media, there are no real consequences to these networks for distributing fake news (except in Germany, [where lawmakers recently passed](#) the much-debated Network Enforcement Act, which allows fines of up to \$57M

against social media companies that don’t remove “obviously illegal” content on their sites within 24 hours). Fake news helps maximize ad-click revenue by keeping users on the platform.¹⁴

This phenomenon has not emerged in a vacuum; decades of declining trust in the mainstream media have created an opening for these alternative sources of news.¹⁵ Troublingly, [studies have shown](#) that it is very challenging to persuade people with facts once they have adopted a belief or position because of confirmation bias¹⁶ and the “backfire effect;”¹⁷ this helps maintain the atmosphere of confusion and competing reality. Findings such as in the API study mentioned above cast doubt on whether current fact-checking models are focused on the right locus of trust and belief.

Indeed, some types of mis- and disinformation have serious ramifications for public dialogue and trust. Facebook has sought to set up a mechanism to allow fact-checkers to indicate when an article has been disputed, though in some cases such a flag has resulted in wider readership.¹⁸

The Omidyar Group has funded extensive research by academics in partnership with the news industry, as well as work on understanding the prevalence and dynamics of such issues, to better prepare responses that will combat this phenomenon.¹⁹ We have also provided support to the International Fact-Checking Network to bring together journalistic fact checkers and advanced practices globally.²⁰ But there is still much work to be done with regard addressing the powerful monetary motivation that drives social media algorithms, seemingly at the expense of civility, truth and constructive conversation.

¹² Samantha Subramanian, *Inside the Macedonian Fake-News Complex*, Wired, Published February 15, 2017, <https://www.wired.com/2017/02/veles-macedonia-fake-news/>

¹³ Laura Hazard Owen, *Republicans seem more susceptible to fake news than Democrats (but liberals, don’t feel too comfy yet)*, NeimanLab, Published May 5, 2017, <http://www.neimanlab.org/2017/05/republicans-seem-more-susceptible-to-fake-news-than-democrats-but-liberals-dont-feel-too-comfy-yet/>; and Zack Beauchamp *Democrats are falling for fake news about Russia*, Vox, Published May 19, 2017, <https://www.vox.com/world/2017/5/19/15561842/trump-russia-louise-mensch>

¹⁴ That is, the underlying AI on platforms like YouTube and Facebook promotes these links because of their popularity, disregarding authenticity or relevance.

¹⁵ Furthermore, it fits well with previous successful strategies used by motivated actors (like the [Tobacco industry](#)) to refute findings that threaten their business model/vote bank/etc.

¹⁶ The Edelman Trust survey found that respondents are four times more likely to deliberately ignore information that does not support

a position they believe in. Furthermore, people often overestimate their ability to identify fake news—a [Pew study](#) found that 23% of Americans had shared fake news, in some cases even after knowing it was fake.

¹⁷ Craig Silverman, *The Backfire Effect*, Columbia Journalism Review, Published June 17, 2011, http://archives.cjr.org/behind_the_news/the_backfire_effect.php

¹⁸ Sam Levin, *Facebook promised to tackle fake news. But the evidence shows it’s not working*, The Guardian, Published May 16, 2017, <https://www.theguardian.com/technology/2017/may/16/facebook-fake-news-tools-not-working>

¹⁹ American Press Institute, *Fact checking research*, <https://www.americanpressinstitute.org/fact-checking-research-studies/>; and Brendan Nyhan, *Why More Democrats Are Now Embracing Conspiracy Theories*, The New York Times, Published February 15, 2017, <https://www.nytimes.com/2017/02/15/upshot/why-more-democrats-are-now-embracing-conspiracy-theories.html>

²⁰ Poynter, *The International Fact-Checking Network*, <https://www.poynter.org/about-the-international-fact-checking-network/>

ISSUES: Focusing on Six Key Risks

Issue Three: Conflation of popularity, legitimacy and user intentionality

Over the last two decades, technology companies have spent a huge amount of money and effort to develop ways to get people to trust each other online, in conversations and transactions, on various platforms and marketplaces. Social media takes this to the next level—doubling down on the age-old locus of trust, reputation, and belief in one's networks. Perhaps it is no surprise that today, a majority of global respondents to the [Edelman Trust Survey](#) claim to believe individuals over institutions. However, one of the arenas in which this has serious consequences is on platforms where all individuals can publish without meaningful editorial insight, and where polarization has led to echo chambers. Crowd-sourced discussion platforms, including ones such as Wikipedia, Quora, and Reddit, further blur the lines between specialists and the layperson, creating false equivalencies. In the U.S., the crowd-sourced information phenomenon is now tied into part of a larger narrative and [growing backlash against experts and elites](#), who are viewed having a self-serving agenda.

Crucial to how users consume information is the algorithmic logic of certain social media platforms and the way they engineer viral sharing in the interest of their business models. The non-neutral algorithms of Facebook and Twitter actively use selection criteria to enhance the visibility of certain information. What's highly problematic about this—apart from documented

instances of [algorithmic bias](#)—is that the criteria attribute legitimacy to popularity; thereby flooding the public with multiple, competing, unverified assertions.²¹ This isn't just restricted to Facebook and Twitter; a variant of the problem exists at Google, where "auto-fill search terms" assume user intentionality and conflate this with interest. For instance, in 2013, [UN Women](#) launched a powerful ad campaign revealing Google's autocomplete suggestions for "Women shouldn't..." "Women cannot..." and "Women need to..." among others. The top results included "Women shouldn't have rights," "Women cannot drive," and "Women need to be put in their place." The algorithms for those phrases have since been updated, and there are [certain terms that Google will not autocomplete](#), including "Bisexuals are..." and "Lesbians are..." Yet plenty of other examples of bigotry, sexism, and racism lurk within other seemingly innocuous searches. Such auto-fill search terms cannot only reinforce prejudices; but when used to analyze user preferences and behavior, they can also reveal [fascinating and hard-to-prove insights](#).

Finally, attributing legitimacy to popular search queries is worrying in a world where it can reinforce—or even leverage—assumptions that code is unbiased. For instance, the Edelman Trust Barometer revealed that 59% of global respondents are prone to believe a search engine over a human editor.^{22,23}

²¹ It may be worth questioning whether this is due to the editorial choices of the platforms, or whether it is a result of platform design responding to reader preferences and prejudice.

²² Edelman Trust Barometer, 2017

²³ Nor, conversely, [can individuals trust supposedly human bylines anymore](#), as automated content increases on the Internet, and raises questions about transparency, transferred trust and ethics.

ISSUES: Focusing on Six Key Risks

Issue Four: Political capture of platforms

Open networks, such as Twitter and YouTube, are particularly vulnerable to political capture by populist leaders—and their armies of trolls and automated bots—for motivated use: e.g. to shut down dissent and minority voices, create the false impression of momentum, intervene across borders, or manipulate public sentiment.²⁴ Of course, it is not just politicians and political parties that can capture platforms—even established democratic governments have spent public money to manipulate opinion over social media.

In their [recent report](#) from the Oxford Internet Institute, Samantha Bradshaw and Philip N. Howard point to the British Army, which announced in January 2015 that one of its brigades would “focus on non-lethal psychological operations using social networks like Facebook and Twitter to fight enemies by gaining control of the narrative in the information age (Solon, 2015).” In other words, “The primary task of this unit is to shape public behavior through the use of ‘dynamic narratives’ to combat the political propaganda disseminated by terrorist organizations.”²⁵

In the same report, Bradshaw and Howard referred to trolls and bots as modern-day cyber troops, saying they are “government, military or political party teams committed to manipulating public opinion over social media.” The report also compared such organizations across 28 countries and found:

“...Cyber troops are a pervasive and global phenomenon. Many different countries employ significant numbers of people and resources to manage and manipulate public opinion online, sometimes targeting domestic audiences and sometimes targeting foreign publics. Looking across

the 28 countries, every authoritarian regime has social media campaigns targeting their own populations, while only a few of them target foreign publics. In contrast, almost every democracy in this sample has organized social media campaigns that target foreign publics, while political-party-supported campaigns target domestic voters. Over time, the primary mode for organizing cyber troops has gone from involving military units that experiment with manipulating public opinion over social media networks to strategic communication firms that take contracts from governments for social media campaigns.”²⁶

In short, trolls and bots disguised as ordinary citizens have become a weapon of choice for governments and political leaders to shape online conversations in many illiberal regimes and movements.

[Governments in Turkey, China, Israel, and Russia](#) are known to have deployed thousands of hired “social media operatives” who run multiple accounts and manage bots to shift or control public opinion—including by having arguments among their various accounts that are settled in favor of the government. [A recent study](#) of the Chinese government’s use of social media revealed that paid government employees generate about 448 million comments every year.²⁷ In Myanmar, several of the government’s official Facebook accounts propagate exclusionary and hateful views about Muslim minorities.²⁸ Second, [several political leaders have a direct line to their electorate](#). Thus, they can share unedited (by media or party), false, or provocative statements, subvert or bypass established communications protocol, normalize hateful or cynical views themselves, or implicitly approve the messages of their social media supporters (this includes

²⁴ An interesting point about the design of Facebook versus Twitter: the former lets you select your community, does not encourage non-reciprocal following, and the upside is safety and protection from trolls; the downside is self-selection into like-minded groups and echo chambers. Twitter is the converse: open community structure, and encouraging of non-reciprocal following, which makes it better for filtering out fake news, but much more vulnerable to both political capture via trolls and bots, and intimidation.

^{25,26} Samantha Bradshaw & Philip N. Howard, *Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation*. Samuel Woolley and Philip N. Howard, Eds. Working Paper 2017.12. Oxford, UK: Project on Computational Propaganda. <http://comprop.oii.ox.ac.uk/>

²⁷ A little over half of these comments are made on government sites, albeit pretending to be comments made by ordinary citizens. The rest are made on commercial sites, mixed into streams with family news, dog photos and the like. The result, as Gary King, Jennifer Pan and Margaret Roberts describe, is that a “large proportion of government website comments, and about one of every 178 social media posts on commercial sites, are fabricated by the government.”

²⁸ Megha Rajagopalan, *Internet Trolls Are Using Facebook To Target Myanmar’s Muslims*, BuzzFeed, Published March 18, 2017, <https://www.buzzfeed.com/meghara/how-fake-news-and-online-hate-are-making-life-hell-for?>

ISSUES: Focusing on Six Key Risks

Issue Four: Political capture of platforms

cross-border and bot-enabled activity). [President Duterte of the Philippines](#) has successfully used social media to control the narrative on the drug war (which has resulted in over 7,000 deaths since mid-2016²⁹) and to silence any questions on his tactics.³⁰ Prime Minister Modi of India offers a [direct-to-your-phone tweet service](#) as part of the Digital India initiative, and he has used Twitter to make significant policy announcements—such as the recent statement on demonetization, which bypassed the conventional route of first presenting a major new policy to Parliament.

The issue is worsened through automation. Automated bots can create the false impression of momentum at scale: [highly automated accounts](#), defined as accounts that tweeted 450 or more times with a related hashtag and user mention, generated close to 18% of all Twitter traffic about the 2016 U.S. presidential election. [Researchers at USC](#) recently estimated that “between

9% and 15% of active Twitter accounts are bots”—which could mean that there are up to 50 million bot accounts.³¹ There are also implications for national sovereignty, as the uncertain geographic origin of bots facilitates their use by foreign governments or groups seeking to intervene in another country’s political process. In some ways this undermines a crucial premise of democracy where processes are set up to give voice to the people. Malevolent bots allow certain voices to be amplified disproportionately to manipulate public sentiment.

Moreover, few regulations or protocols protect users against trolling and online abuse, and social media companies are not always quick to—or equipped to—respond globally. The Myanmar ICT for Development Organization ([MIDO](#)) found that only 10% of the posts it reported as hate speech on Facebook were ultimately removed.³²

Issue Five: Manipulation, micro-targeting, and behavior change

Unprecedented personal data captured by social media platforms enables sophisticated psychographic or behavioral profiling and micro-targeting. Social media companies today have troves of data that any government would like to access, and, in some cases, successfully have.³³ One use case for this micro-targeting is “dark advertising,” in which political advertisements are shown to Facebook users without being subject to traditional regulatory guidelines. As Phil Howard argued in a [Washington Post opinion piece](#), “while we are all talking about ‘fake news,’ we should also be talking about the algorithms and fake accounts that push bad information around.” Technology firms, with large amounts of personal data, support the experiments that can nudge behavior.

Big data analysis, combined with computational psychology and behavioral and demographic analysis, has allowed firms (not just Google and Facebook, but reportedly also [Cambridge Analytica](#), though the claim of the influence of the latter may be overstated) to “decode” users’ personalities and target them with tailored messaging to drive behavioral shifts. Since the fundamental business model of these platforms is based on monetizing user data and attention, we might expect this to continue, if not intensify.

The absence of appropriate regulation that allows widespread surveillance and collection of user data has extended into democratic election campaigns. The continuous surveillance of users through feedback

²⁹ Michael Bueza, IN NUMBERS: The Philippines’ ‘war on drugs’, *Rappler*, Updated April 23, 2017, <https://www.rappler.com/newsbreak/iq/145814-numbers-statistics-philippines-war-drugs> Rappler is an Omidyar Network investee.

³⁰ Interview with Maria Ressa, *Rappler* (Philippines), 2017. See also Rappler’s series on the weaponization of speech in the Philippines: <http://www.rappler.com/nation/148007-propaganda-war-weaponizing-internet>. Rappler is an Omidyar Network investee.

³¹ The researchers used more than one thousand features to identify bot accounts on Twitter, in categories including friends, tweet content and sentiment, and time between tweets. See Social

bots distort the 2016 U.S. Presidential election, online discussion by Alessandro Bessi and Emilio Ferrara. *First Monday*, Volume 21, Number 11 -7 November 2016, <http://firstmonday.org/ojs/index.php/fm/article/view/7090/5653>

³² See <https://www.buzzfeed.com/meghara/how-fake-news-and-online-hate-are-making-life-hell-for?>

³³ Glenn Greenwald and Ewan MacAskill, NSA Prism program taps in to user data of Apple, Google and others, *The Guardian*, Published June 7, 2013, <https://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data>.

ISSUES: Focusing on Six Key Risks

Issue Five: Manipulation, micro-targeting, and behavior change

mechanisms built into the structure of social media (retweets, likes, comments) and the underlying algorithms allow campaigns to adapt strategies and messaging in real time. For instance, the Trump campaign was measuring responses to [40–50,000 different variants of ads](#) every day, then adapting and evolving their messaging based on that feedback. And unlike TV and print, where political ads are declared, online political messaging can be more amorphous, often appearing more like news reporting or user messaging. Only recently, [The New York Times reported](#) on Facebook's disclosure that "it had identified more than \$100,000 worth of divisive ads on hot-button issues purchased by a shadowy Russian company linked to the Kremlin. Most of the 3,000 ads did not refer to particular candidates but instead focused on divisive social issues such as race, gay rights, gun control, and immigration... The ads, which ran between June 2015 and May 2017, were linked to some 470 fake accounts and pages the company said it had shut down."

As [Tim Wu](#) notes in an Op-Ed on the issue: "Perhaps the greatest problem for a democracy is that companies like Facebook and Twitter lack a serious financial incentive to do anything about matters of public concern, like the millions of fake users who are corrupting the democratic process."³⁴

In parallel, [a Berkman Klein Center study](#) found that in recent years, a sustained campaign from Breitbart News has led to the "right-shifting" of American public discourse and expectations; a timely issue, given that the expansion of Breitbart into Europe is already creating similar ripples.³⁵ Even more, as other research, supported partially by The Omidyar Group, suggests, voter manipulation in order to create and maintain inequality in political access and involvement is a distinct strategic goal of such advertising: "digital advertising limits algorithmic opportunity to access and acquire political information. Voters are strategically defined, and information inequality is created between the arbitrarily defined 'strategically important' and 'strategically unimportant.' Discriminately defined by campaigns, different voters receive different information, thereby engaging differently in politics."³⁶

Social media platforms have become a [vastly preferred channel for digital advertising spend](#); this is even more so the case for mobile advertising.³⁷ In turn, this has further separated publisher and journalist, cannibalized the main revenue sources of traditional news organizations, and depleted their ability to produce quality news, hold the powerful to account, or to be gatekeepers of the quality and terms of discourse. As a result, the monetization of our attention is undermining journalism, which is traditionally a contributor to accountability and, ultimately, democracy.

³⁴ Tim Wu, *Please Prove You're Not a Robot*, *The New York Times*, Published July 15, 2017, <https://www.nytimes.com/2017/07/15/opinion/sunday/please-prove-youre-not-a-robot.html>.

³⁵ Interview with David Schraven, Correctiv (Germany), where his group has documented the dis-information campaign that Breitbart and partners are undertaking, focusing on white supremacy, anti-immigration and anti-EU sentiment; and Berkman Klein Center, "Partisan Right-Wing Websites Shaped Mainstream Press Coverage Before 2016 Election, Berkman Klein Study Finds," Published August 16, 2017, <https://cyber.harvard.edu/node/99982>.

³⁶ Kim, Y. 2017, *Algorithmic Opportunity: Digital Advertising and Inequality in Political Involvement*. *The Forum*, 14(4), pp. 471-484. Retrieved 12 Sep. 2017, from doi:10.1515/for-2016-0034. Also, see <http://www.eyeelections.com/>.

³⁷ Google and Facebook accounted for 75% of all new online ad spending in 2015. In the US, \$0.85 of every new dollar spent on digital went to the two companies in the first quarter of 2016. Matthew Garrahan, *Advertising: Facebook and Google build a duopoly*, *Financial Times*, Published June 23, 2016, <https://www.ft.com/content/6c6b74a4-3920-11e6-9a05-82a9b15a8ee7>.

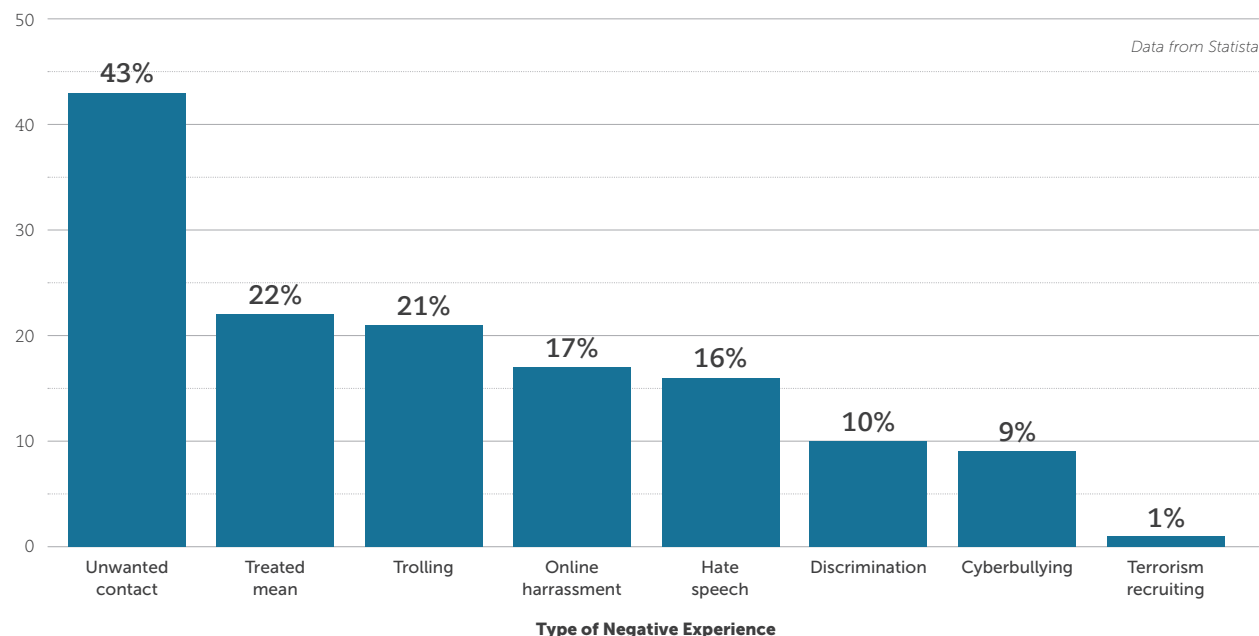
Issue Six: Intolerance, exclusion of disadvantaged or marginalized voices, public humiliation, and hate speech

Much has been written about the anonymity and distance that gives social media its promise and power to promote speech, for example, in repressive contexts such as the Arab Spring, in cases involving whistle-blowers, and in allowing individuals to speak freely about gender, representation, and identity. However, as Figure 2 shows, it is also the case that such design enables hate speech by individuals, public humiliation, and racial and sexual harassment.

In Myanmar, where there is much instability and violence within the country, major portions of the content on Facebook—a significant source of news for locals—are said to be divisive and hateful.³⁸ BuzzFeed has [documented](#) that the issue extends to the political elite. Rappler has found similar evidence in the Philippines.³⁹

With respect to disadvantaged or minority voices, demeaning hashtag labels have become an organizing mechanism to target a certain group, e.g., “snowflake” for liberals in America; “libtard” (liberal+retard) or “sickular” (slur for secularist); “presstitutes” in the Philippines and India are commonly used hashtag slurs to troll those who speak unfavorably of the populist governments. The profusion of such content, and the relative lack of grievance and redress options, can result in significant norm-shifting over time as to what is acceptable and allowable. Constant online abuse can lead members of frequently targeted groups—women and minorities—to self-censor or opt out of participating in public.

Figure 2: % of users reporting negative experiences on social media



³⁸ The Institute for War and Peace Reporting found [565 cases](#) of hate speech on social and broadcast media in Myanmar from August to October last year, with just over half targeted to Muslims and Rohingya. Yi Shu Ng, *Facebook finds it tough figuring out real hate speech from puns*, Mashable, Published May 29, 2017, http://mashable.com/2017/05/29/facebook-myanmar-hate-speech/#M9iVt2v_jkqi

and Hein Ko Soe, *Hitting back against hate speech*, Frontier Myanmar, Published September 16, 2016, <https://frontiermyanmar.net/en/hitting-back-against-hate-speech>.

³⁹ See <http://mashable.com/2017/05/29/facebook-myanmar-hate-speech/#80bBOVXA9kqB>

Looking Forward

Social media platforms are ingrained in our daily lives and provide much of the infrastructure of democratic debate. They have essentially become the modern “public square,” and they have command over both our attention and much of our personal data. As we study this issue more closely, it’s clear that we have much more to learn and many difficult questions to answer.

Among these questions: what responsibility do these companies have in acknowledging their role and taking

accountability for shaping information and news, and affecting public opinion and behavior?⁴⁰ Some argue that just as it would be unconscionable for the editor of a traditional media outlet to not care about what he or she publishes, promotes, or endorses, the designer of a social media site’s non-neutral algorithms has a responsibility to acknowledge his or her code’s role in selecting which content to prioritize or de-prioritize, and the implications that these models have on democratic deliberation.

The Omidyar Group’s Near Term Approach

The Omidyar Group’s approach to the challenges of online hate speech includes a partnership in the U.S. with the Anti-Defamation League to establish a Silicon Valley center that will focus on finding solutions. Indeed, platforms have also struggled to remove content judged to be problematic for several reasons; given the difficulty of setting up normative codes, content standards, and policing capacity that would allow for different outcomes. Such efforts— whether for terrorism, bullying and violent language, or targeted trolling and hate—create challenges for platforms that are reluctant to determine who adjudicates speech, where to draw the line regarding freedom of expression, and how to differentiate legitimate dissent from calls to violence or bullying. Of late, these platforms have formed a shared database to address content that violates their policies.

Additionally, to what extent can we expect these platforms to internally “self-regulate” their content and activity, especially if it has implications for their revenue model? While increased oversight might strengthen long-term trust and engagement, it might also come with nearer-term attention costs that would require mitigation. To an extent, firms like Facebook and Google have announced steps to combat these issues themselves—e.g. Google has founded [First Draft News Coalition](http://firstdraftnewscoalition.com/2016/12/23/zuckerberg-media-company/), and Facebook is joining The Ford Foundation, Mozilla, and Craigslist Founder Craig Newmark in funding the [News Integrity Initiative](http://newsintegrityinitiative.org/).⁴¹ While seen as positive, these steps do not fully answer questions about their commitment.⁴²

Other examples of self-regulation might include:

- Technological/engineering decisions taken by social media firms to change how their algorithms work to address these issues
- Decisions to allocate at-scale resources to community standards, including effective grievance redressal mechanisms
- Adherence to self-generated industry standards and commitments to minimizing the unintended consequences of social media technologies and platforms
- In the interim, measures such as partnerships with institutions to seed and innovate solutions on critical topics, with which the social media platforms acknowledge they require assistance and external expertise, e.g. Facebook’s decision to share
- Information and partnership with the [International Fact Checking Network](http://internationalfactcheckingnetwork.org/) and with Germany’s [Correctiv](http://correctiv.org/) to combat the spread of mis- and dis-information
- Social media platforms could also commit to greater transparency around dark advertising, collaborating with researchers and civil society to provide access to data about how information is being shared and targeted, especially when this is shown to be in the public interest.

⁴⁰ Mathew Ingram, *Mark Zuckerberg Finally Admits Facebook Is a Media Company*, *Fortune*, Published December 23, 2016, <http://fortune.com/2016/12/23/zuckerberg-media-company/>

⁴¹ Democracy Fund works closely with the News Integrity Initiative

⁴² Have the platforms determined they need to make substantial investments to be more than agnostic about their impact on

democracy, and ensure the platform is as pro-democracy as possible? Will they be willing to bear the at-scale costs of doing so and engineer solutions to create disincentives for viral deceptions, less homogeneous networks of thought/echo chambers, etc. at the cost of making the platform less “addictive”? And what can the public do to encourage such efforts?

Thinking Long Term

As we consider how the future could unfold, we believe it is necessary to think about the broader context, the nature of the public square, and how facts and evidence play a role. For example, in prior decades there were campaigns against the effects of smoking on individuals, which changed behavior significantly over time and improved public health overall. Going forward, we need to ensure that the public square is not corrupted and that science and evidence, and the process of making decisions, continues in a manner that is seen to be trusted and is also trustworthy.

Over the long term this could require engagement with journalists across the globe who want to improve their

data savvy, either by improved reporting on these issues or by working with them to involve social data science in their reporting. Many of the problems above require collaboration between social data scientists for discovery and journalists for storytelling.

It may also require engagement on questions of public policy by all parties. Countries around the world are struggling to understand steps they might take in response to these challenges and without proactive engagement from all parties (governments, industry, and advocacy groups) there is a real danger that democratic governments will do things that further foreclose the public square.

Our Commitment

The Omidyar Group recognizes that the public square is in a time of great transition, and we are supporting a variety of organizations aimed at increasing civic engagement, fostering trust in our institutions, and strengthening the role of journalism for the health of our democracy. In the United States, Democracy Fund has invested more than \$18 million in people and organizations working to create a more vibrant public square, including efforts to rebuild trust in journalism and to address the rise of misinformation. In particular, Democracy Fund has partnered with other funders to invest in prototype ideas to improve the flow of accurate information. It also supports the Engaging News Project and the Coral Project, two leading efforts to build a more engaged online public square. Recently, Democracy Fund and First Look Media [announced \\$12M](#) in grants to support investigative journalism and a robust, free press.

Globally, [Omidyar Network announced a \\$100M commitment](#) to fund investigative journalism, combat the spread of mis- and disinformation online, increase citizen engagement, and restore trust in institutions.⁴³

We are committed to seeking and supporting innovative solutions to mitigate the negative consequences of social media participation—and manipulation. We have published this paper to share our initial analysis and outstanding questions and to begin a productive dialogue, among diverse stakeholders, on this important topic. We look forward to continued sense-making and exploring possibilities to partner with others as we search for answers. Please email inquiries@omidyargroup.com to discuss how we might work together.

⁴³ The first recipients include the International Consortium of Investigative Journalists, the group behind last year's release of the Panama Papers, which will receive \$4.5 million.

Works Referenced

Framing/Cross-cutting:

- **Digital News Report 2016**, Reuters Institute for the Study of Journalism, <http://www.digitalnewsreport.org/>
- **News Use Across Social Media Platforms 2016**, Pew Research Center, <http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/>. See also: <http://www.journalism.org/2016/07/07/trust-and-accuracy/>
- **How Millennials Get News: Inside the habits of America's first digital generation**, American Press Institute, 2015, <https://www.americanpressinstitute.org/publications/reports/survey-research/millennials-news/>
- **James Stewart, Facebook Has 50 Minutes of Your Time Each Day. It Wants More.** *The New York Times*, Published May 5, 2016, <https://www.nytimes.com/2016/05/06/business/facebook-bends-the-rules-of-audience-engagement-to-its-advantage.html>
- **Rowland Manthorpe, It's the attention economy, stupid**, *Wired*, Published November 9, 2016, <http://www.wired.co.uk/article/us-president-donald-trump-attention-economy>
- **Daniel Nesbit, The New World of Online Political Advertising** (presentation, Stanford, 2015–16). Available at: <https://law.stanford.edu/publications/the-new-world-of-online-political-advertising/>
- **Nathaniel Persily, The Campaign Revolution Will Not Be Televised**, *The American Interest*, Published October 10, 2015. See also: white papers from Nate Persily's practicum at Stanford Law School: <https://law.stanford.edu/education/only-at-sls/law-policy-lab/practicums-2015-2016/campaign-reform-and-the-future-of-political-communication/>
- **John Danaher, The Threat of Algocracy: Reality, Resistance and Accommodation**, *Philos. Technol.* (2016) 29: 245. <https://doi.org/10.1007/s13347-015-0211-1>
- **Talia Stroud, Attention as a Valuable Resource** (Unpublished)
- **Joseph S. Nye Jr., Will the Liberal Order Survive?** *Foreign Affairs*, January/February 2017. Available at <https://www.foreignaffairs.com/articles/2016-12-12/will-liberal-order-survive>
- **Cass Sunstein, #Republic: Divided Democracy in the Age of Social Media**, Princeton University Press, 2017.
- **Ethan Zuckerman, Self-segregation on social networks and the implications for the Ferguson, MO story**, *EthanZuckerman.com*, Published August 27, 2014, <http://www.ethanzuckerman.com/blog/2014/08/27/self-segregation-on-social-networks-and-the-implications-for-the-ferguson-mo-story/>
- **danah boyd, Hacking the Attention Economy**, *Data & Society*, Published January 5, 2017, <https://points.datasociety.net/hacking-the-attention-economy-9fa1daca7a37>
- **danah boyd, Why America is Self-Segregating**, *Data & Society*, Published January 5, 2017, <https://points.datasociety.net/why-america-is-self-segregating-d881a39273ab>
- **Joshua Bleiberg and Darrell M. West, "Political polarization on Facebook,"** *Brookings Institute*, Published May 13, 2015, <https://www.brookings.edu/blog/techtank/2015/05/13/political-polarization-on-facebook/>
- **Art Swift, Americans' Trust in Mass Media Sinks to New Low**, *Gallup*, Published September 14, 2016, <http://www.gallup.com/poll/195542/americans-trust-mass-media-sinks-new-low.aspx>
- **Dirk Helbing et al, Will Democracy Survive Big Data and Artificial Intelligence?** *Scientific American*, Published February 25, 2017, <https://www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence/>
- **Jamie Condliffe, Facebook Forbids the Use of User Data for Surveillance**, *MIT Technology Review*, Published March 14, 2017, <https://www.technologyreview.com/s/603855/facebook-forbids-the-use-of-user-data-for-surveillance/>

Interviews:

- **Maria Ressa**, *Rappler* (Philippines). 2017
- **David Schraven**, *Correctiv* (Germany). 2017

Issue One: Echo chambers, polarization, and hyper-partisanship

- **Ethan Zuckerman, Self-segregation on social networks and the implications for the Ferguson, MO story**, *EthanZuckerman.com*, Published August 27, 2014, <http://www.ethanzuckerman.com/blog/2014/08/27/self-segregation-on-social-networks-and-the-implications-for-the-ferguson-mo-story/> (repeat citation)

WORKS REFERENCED

- John McDermott, *Why Facebook is for ice buckets, Twitter is for Ferguson*, Digiday, Published August 20, 2014, <https://digiday.com/media/facbeook-twitter-ferguson/>
 - American Press Institute, 'Who shared it?' How Americans decide what news to trust on social media, Published March 20, 2017, <https://www.americanpressinstitute.org/publications/reports/survey-research/trust-social-media/>
 - The Wall Street Journal, "Blue Feed, Red Feed", Published May 18, 2016, <http://graphics.wsj.com/blue-feed-red-feed/>
 - Yochai Benkler, Robert Faris, Hal Roberts, and Ethan Zuckerman, *Study: Breitbart-led right-wing media ecosystem altered broader media agenda*, Columbia Journalism Review, Published March 3, 2017, <https://www.cjr.org/analysis/breitbart-media-trump-harvard-study.php>
 - Political Polarization in the American Public, Pew Research Center, Published June 12, 2014, <http://www.people-press.org/2014/06/12/political-polarization-in-the-american-public/>
 - Ezra Klein and Alvin Chang, 'Political identity is fair game for hatred': how Republicans and Democrats discriminate," Vox, December 7, 2015, <https://www.vox.com/2015/12/7/9790764/partisan-discrimination>
 - Daniel Cox, Ph.D. and Robert P. Jones, Ph.D., *Beyond Guns and God: Understanding the Complexities of the White Working Class in America*, Public Religion Research Institute, Published September 20, 2012, <https://www.pri.org/research/race-class-culture-survey-2012/>
 - Jason Patinkin, *How to Use Facebook and Fake News to Get people to Murder Each Other*, BuzzFeed, Updated January 16, 2017, <https://www.buzzfeed.com/jasonpatinkin/how-to-get-people-to-murder-each-other-through-fake-news-and?>
 - Philip N. Howard, Gillian Bolsover, Bence Kollanyi, et al, *Junk News and Bots during the U.S. Election: What Were Michigan Voters Sharing Over Twitter?* Comprop Data Memo 2017.3. Oxford, UK: Project on Computational Propaganda. <http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/03/What-Were-Michigan-Voters-Sharing-Over-Twitter-v2.pdf>
 - John Markoff, *Automated Pro-Trump Bots Overwhelmed Pro-Clinton Messages, Researchers Say*, The New York Times, Published November 17, 2016, <https://www.nytimes.com/2016/11/18/technology/automated-pro-trump-bots-overwhelmed-pro-clinton-messages-researchers-say.html>
 - Guillaume Chaslot, *YouTube's A.I. was not neutral in the US presidential election*, Medium, Published November 27, 2016, <https://medium.com/the-graph/youtubes-ai-is-neutral-towards-clicks-but-is-biased-towards-people-and-ideas-3a2f643dea9a>
 - Samanth Subramanian, *Inside the Macedonian Fake-News Complex*, Wired, Published February 15, 2017, <https://www.wired.com/2017/02/veles-macedonia-fake-news/>
 - Tim Harford, *The problem with facts*, Financial Times, Published March 8, 2017, <https://www.ft.com/content/eef2e2f8-0383-11e7-ace0-1ce02ef0def9>
 - Michael Barthel, Amy Mitchell and Jesse Holcomb, *Many Americans Believe Fake News Is Sowing Confusion*, Pew Research Center, Published June 15, 2016, <http://www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/>
 - Dan Tynan, *How Facebook powers money machines for obscure political 'news' sites*, The Guardian, Published August 24, 2016, <https://www.theguardian.com/technology/2016/aug/24/facebook-clickbait-political-news-sites-us-election-trump>
 - Craig Silverman and Lawrence Alexander, *How Teens In The Balkans Are Duping Trump Supporters With Fake News*, BuzzFeed, Published November 3, 2016, <https://www.buzzfeed.com/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo?>
- Issue Two:** Proliferation of several types of misinformation and disinformation
- Claire Wardle, *Fake news. It's complicated*, First Draft News, Published February 16, 2017, <https://firstdraftnews.com/fake-news-complicated/>
 - American Press Institute, 'Who shared it?' How Americans decide what news to trust on social media, Published March 20, 2017, <https://www.americanpressinstitute.org/publications/reports/survey-research/trust-social-media/> (repeat citation)
 - Andrew Breiner, *Pizzagate, explained*, Salon, Published December 10, 2016, <http://www.salon.com/2016/12/10/pizzagate-explained-everything-you-want-to-know-about-the-comet-ping-pong-pizzeria-conspiracy-theory-but-are-too-afraid-to-search-for-on-reddit/>

WORKS REFERENCED

- **Emma Grey Ellis, *Fake Think Tanks Fuel Fake News—And the President's Tweets***, *Wired*, Published January 24, 2017, <https://www.wired.com/2017/01/fake-think-tanks-fuel-fake-news-presidents-tweets/>
 - **Nick Bilton, *Fake news is about to get even scarier than you ever dreamed***, *Vanity Fair*, Published January 26, 2017, <https://www.vanityfair.com/news/2017/01/fake-news-technology>
 - **Amar Toor, *Germany passes controversial law to fine Facebook over hate speech.***, *The Verge*, Published June 30, 2017, <https://www.theverge.com/2017/6/30/15898386/germany-facebook-hate-speech-law-passed>
 - **Elizabeth Kolbert, *Why facts don't change our minds***, *The New Yorker*, Published February 27, 2017, <http://www.newyorker.com/magazine/2017/02/27/why-facts-dont-change-our-minds>
 - **Craig Silverman, *The Backfire Effect***, *Columbia Journalism Review*, Published June 17, 2011, http://archives.cjr.org/behind_the_news/the_backfire_effect.php
 - **Sam Levin, *Facebook promised to tackle fake news. But the evidence shows it's not working***, *The Guardian*, Original Publication May 16, 2017, <https://www.theguardian.com/technology/2017/may/16/facebook-fake-news-tools-not-working>
 - **Brendan Nyhan, *Why More Democrats Are Now Embracing Conspiracy Theories***, *The New York Times*, Published February 15, 2017, <https://www.nytimes.com/2017/02/15/upshot/why-more-democrats-are-now-embracing-conspiracy-theories.html>
 - **American Press Institute, *Fact checking research***, <https://www.americanpressinstitute.org/fact-checking-research-studies/>
- Issue Three:** Editorial conflation of popularity, legitimacy and user intentionality
- **Edelman Trust Barometer 2017**, <http://www.edelman.com/global-results/>
 - **Tom Nichols, *How America Lost Faith in Expertise***, *Foreign Affairs*, March/April 2017 issue, <https://www.foreignaffairs.com/articles/united-states/2017-02-13/how-america-lost-faith-expertise>
 - **Claire Cain Miller, *When Algorithms Discriminate***, *The New York Times*, Published July 9, 2015, <https://www.nytimes.com/2015/07/10/upshot/when-algorithms-discriminate.html>
 - **UN Women (website), *UN Women ad series reveals widespread sexism***, October 21, 2013, <http://www.unwomen.org/en/news/stories/2013/10/women-should-ads>
 - **David Auerbach, *Filling the Void***, *Slate*, Published November 2013, http://www.slate.com/articles/technology/bitwise/2013/11/google_autocomplete_the_results_aren_t_always_what_you_think_they_are.html
 - **Seth Stephens-Davidowitz, *Everybody lies: how Google search reveals our darkest secrets***, *The Guardian*, Published July 9, 2017, <https://www.theguardian.com/technology/2017/jul/09/everybody-lies-how-google-reveals-darkest-secrets-seth-stephens-davidowitz>
 - **Nikki Williams, *Bot Journalism: Who's Writing That Piece You're Reading?*** Center for Digital Ethics and Policy, <http://digitalethics.org/essays/bot-journalism-whos-writing-that-piece-youre-reading/>
- Issue Four:** Political capture
- **Samantha Bradshaw & Philip N. Howard, *Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation***. Samuel Woolley and Philip N. Howard, Eds. Working Paper 2017.12. Oxford, UK: Project on Computational Propaganda. <http://comprop.oii.ox.ac.uk/>
 - **Leo Benedictus, *Invasion of the troll armies: from Russian Trump supporters to Turkish state stooges***, *The Guardian*, Published November 6, 2016, <https://www.theguardian.com/media/2016/nov/06/troll-armies-social-media-trump-russian>
 - **Henry Farrell, *Chinese government fakes nearly 450 million social media comments a year. This is why***, *The Washington Post*, Published May 19, 2016, <https://www.washingtonpost.com/news/monkey-cage/wp/2016/05/19/the-chinese-government-fakes-nearly-450-million-social-media-comments-a-year-this-is-why/?>
 - **Megha Rajagopalan, *Internet Trolls Are Using Facebook To Target Myanmar's Muslims***, *BuzzFeed*, Published March 18, 2017, <https://www.buzzfeed.com/meghara/how-fake-news-and-online-hate-are-making-life-hell-for?>
 - **Matthias Lüfkens, *How world leaders build their massive Twitter followings***, *Quartz*, Published May 7, 2015, <https://qz.com/399530/how-world-leaders-build-their-massive-twitter-followings/>

WORKS REFERENCED

- **Michael Bueza, IN NUMBERS: The Philippines' 'war on drugs,'** *Rappler*, Updated Publication April 23, 2017, <https://www.rappler.com/newsbreak/iq/145814-numbers-statistics-philippines-war-drugs>
- **Maria Ressa, Propaganda war: Weaponizing the Internet,** *Rappler*, Updated October 3, 2016, <http://www.rappler.com/nation/148007-propaganda-war-weaponizing-internet>
- **Twitter Blog, Twitter Samvad brings digital governance to the masses,** Posted March 24, 2015, https://blog.twitter.com/official/en_in/a/2015/twitter-samvad-brings-digital-governance-to-the-masses.html
- **Bence Kollanyi, Philip N. Howard and Samuel C. Woolley, Bots and Automation over Twitter during the U.S. Election,** <http://comprom.oii.ox.ac.uk/wp-content/uploads/sites/89/2016/11/Data-Memo-US-Election.pdf>
- **Michael Newberg, As many as 48 million Twitter accounts aren't people, says study,** *CNBC*, Published March 10, 2017, <http://www.cnn.com/2017/03/10/nearly-48-million-twitter-accounts-could-be-bots-says-study.html>
- **Alessandro Bessi and Emilio Ferrara, Social bots distort the 2016 U.S. Presidential election (online discussion),** *First Monday*, Volume 21, Number 11-7 November 2016, <http://firstmonday.org/ojs/index.php/fm/article/view/7090/5653>.
- **Jonathon Morgan, Sockpuppets, Secessionists, and Breitbart,** *Medium*, <https://medium.com/data-for-democracy/sockpuppets-secessionists-and-breitbart-7171b1134cd5>
- **Raghu Karnad, It's Not What Modi Is Tweeting – It's What He Is Reading,** *The Wire*, Published September 2, 2017, <https://thewire.in/107145/narendra-modi-twitter-trolls/>
- **Andrew Marantz, Trolls for Trump,** *The New Yorker*, Published in October 31, 2016 issue, <http://www.newyorker.com/magazine/2016/10/31/trolls-for-trump>
- **Chay F. Hofileña, Fake accounts, manufactured reality on social media,** *Rappler*, Updated October 15, 2016, <http://www.rappler.com/newsbreak/investigative/148347-fake-accounts-manufactured-reality-social-media>
- **Maria Ressa, How Facebook algorithms impact democracy,** *Rappler*, Updated October 9, 2016, <http://www.rappler.com/newsbreak/148536-facebook-algorithms-impact-democracy>
- **Issue Five: Manipulation, micro-targeting and behavior change**
- **Philip N. Howard and Robert Gorwa, Facebook could tell us how Russia interfered in our elections. Why won't it?,** *The Washington Post*, Updated May 20, 2017, https://www.washingtonpost.com/opinions/facebook-could-tell-us-how-russia-interfered-in-our-elections-why-wont-it/2017/05/19/c061a606-3b21-11e7-8854-21f359183e8c_story.html
- **Hannes Grassegger and Mikael Krogerus, The Data That Turned the World Upside Down,** *Vice*, Published January 28, 2017, https://motherboard.vice.com/en_us/article/mg9vvn/how-our-likes-helped-trump-win
- **Berit Anderson, The Rise of the Weaponized AI Propaganda Machine,** *Medium*, Published February 12, 2017, <https://medium.com/join-scout/the-rise-of-the-weaponized-ai-propaganda-machine-86dac61668b>
- **Mathew Ingram, How Google and Facebook Have Taken Over the Digital Ad Industry,** *Fortune*, Published January 4, 2017, <http://fortune.com/2017/01/04/google-facebook-ad-industry/>
- **Matthew Garrahan, Advertising: Facebook and Google build a duopoly,** *Financial Times*, Published June 23, 2016, <https://www.ft.com/content/6c6b74a4-3920-11e6-9a05-82a9b15a8ee7>
- **Young Mie Kim, Algorithmic Opportunity: Digital Advertising and Inequality in Political Involvement,** *The Forum: A Journal of Applied Research in Contemporary Politics*. (December 2016) 4:359. <https://doi.org/10.1515/for-2016-0034>
- **Glenn Greenwald and Ewan MacAskill, NSA Prism program taps in to user data of Apple, Google and others,** *The Guardian*, Original Publication June 7, 2013, <https://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data>
- **Tim Wu, Please Prove You're Not a Robot,** *The New York Times*, Original Publication July 15, 2017, <https://www.nytimes.com/2017/07/15/opinion/sunday/please-prove-youre-not-a-robot.html>
- **Berkman Klein Center, Partisan Right-Wing Websites Shaped Mainstream Press Coverage Before 2016 Election, Berkman Klein Study Finds,** August 16, 2017, <https://cyber.harvard.edu/node/99982>

* May require a subscription to access

Issue Six: Intolerance, exclusion of marginalized voices, public humiliation and hate speech

- **Rajagopalan, Internet Trolls Are Using Facebook To Target Myanmar's Muslims** (repeat citation)
- **Ressa, Propaganda war: Weaponizing the Internet.** (repeat citation)
- **Robert Barnes, Supreme Court case tests the limits of free speech on Facebook and other social media**, The Washington Post, Published November 23, 2014, https://www.washingtonpost.com/national/supreme-court-case-tests-the-limits-of-free-speech-on-facebook-and-other-social-media/2014/11/23/9e54dbd8-6f67-11e4-ad12-3734c461eab6_story.html?utm_term=.44b93e6f40d0
- **Hein Ko Soe, Hitting back against hate speech**, Frontier Myanmar, Published September 16, 2016, <https://frontiermyanmar.net/en/hitting-back-against-hate-speech>
- **Yi Shu Ng, Facebook finds it tough figuring out real hate speech from puns**, Mashable, Published May 29, 2017, http://mashable.com/2017/05/29/facebook-myanmar-hate-speech/#M9iVt2v_jkqi

Looking Forward and Near Term Approach

- **Mathew Ingram, Mark Zuckerberg Finally Admits Facebook Is a Media Company**, Fortune, Published December 23, 2016, <http://fortune.com/2016/12/23/zuckerberg-media-company/>
- **Katie Collins, Facebook spearheads \$14M consortium to counter fake news**, CNet, Published April 3, 2017, <https://www.cnet.com/news/facebook-spearheads-14m-consortium-to-counter-fake-news/>
- **Elle Hunt, 'Disputed by multiple fact-checkers': Facebook rolls out new alert to combat fake news**, The Guardian, Published March 22, 2017, <https://www.theguardian.com/technology/2017/mar/22/facebook-fact-checking-tool-fake-news>
- **Natasha Lomas, Facebook takes its fake news fight to Germany**, Tech Crunch, Published January 16, 2017, <https://techcrunch.com/2017/01/16/facebook-takes-its-fake-news-fight-to-germany/>
- **Davey Alba, YouTube's Ad Problems Finally Blow Up in Google's Face**, Wired, Published March 25, 2017, <https://www.wired.com/2017/03/youtubes-ad-problems-finally-blow-googles-face/>

- **Anthony Faiola and Stephanie Kirchner, How do you stop fake news? In Germany, with a law**, The Washington Post, Published April 5, 2017, https://www.washingtonpost.com/world/europe/how-do-you-stop-fake-news-in-germany-with-a-law/2017/04/05/e6834ad6-1a08-11e7-bcc2-7d1a0973e7b2_story.html?
- **Laura Hazard Owen. Republicans seem more susceptible to fake news than Democrats (but liberals, don't feel too comfy yet)**, NeimanLab, Published May 5, 2017, <http://www.neimanlab.org/2017/05/republicans-seem-more-susceptible-to-fake-news-than-democrats-but-liberals-dont-feel-too-comfy-yet/>
- **Zack Beauchamp, Democrats are falling for fake news about Russia**, Vox, Published May 19, 2017, <https://www.vox.com/world/2017/5/19/15561842/trump-russia-louise-mensch>

Thinking Long Term

- **April Simpson, First Look Media and Democracy Fund to provide \$12M for investigative journalism**, Current.org, Published March 27, 2017, <https://current.org/2017/03/first-look-media-and-democracy-fund-to-provide-12m-for-investigative-journalism/>
- **Anita Singh, eBay founder Pierre Omidyar commits \$100m to fight 'fake news' and hate speech**, The Telegraph, Published April 4, 2017, <http://www.telegraph.co.uk/news/2017/04/04/ebay-founder-pierre-omidyar-commits-100m-fight-fake-news-hate/>



omidyargroup.com

For questions or comments, please contact:
inquiries@omidyargroup.com