1

Introduction

Nathaniel Persily and Joshua A. Tucker

Widespread concern about the effects of social media on democracy has led to an explosion in research over the last five years. This research comes from disparate corners of academia: departments of political science, psychology, law, communication, economics, and computer science, alongside new initiatives in data science and even artificial intelligence. A new field is forming, and it is time to take stock of what we know, what we need to know, and how we might find it out. That is the purpose of this book.

Of course, research on the impact of technology, in general, and the Internet, in particular, on democracy is not new. The early utopianism of the Internet proffered a theory of "liberation technology" – a mode of unimpeded, transnational communication that would disrupt authoritarian regimes and promote freedom around the world. Similarly, research exists on the impact of this new technology focused on phenomena such as small donor fundraising, online community building, and the subversive use of the Internet in protests and campaigns, in both democratic and nondemocratic regimes. However, early research was scant and far from systematic, as it tended to rely on studies of blogs or individual campaigns. Yet, to the extent the research hinted at some normative argument as to the Internet's potential, it largely pointed in a prodemocratic direction.

The 2016 presidential election in the United States and, to a lesser extent, the Brexit referendum earlier that year in the United Kingdom, changed the received wisdom. Looking for an explanation for those surprising results, many turned to the new technology of political communication. Blame was (and continues to be) cast on bots, foreign election interference, online disinformation, targeted ads, echo chambers, and related phenomena. Indeed, since 2016, analysis of any election, social movement, populist victory, or instance of political violence will almost inevitably include some assessment of the role of new technology in determining winners and losers.

As conventional wisdom concerning the effect of the Internet on democracy abruptly shifted, so too did much of the research. That shift was not uniform; in fact, one might say that two camps have emerged. The first emphasizes the rise of social media echo chambers, fake news, hate speech, "computational propaganda," authoritarian governments' online targeting of opponents,

1

threats to journalism, and foreign election interference. The other school challenges the independent significance of the shift in technology (as opposed to other sociological factors) while also suggesting that the magnitude and prevalence of the alleged technology-related problems are overblown.

The goal of this book is to synthesize the existing research on social media and democracy. We present reviews of the literature on disinformation, polarization, echo chambers, hate speech, bots, political advertising, and new media. In addition, we canvass the literature on reform proposals to address the widely perceived threats to democracy. We seek to examine the current state of knowledge on social media and democracy, to identify the many knowledge gaps and obstacles to research in this area, and to chart a course for future research. We hope to advocate for this new field of study and to suggest that universities, foundations, private firms, and governments should commit to funding and supporting this research.

We have also made a deliberate choice, which might be jarring to some readers, to include both scientific analysis and policy discussion in a single volume. We made this choice consciously because we worry that the policy community and the scientific community are not speaking to one another enough. We are concerned that reliance on untested conventional wisdom based on folk theories of technology's impact on democracy is leading to misguided reform proposals that may even worsen the problems they are attempting to solve. Conversely, we think the academics studying online harms are often uninformed about the legal regime in which the internet platforms operate. Rules relating to content moderation, antitrust, political advertising, and other domains of online speech structure the environment in which the alleged online harms of disinformation, polarization, and hate speech manifest. Social scientists need to appreciate the policy context, and policymakers need to understand the current state of knowledge regarding the harms they seek to manage through legislation and regulation.

We should also emphasize that the development of this field has become even more urgent as the Covid-19 pandemic further transforms the online information ecosystem. We undertook the research for this book in the year before the pandemic hit, but the topic has only grown in significance since then. Concerns about Covid-19–related disinformation, as well as how the platforms and governments have responded, have only increased as a result of the pandemic. Forced to stay at home, the mass public has, if anything, become even more dependent on platforms, such as Facebook, Google, and now Zoom, for information and communication services. As we write this, the platforms appear to be taking extraordinary measures against online speech deemed dangerous to public health and safety, but it remains to be seen whether the Covid-specific responses represent a new normal in regulation of disinformation. Either way, the need for empirically grounded understandings of the changing dynamics of online communication to inform public policy in this arena has become, if anything, even more important.

SUMMARY OF CHAPTERS

This book should be read with the goal in mind of providing an empirical foundation for sound public policy. The first half of the volume contains literature reviews on the central empirical questions surrounding social media and democracy: disinformation, polarization/echo chambers, hate speech, political advertising, bots/computational propaganda, and the changing landscape for journalism and mass media. The second half surveys reform proposals for both the platforms and governments: measures to correct misinformation, reforms of intermediary liability rules for platforms, comparative media regulation, and transparency measures. To be sure, the chapters do not completely cover the landscape of either "the problem" or the potential "solutions," but we hope that the volume provides a good introduction for those interested in understanding this emerging field.

In Chapter 2, Princeton professor Andrew M. Guess and University of Utah professor Benjamin A. Lyons survey the literature on online disinformation. As with all scholars in this field, they grapple with the difficulty of defining disinformation. How we define the problem significantly affects the observed prevalence of disinformation. They caution against attributing widespread beliefs in falsehoods predominantly to social media. However, they present their best estimates from available research as to how much disinformation exists, who produces it, and who consumes it.

In Chapter 3, Pablo Barberá, formerly a professor at the London School of Economics (when he wrote this chapter) but now a research scientist at Facebook and a professor at the University of Southern California, examines the topic of echo chambers and polarization. A conventional view of the problem posits that, given the explosion of online media sources, people are now able to opt into homogeneous media ecosystems, preselected to reinforce their prior beliefs. As a result, people today are less likely to share a common narrative of facts and news, because they exist in segregated "filter bubbles" or "information cocoons," particularly on social media. The results for politics are pernicious as compromise becomes less possible and election campaigns rely on mobilizing dramatically different bases rather than attempting to persuade moderate voters. The research Barberá surveys challenges this conventional view, however. Cross-cutting interactions on social media and exposure to diverse sources of news are at least as common as they are in the offline world and, in many cases, more likely. Ranking algorithms, often blamed for serving users what they want to see, do not appear to have as dramatic an effect on polarization as once assumed. Some people may live in segregated online news enclaves, but they appear to be a smaller share of the population than expected, at least in the Western democracies that form the bulk of examples in existing research.

A related issue to polarization is online hate speech, a topic covered in Chapter 4 by Alexandra A. Siegel, a professor at the University of Colorado

Boulder. She, too, grapples with the definition of the problem of concern, as have courts, policymakers, and the internet platforms themselves, which have tried to walk the difficult line between unprotected hate speech and permissible expression. Although a large share of users report experiences with online hate speech or harassment (however defined), for only a small share does it comprise a significant amount of the speech viewed on the mainstream platforms. Of course, for some users, such as journalists or high-profile speakers who are targeted, hate speech and threats will comprise a larger share of the communication they view. Moreover, on some platforms, such as Reddit, 4chan or 8chan, avowedly racist echo chambers can flourish. Siegel concludes by surveying studies that look at how online hate speech leads to changes in attitudes, as well as offline hate crimes, and then examines measures that have been successful in combating such speech.

In Chapter 5, Samuel C. Woolley, a professor at the University of Texas at Austin examines the role of bots and computational propaganda. As he notes, bots are simply "online software programs that run automated tasks." They can be used for good or ill and are responsible for roughly half of online traffic. When it comes to political bots, though, he notes that they are ordinarily developed to deceive – that is, to trick both users who read their messages and algorithms that can be manipulated to grant undeserved popularity to certain topics or accounts. He notes how bots are now used to intimidate elites and social groups, as well as to spread disinformation. The Internet's privileging of anonymity and automation is what gives political bots their power. Perhaps more than any other chapter in the volume, this discussion of bots isolates how new technology, itself, places stress on democracies. Whereas previous generations of media experienced disinformation, polarization, and hate speech, bots are a unique feature of the Internet Age.

Chapter 6, by Wesleyan professor Erika Franklin Fowler, Bowdoin College professor Michael M. Franz, and Washington State professor Travis N. Ridout, covers political advertising. It pays particular attention to the United States, since it is responsible for more political advertising than any other country in the world by orders of magnitude. The authors detail the regulatory vacuum into which online ads fall. As a result, the infamous Russian purchase of ads in the 2016 US presidential campaign should not be seen as such a surprise, given the absence of effective law governing online ads, especially so-called issue ads that discuss controversial topics rather than supporting or opposing particular candidates. The data on online advertising have been scarce until recently. Following the 2016 election, Google, Twitter, and Facebook all developed different ad archives that provide for greater transparency than the law requires and will allow scholars to analyze political advertising going forward. The authors present what data we have from previous elections, while highlighting the need for more detailed data from the platforms.

In Chapter 7, Professor Rasmus Kleis Nielsen and Richard Fletcher of the Reuters Institute at Oxford review the literature on the implications of the

transition to online media and journalism for democracy. They describe the impact of digital and mobile technologies on news organizations as a kind of "creative destruction." They show that the decline of newspapers started well before the rise of the Internet, but digital technologies have accelerated their decline. Websites destroyed the market for classified ads, which had been the lifeblood of local newspapers, but Google and Facebook have gained a duopoly on online advertising. Those firms free ride off the content produced by publishers while competing against those same publishers for advertising dollars. At the same time as the platforms are "disrupting" the business model for news, defining who or what constitutes "the news" or "the media" becomes complicated in the Internet Age, when anyone can blog, tweet, or post. The authors also note that news audiences have moved from a system of "direct discovery," in which audiences intentionally visit or receive the news from the original source, to "distributed discovery," in which the audience receives the news from "search engines, social media, and other platform products." The "automated serendipity" produced by search engines and social media leads online audiences to gain exposure to more sources of news than they would if limited to offline sources. The authors conclude that the rise of online news undermines established institutions of twentieth-century democracy, such as political parties, legacy media, and member-based interest groups, but that a new, more democratic media environment has benefits as well.

In Chapter 8, Chloe Wittenberg and Professor Adam J. Berinsky, both of MIT, discuss the different ways to correct misinformation. Their chapter is humbling, in large part because they describe how difficult it is to correct misinformation. Merely correcting misinformation with disclaimers or counter-speech rarely erases the false belief. Because of motivated reasoning and other factors, countering misinformation may backfire for some people and even reinforce false beliefs. They note that the most effective responses to misinformation require corrections from a source the believers trust, delivered in a way that affirms their worldview. As such, the context in which misinformation arises and different qualities of the person who engages with it will often determine how strategies must be tailored to address false beliefs.

In Chapter 9, the first of the "policy" or "reform" chapters, Stanford professor Francis Fukuyama and Andrew Grotto, director of the Stanford Program on Geopolitics, Technology, and Governance, focus on how different countries regulate legacy media, with an eye to how they might regulate the Internet and social media. Some countries, such as France, Germany, and Great Britain, have a long and robust tradition of public broadcasting. Always suspicious of centralized authority, the United States, in contrast, was late to establish a Corporation for Public Broadcasting, which never attained the power and popularity of its European counterparts. Southern European countries and many former Communist countries have found themselves with an oligarchic model of media

regulation – the most extreme form found in Italy during Silvio Berlusconi's monopolistic reign. The authors note that traditions in Europe with respect to broadcast regulation flow over into regulation of the Internet. In France, for example, the Macron government has established expedited procedures to deal with misinformation and ordered platforms to take down offending content. Germany, quite famously, passed the NetzDG, which makes internet platforms liable for certain illegal speech that occurs on their platform after they have been warned. For similar regulation to arise in the United States, old legal tools, such as the Fairness Doctrine or "must carry" provisions, or new conceptions of antitrust, would need to be developed to rein in the power of the platforms.

In Chapter 10, Daphne Keller, director of the Program on Platform Regulation at the Stanford Cyber Policy Center, and Paddy Leerssen, PhD candidate at the University of Amsterdam, review the literature on government and platform takedown of internet content. The authors point to the available data published by governments, academics, nongovernmental organizations (NGOs) and the platforms themselves as to how much content they take down and for which reasons. They also describe how laws that fail to factor in the operational realities of "notice and takedown" systems can have the effect of causing platforms to overcensor in order to avoid legal liability. Much remains to be learned as to the platforms' takedown of content based on either their community standards or legal obligations; but, from the available literature, Keller and Leerssen warn of high rates of false positives in both filtering and human review of content. Moreover, in the face of vague legal directives, platforms tend to overcensor to avoid liability, a finding that takes on added urgency in view of President Trump's May 2020 Executive Order on Preventing Online Censorship. The authors examine the range of takedowns from hate speech and intellectual property violations to terrorist content and the "right to be forgotten."

Tim Hwang, research fellow at the Center for Security and Emerging Technology at Georgetown University, deals with similar issues in Chapter 11, "Dealing with Disinformation: Evaluating the Case for Amendment of Section 230 of the Communications Decency Act." The chapter considers what, if any, amendments should be made to section 230 of the Communications Decency Act (CDA) in the name of protecting against disinformation. Hailed as a cornerstone of the free Internet, that legal provision largely immunizes platforms from liability for the speech of outsiders present on their sites, while also encouraging platforms to take action against certain categories of objectionable content. Hwang argues that confronting the problem of disinformation does not require undermining the core components of CDA 230. Ancillary regulations concerning transparency, bots, advertising, microtargeting, or a kind of "net neutrality" for platforms would not require changing section 230 but could be legislated independently. Hwang warns about changing the intermediary liability rules in section 230. Like Keller and Leerssen, he worries that platforms might overcorrect, take down more speech than required, and become less transparent.

In Chapter 12, Robert Gorwa and Professor Timothy Garton Ash, both of Oxford University, examine reforms to promote transparency of the platforms. They describe the various voluntary transparency reports regarding takedowns and takedown requests from governments, as laid out by the Global Network Initiative (created by firms and civil society groups). In addition, local regulations or platform-specific legal actions (such as Facebook's 2011 Federal Trade Commission [FTC] Consent Decree) may require additional transparency from the platforms. They detail how different platforms have adopted different transparency rules for their community standards, takedowns, advertising, and other domains. Most notable is Facebook's recent innovation in creating an outside oversight board that will hear appeals from content takedowns. Moreover, third parties have embarked on certain transparency efforts, such as when ProPublica attempted to crowdsource political advertisements on Facebook. They conclude with warnings about how some transparency measures, if poorly tailored, do little to further openness and better understanding of platform practices and can even backfire depending on how companies adapt to these new rules.

Chapter 13 presents a conclusion in which we discuss "The Challenges and Opportunities for Social Media Research." In particular, we stress the importance of access to new forms of data for public-facing research and note the new legal and ethical questions arising from research in this domain. We have entered a new world for research on fundamental questions of political communication and behavior. Although the amount of data now available for research in these areas is unprecedented, large companies control access to most of the data that contain the answers to the questions social scientists are now asking. How social scientists interact with these companies, let alone whether to accept funding and exclusive data access from them, has become a unique challenge for modern research. Moreover, privacy concerns, especially in the wake of the notorious Cambridge Analytica scandal, have led the major internet platforms to become increasingly restrictive of data access for researchers. In the name of protecting privacy, governments have clamped down as well, with laws such as the European General Data Protection Regulation (GDPR). Although GDPR includes an exception for research, lawyers at the platforms have interpreted the exception narrowly and continue to raise privacy objections as a significant barrier to research. Nevertheless, the importance of greater data access for analysts who produce research placed in the public domain (as opposed to internal researchers who work for the platforms) has never been greater. The many policy suggestions discussed in the second half of this volume require rigorous scientific research to inform their advocacy and implementation. How scholars will navigate this new research terrain remains an open question, but we hope this volume serves as a clarion call for regulators, firms, funders, and the research community to provide an answer.

IMPORTANCE OF THE RESEARCH

Despite the limitations to data access all acknowledge, the authors in this volume present certain important insights as to the effect of new communication technologies on democracy. First and foremost, it is now beyond doubt that the way in which citizens consume information about politics – and, consequently, the way in which elites produce information for citizens – has fundamentally changed over the past decade. Moreover, the landscape of political communication is in great flux – so much so that the research presented here will need to be updated in short order as new platforms emerge, existing platforms adopt new policies, and political actors adapt to the dynamics we identify.

Second, we need to update more frequently our substantive understanding of how people are exposed to and process political information. In other words, the questions addressed in the following chapters are important if we want to understand the functioning of politics in the current moment. Several chapters discuss issues that are genuinely new, such as Chapter 5 by Woolley on bots (social media accounts that produce content via automated algorithms) and Chapter 7 by Nielsen and Fletcher on the impact of the digital revolution on the media industry and individuals' news consumption. Others represent new takes on old questions, such as Chapter 4 by Siegel on hate speech and Chapter 2 by Guess and Lyons on political disinformation. Still others consider old questions that need to be considered anew in the digital environment: Chapter 3 by Barberá on political polarization, here considered in the context of social media usage, and Chapter 8 by Wittenberg and Berinsky on correcting misinformation.

Third, we find ourselves in a moment where there has been a radical transformation in the way we can actually study political activity employing both qualitative and, especially, quantitative analysis. The momentous development here has been the emergence of digital trace data – that is, digital records that are left behind from human activity that can subsequently be analyzed. Indeed, it is difficult to think of many aspects of day-to-day life that do not leave behind digital trace data given the ubiquity of smartphones and internet access – to say nothing of electronic locks, credit cards, and digitized transportation records (a feature of modern life that may become all the more important in the "contact tracing" world of Covid-19). For political science, however, the rise of social media may be the most transformative of all. For the first time in human history, we have real time records of millions – if not billions – of people as they discuss politics, share information about politics, and organize politically. Each of these actions simultaneously produces an archived, digitized record. We are also living through a period of time in which great strides have been made in how to extract information from enormous collections of electronic data generally (machine learning) and how to use statistical methods to analyze text (natural language processing and other

text-as-data tools). Taken together, these developments have unlocked whole new methods of studying politics and political behavior. Taking stock of what we can learn, have learned, and should be able to learn from these new methods of analysis is therefore important. Collectively, the research summarized in the ensuing chapters provides a window into these developments insofar as they pertain to social media and politics.

Finally, in the post-2016 US presidential election, post–Cambridge Analytica era, there has been tremendous pressure on policymakers to "do something" about many of the topics discussed in this volume. This pressure, however, presents a serious challenge in view of one of the primary conclusions of this volume: We are only scratching the surface of what we know about many of these phenomena. To present just one example, there is a great desire to design interventions to reduce the spread of fake news. Yet, if we do not know who shares fake news, why they share fake news, or even whether they are sharing fake news because they think it is true or, instead, because they agree with it ideologically and do not care if it is fake, how can we design appropriate policy to reduce its spread? Moreover, if we do not know the effects of exposure to fake news, then we cannot know the "benefits" of reducing exposure to it. Because everyone recognizes the potential harms of empowering companies, such as Facebook and Google, to be "arbiters of truth," the benefits of reducing exposure to fake news must be considerable to justify ceding that kind of power over the speech marketplace to profit-maximizing American companies. Thus, the kind of research that we report throughout this volume has a crucial role to play in informing policy decision-making. We hope that by gathering so much of it in one place, we can make it accessible for policymakers considering reform options. We also hope that by candidly expressing how little we know about the dynamics of social media and democracy in some domains, we issue a note of caution to reformers seeking, hastily and prematurely, "to do something or anything" before we understand the nature of the problems that need solving.