Article

Disinformation as a Threat to Deliberative Democracy

Political Research Quarterly
2021, Vol. 74(3) 703–717

DOI: 10.1177/1065912920938143
journals.sagepub.com/home/prq
SAGE

Spencer McKay[1] and Chris Tenove[1]

Abstract

It is frequently claimed that online disinformation threatens democracy, and that disinformation is more prevalent or harmful because social media platforms have disrupted our communication systems. These intuitions have not been fully developed in democratic theory. This article builds on systemic approaches to deliberative democracy to characterize key vulnerabilities of social media platforms that disinformation actors exploit, and to clarify potential anti-deliberative effects of disinformation. The disinformation campaigns mounted by Russian agents around the United States' 2016 election illustrate the use of anti-deliberative tactics, including *corrosive falsehoods, moral denigration*, and *unjustified inclusion*. We further propose that these tactics might contribute to the system-level anti-deliberative properties of *epistemic cynicism, techno-affective polarization*, and *pervasive inauthenticity*. These harms undermine a polity's capacity to engage in communication characterized by the use of facts and logic, moral respect, and democratic inclusion. Clarifying which democratic goods are at risk from disinformation, and how they are put at risk, can help identify policies that go beyond targeting the architects of disinformation campaigns to address structural vulnerabilities in deliberative systems.

Keywords

disinformation, deliberative democracy, media regulation, systems, social media, political communication

The threat that online disinformation poses to democracy has arguably made it "*the* defining political communication topic of our time" (Freelon and Wells 2020, 145), particularly in the wake of high-profile Russian interference in the 2016 U.S. presidential election. Referring to Russia's role in that election, former Central Intelligence Agency Acting Director Michael Morell stated, "It is an attack on our very democracy. It's an attack on who we are as a people . . . this is to me not an overstatement, this is the political equivalent of 9/11" (Morell and Kelly 2016). In its *Online Harms White Paper*, the United Kingdom similarly warns that there is "a real danger that hostile actors use online disinformation to undermine our democratic values and principles" (United Kingdom Parliament 2019, 5). These concerns have motivated widespread demands for policy changes, although these demands are often unclear about the democratic values and principles at risk (Tenove 2020).

How might disinformation harm democracy? One possibility, which has received extensive attention from policymakers and researchers, is that disinformation may change election outcomes. This emphasis has serious limitations. First, concerns about disinformation changing electoral outcomes may be overblown. There is an ongoing debate about whether disinformation can significantly influence voting preferences (Allcott and Gentzkow 2017; Guess, Nyhan, and Reifler 2018), in part because voting preferences tend to be relatively stable despite campaign messaging (Kalla and Broockman 2018). Second, empirical research shows that disinformation campaigns are frequently designed to achieve goals other than changing election outcomes, such as undermining the institutions and social conditions necessary for democracies to function. Third, from a normative perspective, democracy is not reducible to elections. As a complement to an emphasis on electoral outcomes, we argue that a deliberative democracy framework can help clarify the normative harms of disinformation in ways that make sense of the growing empirical literature on the tactics, aims, and outcomes of disinformation campaigns.

[1]The University of British Columbia, Vancouver, Canada

Corresponding Author:
Chris Tenove, Department of Political Science, The University of British Columbia, C425-1866 Main Mall, Vancouver, Canada BC V6T 1Z1.
Email: cjtenove@mail.ubc.ca

Researchers and policymakers understand disinformation to be *intentionally* false or deceptive communication, used to advance the aims of its creators or disseminators at the expense of others (Wardle and Derakhshan 2017).[1] Disinformation actors employ complex mixtures of fact and fabricated content, manipulated images and videos, false information sources, automated accounts or "bots," and other tactics (Jack 2017; Wardle and Derakhshan 2017). Thus, disinformation may promote false understandings by means other than delivering false content, such as using fake identities or true but misleading content to trigger false inferences (Fallis 2015). Moreover, disinformation campaigns often seek to amplify social divisions and distrust in addition to promoting misperceptions (Arif, Stewart, and Starbird 2018; Marwick and Lewis 2017; Wardle and Derakhshan 2017).

Empirical researchers have increasingly shown that disinformation and its harms need to be examined in the context of media *ecosystems* (Benkler, Faris, and Roberts 2018; Golovchenko et al. 2020; Wardle and Derakhshan 2017). A "systemic" theory of deliberative democracy is therefore appropriate, because it emphasizes the complex interactions between different spaces of communication in a political system (Dryzek 2010; Habermas 1996; Mansbridge 1999; Mansbridge et al. 2012). Through a division of labor among different forums and institutions, a healthy deliberative system realizes *epistemic, ethical*, and *democratic* functions (Mansbridge et al. 2012). The epistemic function promotes the likelihood that opinions and decisions will be informed by facts and logic, the ethical function promotes mutual respect among citizens, and the democratic function promotes inclusion and equal opportunities for participation.

To develop our analysis, we introduce a systemic approach to deliberative democracy and identify new structural vulnerabilities to disinformation in contemporary media systems. Specifically, we argue that social media reduce accountability for speech through anonymity and other means, displace news media and other communication gatekeepers with algorithmic systems, and challenge democratic oversight. Next, we use the paradigmatic cases of disinformation campaigns by Russian agents surrounding the 2016 election to show how they leveraged these structural vulnerabilities to advance *corrosive falsehoods, moral denigration*, and *unjustified inclusion*. We propose that these tactics can contribute to system-level anti-deliberative properties of *epistemic cynicism, techno-affective polarization*, and *pervasive inauthenticity*. Finally, we argue that policies to counter online disinformation should focus on the structural vulnerabilities introduced by social media, and be guided by democratic processes of meta-deliberation.

Deliberative Systems and Social Media Vulnerabilities

Deliberative democracy "is a normative theory of democratic legitimacy based on the idea that those affected by a collective decision have the right, opportunity, and capacity to participate in consequential deliberation about the content of decisions" (Ercan, Hendriks, and Dryzek 2019, 23). From this standpoint, democratic decision making is not just about electing representatives or aggregating preferences but also about the communicative exchanges among citizens that inform their own preferences and their understanding of fellow citizens, and which ultimately ground the legitimacy of public decision making (Habermas 1996; Young 2000).

Deliberative theory has turned to systemic frameworks to pursue more comprehensive and realistic examinations of how deliberation might work in complex democracies. A deliberative system is "a set of distinguishable, differentiated, but to some degree interdependent parts, often with distributed functions and a division of labour, connected in such a way as to form a complex whole" (Mansbridge et al. 2012, 4). In a deliberative system, different forums of communication will play different roles in advancing these normative goods. For instance, Habermas (1996) proposes two general spaces: a peripheral or "wild" general public sphere where informal political discussions take place, and a core where deliberation is focused within decision-making institutions, such as legislatures. Building on this model, Dryzek (2010) emphasizes the role of *transmission* mechanisms, including media engagement and public consultation processes, which are needed to make the *empowered spaces* of discourse (i.e., Habermas' core) responsive and accountable to *public spaces* of discourse (i.e., the wild public sphere). Transmission mechanisms do not simply transfer claims from the wild public sphere but also filter, shape, or contest them (Lyons 2017, 3–4). As Mansbridge (1999, 221) puts it,

> If a deliberative system works well, it filters out and discards the worst ideas available on public matters while it picks up, adopts, and applies the best ideas. If the deliberative system works badly, it distorts facts, portrays ideas in forms that their originators would disown, and encourages citizens to adopt ways of thinking and acting that are good neither for them nor for the larger polity.

To achieve this, transmission mechanisms must help "launder" claims in the wild public sphere "by putting them through an ever more rigorous process of justification" as they move into empowered spaces (Chambers 2017, 270).

A systemic approach to deliberative democracy is particularly well suited to today's hybrid media systems (Chadwick 2017). It foregrounds the complex interconnections between different communicative spaces, rather than focusing on the deliberative quality of individual forums. Furthermore, it can be used to analyze the flows and transformations of political communication as it circulates through media ecosystems, including via social media. Social media platforms introduce novel forms of transmission—such as hyperlinks, hashtags, and memes—and undermine or alter preexisting transmission mechanisms (Lyons 2017). For instance, public officials use social media accounts to make claims about public issues and policies, and do so in ways that may circumvent traditional media gatekeepers. Social media also enable publics to provoke responses from empowered spaces, including through new forms of platform-enabled activism (exemplified by the #metoo and #BlackLivesMatter movements).

Deliberative democrats have long been interested in online deliberation, but few have given sustained attention to disinformation. Several articles focus on epistemic problems posed by social media, including the increased spread of misinformation or "fake news," and the rise of what some call "post-truth politics" (Bimber and Gil de Zúñiga 2020; Chambers 2020; Curato, Hammond, and Min 2019, 138–46; Forestal 2020). However, they do not significantly engage the empirical research on disinformation campaigns, including the important relationships between false and misleading claims, incivility, and unjustified inclusion. Nor do they develop a full normative theory that links vulnerabilities, harms, and policy proposals. Drawing on existing literature on social media, the rest of this section highlights three vulnerabilities that disinformation campaigns exploit in deliberative systems.

Architecture of Engagement and Amplification

Social media platforms disrupted the preexisting institutions and practices that amplified or filtered out claims in media systems. In particular, they have partially displaced journalists as gatekeepers (Graves and Anderson 2020). This development is not entirely negative, because journalistic news selection often excludes diverse perspectives and caters to holders of political or economic power, among other defects. However, social media are not neutral "intermediaries" (Chadwick 2017; Gillespie 2018), and they shape the genres, speed, curation, and dissemination patterns of communication in new and often problematic ways.

One major change is the role of algorithms. For instance, platform algorithms determine the discoverability of content via search engines, set the order of messages in newsfeeds, and provide changing and often micro-targeted lists of "trending topics" to users. Algorithmic curation is opaque: most users do not understand how their individual information feeds are moderated, and independent researchers do not have the data needed to understand dissemination patterns. To serve social media company business models, this "black box moderation" uses "engagement-optimizing algorithms that prioritize enticing content—even if from a source peddling clickbait or political outrage, or promoted by bots, trolls, or . . . agents using false identities" (Kornbluh, Goodman, and Weiner 2020, 8).

There is long-standing concern that social media produce "echo chambers" or "filter bubbles" (Pariser 2011), such that selective exposure to information in like-minded communities increases political polarization and decreases acceptance of shared facts. Evidence suggests that relatively few citizens are confined to these echo chambers (Dubois and Blank 2018; Nelson and Webster 2017), but social media do appear to facilitate the circulation of messages that provoke strong emotions or signal identity affiliations (Vosoughi, Roy, and Aral 2018; Wardle and Derakhshan 2017). For instance, online partisan news articles stoke anger that can encourage users to share material that is polarizing or generates hostility (Hasell and Weeks 2016). It appears that the exposure to and promotion of false claims are mediated by factors, including political partisanship, political interest, media literacy, age, and certain psychological traits (Bail et al. 2020; Dubois and Blank 2018; Grinberg et al. 2019). In the United States, these effects appear to be politically asymmetric, as more right-wing social media users appear to consume and share false news at higher levels (Grinberg et al. 2019; Guess, Nyhan, and Reifler 2018).

Disinformation campaigns exploit these elements of social media design to push disinformation from relatively fringe sites into much broader circulation. In particular, provocative content may generate viral dissemination on social media or may provoke reporting by mainstream news outlets (Phillips 2018; Vargo, Guo, and Amazeen 2018). This strategy is referred to as "the propaganda pipeline" by Benkler, Faris, and Roberts (2018). The social media accounts of Russian-backed operatives often used social media accounts to amplify divisive alt-right messaging that would be picked up by a receptive right-wing media ecosystem in the United States which includes right-wing talk radio outlets and Fox News (Benkler, Faris, and Roberts 2018; Marwick and Lewis 2017).

The transmission of disinformation does not only move into empowered spaces from the wild public sphere. Elites who believe or endorse disinformation remain worrisome, as these false claims may be even more likely to spread through the deliberative system. Notable examples include politicians who advanced the birther movement that questioned whether President Obama was born

in the United States, unsubstantiated claims of voter fraud, and conspiracy theories about billionaire George Soros (Martin 2017; Marwick and Lewis 2017).

Anonymity and Unaccountability

Online communication has introduced new means to disguise or misrepresent the authors and amplifiers of communication. The impact of these developments is complex as anonymity promotes inclusion or exclusion in different contexts (Asenbaum 2018). Anonymity or pseudonymity can enable people to speak their minds without fearing repercussions for expressing views that could face social sanction or government repression. Techniques to disguise one's identity may be important measures to counteract the increased capacities of digital media systems for state and corporate surveillance. Among other benefits for deliberative systems, this can enhance inclusion and expose a "false impression of conformity" of opinion on issues (Moore 2018, 182).

However, the possibility for social media users to disguise their identity can also enable manipulative and unaccountable communication. Disinformation actors regularly misrepresent their identities through the use of "sock puppets" (fake accounts operated by humans) or "political bots" ("algorithms that operate over social media, written to learn from and mimic real people," Woolley and Howard 2016, 4885). As Moore (2018, 182) puts it, the misrepresentation of identity "leaves the listener unable to judge the interests, agendas, and biases of the speaker, and thereby creates opportunities for strategic and deceptive communication." A similar problem applies to fake news organizations, think tanks, and government agencies. Because all sources of content tend to look the same on social media, users often suffer "source blindness" and do not apply the interpretive shortcuts they would otherwise use to assess source credibility (Pearson 2020).

The ability to misrepresent identities on social media is one of many means by which social media can reduce accountability. Most basically, some individuals can push false and insulting messages online, even using a real name, knowing that violations of social norms are unlikely to lead to sanctions in their off-line lives. Disinformation actors can micro-target messages to audiences in ways that avoid detection by more critical publics and they can engage in transnational operations, knowing that it is difficult or impossible for state authorities to enforce their laws outside their jurisdiction.

Inadequate Democratic Oversight

Democracies must protect citizens' opportunities to contribute to public discussion, especially in the wild public sphere; prevent communication from being overwhelmed by coercive, economic, or other forms of dominating power; and support transmission mechanisms that facilitate accountability and responsiveness of empowered sites of discourse. Media regulation is fraught due to threats of capture by the state, political factions, and economic actors, and all democracies regulate mass media in ways that grapple with these tensions. However, social media companies have evaded many forms of accountability and public oversight that apply to other mass media companies.

First, social media companies have long downplayed their role as media companies to avoid attendant regulations (Gillespie 2018). While social media undoubtedly have different patterns of use and editorial control than other types of mass media, they clearly wield social and political power that requires oversight. Second, the influence that social media platforms have on mass communication and political opinion has been opaque, which has limited demands for public justification. Their means of influence—including algorithmic boosting and targeted messaging that is invisible to other users—are hard to detect, particularly because the companies themselves have long avoided transparency (Gillespie 2018; Kornbluh, Goodman, and Weiner 2020). Third, social media platforms structure user interaction at a global level, according to policies and design choices that are primarily made in company headquarters in a single country (predominately the United States). To a large extent, regulation and oversight have been limited to the country where these headquarters are based.

Governments and experts have proposed increased regulation of social media platforms to address disinformation and other forms of "online harms" (United Kingdom Parliament 2019; see also European Commission 2018). As we will argue, a deliberative systems analysis can help identify policies to reduce harms that disinformation poses to democratic debate, while addressing concerns about undemocratic controls on speech and violations of rights to free expression.

Deliberative Systems and Anti-deliberative Communication

A systemic approach to deliberative democracy acknowledges that the quality of political communication will vary greatly in different spaces and at different times. Even the "unrestricted communication" of wild, informal public spheres is valuable because this is where "collective identities and need interpretations can be articulated with fewer compulsions than is the case in procedurally regulated public spheres" (Habermas 1996, 308). Deliberative systems can respond positively to non-deliberative activity, such as raucous protests that prompt public attention

to demands from marginalized voices (Mansbridge et al. 2012, 19). However, it is possible for discourse in public spheres to be too wild, for transmission mechanisms to amplify rather than reduce deception and false claims, for failures of deliberation in empowered sites, and for other problems that undermine the quality of deliberation throughout the system. For instance, extremely "partisan and aggressive" protests can generate "a toxic atmosphere for deliberation and thus is not system enhancing over time" (Mansbridge et al. 2012, 20).

It is therefore necessary to understand when forms of communication may degrade a system. This topic has received relatively little attention from deliberative theorists.[2] A systemic view suggests that communication may harm the functions of a deliberative system even if it is not obviously anti-deliberative at a micro level—that is, "manifestly disrespectful, strategic, and coercive" (Rollo 2017, 592). While intentional falsehoods and insults may be recognized as anti-deliberative by participants or observers of a conversation, other instances of anti-deliberative communication may only be detectable by looking at their effects across forums, later in time, or both.

To explore that possibility, we investigate the potential harms that disinformation campaigns could do to the normative functions of a deliberative system. We pay particular attention to tactics that exploit the vulnerabilities introduced by social media. We focus on disinformation *campaigns*, rather than individual messages, as campaigns are most likely to entail coordinated efforts to disrupt deliberative communication at a system level, including through the mobilization of apparently non-problematic messages to achieve anti-deliberative ends. Empirical research on these campaigns has increasingly focused on how false or deceptive messages are injected into and spread throughout the media systems, but the normative harms associated with these findings remain undertheorized.

Our analysis draws on the influential articulation of deliberative systems theory by Mansbridge et al. (2012), who argue that deliberative systems have *epistemic, ethical*, and *democratic* functions. As we show, disinformation campaigns can harm deliberative systems in ways that are not restricted to undermining the epistemic function by spreading false information. Disinformation campaigns can also undermine the ethical and democratic functions of a deliberative system, such as by decreasing moral respect and by displacing legitimate voices from conversations.

The Harms of Online Disinformation Campaigns

To investigate how disinformation may harm deliberative systems, we analyze the disinformation campaigns mounted by Russian operatives surrounding the 2016 U.S. election. This case has received extensive empirical analysis and has shaped global discussions about online disinformation and its threats to democracy.[3] We do not aim to theorize all of the potential effects of these campaigns, nor do we argue that they exhaust all potential qualities and mechanisms of disinformation. Our analysis is illustrative, showing how disinformation campaigns may use tactics of corrosive falsehoods, unjustified inclusion, and moral denigration, and that these might contribute to broader systemic harms such as epistemic cynicism, pervasive inauthenticity, and techno-affective polarization.

Academic studies and government investigations have concluded that the Russian government pursued a wide-ranging, multi-year disinformation campaign in the United States (DiResta et al. 2018; Howard et al. 2018; Jamieson 2018; U.S. Senate 2019). While there are competing positions on whether the Russian-backed disinformation campaigns were pivotal to the election of Donald Trump (Benkler, Faris, and Roberts 2018; Jamieson 2018), there is broad consensus that Russian government staff and contractors used social media platforms to shape political discourse and social relations (U.S. Senate 2019). The most active social media activities appear to have been undertaken by the Saint Petersburg–based Internet Research Agency (IRA). The IRA used false individual and group identities to post or amplify messages, and it is estimated that this content reached at least 126 million people on Facebook, over 20 million Instagram users, and 1.4 million Twitter users (DiResta et al. 2018), along with other channels. Based on their analysis of data from social media companies turned over to the U.S. Senate Select Committee on Intelligence, DiResta et al. (2018, 4) conclude that Russian operatives conducted a "sweeping and sustained social influence operation consisting of various coordinated disinformation tactics aimed directly at US citizens, designed to exert political influence and exacerbate social divisions in US culture."

This online campaign shares many qualities with decades of information operations by the Russian government, mounted against domestic and foreign populations (Paul and Matthews 2016; U.S. Senate 2019, 11–14). Russian *dezinformatsiya* tactics include the propagation of false or misleading information in the news media, and the creation or funding of domestic front groups to advance Russian aims, with the goals of "widening existing rifts; stoking existing tensions; and destabilizing other states' relations with their publics and one another" (Jack 2017, 9). Perhaps the most notable development is that this and other online disinformation campaigns are "*participatory* in nature. Their messages spread through—and with the help of—online crowds and other information providers" (Starbird, Arif, and Wilson 2019,

emphasis in original). That is, online disinformation campaigns achieve a sense of ubiquity and scale at relatively low cost by exploiting vulnerabilities in the structure of the social media platforms that mediate interactions between citizens.

Corrosive Falsehoods and Epistemic Cynicism

False claims and deceit lie at the heart of disinformation campaigns, and it is unsurprising that attention has focused on what Mansbridge et al. (2012, 11) call the epistemic function of deliberative systems, which "is to produce preferences, opinions, and decisions that are appropriately informed by facts and logic and are the outcome of substantive and meaningful consideration of relevant reasons." Occasional false claims in political discussions are unavoidable and may contribute to deliberation if they prompt responses from other discussants that improve collective and individual understandings. Disinformation campaigns may harm the epistemic function at a systemic level by promoting false claims at a large scale or in empowered spaces. They may also attack processes that help to "launder" the epistemic quality of claims. Decision-making processes, political forums, and everyday talk need to be informed by or accountable to processes with high epistemic demands, such as scientific bodies, judicial systems, or processes of robust citizen debate.[4] Disinformation campaigns often advance *corrosive falsehoods*, which promote misperceptions and undermine sources of higher epistemic quality. These problems are particularly worrisome if citizens become hostile or indifferent to processes with higher epistemic demands, which we call *epistemic cynicism*.

Inconsistent democratic oversight of platforms and other media entities allows disinformation campaigns to leverage algorithmic gatekeeping and opportunities for viral amplification to distribute false claims. The amplification and transmission of misleading claims across sites may undermine a deliberative system's epistemic function because even incidental exposure can increase the probability of belief in false information (Feezell 2018; Pennycook, Cannon, and Rand 2018). These tactics can be seen in Russian campaigns that promote pseudoscientific conspiracies on vaccines and other issues (Broad 2020; Broniatowski et al. 2018; DiResta et al. 2018, 69).

Simply examining the balance of true and false claims may not capture the full range of epistemic harm (Starbird, Arif, and Wilson 2019, 18–19). Disinformation actors not only propagate false and deceptive claims, they often seek to crowd out or devalue contributions from processes or institutions of higher epistemic quality (Hagen et al. 2020; Krafft and Donovan 2020). For instance, Russian disinformation actors routinely spread false news stories and attack professional journalism organizations (DiResta et al. 2018, 66–67; U.S. Senate 2019, 20–21); they also disparage science agencies while amplifying pseudoscientific conspiracies (Broad 2020; Broniatowski et al. 2018). These efforts built on increasing distrust of news media and institutions of expertise in the United States, particularly on the right (Mitchell et al. 2019). Research suggests that false claims are more likely to gain traction in a system where there is antipathy toward intellectuals, expert consensus, and scientific institutions (Stecula, Kuru, and Jamieson 2020).

In addition to attacking the credibility of institutions that claim higher epistemic quality, such as news outlets and think tanks, Russian disinformation campaigns also create faux versions of these institutions, which can promote competing claims on issues (Broad 2020). This tactic has long been used by "merchants of doubt" to sew confusion about climate change or the health risks of tobacco (Oreskes and Conway 2011) but is made easier by the problem of "source blindness" on social media (Pearson 2020). The proliferation of competing claims can contribute to the belief that truth claims, including expert claims, are largely dictated by political commitments. This cultivation of *epistemic cynicism* has been identified as part of the Russian government's "firehose of falsehood" propaganda strategy, which promotes the belief that it is fruitless to seek true accounts of political matters in a media environment replete with conflicting and false claims (Paul and Matthews 2016; U.S. Senate 2019, 16). Online disinformation appears to pose a greater risk than previous attempts at overwhelming citizens with conflicting information due to the possibility of coordinated or algorithmic dissemination of content across platforms (Krafft and Donovan 2020; Wardle and Derakhshan 2017). False and conflicting stories can appear on many different sites and social media pages, creating a situation where individuals conclude, "I don't know what to believe" (Toff and Nielsen 2018, 649–52). Polling has not only found that concerns about false news are pervasive among Americans, but it has prompted a substantial portion (42%) to reduce their information-seeking and simply consume less news (Mitchell et al. 2019).

Moral Denigration and Techno-Affective Polarization

Deliberative systems fulfill their ethical function when they "promote mutual respect among citizens" (Mansbridge et al. 2012, 11). The notion of mutual respect reflects the ethical requirement to treat actors as autonomous agents who can produce their own claims and have those claims taken seriously by others. Mutual respect is also prudential, as it helps to facilitate deliberation among those who disagree. Disinformation campaigns can *morally denigrate*

certain groups or individuals through insults, usually grounded in falsehood; they may also make use of social media platforms to undermine both the inherent and prudential forms of the ethical function at the system level through a process that we refer to as *techno-affective polarization*.

Russian-backed disinformation campaigns used false claims, conspiracy theories, chauvinistic language and visual imagery to stoke moral revulsion toward particular individuals (such as electoral candidates or journalists), political parties, and social groups. For instance, presidential candidate Hillary Clinton was vilified by false stories about her involvement in child sex trafficking and frequent repetition of epithets such as "traitor" (DiResta et al. 2018; Howard et al. 2018). The IRA created a Facebook page that appeared to be curated by nativist Americans and used it to promote content such as an image of a caricatured Mexican head on a boll weevil, which included the text, "About 20 million parasites live in the United States illegally. They exploit Americans and give nothing in return. Isn't it time to get rid of parasites that are destroying our country?" (Neiwert 2018). Such messages can be distributed at large scale both by circumventing gatekeepers and leveraging the tendency for increased engagement and viral spread of provocative content.

The moral denigration of political opponents degrades speech norms and is toxic to mutual respect. Less obviously, the use of false identities makes it possible for fake accounts to misrepresent the views of social groups in a way that denies them the capacity to author their own claims. Fake accounts, bots, and promoted posts can create the appearance of significant citizen support for disrespectful claims that may be premised on false information (Broniatowski et al. 2018; Howard et al. 2018). Such messages appear to be earnest contributions by groups of American citizens, potentially encouraging false inferences about what identifiable social groups believe about other social groups and denying them the capacity to author their own claims (Ahler and Sood 2018). Focusing on IRA accounts that addressed the #BlackLivesMatter movement, Arif, Stewart, and Starbird (2018, 22) observe that they enacted "harsh caricatures of political partisans that may have functioned both to pull like-minded accounts closer and to push accounts from the other 'side' even further away."

The IRA and other disinformation actors have built on significant affective polarization between social groups, particularly political partisans. The IRA attempted to target and inflame more polarized audiences, and more politically polarized social media users indeed appear more likely to have interacted with IRA accounts (e.g., Bail et al. 2020). Studies of partisanship suggest that the emotional intensity of online media can lead to motivated reasoning and opinion polarization (Asker and Dinas 2019; Hasell and Weeks 2016; Tucker et al. 2018, 52–53). We refer to the mutually reinforcing interactions between affective polarization and social media behavior as techno-affective polarization. In addition to decreasing respect between groups, techno-affective polarization may undermine mutual respect's prudential function as the "lubricant of effective deliberation" (Mansbridge et al. 2012, 11), making "good faith discussion and honest engagement with divergent perspectives more difficult" (Strickler 2018, 13). Studies show that disrespectful communication online can trigger emotions that make deliberation more difficult (Gervais 2015).

Unjustified Inclusion and Pervasive Inauthenticity

Communication may be anti-deliberative if it hinders people's ability to participate in political discussions that affect them, whether or not its content appears problematic. The democratic function of deliberation is to include those who will be affected by decisions in the decision-making process on equal terms (Mansbridge et al. 2012, 12). Democratic theorists have long recognized that inclusiveness is not limited to formal entry to discursive forums. People may be internally excluded when they "find that their claims are not taken seriously and [. . .] believe that they are not treated with equal respect" (Young 2000, 55). An alternate problem is *unjustified inclusion*, in which actors who lack normative entitlements to participate may insert themselves in democratic discussions. The discovery of widespread unjustified inclusion by disinformation actors may produce further harm to a deliberative system if it contributes to a belief in *pervasive inauthenticity*, a systemic harm in which democratic discourse is corrupted by the belief that a significant number of interlocutors hold problematic identities, such as fake accounts, foreign agents, and bots.

Foreign interference is an obvious case of unjustified inclusion.[5] Democracies have largely failed to establish norms or laws regarding foreign-based political speech on social media, despite the global reach of these platforms. The Russian IRA exploited this regulatory gap and the misuse of anonymity to facilitate unjustified inclusion. Most IRA messaging appeared to come from American individuals and organizations, using faked identities "to infiltrate political discussion communities on the right and left, including black activist communities, in order to exacerbate social divisions and influence the agenda" (Howard et al. 2018, 39). These kinds of inclusions were mobilized on many issues and in many identity groups to promote social rifts and animosity on issues, including race, religion, immigration, gun regulation, and partisan

affiliation (Bastos and Farkas 2019; DiResta et al. 2018; Hagen et al. 2020; Howard et al. 2018). Tweets from IRA accounts also appeared in news stories as *vox populi* (i.e. tweets supposedly by average citizens), which shows that their reach was not limited to groups on social media (Lukito et al. 2020).

Russian disinformation campaigns to varying degrees displaced the inclusion of legitimate members of the American polity, outcompeting them using bots, multiple fake accounts, and paid-for promotion of content (Arif, Stewart, and Starbird 2018; Dawson and Innes 2019; Hagen et al. 2020; Howard et al. 2018). These forms of unjustified inclusion devalue the contributions of legitimate members of the polity. Revelations of the extensive presence of fake and algorithmic accounts have led to growing skepticism about whether communication online comes from an authentic human being. In fact, "the accusation of an account being a 'Russian bot' [has become] a common dismissive reply online" (Friedberg and Donovan 2019, 3), which can be used to ignore or invalidate a message without engaging its content. These criticisms suggest that the problem is not necessarily widespread inauthenticity but also a widespread *perception* of inauthenticity (Starbird, Arif, and Wilson 2019, 17–18). This blurring of lines between real and fake identities makes it increasingly difficult to tell who is being included in online conversations (Hagen et al. 2020), and thus whether a discussion is informed by the perspectives of people affected by the issue or decision at stake (Tanasoca 2019).

Policy Responses and Meta-deliberation

Public attention to disinformation campaigns in the United States and elsewhere has prompted many policy interventions by governments and social media companies, often explicitly aimed at protecting democracy (Tenove 2020). There are deep disagreements regarding the appropriate responses to democratic threats, particularly those that lead to new regulations on speech and political campaigning. This section examines how the previously identified vulnerabilities might be addressed to minimize harm to deliberative systems.

Improving Democratic Oversight

There is increasing agreement that oversight of social media platforms needs to be democratized. Facebook's Mark Zuckerberg recently declared, "I don't think private companies should make so many decisions alone when they touch on fundamental democratic values" (Zuckerberg 2020). Social media platforms are being called upon to be better "custodians of the internet" (Gillespie 2018), though they face serious legitimacy gaps when they attempt to address false claims or shut down accounts that violate their terms of service. From a deliberative systems perspective, government regulation presents significant risks as well as opportunities. Regulation can promote the development of frameworks to improve the quality and accountability of content moderation by social media companies, but it can also be captured by state imperatives, partisan interests, or market logics. This is most clearly seen by repressive governments' serious constraints on speech through "fake news" laws (Kaye 2018). Regulation is also risky in more democratic countries with political leadership that promotes anti-deliberative policies and norms (Allyn 2020). Here, it must be recognized that U.S. President Donald Trump adopts many disinformation tactics we have identified, such as amplifying false claims, attacking professional journalism and scientific agencies, and promoting the moral denigration of political opponents and social groups (Marwick and Lewis 2017; Ross and Rivers 2018).

These challenges highlight the need for deliberative democratic oversight, or "meta-deliberation," which is "the ability of a [deliberative] system to identify its own shortcomings and further reform itself" (Dryzek 2010, 146). Successful meta-deliberation engages the reflexive capacities of a system's members to recognize problems and identify solutions. Continued experimentation with policy responses to disinformation needs to be explicitly linked to public discussions of the normative goods that policies may advance or undermine. Reforms to minimize the harms of disinformation need not be state-centric. Corporate and civil society actors, as well as broader publics, should play key roles in formulating, enacting, testing, and criticizing policy responses. These complex and sometimes agonistic interactions characterized the development of self-regulatory and co-regulatory regimes for journalism, another area where the state and media corporations are simultaneously stewards and threats to free and democratic deliberation (Pickard 2019).

Processes of meta-deliberation regarding online disinformation have faced several serious obstacles. Most obviously, disinformation by its nature employs falsehood and deception to obscure its existence or aims. Prior to the spectacle of the 2016 U.S. presidential election, there was relatively little public attention to the democratic threat of online disinformation, or to the role of social media in facilitating these campaigns. While we now know more about online disinformation campaigns, our understanding remains hampered by public misdirection and limited transparency by social media platforms. Effective oversight will require increased *transparency* of speech regulation by platforms and governments, and

of potential harms that result from these approaches. The U.S. Senate committee report on Russian-backed disinformation surrounding the 2016 election admitted that, three years later, "the full scope of this activity remains unknown" (U.S. Senate 2019, 4). Moreover, there are many other platforms that are even less accessible to researchers, such as WhatsApp, Telegram, and TikTok. Government regulation or industry codes of conduct could develop requirements for disclosure of content moderation rules and their enforcement.

Democratic oversight is also made more challenging by the transnational nature of social media communication. The European Union (EU) has arguably been the most effective regulator of social media, in part as the strongest multi-national body with significant democratic legitimacy. Global oversight by a democratic body is not currently possible. Instead, different global governance bodies—from formal multilateral organizations like the EU to voluntary organizations like the Global Network Initiative—can adopt approaches in alignment with international human rights frameworks, which include pro-democratic commitments to freedom of expression, rule of law, and non-discrimination (Kaye 2018).

Effective democratic oversight requires meta-deliberation that is inclusive, public-minded, and sincere, so that diverse ideas and claims are effectively represented and tested. Several legislatures initiated inquiries into specific disinformation campaigns or violations of privacy laws via social media platforms, and then expanded their activities to address structural threats to democratic participation and public discourse (House of Commons, Canada 2018; United Kingdom Parliament 2019). At the international level, parliamentarians and civil society groups created the International Grand Committee on Disinformation and "Fake News," which has held consultations with experts and interrogated social media platform representatives.[6]

In addition to increased transparency, international committees, and special inquiries, citizens should play a larger role in informing policy responses. Proposals for social media councils have highlighted the need to enhance public inclusion in the development of codes of conduct for social media platforms (Tworek 2019). The Canadian Commission on Democratic Expression is convening a citizen's assembly to contribute to policy proposals to address harmful speech online.[7] Citizen involvement in public deliberation on these issues is critical because citizens themselves are implicated in disinformation and deliberative harms. Not only do they often participate organically in disinformation campaigns—either being tricked or unconcerned about the veracity of the content that they engage with—but many citizens have developed attitudes that limit their ability to adopt a deliberative stance toward other individuals and institutions. The potential for disinformation campaigns to cultivate epistemic cynicism, techno-affective polarization, and perceptions of pervasive inauthenticity suggests that simply reducing the frequency or spread of anti-deliberative messages is insufficient. Citizens need to develop motivations and capacities to engage in more deliberative political discourse.

Improving the Architecture of Engagement and Amplification

A deliberative system is more vulnerable to disinformation if anti-deliberative communication is preferentially amplified in the wild public sphere, or if it degrades the deliberative quality of empowered spaces of discourse. Policies to address these problems can take two general forms. They can aim to reduce dissemination of and engagement with anti-deliberative communication (*filter*) or promote dissemination of and engagement with more pro-deliberative communication (*curate*). In general, filtering interventions target messages or behaviors that fail to meet a very low bar, including communication that is strongly disrespectful, false, or deceptive. Curation interventions, by contrast, promote communication that is likely to be pro-deliberative generally or in particular contexts, which could include information from high-quality epistemic sources or relevant perspectives that have not been included in public conversation. Filtration and curation policies can of course harm deliberative systems, as they introduce means to control and silence political participation. As a result, there needs to be a mix of "top-down" interventions (including by governments), improved capacities for intermediary organizations (including journalism organizations and social media companies), and "bottom-up" interventions enabling individuals and groups to better develop and implement strategies.

First, countries need to update and enforce laws to address the most extreme forms of false and disrespectful communication on social media. Most countries have laws against false claims about voting processes, hate speech (including through false claims about social groups), violent threats, and defamation (false and harmful claims about individuals or groups), all of which were promoted by foreign and domestic disinformation actors in the 2016 U.S. election. For these regulations to act as filters, states must be able to identify violations, decrease their circulation, and hold to account those who create or disseminate them. However, government regulations on speech raise serious risks to free and democratic expression, requiring that they be imposed in alignment with a human rights framework (Kaye 2018).

Second, attention must be given to intermediary organizations and their capacities to filter and curate, which can amplify or counteract anti-deliberative communication.

Social media platforms might reduce perceived inauthenticity by taking public and transparent steps to reduce the capacity for political actors to buy influence through the use of bots (Tanasoca 2019). The threat of epistemic cynicism may be more effectively addressed through specific fact-checks rather than generalized warnings about misleading content on social media (Vraga, Bode, and Tully 2020). Even if these fact-checks are not effective for citizens who knowingly share disinformation, they might motivate reflection on moral denigration among those who are incidentally exposed (Lu 2019).

New journalistic norms or standards are also needed to avoid unintentionally amplifying corrosive falsehoods, moral denigration, and fake accounts through reporting that attempts to debunk or criticize them (Lukito et al. 2020; Phillips 2018). Media outlets are not the only organizations that significantly shape transmission by filtering and curating communication. Other key institutions—like political parties—could take measures to limit their amplification of disinformation by adopting codes of conduct that prohibit sharing information online without taking steps to confirm its accuracy or authenticity.[8] Media actors, politicians, and influencers who knowingly amplify disinformation are unlikely to adopt these codes or standards, but other reforms should lessen the systemic impact of their activities.

Finally, policies must address the fact that citizens often contribute to the success or failure of disinformation campaigns. They do so through decisions about what to share or like, and through their role in content moderation, such as flagging speech that violates social media policies, editing Wikipedia pages, and moderating Facebook groups or Reddit communities. Social media platforms need to address algorithms and interface designs that amplify and increase engagement with corrosive falsehoods and moral denigration. Platforms could be designed to more clearly promote pro-deliberative norms of communication (Matias 2019), or to create "friction" in user activity, such as delays to facilitate citizen reflection that may reduce the viral spread of false and toxic content (Kornbluh, Goodman, and Weiner 2020, 29–30). For instance, Instagram and Twitter have recently introduced algorithmically generated warnings to users who are about to post comments that are potentially offensive (Statt 2020), in effect encouraging self-filtering. Just as disinformation campaigns seek to leverage the social influence of citizens, platforms should incentivize citizens to act as trustworthy sources of information for each other (Bode and Vraga 2018).

Addressing Anonymity and Unaccountability

Some policymakers have suggested banning anonymity online to decrease false and disrespectful claims. For instance, the President of the German Parliament has proposed legal requirements for the use of real names on social media platforms (Leistner and Hurst 2020). However, such policies would remove the important benefits that can come from anonymity, including the freedom to criticize authorities or dominant opinions, without necessarily reducing insults or threats (Moore 2018).

A better deliberative system maximizes both the benefits that anonymity may sometimes offer, *and* limits potential harms from anonymous users, including increased moral denigration and unjustified inclusion. To this end, fake identities or bots could be labeled rather than banned, so people can respond accordingly. For instance, California's 2018 "blade runner" law makes it illegal to use bots deceptively to mislead people regarding commercial transactions or elections. It remains to be seen whether these labeling policies are enforceable and whether they reduce the capacity for disinformation actors to use false identities to exacerbate divisions within communities. Furthermore, policies to reduce the purchase and use of coordinated networks of fake accounts (Dawson and Innes 2019), a common disinformation technique, can address issues of unjustified inclusion without requiring the use of real identities. These problems might also be addressed by platform designs that encourage the kinds of diverse membership that can enable critical reflection and self-correction (Forestal 2020).

In addition, platforms should increase accountability for false, disrespectful, and otherwise anti-deliberative claims. This need not require exposing voices to government surveillance or punishment. Social media companies can balance different normative goods by enabling anonymity while addressing moral denigration through content moderation (see Fredheim, Moore, and Naughton 2015). In addition, pseudonymous accounts can have their influence determined by the quality of their contributions. This can occur at the level of individual users or at the level of media outlets (e.g., NewsGuard and other services that provide users and platform companies with evidence-based evaluations of news source credibility).

From a deliberative system perspective, these goals may be best achieved by the existence of online spaces that maximize the freedoms that come with anonymity (but which sometimes enable problematic behavior), other spaces that emphasize accountability and the authentic identity of communicators (but without the freedoms that may come from anonymity), and linkage mechanisms that facilitate productive exchanges between spheres with alternate deliberative advantages.

Conclusion

Disinformation may pose a variety of potential harms to democracy. Adopting a deliberative systems approach,

we have argued that disinformation harms the epistemic function when false claims are weaponized to generate epistemic cynicism, the ethical function when moral denigration exacerbates techno-affective polarization, and the democratic function when fears of unjustified inclusion lead to perceptions of pervasive inauthenticity. The online tactics pursued in such campaigns are not wholly new, but they are able to exploit new vulnerabilities in media systems to pursue their ends. The deliberative systems approach clarifies that some harms are systemic and cannot necessarily be easily observed by looking at individual instances of communication. Furthermore, disinformation actors are not solely responsible. Ordinary citizens, journalists, elected officials, and other individuals frequently transmit claims and behavioral norms through the deliberative system in ways that cause harm.

Our normative analysis complements empirical research on disinformation campaigns. While we have theorized the forms of harm that might plausibly be caused by the tactics used in disinformation campaigns, we have not proven that disinformation degraded the U.S. deliberative system in these ways. Nor have we measured the impact that these systemic harms might have on deliberative quality in particular sites or across the system. Instead, our analysis suggests questions for further research and clarifies their normative stakes. One particularly important question is how a deliberative system may be affected by the *asymmetrical* spread and uptake of disinformation. For instance, strongly right-wing publics in the United States are exposed to false and deceptive claims at particularly high rates. Ideally, transmission mechanisms and political leaders would be responsive to the best claims and underlying motivations expressed in spaces with more anti-deliberative communication, without being overwhelmed by them. In practice, what elements of a deliberative system may manage or exacerbate such a dynamic?

Efforts to address disinformation campaigns and the vulnerabilities they leverage are underway, including policies adopted by social media platforms and by governments. Theoretical and empirical work that builds on a deliberative systems approach can help assess such developments and identify both positive and negative effects of attempts to protect democratic communication.

Authors' Note

Spencer McKay is now SSHRC Postdoctoral Fellow, affiliated with New York University.

Acknowledgments

The authors would like to thank Simone Chambers, Jennifer Forestal, John Gastil, Mark E. Warren, and three anonymous referees for their comments on earlier drafts of this paper.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the Social Sciences and Humanities Research Council of Canada via an insight development grant (430-2018-1059) for Chris Tenove, and Spencer McKay's doctoral (752-2015-2502) and postdoctoral fellowships (756-2019-0259).

ORCID iDs

Spencer McKay https://orcid.org/0000-0001-5650-4636

Chris Tenove https://orcid.org/0000-0002-4244-0878

Notes

1. For instance, the European Commission (EC) defines disinformation as "verifiably false or misleading information that is created, presented and disseminated for economic gain or to intentionally deceive the public, and may cause public harm" (EC 2018, 2–3; for a similar definition, see United Kingdom Parliament 2019, 22).
2. Although see Owen and Smith (2015) on the broader topic of "deliberative wrongs."
3. The Russian campaign serves as a "paradigmatic case" (Flyvbjerg 2006), given its prominence in global political debates and empirical research on disinformation campaigns. While it was designed to address the very distinct U.S. political and media systems, Russian disinformation campaigns have pursued similar strategies in European countries (Dawson and Innes 2019).
4. Deliberative theory does not simply equate expertise with epistemic quality and allocates citizens a notable critical role (see Chambers 2017; Moore 2017).
5. Foreign contributions to democratic deliberation may be normatively justifiable, such as by reference to the all-affected interests principle, which suggests that non-citizens may deserve influence (see, for instance, Fung 2013; Young 2000).
6. See https://www.cigionline.org/igc.
7. See https://ppforum.ca/project/canadian-commission-on-democratic-expression/.
8. See, for instance, https://electionpledge.eu.

References

Ahler, Douglas J., and Gaurav Sood. 2018. "The Parties in Our Heads: Misperceptions about Party Composition and Their Consequences." *The Journal of Politics* 80 (3): 964–81.

Allcott, Hunt, and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives* 31 (2): 211–36.

Allyn, Bobby. 2020. "Stung by Twitter, Trump Signs Executive Order to Weaken Social Media Companies." *npr*, May 28. https://www.npr.org/2020/05/28/863932758/stung-by-twitter-trump-signs-executive-order-to-weaken-social-media-companies.

Arif, Ahmer, Leo Graiden Stewart, and Kate Starbird. 2018. "Acting the Part: Examining Information Operations within #BlackLivesMatter Discourse." *Proceedings of the ACM on Human-Computer Interactions* 2:20.

Asenbaum, Hans. 2018. "Anonymity and Democracy: Absence as Presence in the Public Sphere." *American Political Science Review* 112 (3): 459–72.

Asker, David, and Elias Dinas. 2019. "Thinking Fast and Furious: Emotional Intensity and Opinion Polarization in Online Media." *Public Opinion Quarterly* 83 (3): 487–509.

Bail, Christopher A., Brian Guay, Emily Maloney, Aidan Combs, D. Sunshine Hillygus, Friedolin Merhout, Deen Freelon, and Alexander Volfovsky. 2020. "Assessing the Russian Internet Research Agency's Impact on the Political Attitudes and Behaviors of American Twitter Users in Late 2017." *Proceedings of the National Academy of Sciences of the United States of America* 117:243–50. https://www.pnas.org/content/early/2019/11/20/1906420116.

Bastos, Marco, and Johan Farkas. 2019. "'Donald Trump Is My President!' The Internet Research Agency Propaganda Machine." *Social Media + Society* 5 (3): 1–13.

Benkler, Yochai, Robert Faris, and Hal Roberts. 2018. *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. Oxford: Oxford University Press.

Bimber, Bruce, and Homero Gil de Zúñiga. 2020. "The Unedited Public Sphere." *New Media & Society* 22 (4): 700–15.

Bode, Leticia, and Emily K. Vraga. 2018. "See Something, Say Something: Correction of Global Health Misinformation on Social Media." *Health Communication* 33 (9): 1131–40.

Broad, William J. 2020. "Putin's Long War against American Science." *The New York Times*, April 13. https://www.nytimes.com/2020/04/13/science/putin-russia-disinformation-health-coronavirus.html.

Broniatowski, David A., Amelia M. Jamison, SiHua Qi, Lulwah AlKulaib, Tao Chen, Adrian Benton, Sandra C. Quinn, and Mark Dredze. 2018. "Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate." *American Journal of Public Health* 108 (10): 1378–84.

Chadwick, Andrew. 2017. *The Hybrid Media System: Politics and Power*. Oxford: Oxford University Press.

Chambers, Simone. 2017. "Balancing Epistemic Quality and Equal Participation in a System Approach to Deliberative Democracy." *Social Epistemology* 31 (3): 266–76.

Chambers, Simone. 2020. "Truth, Deliberative Democracy, and the Virtues of Accuracy: Is Fake News Destroying the Public Sphere?" *Political Studies*. Published electronically April 2. doi:10.1177/0032321719890811.

Curato, Nicole, Nicole Hammond, and John Min. 2019. *Power in Deliberative Democracy: Norms, Forums, Systems*. New York: Palgrave Macmillan.

Dawson, Andrew, and Martin Innes. 2019. "How Russia's Internet Research Agency Built Its Disinformation Campaign." *The Political Quarterly* 90 (2): 245–56.

DiResta, Renee, Kris Shaffer, Becky Ruppel, David Sullivan, Robert Matney, Ryan Fox, Jonathan Albright, and Ben Johnson. 2018. "The Disinformation Report: The Tactics & Tropes of the Internet Research Agency." New Knowledge, Austin. https://disinformationreport.blob.core.windows.net/disinformation-report/NewKnowledge-Disinformation-Report-Whitepaper-121718.pdf.

Dryzek, John S. 2010. *Foundations and Frontiers of Deliberative Governance*. Oxford: Oxford University Press.

Dubois, Elizabeth, and Grant Blank. 2018. "The Echo Chamber Is Overstated: The Moderating Effect of Political Interest and Diverse Media." *Information, Communication & Society* 21 (5): 729–45.

Ercan, Selen A., Carolyn M. Hendriks, and John S. Dryzek. 2019. "Public Deliberation in an Era of Communicative Plenty." *Policy & Politics* 47 (1): 19–36.

European Commission. 2018. "Communication—Tackling Online Disinformation: A European Approach." European Commission, Brussels. https://ec.europa.eu/digital-single-market/en/news/communication-tackling-online-disinformation-european-approach.

Fallis, Don. 2015. "What Is Disinformation?" *Library Trends* 63 (3): 401–26.

Feezell, Jessica T. 2018. "Agenda Setting through Social Media: The Importance of Incidental News Exposure and Social Filtering in the Digital Era." *Political Research Quarterly* 71 (2): 482–94.

Flyvbjerg, Bent. 2006. "Five Misunderstandings about Case-Study Research." *Qualitative Inquiry* 12:219–45.

Forestal, Jennifer. 2020. "Beyond Gatekeeping: Propaganda, Democracy, and the Organization of Digital Publics." *Journal of Politics*. Published electronically August. doi:10.1086/709300.

Fredheim, Rolf, Alfred Moore, and John Naughton. 2015. "Anonymity and Online Commenting: The Broken Windows Effect and the End of Drive-By Commenting." Paper presented at the Proceedings of the ACM Web Science Conference, June 28. https://dl.acm.org/citation.cfm?doid=2786451.2786459.

Freelon, Deen, and Chris Wells. 2020. "Disinformation as Political Communication." *Political Communication* 37 (2): 145–56.

Friedberg, Brian, and Joan Donovan. 2019. "On the Internet, Nobody Knows You're a Bot: Pseudoanonymous Influence Operations and Networked Social Movements." *Journal of Design and Science* 6:1–21.

Fung, Archon. 2013. "The Principle of Affected Interests: An Interpretation and Defense." In *Representation: Elections and Beyond*, edited by. J. H. Nagel and R. M. Smith, 236–68. Philadelphia: University of Pennsylvania Press.

Gervais, Bryan T. 2015. "Incivility Online: Affective and Behavioral Reactions to Uncivil Political Posts in a Web-Based Experiment." *Journal of Information Technology & Politics* 12 (2): 167–85.

Gillespie, Tarleton. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media*. New Haven: Yale University Press.

Golovchenko, Yevgeniy, Cody Buntain, Gregory Eady, Megan A. Brown, and Joshua A. Tucker. 2020. "Cross-Platform State Propaganda: Russian Trolls on Twitter and YouTube during the 2016 U.S. Presidential Election." *The International Journal of Press/Politics*. Published electronically April 14. doi:10.1177/1940161220912682.

Graves, Lucas, and C. W. Anderson. 2020. "Discipline and Promote: Building Infrastructure and Managing Algorithms in a 'Structured Journalism' Project by Professional Fact-Checking Groups." *New Media & Society* 22 (2): 342–60.

Grinberg, Nir, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. "Fake News on Twitter during the 2016 U.S. Presidential Election." *Science* 363 (6425): 374–78.

Guess, Andrew, Brendan Nyhan, and Jason Reifler. 2018. "Selective Exposure to Misinformation: Evidence from the Consumption of Fake News during the 2016 U.S. Presidential Campaign." European Research Council. http://www.ask-force.org/web/Fundamentalists/Guess-Selective-Exposure-to-Misinformation-Evidence-Presidential-Campaign-2018.pdf.

Habermas, Jürgen. 1996. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. Cambridge: MIT Press.

Hagen, Loni, Stephen Neely, Thomas E. Keller, Ryan Scharf, and Fatima Espinoza Vasquez. 2020. "Rise of the Machines? Examining the Influence of Social Bots on a Political Discussion Network." *Social Science Computer Review*. Published electronically March 17. doi:10.1177/0894439320908190.

Hasell, A., and Brian E. Weeks. 2016. "Partisan Provocation: The Role of Partisan News Use and Emotional Responses in Political Information Sharing in Social Media." *Human Communication Research* 42 (4): 641–61.

House of Commons, Canada. 2018. "Democracy under Threat: Risks and Solutions in the Era of Disinformation and Data Monopoly." Standing Committee on Access to Information, Privacy and Ethics. http://www.ourcommons.ca/DocumentViewer/en/42-1/ETHI/report-17.

Howard, Philip N., Bharath Ganesh, Dimitra Liotsiou, John Kelly, and Camille François. 2018. "The IRA and Political Polarization in the United States." Project on Computational Propaganda, University of Oxford, Oxford. https://comprop.oii.ox.ac.uk/research/ira-political-polarization/.

Jack, Caroline. 2017. "Lexicon of Lies: Terms for Problematic Information." Data & Society Research Institute, New York. https://datasociety.net/output/lexicon-of-lies.

Jamieson, Kathleen Hall. 2018. *Cyberwar: How Russian Hackers and Trolls Helped Elect a President*. Oxford: Oxford University Press.

Kalla, Joshua L., and David E. Broockman. 2018. "The Minimal Persuasive Effects of Campaign Contact in General Elections: Evidence from 49 Field Experiments." *American Political Science Review* 12 (1): 148–66.

Kaye, David. 2018. *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*. A/HRC/38/35. Geneva: United Nations Human Rights Council.

Kornbluh, Karen, Ellen Goodman, and Eli Weiner. 2020. "Safeguarding Democracy against Disinformation." The German Marshall Fund of the United States, Washington, DC. http://www.gmfus.org/publications/safeguarding-democracy-against-disinformation.

Krafft, P. M., and Joan Donovan. 2020. "Disinformation by Design: The Use of Evidence Collages and Platform Filtering in a Media Manipulation Campaign." *Political Communication* 37 (2): 194–214.

Leistner, Alexandra, and Luke Hurst. 2020. "Should Social Media Users Have to Use Their Real Names?" *Euronews*, January 17. https://www.euronews.com/2020/01/17/should-social-media-users-have-to-use-their-real-names.

Lu, Yanqin. 2019. "Incidental Exposure to Political Disagreement on Facebook and Corrective Participation: Unraveling the Effects of Emotional Responses and Issue Relevance." *International Journal of Communication* 13:874–96.

Lukito, Josephine, Jiyoun Suk, Yini Zhang, Larissa Doroshenko, Sang Jung Kim, Min-Hsin Su, Yiping Xia, Deen Freelon, and Chris Wells. 2020. "The Wolves in Sheep's Clothing: How Russia's Internet Research Agency Tweets Appeared in U.S. News as Vox Populi." *The International Journal of Press/Politics* 25 (2): 196–216.

Lyons, Benjamin. 2017. "From Code to Discourse: Social Media and Linkage Mechanisms in Deliberative Systems." *Journal of Public Deliberation* 13 (1): 4. https://www.publicdeliberation.net/jpd/vol13/iss1/art4.

Mansbridge, Jane. 1999. "Everyday Talk in the Deliberative System." In *Deliberative Politics: Essays on Democracy and Disagreement*, edited by Stephen Macedo, 1–211. Oxford: Oxford University Press.

Mansbridge, Jane, James Bohman, Simone Chambers, Thomas Christiano, Archon Fung, John Parkinson, Dennis F. Thompson, and Mark E. Warren. 2012. "A Systemic Approach to Deliberative Democracy." In *Deliberative Systems: Deliberative Democracy at the Large Scale*, edited by John Parkinson and Jane Mansbridge, 1–26. Cambridge: Cambridge University Press.

Martin, Jonathan. 2017. "Dubious Vote-Fraud Claim Gets the Trump Seal of Approval." *The New York Times*, January 27. https://www.nytimes.com/2017/01/27/us/politics/donald-trump-voter-fraud.html.

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

Matias, J. Nathan. 2019. "Preventing Harassment and Increasing Group Participation through Social Norms in 2,190 Online Science Discussions." *Proceedings of the National Academy of Sciences of the United States of America* 116 (20): 9785–89.

Mitchell, Amy, Jeffrey Gottfried, Galen Stocking, Mason Walker, and Sophia Fedeli. 2019. *Many Americans Say Made-up News Is a Critical Problem that Needs to Be Fixed*. Washington, DC: Pew Research Center. https://www.journalism.org/wp-content/uploads/sites/8/2019/06/PJ_2019.06.05_Misinformation_FINAL-1.pdf.

Moore, Alfred. 2017. *Critical Elitism: Deliberation, Democracy and the Problem of Expertise*. Cambridge: Cambridge University Press.

Moore, Alfred. 2018. "Anonymity, Pseudonymity, and Deliberation: Why not Everything Should Be Connected." *Journal of Political Philosophy* 26 (2): 169–92.

Morell, Michael, and Suzanne Kelly. 2016. "Fmr. CIA Acting Dir. Michael Morell: 'This Is the Political Equivalent of 9/11.'" *The Cipher Brief*, December 11. https://www.thecipherbrief.com/fmr-cia-acting-dir-michael-morell-political-equivalent-911-1091.

Neiwert, David. 2018. "Racially Incendiary Russian Ads Apparently Had Their Intended Effect." Southern Poverty Law Center. https://www.splcenter.org/hatewatch/2018/06/06/racially-incendiary-russian-ads-apparently-had-their-intended-effect.

Nelson, Jacob L., and James G. Webster. 2017. "The Myth of Partisan Selective Exposure: A Portrait of the Online Political News Audience." *Social Media + Society* 3 (3): 1–13.

Oreskes, Naomi, and Erik M. Conway. 2011. *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*. New York: Bloomsbury.

Owen, David, and Graham Smith. 2015. "Survey Article: Deliberation, Democracy, and the Systemic Turn: Survey Article: Deliberation & the Systemic Turn." *Journal of Political Philosophy* 23(2): 213–34.

Pariser, Eli. 2011. *The Filter Bubble: What the Internet Is Hiding from You*. New York: Penguin Books Limited.

Paul, Christopher, and Miriam Matthews. 2016. "The Russian 'Firehose of Falsehood' Propaganda Model: Why It Might Work and Options to Counter It." RAND Corporation, Santa Monica. https://www.rand.org/pubs/perspectives/PE198.html.

Pearson, George. 2020. "Sources on Social Media: Information Context Collapse and Volume of Content as Predictors of Source Blindness." *New Media & Society*. Published electronically March. doi:10.1177/1461444820910505.

Pennycook, Gordon, Tyrone D. Cannon, and David G. Rand. 2018. "Prior Exposure Increases Perceived Accuracy of Fake News." *Journal of Experimental Psychology: General* 147 (12): 1865–80.

Phillips, Whitney. 2018. "The Oxygen of Amplification." Data & Society Research Institute, New York. https://datasociety.net/output/oxygen-of-amplification/.

Pickard, Victor. 2019. *Democracy without Journalism? Confronting the Misinformation Society*. Oxford: Oxford University Press.

Rollo, Toby. 2017. "Everyday Deeds: Enactive Protest, Exit, and Silence in Deliberative Systems." *Political Theory* 45 (5): 587–609.

Ross, Andrew S., and Damian J. Rivers. 2018. "Discursive Deflection: Accusation of 'Fake News' and the Spread of Mis- and Disinformation in the Tweets of President Trump." *Social Media + Society* 4 (2): 1–12.

Starbird, Kate, Ahmer Arif, and Tom Wilson. 2019. "Disinformation as Collaborative Work: Surfacing the Participatory Nature of Strategic Information Operations." Proceedings of the ACM on Human-Computer Interaction. https://faculty.washington.edu/kstarbi/StarbirdArifWilson_DisinformationasCollaborativeWork-CameraReady-Preprint.pdf.

Statt, Nick. 2020. "Twitter Tests a Warning Message that Tells Users to Rethink Offensive Replies." *The Verge*, May 5. https://www.theverge.com/2020/5/5/21248201/twitter-reply-warning-harmful-language-revise-tweet-moderation.

Stecula, Dominik Andrzej, Ozan Kuru, and Kathleen Hall Jamieson. 2020. "How Trust in Experts and Media Use Affect Acceptance of Common Anti-vaccination Claims." *The Harvard Kennedy School Misinformation Review* 1 (1): 1–11.

Strickler, Ryan. 2018. "Deliberate with the Enemy? Polarization, Social Identity, and Attitudes toward Disagreement." *Political Research Quarterly* 71 (1): 3–18.

Tanasoca, Ana. 2019. "Against Bot Democracy: The Dangers of Epistemic Double-Counting." *Perspectives on Politics* 17 (4): 988–1002.

Tenove, Chris. 2020. "Protecting Democracy from Disinformation: Normative Threats and Policy Responses" *The International Journal of Press/Politics*. Published electronically May 25. doi:10.1177/1940161220918740.

Toff, Benjamin, and Rasmus Kleis Nielsen. 2018. "'I Just Google It': Folk Theories of Distributed Discovery." *Journal of Communication* 68 (3): 636–57.

Tucker, Joshua, Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan. 2018. *Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature*. Menlo Park: The William and Flora Hewlett Foundation. https://www.hewlett.org/wp-content/uploads/2018/03/Social-Media-Political-Polarization-and-Political-Disinformation-Literature-Review.pdf.

Tworek, Heidi. 2019. *Social Media Councils*. Waterloo: Centre for International Governance Innovation. https://www.cigionline.org/articles/social-media-councils.

United Kingdom Parliament. 2019. "Online Harms White Paper." Secretary of State for Digital, Culture, Media & Sport and the Secretary of State for the Home Department. https://www.gov.uk/government/consultations/online-harms-white-paper.

U.S. Senate. 2019. *Report on Russian Active Measures Campaigns and Interference in the 2016 U.S. Election, Volume 2: Russia's Use of Social Media with Additional Views*. Washington, DC: United States Senate Select Committee on Intelligence. https://www.intelligence.senate.gov/sites/default/files/documents/Report_Volume2.pdf.

Vargo, Chris J., Lei Guo, and Michelle A. Amazeen. 2018. "The Agenda-Setting Power of Fake News: A Big Data Analysis

of the Online Media Landscape from 2014 to 2016." *New Media & Society* 20 (5): 2028–49.

Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. "The Spread of True and False News Online." *Science* 359 (6380): 1146–51.

Vraga, Emily K., Leticia Bode, and Melissa Tully. 2020. "Creating News Literacy Messages to Enhance Expert Corrections of Misinformation on Twitter." *Communication Research*. Published electronically January 30. doi:10.1177/0093650219898094.

Wardle, Claire, and Hossein Derakhshan. 2017. "Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making." Report DGI(2017)09. Council of Europe, Strasbourg. https://firstdraftnews.com/wp-content/uploads/2017/10/Information_Disorder_FirstDraft-CoE_2018.pdf?x56713.

Woolley, Samuel C., and Philip N. Howard. 2016. "Automation, Algorithms, and Politics Political Communication, Computational Propaganda, and Autonomous Agents — Introduction." *International Journal of Communication* 10: 9.

Young, Iris Marion. 2000. *Inclusion and Democracy*. Oxford: Oxford University Press.

Zuckerberg, Mark. 2020. "Big Tech Needs More Regulation." *Facebook Newsroom*, February 18. https://about.fb.com/news/2020/02/big-tech-needs-more-regulation/.