The Internet's Challenge to Democracy

Nathaniel Persily

From its earliest days, the internet has been viewed as an inherently democratic means of communication. As John Perry Barlow described it in his 1996 *Declaration of the Independence of Cyberspace*, "We are creating a world that all may enter without privilege or prejudice accorded by race, economic power, military force, or station of birth. We are creating a world where anyone, anywhere may express his or her beliefs, no matter how singular, without fear of being coerced into silence or conformity."[1] The inequalities and restrictions of the offline world, under this view, would disappear in cyberspace. The technology itself would be liberating and egalitarian, and in turn, remove the distortions in the marketplace of ideas caused by governments or restrictive legacy communication media.

The utopianism that dominated the first few decades of the internet's growth has now taken a decidedly pessimistic turn. Especially in the two years since the 2016 U.S. Presidential election, the internet and social media have emerged as scapegoats for rising populism, political polarization, hate speech, incitement, disinformation campaigns, and foreign interference with elections. The dominant internet and social media firms (all based in the United States) have also been the target of regulatory action and generalized suspicion under both anticompetition law and various legal regimes governing privacy. Further betraying the egalitarian and libertarian vision of its founding generation, the internet and the services provided by the major platforms are now being coopted by governments to target opponents and intimidate online speakers. Previously heralded as a boon to democracy, the internet now is being blamed for its demise.

Of course, in elections as in all areas of social life touched by emerging technologies, the internet is a medium that can be used for good or ill. It still can give voice to the voiceless, serve as an indispensable tool for political organizing and community building, and provide a powerful, near-costless outlet for protest, fundraising, and campaigning against entrenched incumbents. Each revolution in telecommunications (from the printing press to the telephone to radio to television) brings with it anxiety that the basest human impulses and behaviors will be magnified with the aid of new technologies. The internet is no different.

With that said, the internet and the technologies accompanying it pose particular dangers for democracy. While recognizing this general point, it is important not to fall into the familiar trap of concluding that the prevalence of "dangerous speech" is itself the problem. "Fake News" is as old as news, and hate speech is as old as speech. The challenge for anybody analyzing the

[1] John Perry Barlow, A Declaration of the Independence of Cyberspace, Feb. 8, 1996.

particular stresses that the new technologies and platforms pose for democracies is to isolate the unique features of this new form of communication that threaten the core components of elections, campaigns, and democratic decisionmaking.

What follows here is a description of those unique features of the new technologies that place stress on democracy. Those features would include: the speed of online communication, the importance of virality as the currency for gaining an audience, the privileging of anonymity, the emergence of filter bubbles and echo chambers, the solidification of monopolies in the markets of social media and internet search, and the loss of sovereignty for democracies hoping to wall off their elections from foreign influence by nation states, firms, or stateless actors. These are the characteristics of online communication that then facilitate the well-known problems of disinformation, hate speech, incitement and the like. The reform options that follow the description of these features target one or more of them to try to mitigate the newfound dangers to democracy that the internet presents.

I. Features of the Digital Communication Ecosystem that Place Stress on Democracy

A. Velocity

The speed with which information travels is often a defining feature of any transformational communication technology. The transitions from mail to telegraph to telephone to radio and then to television were each characterized by an increase in the velocity of information transfer. An individual could communicate to more people in a shorter period of time with each additional innovation. The internet represented a leap forward of greater magnitude, in that any individual has the capacity to communicate, instantaneously, to the entire world – or at least, to anyone who is online and willing to listen.

Of course, this does not mean that anyone will listen to what a given individual has to say online. The average tweet, blog entry or Facebook post will have a very limited following. The internet merely enhances that capacity to speak instantaneously to a broad audience – it cannot force them to hear. (Although, perhaps even in this respect, the pervasiveness of the internet on mobile and other devices may allow online communication to be more intrusive than other media, and under certain conditions, allow for widespread dissemination of communication even to unwilling recipients.)

The increased speed of online communication, like communication in general, is a great virtue of the internet. It allows more people to be more informed more quickly. Whether one seeks to acquire information helpful to one's health, finances, or lifestyle, one can now receive it with a few clicks. Similarly, notification of any event – whether geopolitical or personal in significance – can occur more quickly now that anyone can post online a message, picture, audio

or video of an event in real time. Few people today would trade the status quo for a world in which they must wait to know the information they consider important and valuable to their lives.

The downside of the increased speed of internet information transfer, of course, is that any "dangerous" communication can reach a broader audience more quickly. Whether what makes a communication dangerous is its falsity, bias, hatefulness, potential for injury to a person's reputation or privacy, or inherent danger in the information itself (e.g., how to build a nuclear bomb) – that danger is exacerbated by the speed with which that communication is disseminated to the public online.

In particular, because mass communication online can be largely unmediated, the obstacles that exist in the offline world do not impede or slow down the dissemination of falsehoods propagated on the internet. To the extent that widespread offline dissemination of falsehoods relied on their adoption and transmission by, for example, a major media network or newspaper, no such elite intervention or permission is required for the "broadcast" of lies online. All that is required is a willing speaker and an audience paying attention.

As bad as the rapid dissemination of falsehoods may be, it is compounded by the inability to timely correct or combat disinformation. An online lie, once disseminated, can be permanently available on the internet. Any correction or competing information necessarily is playing catchup. To be sure, that was always true with false stories in major newspapers, for example, as few might read a later correction to a damaging article. But because of the virality of internet communication (discussed next), an online lie extends well beyond the site in which it was initially featured. Any correction to the lie necessarily competes at a disadvantage in the online marketplace of ideas: it is both late to the game and in many circumstances cannot follow the same viral pathway of the lie itself. A correction is unlikely to reach either the same audience or one of similar size.

The speed of information transfer poses particular challenges for democracy, because elections occur at a certain period in time. In the United States, we know years in advance when we will elect our national, state and local leaders. Even in those countries that do not have regularly scheduled elections, candidates know months in advance when they will stand for the voters. The predictability and finality of elections facilitates strategies tailored to short term, last minute influence. As harmful as "fake news," hate speech, doxing or internet rumors may be, in general, they pose a more serious challenge when weaponized to have their greatest short-term impact right before an election.

"October surprises" are not new to the internet age, of course. Campaign professionals have long worried about last minute news and events that may affect vote choices. However, research on internet communication has found that false news, in fact, travels faster than true news online.[2] Authors of a recent study published in *Science* find that "rumor cascades" form on social media,

[2] Soroush Vosoughi, Deb Roy, & Sinan Aral, The spread of true and false news online, 359 Science 1146, Mar. 9, 2018.

accelerating false claims at about ten times the speed as true stories. In particular, the authors find that political falsehoods travel the fastest: "false political news traveled deeper and more broadly, reached more people, and was more viral than any other category of false information."

B. Virality

The rapid and widespread propagation of lies online represents one manifestation of the larger phenomenon of unmediated communication magnified through peer-to-peer sharing over social media. As that type of communication pathway becomes more dominant, it privileges a certain type of communication over others. In particular, it places a premium on virality as the quality of communication most necessary to determine audience reach.

With virality as the coin of the political communication realm, certain strategies then follow when political and media actors wish to get their message out (and/or to attract the eyeballs that necessarily lead to higher advertising revenue). Those strategies seek to increase the probability that an individual will read the communication and forward it. It may seem simple, but a lot follows from this property that is distinctive to the internet communication environment. To be sure, "word of mouth" has always been an important quality in gauging the popularity of a product, news story, or advertising campaign. But the internet enables all of us to become re-transmitters of communication in ways distinctly different than the offline world.

Virality is, in part, an indicator (or correlate) of popularity. A communication that is forwarded widely is one that a large number of people find interesting and worth sharing or reading (leaving aside, for the moment, the important issue of virality by way of automated accounts or "bots.") Therefore, in order to "go viral" a communication will often appeal to those instincts that would lead one to forward the message or news story to others.[3] We know, for example, that articles and videos that arouse, either by provoking anger or stoking other emotions, are more likely to be shared by an audience.[4] It should also be of no surprise that virality privileges spectacles, novelty, and outrage, as viewers seek to spread content that takes them by surprise.

---

[3] Christin Scholz, Elisa C. Baek, Matthew Brook O'Donnell, Hyun Suk Kim, Joseph N. Cappella, and Emily B. Falk, *A neural model of valuation and information virality*, 114 PNAS 1 (Mar. 14, 2017), http://www.pnas.org/content/pnas/114/11/2881.full.pdf.

[4] Jonah Berger, Katherine L. Milkman, *What Makes Online Content Viral?*, 49 Journal of Marketing Research 2 (Apr. 2012), http://journals.ama.org/doi/abs/10.1509/jmr.10.0353?code=amma-site; Rosanna E. Guadagno, Daniel M. Rempala, Shannon Murphy and Bradley M. Okdie, *What makes a video go viral? An analysis of emotional contagion and internet memes*, 29 COMPUTERS IN HUMAN BEHAVIOR 6 (Nov. 2013), https://www.sciencedirect.com/science/article/pii/S0747563213001192; Liz Rees-Jones, Katherine L. Milkman, and Jonah Berger, *The Secret to Online Success: What Makes Content Go Viral*, SCIENTIFIC AMERICAN (Apr. 14, 2015), https://www.scientificamerican.com/article/the-secret-to-online-success-what-makes-content-go-viral/.

How does this relate to democracy, though? To be sure, emotional or salacious content has always had its place in people's decisionmaking calculus, political or otherwise. What makes internet virality different is that the priorities of the information ecosystem are, in a sense, crowdsourced. The legitimacy of topics, memes, and messages comes from their popularity, not some other quality such as relevance, newsworthiness, or truth. Again, this populist turn in information transfer has benefits and costs, both of which come from the diminished role of establishment (biased) mediators that had constrained the range of topics, the types of images, and the character of the language fed to news-hungry voters. In the internet world, the news (or at least headlines) you see is, in a sense, "voted on" by your peers and others to decide whether it warrants your attention. In that respect, it is a more *democratic* form of news provision, but the key difference is that the *ex ante* popularity of the messages becomes the criterion for whether the reader sees the message to begin with.

Compare that dynamic to that of the preexisting world, in which editors and producers served as gatekeepers for whatever might be characterized as "news." To be sure, popularity and public interest were criteria that factored into the decision on whether to broadcast or print a story, but to some extent even those values required guesswork by media elites as to which stories might be popular ones. Moreover, competing values had their place in the balancing of whether to give the people what they wanted or what was, according to some metric, "good for them."

Finally, and by way of transition to the next topic on echo chambers, it is important to understand that virality is not limited to political information. More to the point, the democratic structure of the internet places all "information" and "communication" on an equal footing. Indeed, one mistake that people make in analyzing the impact of the internet is to assume that political information or "news" is somehow hived off from other types of media in a viewer's newsfeed. But in reality, the forces that lead to viral cat videos or stories related to celebrities are the same as those that popularize news related to an election campaign or discussion of issues of public concern.

The most important decisions social media and search platforms make concern the relative placement and prevalence of information on the screen. Virality operates both to prioritize "popular" communication (literally, by having popularity factor into algorithmic determinations as to which communication appears at the top of the screen) and to barrage the consumer with the same information, again and again. Almost by definition, a viral communication is some text, video, or image that will repeatedly be in front of the face of consumers, as they receive it from friends, publications, and others in their network.

Combatting "virality," if one were to set out to do so, would cut to the heart of the "social" component of social media. If virality is a problem to be solved, then most measures to address it involve slowing down information transfer or otherwise mediating which types of communication should be allowed to "go viral." Any such attempt requires mediation of an inherently unmediated information environment. As discussed later, this can be done, but not

without some loss to the features that give the various social media platforms the character that users have expected.

C. Anonymity

Anonymous speech is not only valuable in some settings, it is often protected by law. The First Amendment to the U.S Constitution, for example, protects anonymous speakers, especially if they legitimately fear retaliation from governmental or non-governmental actors should their identities be made public. Protections against disclosure of speakers' identities or association membership were indispensable to organizations during the Civil Rights Movement that feared disclosure might threaten the participation and even the lives of activists.[5] Indeed, advocates for the U.S. Constitution itself wrote *The Federalist Papers* under the pseudonym "Publius," to ensure that readers would not associate any individual essay with a particular person who participated in the Constitutional Convention.

Similarly, anonymous online speech provides significant benefits to users. Dissidents in totalitarian regimes are able to tweet about human rights abuses or organize protests only if they believe the government will not be able to discover their identities. Similarly, those seeking help or community on sensitive topics – whether suicide prevention, health information, sexual identity, or a range of private topics – gain shelter in anonymity that would evaporate if all internet speech were to take place "in the open."

With all that said, the megaphone that the internet provides to anonymous speakers gives them unprecedented power. We have come a long way from protecting the anonymous pamphleteers whose reach extends only to where their feet and endurance (and a copy machine) might take them.[6] Anonymity shields speakers from responsibility for their speech and liberates them to engage in the kind of trolling, hateful, inciting, obscene, sensationalist, conspiratorial, and generally extremist speech that the norms of face-to-face communication prevent. This is not to say that in the offline or online world some people do not proudly and notoriously engage in such speech – the protests in Charlottesville attest to that, as do the hate speakers who have gained an online following or those wearing QAnon T-shirts at rallies. The point is that some people, who otherwise might have their speech chilled were they held responsible for it, will engage in such speech under the cloak of internet anonymity – and they will potentially do so with a world-wide audience.

Hate speech is merely one species – if perhaps the one most often noted and researched[7] – of unaccountable speech shielded by the anonymity of the internet. Anonymity facilitates the

---

[5] *See* NAACP v. Alabama ex rel Patterson, 357 U.S. 449 (1958).
[6] *See* McIntyre v. Ohio Elections Commission, 514 U.S. 334 (1995).
[7] *See, e.g.*, Alexandra Siegel, Evgenii Nikitin, Pablo Barberá, Bethany Pullen, Joanna Sterling, Richard Bonneau, Jonathan Nagler and Joshua Tucker, *Trumping Hate on Twitter? Online Hate Speech and White Nationalist Rhetoric in the 2016 US Election Campaign and its Aftermath* (New York University SMaPP Lab, Working Paper, 2017), http://textasdata2017.net/wp-content/uploads/2017/08/siegel_princeton_TAD_short.pdf; Anti-Defamation League,

creation and amplification of *all* objectionable content with which speakers and audiences alike seek unaccountable engagement. The same could be said for threats, bullying, and trolling – speech that can overlap with racist or other hate speech but usually reflects direct and individual, rather than group, targeting. Other kinds of extremist speech, ranging from terrorist recruitment to other forms of incitement, also often flourish due to anonymity.

Foreign election interference through online communication, as well, is only made possible because of the difficulty in discerning the origin of anonymous speech. The internet masks not only the identity, but also the location of the speaker. Foreign speakers can pose as domestic ones, and government agents (in content such as their tweets, trolling, and news reporting) can appear as normal members of the internet crowd. The "foreignness" of the speaker need not be limited to the now-familiar Russian-style foreign state actor intervention. Any speakers (such as an out-of-state organization in a local election) who calculate that revealing their residency might lead an audience to discount their speech may gain something from the anonymity that the internet provides.

Not only does internet anonymity conceal the identity and location of the speaker, but it can also obscure even their humanity. The internet's "bot" problem is a consequence of the privileging of online anonymity. Not only can it be impossible to determine *who* is speaking to you online, but it is becoming increasingly difficult to discern whether such speech comes from a human being at all. The millions of bots on Twitter (representing over ten percent of accounts in the U.S. and an even greater share in other countries)[8] create, forward, and publicize content that is often indiscernible to the average user. In fact, estimates suggest that bots are more prolific than human users in sharing links on Twitter.[9] Indeed, much of what they do is simply repeat or repackage messages from others so as to trick algorithms (such as those that determine search engine or newsfeed rankings) into making such content more prominent due to fictitious gains in popularity. The same, of course, can be said for the role of bots in padding the number of followers, likes, clicks, or other measures of engagement so as to misrepresent the popularity of a person, account, or news story. For the internet platforms that care about the bot problem (and some do more than others), they are constantly engaged in a cat and mouse game, as talented adversaries continue to try to make their automated accounts more "human" so as to evade the platforms' bot-detection systems.

"The Online Hate Index" (Jan. 2018), https://www.adl.org/resources/reports/the-online-hate-index; Southern Poverty Law Center, *McInnes, Molyneux, and 4chan: Investigating pathways to the alt-right* (Apr. 19, 2018), https://www.splcenter.org/20170118/google-and-miseducation-dylann-roof; Alice Marwick and Rebecca Lewis, *Media Manipulation and Disinformation Online*, DATA&SOCIETY (May 15, 2017), https://datasociety.net/output/media-manipulation-and-disinfo-online/.

[8] Onur Varol, Emilio Ferrara, Clayton A. Davis, Filippo Menczer, Alessandro Flammini, Online Human-Bot Interactions: Detection, Estimation, and Characterization (March 2017), https://arxiv.org/pdf/1703.03107.pdf .

[9] Stefan Wojcik, *5 things to know about bots on Twitter*, Pew Research Center: Internet Technology (Apr. 9, 2018), http://www.pewresearch.org/fact-tank/2018/04/09/5-things-to-know-about-bots-on-twitter/.

But why does anonymous internet speech, which, as noted above, might aid in critical forms of protest against authoritarian regimes, also create tensions for a democracy? If one believes in the strong form of the marketplace of ideas, for example, the identity (or lack thereof) of the speaker should not matter: those who hear the speech ought to be able to evaluate the truth of the statements themselves, in the context of counter speech that exposes falsehoods and biases. Moreover, the audience should be able to discount the message from anonymous or unfamiliar speakers (assuming they are not impersonating someone else) so as to weight trusted, familiar sources more.

Perhaps to state the obvious, there simply is no support for the strong version of the marketplace of ideas when it comes to anonymous speech in the internet age. That is not to say that anonymity should not be valued and protected in many or even most circumstances. Rather, the suggestion here is that the masking of identity built into the structure of internet communication brings with it inevitable risks of misrepresentation and manipulation. The inability to identify the other person (if it is a person) at the other end of the computer conversation often leads that person to engage in certain types of speech that they would not engage in face-to-face. The norms of civility, the fears of retaliation and estrangement, as well as basic psychological dynamics of reciprocity that might deter some types of speech when the speaker and audience know each other – all are retarded when the speech is separated from the speaker, as it is online. The now well-documented anonymous online threats to journalists, in particular, bring the argument into sharp relief. Such "speech" can chill other speech, much of which is essential to an informed electorate and well-functioning democracy.

For purposes of democratic discourse, then, the pervasiveness of internet anonymity facilitates kinds of speech that are harmful to democracy, hinders audiences' capacity to discount messages by the identity of the speaker, and presents challenges to speech regulators (from either platforms or governments) who seek to punish or deter anonymous speakers for their behavior online. Again, anonymity "protects" speakers, facilitating anti-regime, anti-establishment or anti-majority voices in any society. It protects the Turkish or Egyptian protester seeking to organize online protests against authoritarian behavior, just as it also protects neo-Nazis, those who threaten journalists, and sophisticated Russian trolling operations seeking to divide and destabilize democracies. When it comes to elections, though, the unaccountable speech anonymity facilitates can promote division and deception that hinders the proper functioning of a democracy. It enables extremist voices that seek to undercut the legitimacy of the electoral process and basic constitutional values. Anonymity and pseudonymity (adopting an online persona other than one's own) also facilitate the kind of lying and misrepresentation that undercut a well-informed electorate. In the internet world, anonymous and pseudonymous speakers cannot be held to account for the truth of their electorally relevant statements. Consequently, the speaker bares no cost for repeating lies and promoting false content. Although, to be sure, a great many political actors engage openly in divisive and deceptive speech these days, online anonymity provides cover to anyone who might wish to spread lies and division to a potential world-wide audience.

D. Homophily: Filter Bubbles, Echo Chambers, and Information Cocoons

Even before the 2016 U.S. Election heightened people's awareness of the potential downside of the internet for democracy, a growing set of critics had identified the particular pathology of "echo chambers" as a source of concern. Polarization was then seen as the chief political ill in the United States. The internet, or rather, the way people consumed news and conducted conversations online, was suggested as a partial driver of this polarization. If people live in online information cocoons, the argument went and goes, then they are not exposed to alternative viewpoints and remain fixed in their beliefs.

This oft-made critique of the greater choice, access, and personalization the internet affords over legacy media is really two arguments, somewhat in tension with one another. The first is a lamentation of the decline of the public square. The internet exacerbates polarization, under this view, because people lack a common forum in which they will encounter information and argument different from what they experience in their close social circles. If polarization develops, at least in part, because people opt into news and information sources that reinforce their prior beliefs, then perhaps a space (virtual or real) in which they can be exposed to other points of view will moderate their beliefs. Not only might they be persuaded by arguments they have never entertained, but they will learn facts inconsistent with the stories they are told in their information cocoons. If one believes that the marketplace of ideas provides the best test for truth by allowing arguments to compete against each other, then exposing people to competing ideas would be a necessary, if certainly not sufficient, check on the spread of falsehoods and weak arguments.

The second argument implied by the echo chamber critique is a bit different. It suggests that the balkanization of online media eliminates any common source of information about which arguments can take place. Here, the argument is not based on the absence of a forum for different and competing arguments, but rather on the lack of a common source of authority in the online world that can provide a shared base for truth. As a result, people believe in "alternative facts" based on the information sources they opt into. Whereas the first lament focuses on the lack of exposure to alternative viewpoints, the latter critique raises a contradictory concern: the lack of exposure to a common set of news and information. On this view, the "problem" with the internet is the lack of a common experience that defines the community. These critics have nostalgia for a time when personalities like Walter Cronkite could command the attention of a third of the American population each evening. At the time, a common source of information united the body politic, which shared the experience of tuning in to a limited set of television news networks and (on the local level) reading a limited number of newspapers. Moreover, such news sources obeyed a series of professional norms that shaped boundaries around what counted as news and what was permissible to broadcast. Network television was also subject to certain legal restrictions such as the fairness doctrine and equal time rule, which served to check against political bias.

Of course, the benefits of these limited, but community-building, information sources was also fodder for the well-known critiques against them. Many saw those artificially-dominant (in the sense that the limited broadcast spectrum space required the government to apportion out licenses to only a few entities) mainstream sources as biased against political and racial minorities. Conservatives pointed to the fact that few journalists with nationwide exposure were Republicans,[10] and left-wing critics saw the corporate-controlled mainstream media as motivated by ratings and advertising, and therefore biased in favor of news that would not rock the establishment boat.[11] Similarly, as the main networks and newsrooms were dominated by white males, the lack of diversity was seen as biasing news coverage in favor of the backgrounds of reporters and the majority of their audience.[12]

Under this view, the explosion of news sources, first with cable television and then with the web, liberated populations from the tyranny of editors and broadcasters who would limit "news" to what they considered appropriate and healthy for the audience. As with the web generally, citizen journalism and the multiplicity of news organizations the web enabled allowed a diverse array of voices to be heard or at least to have a platform from which to garner a nationwide or even global audience. It broke down barriers as to what constituted news and when and where you could access it. Audiences were now empowered to select the news that attracted them (or for that matter, no news at all).

As in so many other contexts, the individualism of the web threatened the sense of community (and even reality) forged in the previous information environment. The critique then grew that echo chambers had developed online and people were merely getting the news they wanted, not the news they needed or that a democracy requires. Even if the previous news environment had its authoritarian qualities, the argument goes, the simultaneous cacophony and isolation of the web conflicts with democracy's need for some community defining baseline of reliable information.

But is this conventional critique accurate? Do people self-select into echo chambers and receive news that merely confirms what they already believe? Or perhaps of greater relevance, is their online news consumption and political discussion qualitatively different (that is, more homophilous) than their offline behavior?

The available evidence provides mixed support (at most) for the strong version of this conventional critique and also suggests that the echo chamber argument needs to be

---

[10] Nicole Hemmer, *The Conservative War on Liberal Media Has a Long History*, THE ATLANTIC (Jan. 17, 2014), https://www.theatlantic.com/politics/archive/2014/01/the-conservative-war-on-liberal-media-has-a-long-history/283149/.

[11] Peter Dreier, *Capitalists vs. the media: an analysis of an ideological mobilization among business leaders*, MEDIA, CULTURE AND SOCIETY (1984), http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.832.6911&rep=rep1&type=pdf.

[12] Tom Hrach, *A felonious former Illinois governor's surprising contribution to journalism*, Columbia Journalism Review (Feb. 8, 2017), https://www.cjr.org/the_feature/illinois-governor-kerner-journalism.php.

reconceptualized. First, the conventional critique overstates the amount of time – either online or offline – that people spend acquiring and digesting news or politically relevant information. Most people come to social media to be social: that is, to interact with friends and family in the same ways they do offline.[13] Similarly, most people search the web to find answers to questions that arise in their lives: what restaurant to go to, which movies to see, or whether their mild sickness is indicative of some exotic life-threatening illness. To be sure, politically interested and knowledgeable web users will exhibit different online behaviors and interests than those less interested, just as they do offline. All things being equal, a dedicated and politically interested liberal, for example, might be more likely to have similar friends with similar interests. But most people are not so politically interested, nor are they likely to use the web primarily for political information. This is not to disagree with the fact that most people already or soon will get their political information from social media and online sources. Rather, for most people politics will continue to occupy a small share of their attention, even these days when it seems like politics overwhelms all other news and topics. As Facebook has publicly released, about 4% of the newsfeed of the average Facebook user is comprised of what might loosely be thought of as "news."[14]

Second, when it comes to social media, and the reliance on friend networks for political information, the evidence suggests that, for most of us, our online lives are not as politically homophilous as most critics suggest.[15] They seem to exhibit levels of homophily comparable to our friendship networks in the offline world. Indeed, they are often more politically diverse, because as geographic political segregation grows and people "vote with their feet" into politically homogenous areas, retaining online friendships with old school friends and extended family often exposes one to political views different than those growing from our politically segregated neighborhoods. In short, we all have that crazy uncle who posts crazy or extremist material on Facebook – exposing us to information and communication we might not see from our closest friends. The key to understanding news exposure on Facebook is the outsized importance of "weak ties" in supplying information on social media.[16] Whereas in our work and home life, we tend to

---

[13] Lee Rainie, *Americans' complicated feelings about social media in an era of privacy concerns,* PEW RESEARCH CENTER (Mar. 27, 2018), http://www.pewresearch.org/fact-tank/2018/03/27/americans-complicated-feelings-about-social-media-in-an-era-of-privacy-concerns/.

[14] *Helping Ensure News on Facebook Is From Trusted Sources*, FACEBOOK (Jan. 19, 2018), https://newsroom.fb.com/news/2018/01/trusted-sources/.

[15] Richard Fletcher and Rasmus Kleis Nielsen, *Are people incidentally exposed to news on social media? A comparative analysis*, NEW MEDIA & SOCIETY (Aug. 17, 2017), http://journals.sagepub.com/doi/abs/10.1177/1461444817724170; Maeve Duggan and Aaron Smith, *The Political Environment on Social Media*, PEW RESEARCH CENTER (Oct. 25, 2016), http://www.pewinternet.org/2016/10/25/political-content-on-social-media/.

[16] Eytan Bakshy, Itamar Rosenn, Cameron A. Marlow, Lada A. Adami, *The Role of Social Networks in Information Diffusion*, INTERNATIONAL WORLD WIDE WEB CONFERENCE COMMITTEE (2012), http://delivery.acm.org/10.1145/2190000/2187907/p519-bakshy.pdf?ip=128.12.244.4&id=2187907&acc=ACTIVE%20SERVICE&key=AA86BE8B6928DDC7%2E0AF80552DEC4BA76%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&__acm__=1536730062_9a445a1c41475c28b3d31dd2478d2b11.

talk politics with our closest friends, when it comes to our friends on Facebook, we become exposed to information from a larger group of people, some of whom are politically different than the friends we would select for political discussion if we were more discerning.

Third, it remains the case that mainstream sources remain much more popular than extremist sources among the vast majority of internet users.[17] To be sure, there are times, and pre-election periods may be one of them, when certain extreme sources rival certain mainstream sources or at least certain stories from these sources might.[18] Moreover, for certain searches, extremist news sources might ascend to the top of search results, especially when they are the product of search engine manipulation. But because the number of mainstream sources and the amount of mainstream news far outstrips the amount of extremist content, the overwhelming majority of users will see such mainstream sources more often in their newsfeed or search results than the extremist sources.

The social science as to online echo chambers has moved away from the "strong version" that suggests most people live political homophilous online lives to a set of more complicated questions as to "who" experiences echo chambers and "why." Even if most people have friendship networks that resemble their offline lives, some people might not only have politically more homogenous networks, but they might, in fact, seek them out. And it may be that the effect of echo chambers is different on different people: that is, for people who are otherwise not engaged in politics but have a homogenous group of online friends who are more engaged, the effect of the online echo chamber could be to polarize them toward the median member of the group.

No one can doubt that the internet and social media make echo chambers more available to those who seek them. Indeed, that is the beauty of the internet: you can find like-minded people anywhere in the world, whatever the peculiar connection you might have with them. As said above, that is true for knitting enthusiasts as it is for neo-Nazis and terrorist sympathizers. However, the norm erosion that occurs due to viral or anonymous speech is exacerbated by the lack of friction of the online world in finding political comrades-in-arms. More to the point, among online groups of like-minded partisans, individual speakers do not need to moderate their positions or speech to be acceptable to a larger, more diverse audience. In groups self-selected for their political stances, speakers can compete to be the most outrageous and extreme, and they will be unlikely to confront any sanctions.

Of course, echo chambers vary considerably in the degree to which they harden people's extremist beliefs, promote violence or otherwise threaten democracy. At one extreme are the dark

---

[17] Jeffrey Gottfried, Michael Barthel and Amy Mitchell, *Trump, Clinton Voters Divided in Their Main Source for Election News*, Pew Research Center (Jan. 18, 2017), http://www.journalism.org/2017/01/18/trump-clinton-voters-divided-in-their-main-source-for-election-news/.

[18] Samantha Bradshaw and Philip N. Howard, *Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation* (Oxford Institute, Working Paper 2018.1), http://comprop.oii.ox.ac.uk/research/cybertroops2018/.

corners of the internet – discussion groups and bulletin boards on sites like 4chan and 8chan or avowedly racist subreddits. These forums often give birth to and cultivate conspiracy theories, like the famous Pizzagate, in which a believer shot up a Washington, DC, pizza parlor described as the site of a child trafficking ring involving Hillary Clinton and other top Democrats.[19] Most recently, another strange conspiracy initiated on 4chan involving sex trafficking and the so-called Deep State, adopting the moniker QAnon, has made its way into mainstream circles as adherents have attended rallies with the President of the United States.[20]

While those conspiracy theories might occupy one end of the spectrum, they are emblematic of what happens among intense ideological adherents in online communities defined by tribal allegiance. Whether defined by race, religion, party, or interest, online groups can facilitate a sense of group cohesion and tribalism. When arguments or conspiracies go unchallenged, let alone become the stuff of cheerleading among the group, weak ties become stronger and soft attitudes harden. It is very difficult to get a handle on how big a phenomenon these extremist groups are – that is, what share of internet users spend considerable time in online groups or with online sources characterized by this type of homophily. Media attention, particularly after a group member commits violence, is a poor indicator of the scale and representatives of the phenomenon. Nevertheless, numerous studies of the alt-right in the United States and Europe give us a sense of the power of these groups and their ability to organize both online and offline, targeting opponents and orchestrating sophisticated social media campaigns.[21]

Finally, although we tend to think of homophily as a demand-side phenomenon (with people opting into echo chambers), the flip-side of echo chambers is microtargeting and the emergence of tools and strategies to deliver messages to consumers designed to appeal to their identity, experience and beliefs. While targeted advertising is as old as advertising, microtargeting in the digital age represents an extreme difference in degree if not in kind. More to the point, the internet enables unprecedented gathering of information on individuals (including search histories, friendship networks, and buying habits) and therefore the crafting of messages designed to appeal to their particular preferences and prejudices.

Of course, microtargeting is just another tool or feature of life in the era of big data and the internet; it can be used for good or ill. Indeed, the dark shadow cast over microtargeting since the 2016 U.S. election differs considerably from the fascination with it following the 2008 and 2012 U.S. elections.[22] In those elections, the campaign of Barack Obama was roundly praised for its capacity to craft targeted message to raise money and mobilize its supporters online. In 2016, however, microtargeting took a darker turn as the scandal surrounding Cambridge Analytica

[19] Matthew Haag and Maya Salam, *Gunman in 'Pizzagate' Shooting is Setenced to 4 Years in Prison*, NYTIMES (June 22, 2017), https://www.nytimes.com/2017/06/22/us/pizzagate-attack-sentence.html.
[20] Kyle Feldscher, *QAnon-believing 'conspiracy analyst' meets Trump in the White House*, CNN (Aug. 25, 2018), https://www.cnn.com/2018/08/25/politics/donald-trump-qanon-white-house/index.html.
[21] Joan Donovan, Becca Lewis and Brian Friedberg, *Networked Nationalisms: White Identity Politics at the Borders*, MEDIUM (July 11, 2018), https://medium.com/@MediaManipulation/networked-nationalisms-2983deae5620.
[22] See Sasha Issenberg, *The Victory Lab: The Secret Science of Winning Campaigns* (2012).

presaged a future in which psychographic profiling could be employed to craft individualized messages that manipulate subconscious motivations to achieve political ends.[23] To be clear, few people think that Cambridge Analytica was successful, this time, in using such psychographic profiling methods.[24] However, they, along with other organs of the Trump campaign, used the advertising tools made available by Facebook and other platforms that allow the construction of custom audiences – that is, a group of Facebook users defined by certain characteristics, tastes, and behavior. With these tools, the campaign was able not only to target supporters, but also to send demobilizing (and at times, racially tinged) messages to potential supporters of its opponent.[25]

Microtargeting represents an extension of the homophily argument because it exists as a tool that both the platforms and political actors can use to construct communities and deliver messages or advertisements to achieve political goals. Facebook, for example, not only allows advertisers to target based on demographic characteristics such as age, gender, education, and location, but also enables the creation of a "custom audience." A purchaser creates a custom audience by assembling a list of email addresses and delivering them to Facebook for ad targeting. Often, such groups are created by third party consultants or marketers, who themselves have used available big data to envision the types of people that will be susceptible to the desired message. Once that custom audience is created, Facebook also offers a service of creating a "lookalike audience", which draws conclusions from the custom audience to extend the advertisement to a group of people that shares similar characteristics, which includes not only demographic attributes but also shared interests and political views. Although the platforms facilitate the individualized delivery of these targeted messages, it is important to understand that an entire outside industry has developed to use big data (often even from public sources) to enable targeted of audiences over the internet.

The rising concerns surrounding microtargeting, like critiques of propaganda throughout history, arise from a basic distrust of individuals' abilities to resist the manipulative messages that play on their emotions. In the context of political advertising and election campaigns, we worry about the unfair advantage in the attainment of political power that goes to the best manipulator with the best data. In an idealized version of democracy, voters' evaluate candidates and parties on their merits and make an informed decision based on available public information. Many people have considered political advertising, in general, to violate this idealized conception, but all the more so, as microtargeting has become increasingly sophisticated, people lose confidence in the marketplace of ideas as the test for democracy-relevant truths.

---

[23] Sue Halpern, *Cambridge Analytica and the Perils of Psychographics*, NEW YORKER (Mar. 30, 2018), https://www.newyorker.com/.../cambridge-analytica-and-the-perils-of-psychographics.
[24] Evan Halper, *Was Cambridge Analytica a digital Svengali or snake-oil salesman?*, L.A. Times (Mar. 21, 2018), http://www.latimes.com/politics/la-na-pol-cambridge-analytica-20180321-story.html.
[25] Joshua Green and Sasha Issenberg, *Inside the Trump Bunker, With Days to Go*, BLOOMBERG (Oct. 27, 2016), https://www.bloomberg.com/news/articles/2016-10-27/inside-the-trump-bunker-with-12-days-to-go.

E. Monopoly

The contemporary media landscape differs markedly from its predecessors in the power and reach of the major internet platforms. This is not to say that media monopolies – local or national – have not existed before. An oligopoly of the three major television networks in the U.S. existed for generations, and in other countries, consumers often had even less choice, especially when the state controlled the media. Newspapers often had local monopolies, with chains that had national reach. Both now and previously, media conglomerates assemble together multiple media properties, as well as the modes of delivery (such as cable TV providers), under one roof. Concentrated power in media markets is not a new phenomenon.[26]

The online media environment is qualitatively different. To some extent, today is an age of unprecedented media pluralism and diversity. There are more news sources than ever before, and anyone with access to the internet can attain information from more sources of various ideological predispositions than during any previous age. Indeed, in this day and era, it becomes difficult to define "the media" as almost anyone can tweet, post, or blog.

Alongside this balkanization of the media, concentration has occurred among the major internet platforms.[27] Facebook is the dominant social media platform, and along with its properties, WhatsApp and Instagram, comprises an unrivaled position in its share of online social interaction. Google is functionally a monopoly when it comes to search, and its property, YouTube, is functionally a monopoly when it comes to online user-produced video. Both companies would be quick to describe themselves as something other than "media" companies – in part, so as to distinguish themselves from publishers, who under U.S. law would be liable for the content on the platforms. Nevertheless, no one can doubt the power and omnipresence of these platforms in their specific domains.

From a traditional antitrust (or antimonopoly) perspective, though, these platforms represent a bit of a categorization challenge. In general, monopolists exert their unfair power by increasing price and, perhaps, decreasing quality. But these firms offer their products for free. Consumers are not exploited in the traditional way that monopolies might take advantage of them. Their market power derives from their popularity and the amount of time people spend on the sites.

One might not say the same for advertisers. Roughly 73 percent of new ad dollars in recent years have flown to Google and Facebook.[28] As a result, those platforms (and other internet innovations, such as Craigslist, that have made classified ads profitless) have drained revenue from certain classes of media properties, particularly local journalistic institutions. To the extent they

---

[26] Tim Wu, *The Attention Merchants*: *The Epic Scramble to Get Inside Our Heads* (2016); Tim Wu, *The Master Switch: The Rise and Fall of Information Empires* (2010).
[27] Tim Wu, *The Curse of Bigness: Antitrust in the New Gilded Age* (2018).
[28] John Koetsier, Digital Duopoly Declining? Facebook's, Google's Share of Digital Ad Dollars Dropping, FORBES (Mar. 19, 2018), https://www.forbes.com/sites/johnkoetsier/2018/03/19/digital-duopoly-declining-facebooks-googles-share-of-digital-ad-dollars-dropping/#5561b5b760a8.

have power over a *market*, then, it is the advertising market, and they derive this power merely from their capturing of people's attention.

As a result of these unique monopoly qualities, the traditional tools of antitrust or competition law fit uncomfortably. To be sure, the firms could be broken up into their constituent parts, with WhatsApp and Instagram being severed from Facebook, and YouTube (as well as the Android operating system) from Google. Moreover, in some instances, the platforms could be reined in by traditional rules prohibiting vertical integration, along the lines of European enforcement actions against Google for favoring of its own products in search results or requiring its browser and search engine to be given priority on Android phones.[29] These actions, and others like it, can take some money from those corporations and might be desirable with respect to diminishing their overall value and size, but they will not do much to constrain the most important sources of their power over communication.

Most of the power of these platforms – at least from the perspective of their impact on democracy – derives from simple features of search results or the newsfeed. In other words, Google's power derives from the fact that virtually everyone turns to it as the authoritative index of the web. No severing of YouTube or other properties will diminish that dimension of the company's popularity and power, or most importantly, its capacity to exploit its power over search to direct eyeballs toward certain products and websites. Similarly, most of the democracy-relevant power of Facebook comes from its newsfeed – that is, its capacity to direct and maintain user attention to its particular packaging and hierarchy of communication and advertisements. A corollary to that power, of course, is its ability to decide the relative priority of certain types of information (or disinformation) and publications. The more important that the newsfeed becomes as the conduit for politically relevant information (or, again, disinformation), the more critical the decisions that Facebook makes as to what types of information appear on the platform and in what order.

Herein lies the particular monopoly power of the platforms that seems most relevant to democracy and elections. In many respects, decisions as to which communications to allow on these platforms are more important than government speech restrictions. Their rules as to disinformation, hate speech, incitement, or threats, for example, may "govern" more speech than the laws on the books, especially given that their automated filters have capacity to "preemptively regulate" in ways unavailable to government speech restrictions. Their procedures for filtering and taking down content determine the boundaries of acceptable speech in the communication environment most used by candidates, journalists, and voters.

[29] Indeed, one of the most understudied subjects in the economics of these platforms is the relationship between the telecommunications industry and services they provide. Their monopoly position is often solidified by discriminatory pricing by providers of mobile phones and telecommunications services. This can occur, for example, when certain apps automatically come with the purchase of a phone or when certain apps are favored in data plans the mobile phone carriers provide.

Similarly, the algorithms themselves – whether for search or for the newsfeed – translate into unique power over decisions as to what people see and read. Whenever the platforms deprioritize certain classes of publications (*e.g.*, because of their ideology, authoritativeness, novelty, likely engagement, or clickbaitish-ness) or even certain types of communication over others (for instance, content from friends as opposed to news sources, as Facebook announced earlier this year[30]), they make decisions with extensive repercussions for the flow of political information. In many ways, these less transparent decisions as to the prioritization of communication are even more important than the more notorious decisions as to what speech finds a place on the platform.

It has become commonplace, for example, in the United States for conservative publishers to decry "shadowbanning" by Twitter and other platforms. The term refers to the demotion of content to the point where few people are exposed to it. The platform does not remove the content, but neither does it serve it high up to viewers in their newsfeeds or search results. It requires, instead, that users specifically seek out the content. President Trump made a similar claim recently when he erroneously accused Google of biasing search results for "Trump news"[31] against conservative media. In all of these cases, the information is still available on the platform, but it is (allegedly) placed so low in the relevant list that exposure will be greatly reduced.

Intentional political discrimination is only the most blatant danger of algorithms structuring political discourse. The platforms' monopoly power presents dangers to democracy precisely because some type of discrimination is inherent in the products themselves. Google orders websites in its search results, and Facebook and Twitter organize communication in their newsfeeds. Something goes at the top and something is pushed off the page. Whether or not this is done explicitly for "partisan" reasons, the algorithm, by its nature, determines priorities and hence, "discriminates" among different types of communication. The more important the platform is for a given communication ecosystem (and in some areas of the developing world, Facebook *is* the internet), the more powerful it will be in setting the priorities for political communication in the country.

F. Sovereignty

Election manipulation by foreign actors is not a phenomenon original to the internet age. During various periods of international conflict, governments have attempted regime change in others, and if the subject country is a democracy, one way to do so was to assist in the election of new leaders friendly to the intervening power. Nevertheless, as ongoing investigations of Russian

[30] Mike Isaac, *Facebook Overhauls News Feed to Focus on What Friends and Family Share,* NYTIMES (Jan. 11, 2018), https://www.nytimes.com/2018/01/11/technology/facebook-news-feed.html.
[31] Donald J. Trump (@realDonaldTrump), TWITTER (Aug. 28, 2018, 12:02 PM), https://twitter.com/realDonaldTrump/status/1034456281120206848.

influence on the 2016 U.S. Election and the Brexit referendum have demonstrated, the internet supplies new tools for foreign electoral manipulation.

The "sovereignty" issues that the internet poses for democracies go well beyond electoral manipulation, as serious as that is. Deeper dives into the character of Russian advertisements, organic content, and amplified domestic communication have demonstrated how a foreign government can foster division and confusion in a democracy, both during an election period and beyond.[32] Deploying bots, trolls, and cyborgs to pollute the information ecosystem of the target democracy, aggressors can take advantage of the anonymity and pseudonymity of online communication to behave like domestic political speakers and campaigns. Even hiding in plain sight, they can use state-sponsored press, as with *RT* and *Sputnik*, to build foreign audiences, to amplify memes and stories, and to activate a network of supporters during elections or other critical democratic moments.

In the pre-internet age, information warfare might involve governments dropping leaflets on unsuspecting populations or secretly manipulating elites in campaigns and the media. Now, the worldwide nature of the web allows for coordinated manipulation without physically venturing beyond one country's borders. Intelligence services can "work from home," as it were, by exploiting the web's inherent anonymity, which (with some level of sophistication) can mask the origin of communication as well.

Non-state actors also take advantage of the uncertain origin of internet-based communication. The well-known use of the internet by terrorist organizations for recruitment or messaging makes clear how a lack of state affiliation or sponsorship does not serve as a barrier for using the internet to target and persuade vulnerable populations. Similarly, in the electoral context, international "consulting groups" – some with defined ideologies and objectives and others that sell services to the highest bidder – can serve as a one-stop shop for assistance and tools for those seeking to exploit the vulnerabilities of the internet to target populations for messages of persuasion, demobilization, and division.

II. Agents of Reform: Governments, Platforms, Civil Society

There are a limited number of institutions in a position to address the challenges that the internet poses for democracy. For each of the categories of reform discussed next, governments, the major internet platforms, and civil society can play a role. In an ideal world, they would work together with common purpose. Interventions in this arena, however, often confront significant political and legal obstacles, as almost all of them involve some kind of restriction or

[32] *See* Kathleen Hall Jamieson, *Cyberwar: How Russian Hackers and Trolls Helped Elect a President What We Don't, Can't, and Do Know* (2018).

reorganization that affects political speech. As a result, some agents of reform are better positioned than others to tackle the different challenges the digital environment poses for democracy.

A. Government Regulation

When it comes to government regulation, three models are competing for popularity. As Timothy Garton Ash has put it well,[33] China, Europe, and the United States provide different models for regulating internet speech and internet platforms. They create a spectrum of censorship and state involvement that other countries are considering now as well. Given the widespread concern that a free internet is posing unique challenges for democracy, the full panoply of options are on the table as countries consider protecting their populations from "dangerous speech." Of course, countries span this spectrum as to how much speech they allow – on the internet or elsewhere – with some shutting down the internet or punishing speakers offline for what they say and do online. But as governments look for models to emulate these three archetypes provide some direction.

The Chinese "walled garden" approach represents the most extreme example of government regulation and involvement, at least among countries with widespread access to the internet.[34] China censors and punishes online speech, bans platforms like Google and Facebook from operating in the country, and maintains a million-person surveillance team to observe and guide discussions online. China may occupy the authoritarian extreme of the regulatory spectrum, but a great deal of online interaction and commerce still occurs in China nonetheless. Moreover, Chinese discrimination against foreign firms and platforms provides a model for sovereign control of the internet that other countries, which feel "invaded" and helpless in the face of American platform power, find attractive. Especially given the success of Chinese internet firms, such as Alibaba and Weibo, the Chinese model of internet regulation, despite its authoritarianism, has considerable appeal to those countries that have not fully bought into Western notions of free speech.

At the other extreme is the United States with its libertarian view of speech and generally deregulatory posture toward the internet. It is worth remarking that, even apart from internet regulation, the United States occupies the far end of the spectrum when it comes to speech regulation. The American First Amendment protects some categories of speech that are widely regulated around the world. The U.S. Supreme Court's doctrine regarding obscenity, hate speech, libel, incitement, fighting words, commercial speech, campaign finance, and a host of other free speech domains distinguish the U.S. in its extensive protection of most categories of speech.

---

[33] *See* Timothy Garton Ash, *Free Speech: Ten Principles for a Connected World* (2016).

[34] *See* Gary King, Jennifer Pan, and Margaret Roberts, *How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argumen*t, AMERICAN POLITICAL SCIENCE REVIEW (2017), available at http://jenpan.com/jen_pan/50c.pdf.

Similarly, U.S. legal treatment of platforms is the most protective in the world. Section 230 of the Communication Decency Act[35] immunizes platforms (in most situations) from liability for other parties' speech that occurs on their platforms over which they do not exercise editorial control. This legal protection is often given credit for the rapid growth of Google and Facebook, as well as other platforms for internet commerce. Indeed, it is the lack of liability for customer speech that allows these platforms to adopt lean organizational structures, rather than employ a greater number of moderators who would monitor and take down illegal or tortious content.

For the most part, Europe adopts a model of greater intermediary liability and greater restriction of online speech. Most notoriously, the German NetzDG law[36] provides for fines of internet platforms up to fifty million Euros for illegal speech that remains on the platform after they have been notified (with some exceptions). That law piggybacks onto greater restrictions present in the German law on defamation and hate speech (e.g., a ban on Holocaust denial). The law does not itself provide guidance to the platforms as to what speech per se they should take down. Rather, it specifies that they, in effect, look to the law and precedent to make determinations as to whether speech that is identified as problematic is actually illegal. This offloading of legal responsibility has been copied by governments around the world, including Russia.[37]

Even beyond content restrictions, Europe has been at the forefront of platform regulation. In both antitrust and privacy protection, the European Commission has levied stiff (multi-billion and multi-million dollar) fines against Google and Facebook.[38] The impact of European regulation is so pronounced that in some areas, such as privacy, the platforms have decided to adopt the European regulation around the world. Some have pointed to this as an example of the Brussels effect[39] – the power of Europe, given the size of its market, to force a race to the top (or bottom, depending on your point of view) in areas of internet regulation. For now, because Europe-wide regulation of disinformation or hate speech has not yet emerged, the platforms have not had to decide whether such rules would have worldwide impact or whether geo-fencing of content to European consumers is preferable. But such moves might be on the horizon. (One should also note that the increased regulation of platforms in Europe likely has the effect of hurting startup platforms more, given that they do not have the resources to comply with many such regulations.)

Highlighting the difference between the European Union and its members also points to the potential role of international organizations in "regulating" or at least establishing norms and best practices for both platform and national regulation of the internet. Several have suggested, for example, that the International Covenant on Civil and Political Rights should inform platform

---

[35] 47 U.S.C.A. § 230 .
[36] *Network Enforcement Act*, GERMAN LAW ARCHIVE (Sept. 1, 2017), *available at* https://germanlawarchive.iuscomp.org/?p=1245.
[37] *Russian bill is copy-and-paste of Germany's hate speech law*, REPORTERS WITHOUT BORDERS (July 19, 2017), https://rsf.org/en/news/russian-bill-copy-and-paste-germanys-hate-speech-law.
[38] Ian Bogost, *Europe's Smack to Google May Only Be the Beginning*, THE ATLANTIC (July 18, 2018).
[39] Anu Bradford, The Brussels Effect, 107 Nw. U. L. Rev. 1 (2012).

regulation of speech.[40] As with statements of constitutional rights, in general, it is far from clear whether the Covenant or similar international agreements, such as the U.N. Declaration of Human Rights, are sufficiently precise to assist in concrete questions, such as how and when Facebook should downrank misinformation or what the bright lines should be with regard to hate speech. Nevertheless, given the need for regional or international consistency in the treatment of similar speech on a platform that extends beyond national borders, this may be an area where multi-state cooperation can play a role.

B. Platform Self-Regulation

For the most part and in most countries, the major internet platforms enjoy a large degree of autonomy to decide what speech to permit and how it should be presented on line. In considering the effect of certain technology companies' influence on democracy, however, what sets platforms apart from a run-of-the-mill website is the capacity to influence and structure political conversation on a national or international scale. In areas where a large share of the population primarily gets its news from online sources, the decisions that platforms make as to what speech is allowed and how it shall be organized can often determine the flow of information critical to politics and elections.

As a result, the platforms' terms of service and community guidelines in such regions can be as important, if not more so, than formal law in determining the boundaries of political conversations. How they define hate speech and incitement, whether (and how) they take action against disinformation, and what types of advertising services they offer to political actors provide a structure for online messaging and political competition. Especially in countries without sophisticated enforcement schemes (or even rules) for campaign finance or campaign-related speech, the platform's rules fill a legal void.

Although the web is often portrayed as a state of nature for political speech, the platforms are highly regulated environments. Most of the major platforms have rules governing nudity and obscenity, harmful and violent content, harassment, threats, bullying, impersonation, and hate speech, as well as policies against spamming or copyright violations.[41] They take down millions of pieces of content each year. Most such rules from the platforms go well beyond what is required by national laws. Indeed, if such rules were legislated by the government in the United States, almost all would be declared unconstitutional by the courts.

---

[40] *See* Eilieen Donahoe, *So Software Has Eaten the World: What Does It Mean for Human Rights, Security & Governance?*, JUST SECURITY (Mar. 18, 2016), https://www.justsecurity.org/30046/software-eaten-world-human-rights-security-governance/; Evelyn Aswad, *The Future of Freedom of Expression Online* (August 15, 2018), DUKE L. & TECH. R., Forthcoming, available at https://ssrn.com/abstract=3250950.

[41] *See, e.g., Policies and Safety*, YOUTUBE (2018), https://www.youtube.com/yt/about/policies/#community-guidelines.

For the most part, the criticism of the platforms in the last two years comes from those who believe that they have done too little to address speech that undermines democracy, although some worry about the costs to speech about them doing too much. Polarization, hate speech, disinformation, foreign intervention, fraudulent advertising, and computational propaganda (bots) are on the list of dangerous speech that governments and critics argue should be confronted. And since the 2016 U.S. Election, the platforms have aggressively experimented with a number of policy changes to address these phenomena. As discussed in greater detail later, they have removed fake accounts, demoted false or polarizing content, moved toward greater transparency for political advertising, required greater disclosure in certain contexts, deprived fake news sites of advertising dollars, and tried to use machine learning to identify threats before they materialize. The criticisms rightfully continue, but as the low hanging fruit has been picked, proposals for self-regulation to address these dangers to democracy often turn more specifically to removing speech from platforms. In some contexts, as with the German NetzDG law, it comes in the form of forced self-regulation – that is, requiring platforms to take down certain legally defined categories of speech.

Given the way critics and governments talk about the influence of "the platforms" on democracy, you might think that everyone agrees as to which companies fall within that category. Any such definition begins with Google and Facebook (and their subsidiaries), of course, but after them it becomes someone challenging to fill out the rest. To be more precise, the relevant category depends on which democracy-related problem one seeks to address. If one is focused on social media, in fact, then Google is excluded. The search engine may be a powerful force in delivering information, but Google is not a social media company. If the problem is *social* media, then Twitter would certainly be included, and perhaps LinkedIn. But what about smaller platforms such as Reddit, 4chan, and 8chan, or similar platforms predominant in Asia, such as Line, Kakao Talk, or WeChat? The latter group of sites is often accused of being a repository for hate speech, disinformation, and conspiracy theories. But because their reach and power is not comparable to major platforms, they rarely are included among the chief offenders. However, as governments consider regulation of "platforms," depending on how such platforms are defined, any regulation could sweep up these smaller sites, as well as startups trying to break through. Moreover, if size, power or potential monopoly position is the touchstone, should Apple, Microsoft and Amazon be included, let alone traditional telecom firms or media companies?

The categorization exercise is important because it forces one to focus on which types of problems are prominent on which types of platforms, and how to address them. A search engine presents different challenges and opportunities than a newsfeed or a messaging application, for example. Indeed, even within platforms, only certain products may be the locus of a particular type of problem.[42] For example, outside of YouTube, Google cannot "take down" content from the web. Rather, if it wishes to address dangerous content reached through its search engine, it

[42] Eileen Donahoe, *Don't Undermine Democratic Values in the Name of Democracy,* AMERICAN INTEREST (Dec. 12, 2017), https://www.the-american-interest.com/2017/12/12/179079/.

needs to alter the algorithm so that users are not directed to it. In contrast, Facebook and Twitter have the capacity to take down accounts or delete content from their platforms, as well as demote content so that it is less likely to appear in someone's newsfeed. However, on an encrypted service like WhatsApp, which serves as a messaging device, social media platform of sorts with WhatsApp groups, and functionally as a telephone, the firm may be unaware of the scale and source of the dangerous speech and have fewer tools to address it.

Another reason to focus on the question as to which firms have a special obligation to address the democracy-harming effects of their platforms concerns recent proposals to form a tech consortium focused on common challenges related to content policy and perhaps, threats to elections and democracy. Many different models have been proposed, such as the Motion Picture Association of America (MPAA), British Press Councils, or the Financial Industry Regulatory Authority (FINRA). If such a consortium were to emerge (assuming it could do so consistent with applicable antitrust laws), particularly to offer common standards on self-regulation, who should be included?[43] Could a common set of standards be developed for search engines, video services, messaging apps, and social media companies? Given that only a few companies (namely, Facebook, Google, and Twitter) have received the brunt of the criticism, would other companies have an interest in joining them? (Indeed, even within that group, there is good reason for one company to let the other become the "face of fake news," for example.) And if there are really only two or three platforms of concern, perhaps a consortium is not really necessary, but rather policy should focus on those few firms themselves.

To be clear, the platforms do cooperate in certain contexts. Child endangerment and terrorist recruitment are the most well-known examples. In those domains, the major platforms share information about emerging threats and dangerous actors. The Global Internet Forum to Combat Terrorism is a coalition between Facebook, Microsoft, Twitter, and YouTube dedicated to "leveraging technology, conducting research on patterns of radicalization and misuse of online platforms, and sharing best practices to accelerate our joint efforts against dangerous radicalization." [44] They have also begun, in an informal way, to start exchanging information on efforts by foreign actors to manipulate elections. The platforms could do more, however, especially if they harmonized their policies toward hate speech and disinformation, particularly as they pertain to "watchlists" for known bad actors. However, when it comes to content moderation policies, a consortium like this runs the risk of determining speech rules not only for the United States in which the platforms are headquartered, but also for political debate around the world. Moreover, if a public-private partnership or system of co-regulation were to emerge between these U.S. companies and the U.S. government (akin to FINRA, above), other countries would necessarily feel left out. Yet, at the same time, it is difficult to see how over one hundred such

[43] To be clear, if such a consortium is going to go beyond content moderation toward other issues, such as privacy or data protection, then a whole host of other companies should be included, from Amazon and Apple to cell phone companies, banks and other large holders of data.

[44] Kent Walker, *Working together to combat terrorists online,* GOOGLE (Sept. 20, 20170, https://www.blog.google/outreach-initiatives/public-policy/working-together-combat-terrorists-online/.

partnerships could emerge to tailor the speech regulations and adjudication to the needs of individual countries.

C. Civil Society and Consumers

For the most part, the fight for the future of the internet – and the rules for online engagement over politics – will take place between governments and platforms. However, "the rest of us" are not completely powerless in the face of the democratic stresses due to technological developments. Outsiders can use both traditional modes of pressure toward corporations (and governments), as well as tools uniquely suited for the digital age. Moreover, given that the digital harms related to democracy afflict citizens – indeed, in their capacities as citizens – they have a role to play, too, independent of governments and platforms.

First, as with any other social ill to which corporations contribute and governments might ignore, consumers can use their economic and organizational clout to pressure and shame bad actors. The same tactics of lobbying, shaming, and boycotting that consumer groups use to target oil, tobacco, or financial firms could be used against the internet companies. Movements to "delete accounts" come and go with little success to date,[45] in part because such accounts have become increasingly indispensable to daily life. But pressure on the tech firms from the media and an array of interest groups has reached a fever pitch in recent years, and they certainly have responded to it.

Pressure is both felt by and comes from the employees themselves at these firms, as well. In the wake of the 2016 U.S. Election, Facebook employees notoriously met to complain about the company's role in contributing to the disinformation in that election. In recent months, Google employees have similarly worried and blown whistles on their companies' planned accommodation of censorship in China.[46] Moreover, it is not uncommon for conservative voices inside these firms to complain to the press about ideological bias or to leak evidence of censorship (as famously happened with respect to Facebook's Trending News feature in 2015[47]), which later becomes fodder for arguments leveled by political elites. Of course, employees can vote with their feet as well, and complaints about corporate culture (let alone politics) are a frequent cause for employee exits in Silicon Valley.

The citizen's responsibility for protecting democracy from online threats extends beyond threatening and even influencing the firms themselves, however. Social media gains its force and

---

[45] Jack Nicas, *They Tried to Boycott Facebook, Apple and Google. They Failed.*, NYTIMES (Apr. 1, 2018), https://www.nytimes.com/2018/04/01/business/boycott-facebook-apple-google-failed.html.

[46] Ryan Gallagher, *Google Plans to Launch Censored Search Engine in China, Leaked Documents Reveal*, THE INTERCEPT (Aug. 1, 2018), https://theintercept.com/2018/08/01/google-china-search-engine-censorship/.

[47] Michael Nunez, *Former Facebook Workers: We Routinely Suppressed Conservative News*, GIZMODO (May 9, 2016), https://theintercept.com/2018/08/01/google-china-search-engine-censorship/.

magnifies its dangers to democracy through the repeated forwarding of content by consumers. If the users of the platforms collectively stood up against disinformation and hate speech, those problems might not be eliminated, but they would be significantly reduced. A Pew Research Center poll shows that roughly 25 percent of Americans admit to forwarding fake news.[48] The fight against disinformation begins at home, as it were, with users refusing to participate in the viral game of forwarding the kind of speech that destabilizes democratic norms.

Of course, sudden changes in mass behavior on the scale necessary here are not often realized, but recognizing citizen responsibility turns the lens back on users to open up opportunities for intervention. Enhancing digital literacy, discussed below, represents one popular category of reforms. The intelligence community speaks of building resilience,[49] specifically to dangerous narratives pushed by foreign actors, but the logic potentially extends to all kinds of online speech and activity that could harm the national interest. New norms of healthy social media use should be developed and pushed by all stakeholders with an interest in promoting the upside and reducing the democratic downside of social media.

Finally, new technologies can empower users to take action on their own screens to mitigate the dangers to democracy coming from internet communication. A series of apps, browser extensions, and programs have been developed to assist users who worry about the information they are receiving from online sources. For example, tools are now widely available to detect whether an account is a bot or not.[50] Other tools also attempt to deal with homophily, by showing users the political bias in their newsfeeds and what a more balanced feed might look like.[51] Finally, a great number of institutions have sprung up to detect, for example, Russian social media intelligence activity and to disclose what types of stories those websites are promoting.[52]

One can think of consumer activity of this ilk as trying to get at the "demand" side of the internet speech economy. Government and platform regulation tend to go after the "supply" of problematic content. But nothing will change if the market for fake news or hate speech remains robust due to consumer demand. The prohibited speech will simply move from platform to platform until it reaches the susceptible user. Especially as such online speech moves toward encrypted platforms, there will be very little that either the government or the platforms can do. Users will ultimately be responsible for the content they share and consume.

---

[48] Michael Barthel, Amy Mitchell, and Jesse Holcolm, *Many Americans Believe Fake News Is Sowing Confusion*, PEW RESEARCH CENTER (Dec. 15, 2016), http://www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/.

[49] *See Can Public Diplomacy Survive the Internet?*, U.S. DEP'T. OF STATE (edited by S. Powers and M. Kounalakis) (May 2017), https://www.state.gov/documents/organization/271028.pdf.

[50] *See Botometer*, https://botometer.iuni.iu.edu/#!/ (last visited Oct. 30, 2018).

[51] *See Blue Feed, Red Feed*, WALL STREET JOURNAL, http://graphics.wsj.com/blue-feed-red-feed/ (last visited Oct. 30, 2018).

[52] *See HAMILTON68,* GMF: ALLIANCE FOR SECURING DEMOCRACY, https://dashboard.securingdemocracy.org/ (last visited Oct. 30, 2018).

III. Categories of Reform: The Seven "D"s

Reforms to address "democracy endangering" speech online can take many forms. At their essence, most of them are, in fact, regulations of speech: that is, they involve preventing, removing, altering, or punishing the communication deemed to be dangerous. Reforms, such as those described here, can be imposed by government or the platforms, and in some cases may inspire innovations from the outside (as, for example, with digital literacy or bot-detection programs). To be clear, many of these could also be imposed by authoritarian governments seeking to squelch online speech. As such reforms – initiated either by democratic governments or the platforms – become popular, we should expect more authoritarian governments to push for similar measures that might take a more extreme form. To that end, to the extent some of these measures require machine learning to identify and minimize certain categories of speech, we should not expect that once invented, the artificial intelligence used to identify and prevent one category of speech seen as dangerous to democracies might not be used also against regime-threatening speech, in general.

A. Deletion

Censorship is the least ambiguous and most direct form of speech regulation, of course. All societies (democratic or authoritarian) ban certain types of speech – such as incitement, threats, blackmail, obscenity, fraud, and libel. Online speech that runs afoul of these prohibitions is similarly regulated. But it may be more difficult for the government to enforce these speech regulations online, given the protection for anonymity and the fuzziness of sovereignty. However, all the major internet platforms also follow suit and usually go beyond what the formal law requires in several of these areas.

Indeed, if the U.S. government were to legislate the community guidelines or terms of service of the major platforms, almost all such policies as they currently exist would be deemed unconstitutional under the free speech protections of the First Amendment. Most hate speech is constitutionally protected in the United States, for example. However, to take one typical firm's statement of the category, YouTube defines hate speech as content that "promotes violence against or has the primary purpose of inciting hatred against individuals or groups based on certain attributes, such as: race or ethnic origin, religion, disability, gender, age, veteran status, sexual orientation/gender identity." Inciting hatred against such a broad array of groups would be an impermissibly overbroad standard under the U.S. Constitution. Similarly, Facebook's presumptive prohibition on depictions of nudity goes well beyond the bounds of what would be a permissible law regulating obscenity.

The fact that the terms of service and community guidelines of the major internet platforms go beyond what is required by legislation or permitted by a country's constitution is, in

itself, unremarkable. These are private companies, and like other websites, they have the capacity and freedom to determine the boundaries of speech that occurs on their sites. No one would plausibly suggest that websites, in general, must obey the same strictures as governments. Doing so would itself seriously constrict free expression, as partisan websites would then need to be viewpoint neutral and online speakers would be less able to set up portals with a particular point of view.

Of course, Facebook and Google/YouTube are not just another pair of websites. Arguably, they are the modern public square.[53] Their decisions as to what speech to allow on their sites and the procedures used for takedowns and appeals are as important, if not moreso, as the formal legal rules enacted by governments. When political bias taints their removal decisions, it skews the free flow of information to the citizenry.

As such, these platforms arguably incur certain "state-like" responsibilities when it comes to speech that occurs on their platforms. What these responsibilities entail, however, is far from clear and requires further thinking. No one argues they should be powerless to allow any firm to advertise any goods on their sites, for example. And surely they can be more restrictive than the state when it comes to obscenity or maintaining a certain level of decorum. Moreover, because filtering systems must be done algorithmically at scale often with the benefit of machine-learning, they do not (and cannot) take the form of actual "law." Perhaps more specifically, the general guidelines that appear in community standards and terms of service may be capable of human definition, but the algorithmic decisions that automatically block or prioritize certain content, given that they may be based on an evolving training set, may not be expressible in ordinary language.

Finally, precisely because platforms have greater capacity and flexibility to regulate public debate and the speech environment, governments turn to the platforms to regulate speech that the state often cannot. In other words, governments are quick to offload to the platforms the politically sensitive and complicated decisions over what online speech to permit. Doing so reserves to politicians the right to complain about the political bias of the platforms, as well as to blame them for dangerous speech that slips through (intentionally or otherwise). A bureaucratic architecture able to adjudicate and respond to dangerous online speech in near-real time would require the supervision of the internet seen in authoritarian regimes. As such, legal formulations along the lines of the German NetzDG law that make platforms more responsible for speech that occurs on their sites or requires quick takedown of speech once notified are becoming increasingly popular.

Also, it should be noted that the "deletion" power described here, exists beyond platforms. That is, several different components of the internet architecture are in a position to delete or prevent content from reaching users. Those would include:

---

[53] Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 Harv. L. Rev. 1598, 1611 (2018).

- Platforms (e.g., Facebook, WordPress, etc.), where the content is published.
- Hosts (e.g., Amazon Web Services, DreamHost, etc.), that provide infrastructure on which the platforms live.
- Transit Providers (e.g., Level(3), NTT, etc.), that connect the hosts to the rest of the Internet.
- Reverse Proxies/CDNs (e.g., Akamai, Cloudflare, etc.), that provide networks to ensure content loads fast and is protected from attack.
- Authoritative DNS Providers (e.g., Dyn, Cloudflare, etc.), that resolve the domains of sites.
- Registrars (e.g., GoDaddy, Tucows, etc.), that register the domains of sites.
- Registries (e.g., Verisign, Afilias, etc.), that run the top level domains like .com, .org, etc.
- Internet Service Providers (ISPs) (e.g., Comcast, AT&T, etc.), that connect content consumers to the Internet.
- Recursive DNS Providers (e.g., OpenDNS, Google, etc.), that resolve content consumers' DNS queries.
- Browsers (e.g., Firefox, Chrome, etc.), that parse and organize Internet content into a consumable form.
- Search engines (e.g., Google, Bing, etc.), that help you discover content.
- RIRs (e.g., ARIN, RIPE, APNIC, etc.), which provide the IP addresses used by Internet infrastructure.[54]

At many nodes in the network that forms the Internet, different choke points have the capacity to make decisions as to which content can make it through. For example, when Cloudflare, a content delivery network that handles 10% of Internet requests, removed the Nazi website, *Daily Stormer*, from its service, it effectively made the site inaccessible for a period of time.[55] It predictably received criticism from free speech advocates, who argued about the line that should exist between impermissible and permissible websites for the service.

In addition to threatening the platforms, authoritarian governments sometimes exploit eachof these choke points to control the transmission of content on line. Platforms, such as YouTube and Facebook, may be in the best position to monitor content on their services, especially given their role in prioritizing and targeting certain content to users. However, if certain speech

[54] Matthew Prince, *Why We Terminated Daily Stormer*, CLOUDFLARE (Aug. 16, 2017), https://blog.cloudflare.com/why-we-terminated-daily-stormer/
[55] *Id*.

and speakers can be identified with precision, it can be choked off in many different places on the Internet.

B. Demotion

One of the reasons platforms, despite their size and power, cannot be perfectly analogized to states is that they do not merely host content, they prioritize it. Tempting as the analogy to a public square might be, it falls apart when one dives into the details of what these platforms actually do. They do not merely provide a forum, like the town square, upon which all speakers can engage on a first-come-first-serve basis. They inevitably make decisions about what content comes first and what content comes last. They serve content to their users; they do not merely host it.

The choices platforms make as to the relative priority of certain types of content are, in many respects, more important than the decisions as to what content to take down. The algorithms that determine these priorities are not value neutral. Sometimes the business interests of the platform may take precedence, as for example when it privileges advertising or content more likely to keep users on the site. At other times, popularity might be prioritized, in which case virality becomes an important ingredient as to which content more users have a greater probability of seeing. At still other times, the priority of content, say in the Facebook newsfeed, may vary based on where a user logs on or how good the mobile internet connection is.

Demotion remains a powerful tool for platforms to address problematic content without taking the more extreme step of deleting it from the site. Signals from users or other sources can provide information about certain communications that then factor into the algorithm so as to minimize the reach of the problematic content. For example, Facebook has taken the step of prioritizing forwarded content with which a user has engaged over other content as to which the user has only read the blurb that appears in the newsfeed. In other words, to combat virality and clickbait headlines, Facebook favors forwarded content that someone has actually read, as opposed to just a link with a catchy title that might provoke knee-jerk forwarding. Similarly, when factcheckers have determined a piece of content to be false, Facebook keeps the content on the site (albeit with related and contradictory articles next to it). However, the false content is demoted so that its reach is reduced by eighty percent.[56] These are just two of the many changes to the newsfeed algorithm in the past two years intended to prioritize "healthier" over problematic content.[57]

[56] Tessa Lyons, *Hard Questions: What's Facebook's Strategy for Stopping False News?*, FACEBOOK (May 23, 2018), https://newsroom.fb.com/news/2018/05/hard-questions-false-news/.

[57] Others include prioritizing sources that have received high marks according to "trust" surveys that Facebook gives to its users, as well as a general preference for content from friends and family, as opposed to news sources.

The Google search engine, likewise, prioritizes certain results so as to surface content that might be more informative rather than more relevant. Google has as its mission to "[o]rganize the world's information and make it universally accessible and useful."[58] For the most part, the search engine returns results for a search query that most closely match the information the user is likely seeking. At times, however, Google must "choose" between returning results the user likely wants to see and those that Google determines might be "best" for them.[59] The most notable instance concerns Google News' prioritization of "authoritative content" during crisis situations. As the head of Google News put it: "To reduce the visibility of this type of content during crisis or breaking news events, we've improved our systems to put more emphasis on authoritative results over factors like freshness or relevancy."[60] Among other factors to judge authoritativeness, Google relies on eight factors developed by the "Trust Project":

- **Best Practices:** What are the news outlet's standards? Who funds it? What is the outlet's mission? Plus commitments to ethics, diverse voices, accuracy, making corrections and other standards.
- **Author/Reporter Expertise:** Who made this? Details about the journalist, including their expertise and other stories they have worked on.
- **Type of Work:** What is this? Labels to distinguish opinion, analysis and advertiser (or sponsored) content from news reports.
- **Citations and References:** What's the source? For investigative or in-depth stories, access to the sources behind the facts and assertions.
- **Methods:** How was it built? Also for in-depth stories, information about why reporters chose to pursue a story and how they went about the process.
- **Locally Sourced?** Was the reporting done on the scene, with deep knowledge about the local situation or community? Lets you know when the story has local origin or expertise.
- **Diverse Voices:** What are the newsroom's efforts and commitments to bringing in diverse perspectives? Readers noticed when certain voices, ethnicities, or political persuasions were missing.
- **Actionable Feedback:** Can we participate? A newsroom's efforts to engage the public's help in setting coverage priorities, contributing to the reporting process, ensuring accuracy and other areas. Readers want to participate and provide feedback that might alter or expand a story.[61]

Demotion and prioritization are not merely ancillary features of search results and newsfeeds. They are designed precisely to create a hierarchy that favors some communication over others. When it comes to the kinds of speech that undermine democracy, then, the question becomes which signals from content or sources indicate some democratic danger such that the

[58] *Our mission*, GOOGLE, https://www.google.com/search/howsearchworks/mission/ (last visited Nov. 11, 2018).
[59] *Webmaster Guidelines,* GOOGLE, https://support.google.com/webmasters/answer/35769 (last visited Nov. 11, 2018).
[60] Richard Gingras, *Elevating quality journalism on the open web,* GOOGLE (Mar. 20, 2018), https://blog.google/outreach-initiatives/google-news-initiative/elevating-quality-journalism/.
[61] *Frequently Asked Questions: What is a Trust Indicator,* THE TRUST PROJECT, https://thetrustproject.org/faq/#indicator (last visited Nov. 11, 2018).

algorithm should minimize their reach. The lack of transparency that is essential to these algorithms functioning – that is, so that they cannot be gamed by strategic actors – is one reason why these strategies are often more effective and less notorious than overt filtering or takedowns.

C. Disclosure

If online anonymity is the cause of many of the democracy-related ills of social media, then disclosure might be the best disinfectant. Disclosure can take many forms, though. It could refer to generalized transparency for all sorts of features and business decisions of the platforms, such as the results of specific takedown requests, the ingredients of an algorithm, or even the privacy policy of a website. For the most part, though, when we think of disclosure as a measure to address dangerous online speech, we refer to the provision of additional cues alongside information so that the user can better evaluate the character and source of the communication.

One of the distinctive features of social media platforms is their homogenous packaging of very different types of information. On Facebook and Twitter, for example, a picture from a friend, a *Breitbart* article, an advertisement, a late-night comedy video, and a *New York Times* editorial are all presented in roughly the same way. They each have a blurb, usually a picture, and then a link to click through. As a result, many of the cues we have in the offline world as to veracity and progeny are stripped away as information is reorganized and repackaged in a particular, uniform format. For example, if one were to approach a supermarket checkout counter and see publications talking about crazy political conspiracies, one would discount them as tabloid fiction, because one knows from experience what kinds of publications end up next to the checkout counter and what types of stories those publications concoct.[62] However, if the same conspiracy story is fed to users over Facebook or Twitter, it comes alongside legitimate publications, entertainment, and personal messages. The "packaging" is stripped away and the source and reliability of the information becomes unclear.

Disclosure, in this respect, can supply online cues to make up for the loss that comes from uniform packaging. Additional information or signals can be placed around or within the communication that would help users discount it based on newly supplied knowledge as to its source, author, or character. Because social media and search results necessarily truncate communication for space reasons, disclosure serves principally to counteract the information loss that comes once information available in full elsewhere on the web becomes reformulated into a newsfeed blurb or search result.[63]

---

[62] The same might generally be true of television – we opt into channels with an expectation as to the type of information or entertainment we will find there, and based on the hour, we would expect broadcast stations to deliver "news" at certain times and entertainment at others.

[63] The web itself homogenizes information, as compared to the off line world. The relatively uniform way browsers present information, while much more diverse in its packaging than social media feeds, nevertheless removes some

The platforms have made several changes to provide more information about the source of a communication. To prevent impersonation, Facebook, Twitter, and YouTube place checkmarks next to verified accounts: that is, accounts of "public interest"[64] for which the platform has verified the identity of the account holder. Facebook also now places an "*i*" button next to certain publishers. When users click it, a page appears with more information about the publisher as well as a map of where the link has been shared.[65] In addition, those platforms identify advertisements (to a greater or lesser degree, and sometimes with mixed success) to distinguish paid from organic content. Google also identifies ads at the top of search results as "sponsored" and provides an "*i*" button that, when clicked, explains why the user was targeted with these ads.[66] Since 2016, in the wake of undisclosed Russian-purchased ads in the presidential campaign, both Google and Facebook have also adopted disclosure and disclaimer regimes specifically for political ads.[67]

The platforms have also used disclosure as a tactic to combat false content. Most notably, Facebook has attempted to disclose the results of factchecking alongside false articles. In its first attempt to tackle the problem, Facebook identified false claims with a "DISPUTED" flag. However, Facebook (and independent analysts) then learned that doing so led to greater engagement with the false articles,[68] as well as an erroneous level of trust being attributed to unflagged content, much of which might also be false. Still, factchecks have remained a staple of Facebook's attempt to confront false content, although now Facebook presents related articles that dispute the underlying claim, instead of a flag that might draw attention and greater engagement with the false claim. They also use factchecks to downrank the content so it is less likely to be seen and served to users through the newsfeed algorithm.

Facebook's experience with disputed flags for false stories is a case study in the difficulty of confronting false claims through mere identification as such. Little evidence exists to support the notion that leaving it up to users to reject propositions, once identified as false, will be enough

---

of the locational and tangible bases for heuristics that assist in source attribution and validation. In other words, it is hard for someone to print a newspaper with a wide circulation, but it is relatively easy to set up a webpage with (at least the potential) for worldwide reach.

[64] *See, e.g., About verified accounts,* TWITTER, https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts (last visited Nov. 11, 2018); *How do I request a blue verification badge?,* FACEBOOK, https://www.facebook.com/help/1288173394636262 (last visited Nov. 11, 2018).

[65] Taylor Hughes, Jeff Smith, and Alex Leavitt, *Helping People Better Assess the Stories They See in News Feed with the Context Button,* FACEBOOK (April 3, 2018), https://newsroom.fb.com/news/2018/04/news-feed-fyi-more-context/.

[66] *Why you're seeing an ad*, GOOGLE, https://support.google.com/ads/answer/1634057?hl=en (last visited Nov. 11, 2018).

[67] *What is the Facebook Ads Archive and how do I search it?,* FACEBOOK, https://www.facebook.com/help/259468828226154?helpref=faq_content (last visited Nov. 11, 2018); *Political advertising on Google,* GOOGLE, https://transparencyreport.google.com/political-ads/overview (last visited Nov. 11, 2018).

[68] Tessa Lyons, *Replacing Disputed Flags With Related Articles,* FACEBOOK (Dec. 20, 2017), https://newsroom.fb.com/news/2017/12/news-feed-fyi-updates-in-our-fight-against-misinformation/; Alexander Burgoyne and David Hambrick, *Flagging Fake News or Bad Sources Won't Work*, SLATE (Jan. 12, 2017), https://slate.com/technology/2017/01/educating-people-about-sources-wont-stop-fake-news.html.

to shake their belief in the false content. All the more so is this true if evaluation of the asserted claim requires a user to click a button, such as the "i" button to get more information about it. Most people come to the internet and social media for social reasons; newsgathering is a subsidiary pursuit. The greater the cognitive burden the platform places on users to investigate the truth of an asserted claim, the less likely are users to do so. Moreover, mere identification – especially when it thereby distinguishes the news item from the homogeneously packaged items nearby – only draws attention to the highlighted content, without successfully convincing the user that the content is otherwise dangerous or of low value.

D. Delay

If the privileging of viral communication is the distinctive democracy-endangering feature of the internet, then adding friction to the viral transmission of information could constitute one step toward a solution. Friction could be added in many different ways. All such measures, however, slow down the forwarding of problematic content (or perhaps all content) to put the brakes on peer-to-peer transmission of information.

Krishna Barat, the founder of Google News, has proposed a series of steps to tamp down on virality.[69] The first critical step involves detection of stories that reach a certain level of popularity over a certain period of time. He analogizes this to a wave detection system in the ocean that warns of a tsunami forming far away from shore. This detection could be done algorithmically as the program detects common traits among new stories ricocheting across the internet. The traits Barat describes would include:

1. Is the wave on a topic that is politically charged? Does it match a set of hot button keywords that seem to attract partisan dialog?
2. Is engagement growing rapidly? How many views or shares per hour?
3. Does it contain newly minted sources or sources with domains that have been transferred?
4. Are there sources with a history of credible journalism? What's the ratio of news output to red flags?
5. Are there questionable sources in the wave
6. Sources flagged for fake news by fact checking sites (e.g., Snopes, Politifact)
7. Sources frequently co-cited on social feeds with known fake news sources.
8. Sources that bear a resemblance to known providers of fake news in their affiliation, web site structure, DNS record, etc.

[69] Krishna Bharat, *How to Detect Fake News in Real-Time,* MEDIUM:NEWCO SHIFT (Apr. 27, 2017), https://medium.com/newco/how-to-detect-fake-news-in-real-time-9fdae0197bfd.

9. Is it being shared by users or featured on forums that have historically forwarded fake news? Are known trolls or conspiracy theorists propagating it?
10. Are there credible news sites in the set? As time passes this becomes a powerful signal. A growing story that does not get picked up by credible sources is suspicious.
11. Have some of the articles been flagged as false by (credible) users?

Just because a story or video has these traits, however, does not mean it is necessarily false or dangerous. Rather, these traits serve as a trigger for human review and for a pause in retransmission. Human review should only take a few hours – or at any rate, less than a day – to verify the story or evaluate the potential danger. A disclosure regime that reveals (well after the fact) which stories were subject to this early warning system could prevent abuse and bias by the platforms.

Other types of friction can also slow down viral transmission of disinformation. For example, both Twitter and Facebook can make it more difficult to quote another person's post or content. To the extent that "likes" also lead to viral transmission, their algorithms can be less responsive to content that receives a lot of likes or to users who are "serial likers" or "serial forwarders." And of course, limiting the capacity to use automation (i.e., bots) to create the appearance of popularity and to manipulate search engines and news feeds could go a long way to constraining "artificial virality" – that is, virality that is disconnected from actual popularity. Indeed, the state of California recently passed a law that bans the use of bots to influence elections, unless they are designated as such.[70]

E. Dilution and Diversion

In addition to preventing users from seeing "bad" content, platforms, governments and civil society can take measures to overwhelm users with "good" content or at least steer them toward it. As with all the measures discussed up till now, these moves require a determination of what is good and bad content. However, as described in the previous discussion of demotion, the algorithms inevitably make determinations about the relative priority of communication. They "choose" to elevate content with certain properties; the question is whether other values, such as those that support democracy, should also be included in the mix.

"Dilution" refers to alterations of the "mix" of good and bad content, with the goal of overwhelming bad content so as to mute its potential effect. Governments with robust institutions of publicly funded journalism are in a favored position to take on such a role. If a country has a

[70] Steven Musil, *California bans bots trying to sway elections,* CNET (Oct. 1, 2018), https://www.cnet.com/news/california-bans-bots-secretly-trying-to-sway-elections/; 2018 Cal. Legis. Serv. Ch. 892 (S.B. 1001), *available at* https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201720180SB1001.

non-partisan, trusted, and popular news source, it has the capacity to confront disinformation with truthful content. Depending on the reach and popularity of state-sponsored news outlets, it can combat disinformation both online and through legacy media. On the other hand, a country like the United States, which has poorly funded public broadcasting and widespread distrust in any official state-sponsored news service, is not well-positioned to engage in state-authorized measures to flood the information zone with truthful content. Nordic countries, however, spend a considerable share of public money on such news services, insulate them from political control, and receive broad support from the public. They can engage in a coordinated attempt to respond especially to foreign efforts to propagandize during election periods.

Of course, the same state-sponsored tools that can be used to combat disinformation can also promote it. Recent evidence points to the increased state use of bots and trolls to target their own citizens with disinformation campaigns.[71] Indeed, China maintains a million-person army – the so-called "50-cent army"—to promote pro-regime sentiment online and to infiltrate groups to steer their conversations away from touchy political subjects.[72] By one account, the Chinese government adds close to 450 million comments per year on social media.[73] Combined with a strict regime of filtering and censorship, this "cheerleading" also serves to distract from collective action efforts to organize against the government. Although China may exist at the extreme end of the continuum, Freedom House reports that over thirty countries now engage in efforts to manipulate public opinion through social media.[74]

The platforms also use distraction and dilution to push users away from bad content. When Facebook's "disputed" news flags proved counterproductive, the platform adopted a different tactic that attempted to counteract false content with "related articles" demonstrating the falsity of the claims.[75] As many as three additional articles (often from factcheckers) provide evidence contradicting the claim in the main article. By attaching the related articles to the false story, Facebook also shrinks the "real estate" on the screen available for the false story. The platform

[71] Carly Nyst and Nick Monaco, *State-Sponsored Trolling,* INSITUTE FOR THE FUTURE (2018), http://www.iftf.org/fileadmin/user_upload/images/DigIntel/IFTF_State_sponsored_trolling_report.pdf
[72] Henry Farrell, *The Chinese government fakes nearly 450 million social media comments a year. This is why.*, WASHINGTON POST (May 19, 2016), https://www.washingtonpost.com/news/monkey-cage/wp/2016/05/19/the-chinese-government-fakes-nearly-450-million-social-media-comments-a-year-this-is-why/?utm_term=.af441bff22b2.
[73] Gary King, Jennifer Pan, and Margaret Roberts, *How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, not Engaged Argument*, AMERICAN POLITICAL SCIENCE REVIEW (Apr. 9, 2017), http://gking.harvard.edu/files/gking/files/50c.pdf?m=1463587807.
[74] *Manipulating Social Media to Undermine Democracy,* FREEDOM HOUSE, https://freedomhouse.org/report/freedom-net/freedom-net-2017 (last visited Nov. 11, 2017).
[75] Sara Su, *New Test With Related Articles,* FACEBOOK (Apr. 25, 2017), https://newsroom.fb.com/news/2017/04/news-feed-fyi-new-test-with-related-articles/.

therefore dilutes the impact of the false story by shrinking it next to others, and also diverts attention to the contradictory claims of related articles.[76]

YouTube has attempted a similar tactic of diversion when it comes to terrorist content. In a project called "The Redirect Method"[77] developed by Jigsaw, YouTube has attempted to redirect those seeking terrorist propaganda to content more likely to deradicalize them. Like any advertising strategy, the Redirect Method seeks to find a target audience and then deliver content that persuades them to "buy into" a different product – in this case, rejection of Islamic terrorism. Jigsaw developed this method after talking with ISIS defectors and experts in terrorist recruitment. The firm first compiled a list of search terms (Adwords targeting) for people who were likely searching for ISIS propaganda.[78] It then curated a library of videos, channels, and playlists that would both lead to high engagement from this selective audience, and also steer them away from radical messages. These were not necessarily anti-ISIS or anti-terrorism videos. Rather, they were videos that the research suggested might reduce the attractiveness of the narratives ISIS promoted.

Both the "Related Articles" and "Redirect Method" seem like "soft touch" interventions to address harmful content. They steer viewers away from the bad to the good. As similar methods expand beyond provably false stories and clear terrorist propaganda, though, they are open to the same charge of manipulation of public opinion that has been lodged against states. Indeed, for this reason, some scholars warn of the "search engine manipulation effect"[79] (or SEME) which refers to the ability of search engines, like Google's, to shift voting preferences among undecided voters because the algorithm and search results favor one candidate over another. To be sure, because newsfeeds and search results necessarily place some content above others, some favoritism seems inevitable. But the more that political variables (including those related to disinformation and polarization) feed into the algorithm, the greater the risk of systematic bias in favor of one party over another.

F. Deterrence

Governments and platforms have a variety of tools at their disposal to punish or deter purveyors of harmful content from gaining an audience. These measures can target the producer of such content both online and offline. Removing content or suspending accounts are the most

---

[76] Leticia Bode and Emily Vraga, *In Related News: hat Was Wrong: The Correction of Misinformation Through Related Stories Functionality in Social Media,* 65 JOURNAL OF COMMUNICATION 4 (June 23, 2015), https://onlinelibrary.wiley.com/doi/abs/10.1111/jcom.12166.

[77] THE REDIRECT METHOD, https://redirectmethod.org/ (last visited Nov. 11, 2018).

[78] *The Redirect Method Blueprint,* THE REDIRECT METHOD, https://redirectmethod.org/blueprint/ (last visited Nov. 11, 2018).

[79] Robert Epstein and Ronald E. Robertson, *The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections*, PNAS (Aug. 18, 2015), http://www.pnas.org/content/112/33/E4512.

obvious ways to target speakers. But other strategies can go after the finances or other sources of power of bad actors on the internet.

To take the most obvious example, the United States government punished Russia for its cyber-meddling in the 2016 U.S. election.[80] The Obama administration imposed sanctions on two Russian intelligence agencies, three companies that supported the election interference, and four individuals. It also expelled 35 Russian officials and shut down Russian outposts in New York and Maryland. These measures are just examples of how governments can use traditional measures of diplomacy and even warfare to go after actors viewed to commit mischief in the online world.

The same applies to domestic actors who use the internet or social media to break the law. The fact that a crime is one that necessarily involves "speech" does not make it outside the scope of government regulation. Plenty of crime is limited to mere speech acts, such as blackmail, threats, fraud, child exploitation or pornography, solicitation, conspiracy, and incitement. The same crimes committed through offline speech are punishable when they occur through the use of a computer and internet connection. In part, this was the approach of the German NetzDG law – to force platforms to take down speech that would otherwise be punishable if it occurred offline.

To be sure, in a recent case, the U.S. Supreme Court emphasized the importance of speech in the digital sphere. In the case of *Packingham v. North Carolina*, 137 S. Ct. 1730 (2017), the Court struck down on free speech grounds a state law that prohibited registered sex offenders from accessing certain websites, including Facebook, to prevent them from having access to children. The Court explained, "While in the past there may have been difficulty in identifying the most important places (in a spatial sense) for the exchange of views, today the answer is clear. It is cyberspace—the 'vast democratic forums of the Internet' in general, . . . and social media in particular."[81] But while the statute, in that case, was overbroad, the Court recognized that "[s]pecific criminal acts are not protected speech even if the speech is the means for their commission."[82] That principle applies to online speech, just as it applies to speech in the physical world.

Besides deleting or demoting content or accounts, the platforms have other tools at their disposal to deter bad actors on line. The case of the Macedonian teenagers in the 2016 U.S. election provides a case in point. As is now well-known, a group of teenagers placed pro-Trump

[80] Missy Ryan, Ellen Nakashima, and Karen DeYoung, *Obama administration announces measures to punish Russia for 2016 election interference*, WASHINGTON POST (Dec. 29, 2016) https://www.washingtonpost.com/world/national-security/obama-administration-announces-measures-to-punish-russia-for-2016-election-interference/2016/12/29/311db9d6-cdde-11e6-a87f-b917067331bb_story.html?utm_term=.5ccecd0aa99a.

[81] Packingham v. North Carolina, 137 S. Ct. 1730, 1735, (2017).

[82] *Id*. at 1737.

fake news websites online during the 2016 campaign.[83] They did so not because they were Trump supporters. In fact, they had launched some pro-Clinton sites as well. They simply realized early on that pro-Trump websites created greater traffic and engagement, which translated into more advertising dollars, as served through the Google and Facebook ad serving services. In the wake of the 2016 election, though, Google and Facebook shut down advertising on those sites, and once drained of revenue, the sites were taken down.

Finally, new research has suggested potentially fruitful ways of using bots to target hate speech and polarization. Research by Kevin Munger[84] and Alexandra Siegel[85] describes an approach of sending targeted, automated messages to people who engage in trolling behavior or promote hateful content online. Munger used this approach against people who tweeted racist or extremely partisan speech and Siegel used it against Arabic-speaking Twitter accounts that engaged in sectarian anti-Shia speech. The scholars altered the race and number of followers of bots and tried different types of counter-speech to try to reduce destructive online behavior. They found some promising results that might provide some hints as to mild sanctions platforms could impose on those who break important norms of behavior online.

G. Digital Literacy

In an age when so much of the human experience takes place online and new risks emerge every day, almost everyone is in favor of expanding digital literacy. What proponents mean by digital literacy is far from uniform, however. As with disclosure advocates, moreover, the drive for digital literacy grows out of an assumption that pathological communication and attitude formation in the internet age grow from a curable ignorance as to the reliability of online information. On this view, internet users simply need the right tools to critically evaluate communication to assess its reliability. In the context of resisting foreign-sponsored disinformation campaigns, intelligence professionals refer to this strategy of building resilience as "inoculation" against information operations.

Those who hold out hope for digital literacy usually focus on incorporating such skill-development in primary school curricula. The Stanford Education Department, for example, has developed materials that can be used by high school teachers to educate students how to read

---

[83] Samanth Subramanian, *Inside the Macedonian Fake-News Complex,* WIRED (Feb. 15, 2017), https://www.wired.com/2017/02/veles-macedonia-fake-news/.

[84] Kevin Munger, *Tweetment Effects on the Tweeted: Experimentally Reducing Racist Harassment*, 39 POLITICAL BEHAVIOR 3 (Sept. 2017), https://link.springer.com/article/10.1007/s11109-016-9373-5; Kevin Munger, *Experimentally Reducing Partisan Incivility on Twitter*, (Working Paper Sept. 7, 2017), http://kmunger.github.io/pdfs/jmp.pdf.

[85] Alexandra Siegel, *Online Hate Speech* (Working Paper Sept. 2018), https://alexandra-siegel.com/wp-content/uploads/2018/09/Siegel_Online_Hate_Speech.pdf.

critically and assess whether stories are reliable and fact-based.[86] The materials also educate students on how to distinguish between advertisements and news – a task that can be quite challenging at a time when "sponsored content" is often designed to blur the difference with actual journalism often placed right next to it. The bottom line for these strategies is to imbue age-old lessons of critical thinking adapted for the new information ecosystem. Facebook itself has developed a Digital Literacy Library "to help young people think critically and share thoughtfully on line."[87]

Governments have begun to heed the call on digital literacy. As a case in point, the Swedish government has approved a program to strengthen "digital competency." A government report on the effort describes it as follows:

> The national curriculum now states that schools have a responsibility to 'contribute to pupils developing an understanding for how digitalisation affects the individual and society's development' and that pupils 'shall be given the possibility to develop a critical and responsible approach to digital technology, in order to be able to see possibilities and understand risks, as well as to be able to rate information'.
>
> However, while the curriculum mentions critical thinking with regards to sources, no dedicated subject has been created for the broader set of knowledge and skills which have been referred to as digital citizenship or digital resilience. This includes traditional critical thinking skills – questioning authorial bias, triangulating data sources, and using information selectively – alongside more specific knowledge about how the internet works and how online content can be manipulated. Topics include identifying fake news, learning about the impact of algorithms in creating echo chambers and what filter bubbles are, and finding out what do to if you encounter hate speech or extremist content online. [88]

---

[86] STANFORD HISTORY EDUCATION GROUP, *Evaluating Information: The Cornerstone of Civic Online Reasoning* (Nov. 22, 2016), https://stacks.stanford.edu/file/druid:fv751yt5934/SHEG%20Evaluating%20Information%20Online.pdf.

[87] Antigone Davis and Karuna Nain, *A New Resource for Educations: Digital Literacy Library,* FACEBOOK (Aug. 2, 2018), https://newsroom.fb.com/news/2018/08/digitalliteracylibrary/; *Digital Literacy Library,* FACEBOOK, https://www.facebook.com/safety/educators (last visited Nov. 11, 2018).

[88] Chloe Colliver, Peter Pomerantsev, Anne Applebaum, and Jonathan Birdwell, *Smearing Sweden" International Influence Campaigns in the 2018 Swedish Election*, MSB (Oct. 2018), https://www.isdglobal.org/wp-content/uploads/2018/10/Sweden_Report_October_2018.pdf.

Digital literacy efforts directed toward the young make sense, given that the government can have its greatest influence on public education. However, emerging research on disinformation suggests that older people, especially those new to the internet, are more susceptible to spreading, consuming and believing false content.[89] Perhaps younger users, who are digital natives, are more experienced, savvy, and skeptical of online content. Or perhaps older users, particularly of Facebook, tend to have fewer "friends" delivering content, such that the demotion algorithms that push down disinformation are less effective for users with a limited inventory of stories. In other words, demotion works well for people with a lot of content potentially in their feed, but for a person with just a few friends and a few stories, they might see the whole universe of stories that their friends are posting and liking. Whatever the reasons for the prevalence of digital misinformation among older users, digital literacy programs need to be directed toward that slice of the population perhaps even more than to younger people in primary schools.

Finally, digital literacy can mean something more than evaluating communication for truth or developing critical thinking and civility skills. The concept could include, as well, skills development surrounding specific platforms and apps. People need to understand the basics about how to change settings, how to report a terms of service violation, how to flag stories to be fact checked, or what a verified account looks like. This may seem like minor skills development as compared to learning critical thinking (and it is). But it can be important in areas of the world where, for example, Facebook essentially is the internet, and it relies heavily on users to report problematic content or violations of the community guidelines. Better understanding as to how content eventually appears on one's own screen and how to regulate it oneself represents the first step toward more sophisticated consumption of on-line information.

IV. Emerging Challenges

Contemporary discussion of the challenges the internet poses for democracy focuses principally on the problems of disinformation and different types of dangerous speech, such as incitement and hate speech. Governments, in turn, consider the different models described above to regulate these categories of online speech explicitly or to direct the major platforms to do the dirty work for them. By all accounts, these different measures have made a dent in the problems (often with collateral damage to other speech), but the relevant adversaries, both foreign and domestic, adapt to and evade each intervention with new strategies.

The online communications environment is evolving rapidly, and with it has come a distinct new set of challenges and others clearly visible on the horizon. Several of these arise from different platforms gaining prominence or new technologies shaping the communication

[89] Andrew Guess, Brendan Nyhan, & Jason Reifler, *Selective Exposure to Misinformation: Evidence from the consumption of fake news during the U.S. presidential campaign*, Jan. 9, 2018, available at https://www.dartmouth.edu/~nyhan/fake-news-2016.pdf.

ecosystem in different ways. Others indicate the rise of new actors and strategies to cause harm or pollute the information environment.

A. Encrypted peer-to-peer platforms

Although plenty of criticism has been leveled at Twitter and Google, Facebook has received the brunt of blame when it comes to election interference and disinformation. However, a new species of platforms is competing with the "big three" when it comes to the perceived spread of disinformation or hate speech. These platforms are almost impossible for the government to regulate effectively. Because they rely on encrypted peer-to-peer messaging, they also pose difficult self-regulatory challenges for the companies that invented them.

WhatsApp is the first among equals when it comes to encrypted peer-to-peer messaging platforms. Although owned by Facebook, WhatsApp is the most popular messaging app in 104 countries[90] and has more than 1.5 billion monthly active users on its own, which is more than Facebook's own messaging service.[91] Especially in the developing world, where data plans are often more expensive, WhatsApp has particular dominance. It is used not only to send messages but also to make voice calls, and of particular importance to democracy and elections, it is used to build groups and communicate among them.

One of the reasons to believe that the disinformation and dangerous speech problems attributed to the existing dominant platforms may not be unique to those technologies is that all of these problems are migrating and even exploding on WhatsApp. Extensive use of WhatsApp to spread false rumors and candidate attacks was reported in the recent elections in Brazil and Mexico, as well as in preparation for the upcoming Indian election.[92] In India, the government even blamed a spate of lynchings on "irresponsible and explosive messages filled with rumours and provocation . . . circulated on WhatsApp."[93] The scale of these problems on WhatsApp or any similar service is difficult to measure, though, because even the company itself cannot assess the reach of any given story on the platform.

Several of the unique features of internet communication that pose challenges for democracy are accentuated on these platforms. Anonymity is not only protected, but with the

---

[90] Joseph Schwartz, *The Most Popular Messaging App in Every Country,* MARKETWATCH (Feb. 2018), https://www.similarweb.com/blog/mobile-messaging-app-map-2018.

[91] *Mobile Internet & Apps,* STATISTA (Oct. 2018), https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/; Josh Constine; *WhatsApp hits 1.5 billion monthly users. $19B? Not so bad.,*TECHCRUNCH (Jan. 31, 2018); https://techcrunch.com/2018/01/31/whatsapp-hits-1-5-billion-monthly-users-19b-not-so-bad/.

[92] Elizabeth Dwoskin and Annie Gown, *On WhatsApp, fake news is fast – and can be fatal,* WASHINGTON POST (July 23, 2018); *Nationalism a driving force behind fake news in India, research shows,* BBC (Nov. 12, 2018), https://www.bbc.com/news/world-46146877.

[93] Press Release, Government of India: Ministry of Electronics & IT, WhatsApp warned for abuse of their platform (July 3, 2018), http://pib.nic.in/newsite/PrintRelease.aspx?relid=180364.

addition of encryption, it is even more difficult to discern the origin of certain stories, rumors or memes. Virality, in particular, seems to be an uncontrollable feature of these platforms, leading WhatsApp, for example, to try to reduce the permissible size of WhatsApp groups and the ability to distribute the same message to multiple groups. Even more than Facebook itself, political WhatsApp groups are by nature homophilous, as people usually opt into them to receive messages from their friends with similar views or political leaders they support. Finally, as WhatsApp dominates many different facets of the telecommunication environment in developing countries, its monopoly position means abuse on the platform has outsized significance, as compared to other countries with a more pluralized information and telecommunications environment.

B. Deep Fakes

In the rush to identify the highest-tech online innovation to threaten democracy, many commentators have focused on so-called "Deep Fakes." Deep Fakes refers to the use of artificial intelligence and image synthesis to create video that appears so real that viewers might mistake it for authentic footage. University researchers and entertainers have demonstrated how to use artificial video techniques to put words in our political leaders' mouths.[94] For those who worry about the impact of "fake news" as a tool of disinformation, artificial video seems like the next, giant leap into an abyss in which we no longer will be able to "trust our lying eyes."

Like disinformation generally, Deep Fakes pose two interrelated problems for democratic discourse and decisionmaking. First, any given deep fake can be used strategically to lie to viewers about a particular act. Artificial videos could portray political leaders in compromising positions or cause them to appear to say something that would damage their credibility or electability. Moreover, Deep Fakes could even fabricate events themselves as creators seek to change the apparent facts on the ground in a war or conflict.[95] Simple code is already available on line to assist users in placing one person's face on another person's body. As with so many internet innovations, the pornography industry has led the way here, enabling celebrity faces to be placed on the naked bodies of movie actresses.[96]

However, the greater danger from artificial video is the decline in trust in video generally. If Deep Fakes become widespread, then confidence in true video footage will decline. Just as the

---

[94] Supasorn Suwajanakorn, *Synthesizing Obama: Learning Lip Sync from Audio*, YOUTUBE (July 11, 2017), https://www.youtube.com/watch?v=MVBe6_o4cMI; James Vincent, *Watch Jordan Peele use AI to make Barack Obama deliver a PSA about fake news*, THE VERGE (Apr. 17, 2018), https://www.theverge.com/tldr/2018/4/17/17247334/ai-fake-news-video-barack-obama-jordan-peele-buzzfeed

[95] Robert Chesney and Danielle Citron, *Deep Fakes: A Looming Crisis for National Security, Democracy and Privacy?*, LAWFARE (Feb. 21, 2018), https://www.lawfareblog.com/deep-fakes-looming-crisis-national-security-democracy-and-privacy.

[96] Samantha Cole, *AI-Assisted Fake Porn Is Here and We're All Fucked,* VICE (Dec. 11, 2017), https://motherboard.vice.com/en_us/article/gydydm/gal-gadot-fake-ai-porn.

high prevalence of false news makes more credible the claim that any given news item is false, so too with video does the prevalence of Deep Fakes bring plausible deniability to the truth that any given video is real. For example, when President Trump suggested (several months after its release) that the Access Hollywood video was fake, that lie was easily contradicted by both the video itself and the claims of others who were featured in it. But in a world where public figures are frequently denying the veracity of video, sometimes with good cause because they are subject to Deep Fakes, these types of denials will be believed by viewers looking for a reason to deny the truth of what they see on the screen before them.

Despite the attractiveness of Deep Fakes as the "shiny new object" in the disinformation wars, most people in the industry warn that shallow fakes – or garden variety manipulation of still images – pose a greater threat for the time being. Successful Deep Fakes are time-intensive to create and relatively difficult to escape undetected. More prevalent are conventional alterations of video, audio, and images. Most recently, for example, the White House Press Secretary tweeted an altered (and selectively sped-up) video of a CNN reporter, that misrepresented him as quickly chopping his arm on a female White House staffer attempting to take his microphone away.[97] For alterations like that, no sophisticated artificial intelligence is required. The same can be said for the many images that are cropped or taken out of context (as when an old image is repurposed for a new crisis) to misrepresent underlying facts. Such images themselves are not really "fake" at all. As with disinformation generally, they are selectively altered so as to mislead viewers into believing something occurred that actually had not.

Although Deep Fakes might not at present create an existential threat to the information ecosystem, we are at the beginning of a technological arms race between the creators of Deep Fakes and those that hope to detect them. For the immediate future, different tools and video libraries can be developed to bolster our ability to detect Deep Fakes. However, the time will soon arrive – perhaps in the next five years or so, according to experts – when it will be impossible to distinguish between real and artificial video. At that point, it will become especially important that nonpartisan sources of news be in a position to "vouch" for the video footage they present and that viewers can trust.

The challenge Deep Fakes pose to confidence in video reporting is emblematic of a more general problem on the horizon concerning technology's blending of the offline and online worlds. Quite apart from news and journalism, the rise of virtual and augmented reality breaks down old categories as to what is real and what is artificial. As those technologies gain prominence in the coming decades, we will become accustomed to experiences that are, in whole or part, man-made

[97] Paul Farhi, *Sarah Sanders promotes an altered video of CNN reporter, sparking allegations of visual propaganda*, WASHINGTON POST (Nov. 8, 2018), https://www.washingtonpost.com/lifestyle/style/sarah-sanders-promotes-an-altered-video-of-cnn-reporter-sparking-allegations-of-visual-propaganda/2018/11/08/33210126-e375-11e8-b759-3d88a5ce9e19_story.html?utm_term=.f34865fc479c; Charlie Warzel, *People Are Arguing About Whether This Trump Press Conference Video is Doctored,* BUZZFEED (Nov. 8, 2018), https://www.buzzfeednews.com/article/charliewarzel/acosta-video-trump-cnn-aide-sarah-sanders?bftwnews&utm_term=4ldqpgc#4ldqpgc.

but seem "real." With augmented reality, we will begin to have information superimposed on our everyday observations, with the use of technologies like Google Glass, which allows for computer generated messages to be integrated into our field of vision on an eyeglass-like device. As for virtual reality, the more time we spend in a world thoroughly constructed for us, the less discerning we might become between our experiences in such a world with ones on the outside. The lack of trust we may begin to have in our own senses to determine what is real will, in part, be a function of how much of our lived-experience takes place in a world free of computerized alteration. Although these transformations are far beyond the horizon, they portend a whole new set of challenges for the consumption and trust of information relevant to democracy and elections.

C. Home Assistants, Wearables, and the Internet of Things

The danger that the online information monopolies pose for democracy arises from their powerful ability to determine what a large share of a country's population sees and believes. The Google search engine provides a definitive list of answers to questions, or at least suggested places to find those answers. Facebook organizes interpersonal communication so as to prioritize information for close to two billion people. The platforms' monopoly status varies by country, but their power comes from the eyeballs they attract to their sites and apps and from the impact that their algorithms have on the kind of information to which they expose users.

As we move away from our screens toward technological interfaces that provide a single answer to user queries, the power of a platform monopoly to organize information can grow even further. In particular, home assistants, such as Google Home, Alexa, and Siri, go even beyond a search engine. Their "voices" respond to questions with a single answer, rather than a few dozen suggested blue links. As important as the first Google search result or the top story in a newsfeed might be, at least in those environments, any given algorithmically generated suggestion occurs amongst a group of other similar suggestions.

Not so with the voice assistants. People interact with them like they do with other people: asking questions and expecting to receive a single answer. As a result, the stakes for that answer are quite high. The sources chosen from the internet to respond to those questions need to be accurate and unbiased, especially when called upon to answer questions of political relevance. If not, then the biases in the algorithm that whittles away possible responses to arrive at "the" answer, will have decisive significance in delivering knowledge to consumers.

Once again, the example of home assistants is merely emblematic of the larger challenge posed by an omnipresent information ecosystem with technologies seeking to provide relevant answers to any question at any time in the most convenient form possible. In a relatively short period of time we have moved from searching for answers in libraries to sitting at a home desktop to "carrying" the internet with us on our phones wherever we go. Now, the internet is beginning to "move" with us, becoming even more ubiquitous as the machines around us all go online. In

turn, the ways in which these other devices organize information become especially important. As these new machines "learn" how to answer user questions on everything from medical diagnoses to voting information, the risk grows that a few companies or a few algorithms might be relied upon to provide a growing share of the information relevant for civic engagement.

D. Professionalization of election interference

Russian interference in the 2016 U.S. Election established a paradigm for thinking about outsider manipulation of democratic decisionmaking. That model, which centered on a nation state acting to destabilize an adversary, has quickly been replaced by more complicated modes of election interference. Indeed, the Russian "playbook" has now been professionalized by state and non-state actors alike. A veritable industry has now developed, which sells the various commodities of election interference (bots, trolls and the like) to those interested in these services.

The scandal involving Cambridge Analytica has become more of a metaphor for an array of problems related to election interference. The scandal itself grew out of the misuse of Facebook data by a Cambridge researcher, who transferred social graph data garnered from personality surveys to a consulting firm that eventually would work for Donald Trump's campaign. Most observers in the field do not believe Cambridge Analytica, itself, was very successful in using these data. But the scandal has come to refer to the more general phenomenon of political consultants (even those based in a foreign country) exploiting massive amounts of private social media data to craft targeted (even secret) messages of persuasion and demobilization to affect election outcomes. Although Cambridge Analytica, itself, may have been more bark than bite, other firms have perfected what they promised to do and gone a step further. Not only can governments and political parties now purchase outside expertise to conduct opposition research and targeted social media campaigns, but a whole range of influence operations previously "owned" by state intelligence services are now available to candidates and parties.[98]

At the same time that election interference has become "professionalized," it has also become, like other arenas of internet activity, vulnerable to gang-like actions. The statelessness and disorganization of online associational life enables international coalitions of hackers, trouble makers, anarchists, and criminals to find solidarity in wreaking havoc against the establishment both within and beyond the electoral context. At one end of the spectrum might be "transparency" (very loosely defined) groups, such as Anonymous and Wikileaks, that seek to use the internet to

[98] Mark Mazzetti, Ronen Bergman, David D. Kirkpatrick & Maggie Haberman, *Rick Gates Sought Online Manipulation Plans From Israeli Intelligence Firm for Trump*, NY Times, Oct. 8, 2018, available at https://www.nytimes.com/2018/10/08/us/politics/rick-gates-psy-group-trump.html; PSY Group, *Project "Rome," Campaign Intelligence & Influence Services Proposal*, April 2016, available at https://int.nyt.com/data/documenthelper/360-trump-project-rome/574d679d1ff58a30836c/optimized/full.pdf#page=1 .

expose and counteract what they consider elite wrongdoing. At another end, loosely knit quasi-terrorist or gang groups, such as Legion Holk in Mexico or Seguidores De La Grasa, have incited off-line violence as well as propagated viral disinformation campaigns.

It is becoming increasingly difficult to distinguish between "normal" campaign activity by official arms of domestic political actors and anti-democratic information operations by foreign governments or transnational groups. Even the modern archetype coming out of the 2016 U.S. election of active measures by a foreign government to influence an election outcome has given way to much more diffuse efforts by combinations of domestic and foreign actors, both inside and outside government, with varied motivations ranging from crime, anarchism, and fostering division to actually affecting who wins elective office. As a result, it becomes especially challenging to draw traditional lines between foreign and domestic political activity, government and nongovernmental organizations (including the "media"), and information operations and permissible campaign activity.

Conclusion

This chapter has attempted to identify the challenges digital technologies pose for democracy and to canvass reforms that may help in overcoming them. At the same time that governments, platforms, and civil society seek to overcome these challenges, new problems will undoubtedly arise and new policy interventions will need to be tested to magnify the benefits and minimize the costs of digital technologies for democracy. The internet, after all, is here to stay. The question for efforts underway is how best to realize the original egalitarian, freedom-enhancing, and pro-democracy vision of the internet, while cabining the influence of actors that seek to use these new technologies to undermine democracy itself.