Social Media Freedom of Speech and the Future of our Democracy

LEE C. BOLLINGER

and GEOFFREY R. STONE

Social Media, Freedom of Speech and the Future of Our Democracy

Social Media, Freedom of Speech and the Future of Our Democracy

EDITED BY LEE C. BOLLINGER
and
GEOFFREY R. STONE

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide. Oxford is a registered trade mark of Oxford University Press in the UK and certain other countries.

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America.

CIP data is on file at the Library of Congress
ISBN 978–0–19–762109–7 (pbk.)
ISBN 978–0–19–762108–0 (hbk.)

DOI: 10.1093/oso/9780197621080.001.0001

1 3 5 7 9 8 6 4 2

Paperback printed by Lakeside Book Company, United States of America
Hardback printed by Bridgeport National Bindery, Inc., United States of America

To Jean and Jane
And to the next generation:
Katelyn, Colin, Emma, Cooper, and Sawyer
Julie, Mollie, Maddie, Jackson, and Bee

CONTENTS

ACKNOWLEDGMENTS

This is our fourth major book project together, all centered around major turning points or seminal issues involving the First Amendment. We are always profoundly impressed, first, by the importance and complexity of the role played by the principles of freedom of speech and press in American society and, second, by just how illuminating it can be to draw together some of the most informed and original minds of the time to try to understand these matters better. We, therefore, begin our thanks to our authors and to members of the Commission for their commitment to this project. We are also grateful to Andrew Ceresney, Courtney Dankworth, Morgan Davis, and their colleagues at Debevoise & Plimpton for the extraordinary help in providing the background paper and in shepherding the work of the Commission to arrive at its recommendations. We believe that the contributions of all of these individuals have collectively yielded a valuable volume of ideas.

There are others who deserve recognition. Jane Booth, General Counsel of Columbia University, provided the bridge with Debevoise & Plimpton and her own substantive interventions. The Office of the President at Columbia, under the direction of Christina Shelby, was simply outstanding in organizing just about everything. We make special note of Joanna Dozier in this regard and freelance editor Lori Jacobs.

We continue to be thankful for the deep relationships we have with Oxford University Press, and especially with Dave McBride and his entire staff.

As always, we benefit in inexpressible ways from Jean Magnano Bollinger (Lee's spouse) and Jane Dailey (Geof's spouse). Nothing happens without their counsel and support (and constructive criticism).

And, lastly, we feel beyond fortunate to have Carey Bollinger Danielson as our general editor, on this as well as our prior two books. As we have said each time and do so again, Carey has been the perfect colleague in these ventures.

LIST OF CONTRIBUTORS

Katherine Adams, Senior Vice President and General Counsel, Apple

Jack M. Balkin, Knight Professor of Constitutional Law and the First Amendment, Yale Law School, and founder and director of Yale's Information Society Project

Martin Baron, former Executive Editor, the *Washington Post*

Emily Bazelon, staff writer for the *New York Times Magazine* and Truman Capote Fellow for Creative Writing and Law, Yale Law School

Yochai Benkler, Berkman Professor of Entrepreneurial Legal Studies, Harvard Law School

Lee C. Bollinger, President and Seth Low Professor of the University, Columbia University

Alex Chemerinsky, law clerk, United States District Court

Erwin Chemerinsky, Dean and Jesse H. Choper Distinguished Professor of Law, Berkeley Law

Hillary Clinton, former United States Secretary of State

Jelani Cobb, Ira A. Lipman Professor of Journalism, Columbia Journalism School

Renée DiResta, technical research manager, Stanford Internet Observatory

evelyn douek, lecturer on law, Harvard Law School, and associate research scholar, Knight First Amendment Institute at Columbia University

Russ Feingold, President, American Constitution Society

Mary Anne Franks, Professor of Law and Michael R. Klein Distinguished Scholar Chair, University of Miami School of Law; President and Legislative & Tech Policy Director, Cyber Civil Rights Initiative

Jamal Greene, Dwight Professor of Law, Columbia Law School

Amy Klobuchar, United States Senator for Minnesota

Larry Kramer, President of the William and Flora Hewlett Foundation, formerly Richard E. Lang Professor of Law and Dean, Stanford Law School

Genevieve Lakier, Professor of Law, University of Chicago Law School; Senior Visiting Research Scholar, Knight First Amendment Institute at Columbia University (2021-2022 academic year)

Lawrence Lessig, Roy L. Furman Professor of Law and Leadership, Harvard Law School

Martha Minow, 300th Anniversary University Professor and former Dean of Harvard Law School, board member of public broadcaster WGBH

Mary Minow, presidential appointee under three administrations to the Institute of Museum and Library Services and library law consultant

Nell Minow, expert in corporate governance, an American film reviewer and author who writes and speaks frequently on film, media, and corporate governance and investing, and served on the board of public broadcaster WETA

Newton N. Minow, Senior Counsel at Sidley Austin LLP, the former Chairman of the FCC, and active participant in public broadcasting and media governance

Christina Paxson, President, Brown University

Nathaniel Persily, James B. McClatchy Professor of Law, Stanford Law School

Kate Starbird, Associate Professor, Human Centered Design & Engineering, University of Washington

Geoffrey R. Stone, Edward H. Levi Distinguished Service Professor of Law, University of Chicago

David A. Strauss, Gerald Ratner Distinguished Service Professor of Law, University of Chicago Law School

Cass R. Sunstein, Robert Walmsley University Professor, Harvard University

Sheldon Whitehouse, United States Senator for Rhode Island

Debevoise & Plimpton attorneys: Andrew Ceresney, Jeffrey P. Cunard, Courtney Dankworth, and David A. O'Neil

OPENING STATEMENT

LEE C. BOLLINGER AND GEOFFREY R. STONE

One of the most fiercely debated issues of the current era is what to do about "bad" speech on the internet, primarily speech on social media platforms such as Facebook and Twitter. "Bad" speech encompasses a range of problematic communications—hate speech, disinformation and propaganda campaigns, encouragement of and incitement to violence, limited exposure to ideas one disagrees with or that compete with preexisting beliefs, and so on. Because the internet is inherently a global communications system, "bad" speech can arise from foreign as well as domestic sources. No one doubts that these kinds of very harmful expression have existed forever, but the premise of the current debate is that the ubiquity and structure of this newest and most powerful communications technology magnifies these harms exponentially beyond anything we have encountered before. Some argue that, if it is left unchecked, the very existence of democracy is at risk.

The appropriate remedies for this state of affairs are highly uncertain, and this uncertainty is complicated by the fact that some of these forms of "bad" speech are ordinarily protected by the First Amendment. Yet the stakes are very high in regard to how we answer the question because it is now evident that much of public discourse about public issues has migrated onto this new technology and is likely to continue that course into the future.

Current First Amendment jurisprudence has evolved on the premise that, apart from certain minimal areas of well-established social regulation (e.g., fighting words, libel, threats, incitement), we should place our trust in the powerful antidote of counter-speech to deal with the risks and harms of "bad" speech.[1] Of course, that may well turn out to be the answer to our contemporary dilemmas. Indeed, one can already see the rise of public pressures on internet companies to increase public awareness of the dangers of "bad" speech,

and there are discussions daily in the media raising alarms over dangerous speech and speakers.[2] Thus, it may be that the longstanding reliance on the self-correcting mechanisms of the marketplace of ideas will work again.

But perhaps not. There is already a counter risk—that the increase in "editorial" control by internet companies will be biased against certain ideas and speakers and will effectively censor speech that should be free.[3] On the other hand, even those who fear the worst from "bad" speech being uninhibited often assert that internet company owners will never do enough on their own to initiate the needed controls because their basic, for-profit motivations are in direct conflict with the public good and the management of civic discourse.[4] There is understandable concern that those who control the major internet companies will have an undue and potentially dangerous effect on American democracy through their power to shape the content of public discourse. On this view, public intervention is necessary.

It is important to remember that the last time we encountered a major new communications technology we established a federal agency to provide oversight and to issue regulations to protect and promote "the public interest, convenience, and necessity."[5] That, of course, was the new technology of broadcasting, and the agency was the Federal Communications Commission. The decision to subject private broadcasters to some degree of public control was, in fact, motivated by some of the very same fears about "bad" speech that we now hear about the internet.[6] People thought the risks of the unregulated private ownership model in the new media of radio and television were greater than those inherent in a system of government regulation. And, like today, those who established this system felt unsure about what regulations would be needed over time (in "the public interest, convenience, and necessity"), and they therefore set up an administrative agency to review the situation and to evolve the regulations as circumstances required. For a fuller description and analysis of these issues, see *Images of a Free Press* and *Uninhibited, Robust, and Wide-Open: A Free Press for a New Century*, by Lee C. Bollinger.[7]

On multiple occasions, the Supreme Court has upheld this system under the First Amendment. The formal rationale for those decisions may not apply to the internet, but there is still plenty of room for debate about the true principles underlying that jurisprudence and their continued relevance. In any event, the broadcasting regime stands as arguably the best example in our history of ways to approach the contemporary concerns about new technologies of communication. But, of course, it may be that government intervention in this realm is so dangerous that social media platforms should be left to set their own policies, just as the *New York Times* and the *Wall Street Journal* are free to do.

Section 230 of the Communications Decency Act of 1996 famously shields internet companies from liability for speech on their platforms.[8] Many critics of internet companies have advocated the repeal of this law and have used the idea of its repeal as a threat to get these companies' owners to change their editorial policies (either to stop censoring or to censor more).[9] Another approach would be to enforce existing laws that forbid foreign states and certain actors from interfering in US domestic elections and politics.

Everyone accepts the proposition that efforts by Russia to spread disinformation in order to foster civil strife in America is highly dangerous and properly subject to criminal prohibitions.[10] But, in a much more integrated world, especially one facing global problems (climate change, and so on), it is also true that the American public has a vital First Amendment interest in hearing and communicating with the broader international community. The problem, therefore, will be in finding the right balance between improper foreign interference and the healthy and necessary exchange of ideas on the global stage.

We also need to take stock of the precise nature of the problems we are facing with "bad" speech on social media platforms, as well as what means other than legal intervention might be available to address the problems. Public education, changes in algorithms, the development of a more journalistic culture within the management of these platforms, government pressures on "bad" actors abroad, and other non-legal solutions all need to be explored.

It is also possible that the constraints in existing First Amendment jurisprudence should themselves be amended, not only because the circumstances and contexts are different today but also because experience over time with those doctrines and principles might lead some to doubt their original or continuing validity. Overall, we need to imagine as best we can what a new equilibrium should look like as we experience the impacts on our democracy of this new technology of communication.

Every now and then in the history of the First Amendment an issue comes along that not only poses a perplexing and challenging question about some aspect of First Amendment doctrine or some incremental move, but also calls into question the whole edifice of freedom of speech and press as we have come to know it in the United States. The current debates about the threats to free speech, and even to democracy itself, triggered by the evolution of our newest technology of communication—the internet, and especially social media platforms—constitute such an occasion. The extraordinarily rapid embrace of this method of communication (in less than two decades), together with its pervasive presence in our lives, is both astonishing and revolutionary. This is true especially because the internet and social media are controlled by a few corporations which are structurally designed to reserve to them primary control

of this powerful new means of communication. It is now a central question in the United States and around the world whether this new means of communication strengthens what freedom of speech has marked as the ideal or threatens everything we have so painstakingly built.

This book is dedicated to exploring that question and what follows from the answers we give to it. At this moment in the history of the United States, there is arguably no conundrum of greater importance. When an overwhelming majority of citizens communicates, receives information, and forms political alliances in a single place, and when that place is effectively controlled and curated by a single person or entity (or mathematical model),[11] alarms built over decades of thought about freedom of speech and democracy are triggered. Too much censorship? Or too little? Those, in a sense, are the central concerns. The balance struck is always the test of a free and democratic society, because it is ultimately through speaking and listening that human beings become who they are and decide what to believe. Put simply, do entities like Facebook, Twitter, and YouTube have too much power under existing law to determine what speech we will or will not have access to on social media? Are there changes that can constitutionally be made to the current system that will improve rather than worsen the current state of affairs? And how should we think about the multinational implications of the internet and about how policies adopted in other nations affect freedom of speech in the United States?

We have invited twenty-two authors, all experts on this general subject, to write eighteen essays describing how they see the current and future state of affairs in this new communications universe and what they propose should be done, if anything, to address these challenges. We frame the inquiry with this brief introductory statement and a concluding one. A team from the distinguished law firm of Debevoise & Plimpton (led by Andrew Ceresney, partner) has contributed an essay providing a comprehensive legal background to the current state of play in the area we are investigating. And, finally, we include in this volume the Report of a Commission we convened to review the assembled essays and to make a set of specific recommendations about what we should and should not do in this highly controversial area of policy and constitutional law.

This is our fourth project together of this kind. We have come to this way of exploring highly controversial and complex problems by experience. We have found that giving individuals with deep expertise the opportunity to write freely about their concerns and to explore their suggestions for reform provides a depth of engagement and education for the reader that is hard to achieve in a single-author volume. We believe this is a productive way of exploring a difficult problem and of driving the discussion toward recognizing new realities.

Before turning to the essays, we thought it would be useful for us to set forth a few points we believe to be central to the analysis:

1. *The recency of our First Amendment jurisprudence.* In 2019 we published a volume of essays titled *The Free Speech Century*.[12] Our purpose was to mark, indeed to celebrate, the one hundredth anniversary of the first Supreme Court decisions interpreting the words of the First Amendment: "Congress shall make no law . . . abridging the freedom of speech, or of the press."[13] Although the First Amendment had been added to the Constitution in 1789 as part of the Bill of Rights, the Supreme Court did not weigh in on its meaning until 1919, during World War I.[14] The complex jurisprudence about free speech and free press that we rely on today is, therefore, of fairly recent vintage—we are still learning.[15]
2. *The current jurisprudence of the First Amendment affords extraordinary protection to the freedoms of speech and press.* No other nation in the world today, nor, for that matter, in history, has evolved a system of constitutional protection for the freedoms of speech and press to the degree, or with the level of analytical refinement, that we have today in the United States. In the last half-century, starting with the Court's famous decision in *New York Times v. Sullivan*, involving a defamation lawsuit brought by a Montgomery, Alabama, city official against the *New York Times*, the Court has created the most speech-protective system in the world.[16] Using the premise of a deep distrust of government involvement in speech and a concomitant fear of the natural human impulse to be intolerant and to suppress the thoughts we hate, the judiciary has embraced a high level of protection for free speech and a free press, especially when the speech concerns public issues. The upshot is that we now live in a society committed to a level of public discourse that, in the Supreme Court's words, is "uninhibited, robust and wide-open."[17] As a result, the opportunity for individuals to engage in potentially "bad" or "harmful" speech is very, very wide.
3. *Our contemporary First Amendment jurisprudence is more complex and fluid than many might imagine.* There are several reasons for this. For example, on many of the key issues we will take up in this book, the Supreme Court has taken different, sometimes opposing, positions at various times. To cite just one of many possible examples, in *Beauharnais v. Illinois*, decided in 1952, the Court upheld the constitutionality of a law prohibiting what we would today call "hate speech," whereas in *Brandenburg v. Ohio*, decided in 1969, the Court held that the government cannot punish such speech consistent with the First Amendment unless the government can prove that the speech is "directed to inciting or producing imminent lawless action and is likely to incite or produce such action."[18] Often in these situations, the earlier decision has not been explicitly overruled, thus leaving open the possibility for its later revival. Moreover, many decisions are vague in their implications, leaving room for a broad range of possible interpretations. As a result, one

has to be careful in declaring "what the First Amendment means" in today's world—the First Amendment is often much more varied and complicated than many experts proclaim, which leads to the next point.

4. *The internet and all that comes with it is not the first instance in which the First Amendment has had to come to terms with a revolutionary new technology of communication.* As noted previously, in the last century, the "new" technology was broadcasting—first radio and then television. People were anxious about broadcasting, and for exactly the same reasons they are now fearful of the power of social media platforms: It was clear from the outset that only a small number of "owners" would control this new national means of communication because of the physical limits of the electromagnetic spectrum. Indeed, in order for there to be effective communication, there would typically be only three television stations per community and they would each be linked to national programming through the emerging "networks" (ABC, NBC, and CBS). This concentration of power led to fears that these owners would dominate and censor content, both limiting discussion to only the perspectives and views they embraced and empowering them to manipulate public opinion through propaganda. This was not a hypothetical problem, as the Nazi regime's uses of radio to advance its evil ends vividly demonstrated.[19] Faced with this challenge, the federal government stepped in, established a new federal agency to exercise oversight (the Federal Communications Commission), and charged it with the authority to issue licenses and establish regulations consistent with "the public interest, convenience, and necessity."[20] Censorship by the government was explicitly forbidden, but otherwise the government asserted its authority to oversee this new public resource of communication. This was first established in 1927 and then more comprehensively and permanently in 1934.[21]

   In 1969, in Red Lion Broadcasting v. FCC, the Supreme Court unanimously upheld this system of government regulation of the broadcast media as constitutional under the First Amendment.[22] At issue was an FCC regulation known as the fairness doctrine, under which every broadcaster was required to cover issues of public importance and to do so fairly, by giving the public an array of different points of view about those issues. The Court held that this was constitutional because the "scarcity" of the airwaves necessitated government intervention to ensure that the interests of the public were served.[23] Broadcasters were "fiduciaries," and the government had a legitimate, perhaps even mandatory, role in protecting the public's interest in hearing all points of view.

   Five years later, though, in Miami Herald v. Tornillo, the Court held that similar regulations of newspapers violated the First Amendment even though those regulations were also designed to ensure a well-informed public and to restrain the potentially distorting motives of the increasingly powerful and oligopolistic owners of newspapers.[24] The Court's decision in Miami Herald

did not offer any persuasive explanation as to why newspapers should be treated differently from radio and television networks. As far as the Court was concerned, they existed in parallel universes. The problem, as economists observed at the time, was that the scarcity thesis underpinning the broadcast regime was insufficient to distinguish broadcasting from daily newspapers, which by the 1960s had evolved into an operative monopoly in over 90 percent of American cities.[25]

That fact, however, does not necessarily make the system chosen for broadcasting wrong, constitutionally or otherwise, just as it does not necessarily make the alternative system chosen for daily newspapers wrong. One may see this effectively dual system as wrong because each side reflects a philosophy of the First Amendment that is inconsistent with the other. Or one may see the inconsistency as a reasonable, perhaps even wise, experimental approach to a growing common problem, namely, that of rising monopolistic private power over the marketplace of ideas and democratic discourse. The pressing question at the moment, and for the purposes of this volume, is what any of this tells us about the future of social media.

5. What is the right thing to do when we are uncertain what to do? Without prejudging the questions on the table about regulation of the internet and social media platforms, it is worth bearing in mind that uncertainty does not always necessitate inaction. There is value in considering the range of options open to us as we assess the risks and benefits of the current arrangements, including those of taking small or large steps that leave open decisions about other concrete interventions. The Communications Act of 1934 was of its time, in the sense that it relied on an executive branch administrative structure to develop specific regulations over time, with the benefit of experience and under the vague mandate of serving "the public interest, convenience, and necessity."[26] Similar thinking may be called for as we contemplate today's challenges.

It has been a joy and an education to work with all the participants in this project, and we hope the sum of what is presented here will be as helpful, as challenging, and as stimulating to our readers as it has been to us.

REGULATING HARMFUL SPEECH ON SOCIAL MEDIA: THE CURRENT LEGAL LANDSCAPE AND POLICY PROPOSALS

ANDREW J. CERESNEY, JEFFREY P. CUNARD,
COURTNEY M. DANKWORTH, AND DAVID A. O'NEIL

Social media platforms have transformed how we communicate with one another. They allow us to talk easily and directly to countless others at lightning speed, with no filter and essentially no barriers to transmission. With their enormous user bases and proprietary algorithms that are designed both to promote popular content and to display information based on user preferences, they far surpass any historical antecedents in their scope and power to spread information and ideas.

The benefits of social media platforms are obvious and enormous. They foster political and public discourse and civic engagement in the United States and around the world.[1] Social media platforms give voice to marginalized individuals and groups, allowing them to organize, offer support, and hold powerful people accountable.[2] And they allow individuals to communicate with and form communities with others who share their interests but might otherwise have remained disconnected from one another.

At the same time, social media platforms, with their directness, immediacy, and lack of a filter, enable harmful speech to flourish—including wild conspiracy theories, deliberately false information, foreign propaganda, and hateful rhetoric. The platforms' algorithms and massive user bases allow such "harmful speech" to be disseminated to millions of users at once and then shared by those users at an exponential rate. This widespread and frictionless transmission of harmful speech has real-world consequences. Conspiracy theories and false information spread on social media have helped sow widespread rejection of COVID-19 public-health measures[3] and fueled the lies about the 2020 US presidential

election and its result.[4] Violent, racist, and anti-Semitic content on social media has played a role in multiple mass shootings.[5] Social media have also facilitated speech targeted at specific individuals, including doxing (the dissemination of private information, such as home addresses, for malevolent purposes) and other forms of harassment, including revenge porn and cyberbullying.

These negative effects of social media discourse are profound and endemic, and they are not going away. Policymakers and stakeholders around the world have called for reform, including by regulating social media companies' approaches to the dissemination of harmful speech on their platforms. This introductory essay sets the context in which policymakers might think about how to approach the challenges of harmful speech on social media platforms, including by making such reforms. It first briefly summarizes the US statutory and constitutional framework for approaching the regulation of social media platforms. It then reviews potential US policy approaches, with a particular focus on proposed and implemented legislation concerning social media platforms in the European Union.

Statutory and Constitutional Barriers to Addressing Harmful Speech on Social Media

There are significant barriers in the United States to limiting, regulating, or prohibiting speech on social media that is harmful but lawful. To begin with, much "harmful speech" is lawful, including hateful or racist speech and noncommercial speech that is false or misleading. Users cannot be held legally liable for posting such content. Likewise, social media platforms do not have legal liability for their role in the spread of lawful speech or for removal or moderation of such content. Although the victims of unlawful speech (such as defamation) on social media can sue the speaker, private suits offer little recourse for victims when such speech is instantaneously disseminated across the internet. The victim also cannot sue the platform because of Section 230 of the Communications Decency Act (47 U.S.C. § 230, "Section 230"), which now immunizes social media platforms from most legal claims arising from user content posted on their sites and the platforms' moderation of that content (or lack thereof). This immunity has been widely credited with allowing the internet to flourish but also criticized for allowing social media giants to escape liability for harmful speech on their platforms. Even if Section 230 were repealed, however, social media platforms still could not be held legally liable for harmful but not unlawful speech. Going beyond existing statutes and common-law doctrines, policymakers could try to regulate lawful speech, but that regulation must comport with the First Amendment. As described below, the First Amendment, as

applied to the users who post on social media platforms and the social media platforms themselves, has been interpreted to substantially limit government's ability to regulate lawful speech.

Section 230 of the Communications Decency Act

Section 230 provides broad immunity for online platforms that host or republish speech against a range of claims that might otherwise be used to hold them legally responsible for what their users say or do. Specifically, Section 230 states that providers of "interactive computer services," including social media companies, cannot be held liable both for their decisions to host and disseminate content created by others and for actions they take "voluntarily" and "in good faith" to restrict access to certain types of material.[6] There are certain limited exceptions related to federal criminal activity (such as obscenity and sexual exploitation of children), electronic privacy laws, and intellectual property protection.[7]

Congress passed Section 230 in 1996 in response to *Stratton Oakmont, Inc. v. Prodigy Services Co.*[8] In that case, a New York trial court held that an online service provider could be held liable for defamatory material posted by its users to the extent that the provider was "republishing" the material rather than simply distributing it.[9] The court relied on well-settled principles of defamation law, under which publishers and republishers of defamatory material can be held liable for the defamation, whereas mere distributors of such content (such as a library) may not be.[10] Congress was concerned that, absent statutory protection, online service providers would be disincentivized to remove defamatory content because, in doing so, they could be regarded as a "publisher" of the content and thus subject to suits for defamation.

To address this concern, Congress enacted Section 230 to grant immunity to any provider of an "interactive computer service" for the dissemination of any unlawful speech—not limited to defamation—that its users provide. "Interactive computer service" is defined as "any information service, system, or access software provider that provides or enables computer access by multiple users to a computer server, including specifically a service or system that provides access to the Internet."[11] Immunity is not available, however, to "information content provider[s]"—that is, those "responsible, in whole or in part, for the creation or development" of the content.[12] Social media platforms, including Facebook and Twitter, generally have been found to meet the definition of an "interactive computer service."[13] If, however, they have a role in generating the allegedly unlawful content or are engaged in activities that go beyond publishing third-party content, Section 230 immunity may not apply. Drawing the line between publishing and other acts by platforms has been, in some contexts, an area of dispute.[14]

Section 230 immunizes "interactive computer service[s]," including social media companies, from suit in two respects. First, Section 230(c)(1) is a barrier to liability for hosting, republishing, and disseminating content furnished by third parties. Specifically, it provides: "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider." Courts have interpreted this provision broadly to confer immunity in both the publisher and distributor contexts, including for decisions to review, edit, format, remove, or omit certain content.[15] Section 230(c)(1) effectively bars most state and federal claims by private parties as well as a range of actions by state prosecutors. These include defamation or libel claims based on user-generated content; design defect and website design claims, including that social media platforms' algorithms facilitate widespread sharing of harmful content, such as content that facilitates real-world violence or harassment;[16] other tort claims, such as negligence or invasion of privacy;[17] claims under state and federal anti-discrimination statutes;[18] and some state criminal prosecutions.[19]

Second, Section 230(c)(2) immunizes an "interactive computer service" from liability for removing or moderating content generated by users. Specifically, it states:

> No provider or user of an interactive computer service shall be held liable on account of (A) any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected; or (B) any action taken to enable or make available to information content providers or others the technical means to restrict access to material described in paragraph (1).

The rationale for Section 230(c)(2) (the so-called Good Samaritan provision) was to obviate any legal objection to interactive computer services removing such materials, lest they be deemed a publisher (or otherwise have some legal responsibility for the distribution of content) and, as such, be liable for removing the content. This immunity for removing all manner of material, including any that is "otherwise objectionable," gives social media companies broad authority, if exercised in good faith, to remove content that is potentially defamatory, harassing, bullying, invasive of privacy, inflicting of intentional emotional distress, misinforming, or originating from a foreign government or other source. Although some providers of content have objected to social platforms' decisions to delete or moderate such content (which they are otherwise generally permitted to do under their own terms of service), courts generally have

been deferential to companies' decisions to remove content under this provision.[20] Section 230(c)(2) also has the effect of precluding state government regulation of social media platforms' moderation decisions. In *NetChoice v. Moody*, for example, a district court granted a preliminary injunction enjoining enforcement of a Florida law that, among other things, prohibited certain social media platforms from suspending or removing political candidates' accounts, holding that the petitioners were likely to succeed on the merits of their claim that this prohibition was preempted by Section 230(c)(2).[21]

Section 230 arose from Congress's understanding at the time of enactment that interactive computer service companies host and disseminate millions of messages. It was thought that they would not realistically have any means of monitoring or moderating, on anything like a real-time basis, the lawfulness of content provided by their users—for example, determining whether the content was defamatory or otherwise tortious. That their involvement in publishing those messages could give rise to numerous individual suits was seen as detrimental to the growth of the internet, given both the potentially substantial time and expense of defending against the litigations and the likelihood that the companies would, in the interest of avoiding litigation, become more conservative—deleting more speech than was required. In practical terms, Section 230(c)(1) has, indeed, been a powerful tool for social media companies in that it allows them to dispose of suits asserting that they have liability for unlawful content.[22] Whether the legislative rationales that animated the enactment of Section 230 in 1996 hold true today, given more recent technological developments and the platforms' own role in using their algorithms to amplify the reach of user-generated content, has been the subject of more recent discussion.

Further, with respect to platforms' moderation or removal decisions, litigants have increasingly invoked the "good faith" requirement in Section 230(c)(2) to avoid dismissal of their claims under that section.[23] For example, one court held that Google's defense of Section 230 immunity could not be resolved on a motion to dismiss because the plaintiff had adequately alleged facts showing that Google's explanation for removing the plaintiff's app from its Android app store was pretextual and therefore in bad faith.[24]

The First Amendment

Even if the statutory immunity conferred by Section 230 were repealed or limited by Congress, the First Amendment would limit the permissible scope of regulation of social media speech. The First Amendment would apply to attempts by government to limit or regulate directly users' speech that is harmful, but is otherwise lawful. It would also apply to attempts by the government to regulate social media platforms by requiring them to remove such speech.

The First Amendment may well protect from regulation social media companies' moderation of user-generated content, including deciding what content to feature or promote and how to rank or filter user-generated content, if they are characterized as editorial decisions protected by the Constitution. Whether such regulations could withstand challenge under the First Amendment will depend on (*i*) the type of speech being regulated; (*ii*) if the platform is regulated, the type of activity by the platform; (*iii*) whether the regulation discriminates based on type of speech or can be characterized as content-neutral; and (*iv*) whether courts view content moderation by social media platforms as more closely akin to editorial decisions by print media (which have historically been accorded the strongest First Amendment protection), broadcasters (which have received weaker protection), or others such as cable networks or broadband internet providers.

What Types of Speech Can the Government Regulate?

The First Amendment protects speakers from government prohibition or regulation of most types of speech, subject to limited and time-tested exceptions. Certain forms of speech, such as incitement to imminent unlawful action[25] and obscenity,[26] are not protected at all. Other forms of speech, such as defamation, can be regulated subject to court-developed frameworks that harmonize First Amendment interests with state law–created torts. Whereas political speech is thought to be at the core of First Amendment protection, purely commercial speech, such as advertisements, is accorded somewhat lesser protection,[27] though political advertising may receive greater protection.[28] Courts generally have been reluctant to create new exceptions to the protection afforded by the First Amendment.

In general, the First Amendment is likely to shield from regulation many forms of harmful speech, including false or misleading information[29] and hate speech,[30] that are not currently prohibited or regulated. That means that users cannot be sanctioned for creating and uploading such speech to social media platforms. Accordingly, it is likely that platforms cannot be forced to remove it unless it is otherwise found unlawful.

When Are Platforms' Activities Protected Under the First Amendment?

In general, and subject to the permitted limitations on speech described above, there is full First Amendment protection for speech that the platforms themselves generate. Text, graphics, images, or videos created by the platforms plainly fall within the ambit of constitutional protection.

Where, however, a social media platform simply transmits or displays users' messages (such as on a message board or "wall" on Facebook), that act is not likely to be seen as the platform's own speech. The relatively passive transmission of messages may not be protected by the First Amendment, just as the First Amendment does not safeguard a telephone company's common carrier transmission of voice or data.[31]

Whether the platforms' activities of moderating, filtering, curating, or promoting content are protected by the First Amendment is a closer question.[32] Those activities, including the use of algorithms to edit, rank, and filter content, are arguably analogous to the editorial activities of media outlets or other entities—selecting which content to commission or which letters to the editor to publish. The First Amendment affords broad protections to those types of editorial decisions, whether by print media, broadcasters, cable operators, or even parade organizers.[33] The First Amendment may not protect platforms' development and deployment of algorithms that identify content that might be of interest to users insofar as the use of algorithms could be considered entirely software-facilitated conduct; conduct that does not have an expression-related element is not protected under the First Amendment.[34]

Courts are likely to find that some of the platforms' content-moderation activities are protected, given the inclination to find that the First Amendment protects speech-promoting conduct. In *Sorrell v. IMS Health Inc.*, for example, the Supreme Court rejected the argument that the act of selling or transferring content generated by a third party is only conduct and is not, therefore, a speech-related activity protected by the First Amendment.[35] In a closer analogy, several courts have afforded First Amendment protections to search engines' ranking and filtering content on the basis that their search algorithms make editorial judgments about what information to display, how to organize it, and where to present it.[36]

In the social media context, at least one court has held that Facebook's failure to remove a user's post was protected under the First Amendment.[37] Further, at least two federal district courts have held that state laws designed to prevent social media platforms from removing or de-prioritizing certain content likely violated the platforms' First Amendment rights to engage in such editorial activities.[38] In *NetChoice v. Moody*, the court held that the Florida law likely violated social media platforms' First Amendment–protected rights to remove or moderate content related to or posted by political candidates.[39] In *NetChoice v. Paxton*, the court reviewed a Texas statute prohibiting social media platforms from removing or "de-boosting" user-generated content based on the users' viewpoint, among other things.[40] In granting the plaintiffs' motion for a preliminary injunction, the court recognized the First Amendment rights of social media platforms when they "curate both users and content to convey a message

about the type of community the platform seeks to foster and, as such, exercise editorial discretion over their platform's content."[41]

What Kinds of Regulations Can Survive Constitutional Scrutiny?

Notwithstanding that the text of the First Amendment appears to bar any law that regulates expression, certain categories of speech are subject to permissible government regulation, depending on context. Regulation of social media companies' content moderation activities will be reviewed by courts in light of the applicable standard of scrutiny, depending on whether the law is targeted at particular types of content or is content-neutral.

Courts apply a more deferential standard of review to regulations that are content-neutral.[42] Content-based restrictions, by contrast, target particular speech based on its expressive content and are presumptively unconstitutional.[43] Such restrictions are subject to strict scrutiny, meaning that the law must further a "compelling interest" and be "narrowly tailored" to further that interest.[44] Content-neutral regulations—such as a city ordinance prohibiting posting of any sign on public property[45]—are, however, subject to intermediate scrutiny. Under that test, the regulation will survive constitutional challenge only if it is narrowly tailored to serve a significant government interest that is unrelated to the suppression of free expression and leaves open ample alternative channels for communication.[46]

In the social media context, the courts in *Moody* and *Paxton* held that the Florida and Texas laws, respectively, were impermissible content-based regulations and failed strict scrutiny. In *Moody*, that was because the Florida statute favored particular types of speech (i.e., speech by or about political candidates) over all other types of speech and because it treated identical speakers differently based on size.[47] In *Paxton*, the Texas statute was problematic because it exempted both certain types of content (related to criminal activity or threats of violence against specified groups) and certain types of platforms (those with fewer than fifty million monthly active US users).[48]

To What Extent Do Social Media Fit Into Regulatory Frameworks That Have Been Applied to Other Media?

In considering whether government regulation of social media platforms' content-moderation practices comports with the First Amendment, courts also may look to case law concerning regulation of older forms of media such as newspapers, television and radio broadcasters, cable networks, or common

carriers like telephone service providers. The Supreme Court has treated regulation of each of these forms of media differently under the First Amendment, and thus courts' analysis of social media regulations may depend on which older media they find most closely analogous to social media. Courts would not be writing on a blank slate, however: In 1997, the Supreme Court in *Reno v. ACLU* held that a content-based restriction on the internet (there, the regulation of indecency) could not survive strict scrutiny.[49]

Supporters of enhanced regulation of social media platforms might urge broadcast media as an apt analogy for social media because broadcasters historically have been subject to content-based regulation that would not be constitutional if applied to traditional print media and the editorial decisions made by cable television operators.[50] Beginning in 1927 with the creation of the Federal Radio Commission, which was replaced in 1934 by the Federal Communications Commission (FCC), Congress empowered a new federal agency to license and regulate broadcast media to promote the "public interest, convenience, and necessity."[51] The system resulted in a number of regulations, including the so-called fairness doctrine, which, until 1987, required broadcasters to cover public issues and to give balanced coverage to competing viewpoints on those issues.

In 1969, the Supreme Court unanimously held that the fairness doctrine was constitutional in *Red Lion Broadcasting Co. v. FCC*.[52] The Court reasoned that the "scarcity of broadcast frequencies" made the spectrum a "resource" that justified government intervention to "assure that a broadcaster's programming ranges widely enough to serve the public interest."[53] In doing so, the Court recognized that the government had a proper role in securing the public's right to be informed.[54] Nine years later, the Court elaborated in *FCC v. Pacifica Foundation* that the "uniquely pervasive" nature of broadcasting, which may be seen or heard without viewer consent, supported the constitutionality of the regulation of indecent speech on broadcast media.[55] During that same period, the Court repeatedly rejected similar government regulations as applied to other media, including newspapers[56] and cable television.[57] Only a few years after *Red Lion*, the Court in *Miami Herald Publishing Company v. Tornillo* struck down a state law that granted political candidates who were criticized in a newspaper the right to have the newspaper publish their response to that criticism.[58] The Court also declined to extend *Red Lion* to cable television in *Turner Broadcasting Systems v. FCC*, reasoning that cable television lacks the scarcity of frequencies that characterizes broadcast media.[59]

The rationales behind the Court's deference to regulation of broadcasters—the scarcity of frequencies and the unconsented-to pervasiveness of the medium—do not map cleanly onto social media platforms. In *Reno*, the Court considered the broadcast analogy and its prior precedents and expressly declined to extend *Red Lion* to cyberspace. Its decision was, in part, based on its view that

unlike broadcast television, "communications over the Internet do not 'invade' an individual's home or appear on one's computer screen unbidden," and there are no significant capacity or bandwidth constraints on the free expression of ideas on the internet.[60]

As some have pointed out, the *Red Lion* Court's scarcity-based rationale is difficult to square with the economic realities of both broadcast networks and print media, which was true even when the decision was issued. The Court itself acknowledged that broadcast media had developed since their inception such that more stations could take advantage of the frequency spectrum.[61] Moreover, even at the time, monopolies made print media arguably an even "scarcer" medium than broadcast television.[62] On this view, the scarcity rationale was invoked by the Court to uphold what it viewed as necessary governmental intervention in mass media to promote the quality of public discourse.[63] Given the significance of social media to current public discourse, which arguably supersedes that of radio and television in the last century, a similar public discourse–enhancing motivation could be invoked in support of sustaining limited regulation of social media platforms' content-moderation decisions, consistent with the standard of constitutional review.

Nevertheless, *Reno* would present an obstacle to content-based regulation of social media platforms. Some commentators have asserted that *Reno*'s reasoning is outdated given the development of the internet since 1997. They argue that the pervasiveness and scarcity rationales that the Court concluded were not applicable to the internet have actually become more compelling since *Reno* was decided. Today, they would argue, the internet has become at least as invasive as traditional broadcast media. Social media are accessible anywhere from any internet-connected smartphone, and the largest social media platforms suggest content to users, making it more likely that users will be presented with the opportunity to view some content they have not sought out. Moreover, the social media landscape is now dominated by relatively few companies that control access to large audiences, which creates a kind of scarcity of social media communications platforms through which one can reach large audiences of users (although, of course, individuals may switch to smaller platforms or publish their own content through their own websites).[64]

Yet social media platforms remain in some respects quite different from traditional media. Broadcasters and newspapers operate in a one-to-many ecosystem and are responsible for generating their own content or selecting and editing content for dissemination to the public.[65] Cable and satellite operators also exercise editorial functions in deciding which channels to carry, and some have local channels for which they are, like broadcasters, fully responsible for the content. That is not true of social media platforms, which generate little of their own original content and do not, by and large,

substantially filter user content. Instead, the content is largely generated by users, and social media platforms' contributions are largely confined to algorithms that determine when and how frequently to display that content to other users. The volume of social media content uploaded to and disseminated through platforms is also orders of magnitude greater than the content printed in newspapers or aired on television.

To the extent that social media platforms display or disseminate large volumes of information without monitoring or mediation, they arguably resemble telecommunications common carriers such as broadband or telecom providers. Under Title II of the Communications Act of 1934, telecommunications common carriers—a regulatory classification that grew out of transportation carriers such as railroads and buses—may not refuse to provide service or carry a communication, except where the service is used for clearly illegal purposes.[66] In 2015, the FCC applied that Title II common carrier classification to providers of internet broadband access service, a category that does not include social media platforms.[67] The DC Circuit held that the FCC's classification decision did not violate the First Amendment rights of the service providers to the extent that they are acting as "mere conduits for the messages of others" as opposed to offering their own content.[68]

Social media platforms, however, differ significantly from common carriers or internet broadband access service providers in many respects, as the court in *Paxton* found when it expressly considered and rejected the analogy to common carriers.[69] Messages carried by common carriers and service providers are largely one-to-one (i.e., from a user to another individual or to access a website) and are kept in strict confidence by the provider. By contrast, social media platforms are largely designed to be one- or many-to-many, and their avowed purpose is to enable public exchanges among many users simultaneously or at different points in time. Moreover, unlike common carriers, social media platforms have terms of service that expressly prohibit certain types of messages and allow platforms to moderate or remove messages.

Notwithstanding these distinctions, some have called for social media platforms to be treated more like common carriers. In his 2021 concurrence in *Biden v. Knight First Amendment Institute*, for example, Justice Thomas suggested that common-carrier regulations may be warranted to address social media platforms' "private, concentrated control over online content and platforms available to the public."[70] He contended that like traditional common carriers, social media platforms "'carry' information from one user to another," and the largest platforms, like the largest utility companies, "derive much of their value from network size."[71] Classifying social media companies as common carriers seemingly would preclude regulations that would require social media platforms to actively monitor or moderate the messages they transmit on the basis of

content. For example, common-carrier treatment arguably would have prohibited Facebook from suspending Donald Trump's account in the wake of the Capitol riot and from deciding on an appropriate sanction by considering and weighing individualized, discretionary factors. Conversely, treating social media platforms as analogous to common carriers might require the platforms to carry and display all user messages without discrimination or moderation, which could even extend to voiding certain content-based limitations embedded in their terms of service.

Potential Policy Solutions

Self-Regulation by Social Media Platforms

Many major social media platforms have voluntarily adopted and applied policies and procedures to moderate harmful content, typically in response to current events and public pressure. Social media platforms that moderate content generally do so through three mechanisms: (*i*) removing harmful content and suspending or de-platforming repeat offenders; (*ii*) flagging inaccurate or harmful content; and (*iii*) implementing functionality or algorithmic changes to decrease the visibility of harmful content while making accurate and authoritative content more visible or accessible. These policies are generally implemented through a combination of artificial intelligence, employee review, and user reporting mechanisms. Because social media platforms are private actors, their self-imposed content moderation is not subject to challenge under the First Amendment.

Although many platforms have for years[72] applied mechanisms to stem harassment, bullying, and sexual exploitation, until more recently, they did not have significant policies aimed at curbing protected false or misleading speech or hate speech in the United States. The shift toward moderation of such speech came in response to events such as Russian interference in the 2016 US presidential election, mass protests for racial justice, and the spread of conspiracy theories and disinformation concerning the 2020 US presidential election and COVID-19.[73]

These policies have been criticized both for suppressing or chilling too much speech[74] (particularly speech by political conservatives)[75] and for doing too little to curb harmful speech.[76] The platforms' enforcement of such policies has been criticized as inconsistent and opaque.[77] At the same time, several smaller platforms—most notably Gab, Parler, and 4chan—have carved out a market niche by offering little to no content moderation; as a result, they have become havens for conspiracy theories, false information, and hate speech.

Establishing a Self-Regulatory Organization

The social media industry has generally advocated for maintaining an approach that relies wholly on self-regulation, including through a self-regulatory organization (SRO) that it establishes.[78] A social media SRO could have a number of benefits. To begin with, it could be established without congressional action. It could be nimble and adapt to changing technologies and practices more quickly than regulators could. In addition, inasmuch as participation in the SRO would be voluntary, it could regulate its members' content-moderation practices without raising First Amendment issues—unless the SRO is deemed to be a state actor because its activities are supervised by a government agency.

Insofar as participation in an SRO would not be mandatory, however, the adoption and enforcement of SRO-promulgated principles across the industry may not be uniform. If the larger players joined the SRO, that might address the largest volume of user-generated content. But many smaller players (which, like Gab and Parler, include havens for hate or other harmful speech) may choose not to do so, and their refusal to adhere to content-moderation standards could even help them grow their user bases.

There are several examples of successful self-regulatory bodies in other industries, with varying degrees of authority, that could provide useful models for social media self-regulation. The National Advertising Division of the Better Business Bureau resolves disputes over advertising.[79] The Digital Advertising Alliance establishes privacy principles for the digital advertising industry, thereby enhancing transparency for consumers and allowing them to control their privacy settings. Accountability and enforcement are carried out by the BBB National Programs and its Digital Advertising Accountability Program and by the Association of National Advertisers.[80] Finally, the Financial Industry Regulatory Authority, under the supervision of the US Securities and Exchange Commission (SEC), oversees the broker-dealer industry, creating rules, conducting examinations and market surveillance, bringing disciplinary proceedings, and referring matters to the SEC.[81]

An SRO for the social media industry might, for example, establish rules with respect to transparency of content-moderation policies; recommend approaches for assessing, mitigating, and reporting on risks from harmful content; impose reporting requirements regarding the type and volume of harmful content and the platforms' plans for mitigating it; require disclosures with respect to algorithms, the influence of foreign money, handling of user data, approaches to anonymity, and instances when users' identities might be disclosed; and facilitate dispute resolution between users and platforms with respect to grievances, including failures to adhere to stated policies. It might also issue penalties for noncompliance with terms of use, transparency requirements, and mitigation plans.

Legal and Regulatory Approaches

Policymakers and commentators have argued that industry self-regulation is inadequate to address the problem of harmful speech on social media and that comprehensive government regulation is needed. Any such regulations, however, must balance the impetus to rein in harmful content with free-speech rights, which are protected under international human rights law and domestic law in most advanced democracies.[82] In the United States, proposals have focused on either repealing or amending Section 230 immunity—which would allow individuals to sue platforms for unlawful content—or delegating regulatory authority to a new or existing federal agency to regulate social media platforms. The first of these approaches would have little to no impact on speech that is not unlawful, including hate speech and disinformation. The second could run afoul of the First Amendment if it limits or restricts First Amendment–protected speech or the platforms' protected editorial activities. The statutes need not target social media companies' content-moderation practices in order to raise First Amendment concerns; in *Paxton*, for example, the district court held that requiring platforms to make disclosures about how they curate content and publish transparency reports was unconstitutionally burdensome and likely amounted to compelled speech in violation of the First Amendment.[83] Actual and proposed legislation in the European Union goes somewhat further in regulating content. Such legislation may affect speech in the United States by creating soft norms that social media companies could choose to apply voluntarily to their US users.

Repealing or Amending Section 230

Repealing or amending Section 230 would permit actions against platforms arising from user-generated content that are now barred by Section 230. Put another way, repeal of Section 230 would allow suits against platforms for speech or conduct that is already unlawful or could, by reason of statute or the common law, be made unlawful in the future. Although it would not allow suits against platforms based on their dissemination of lawful content generated by their users, eliminating Section 230 immunity likely would have a significant chilling effect on social media speech. The scope of cognizable claims against social media platforms arising from unlawful content would be left to statutes and the courts, with First Amendment protections as the remaining constraint. In the short term, there could be widespread uncertainty and inconsistency across jurisdictions, leading social media companies to adopt the most stringent common denominator for content moderation in an effort to reduce their civil liability for content that is currently unlawful, such as defamation or invasion

of privacy.[84] The high cost of liability and of monitoring and removing content potentially could drive smaller platforms out of business or deter new entrants to the market, perhaps contributing to an increase in the dominance of large platforms.

Apparently recognizing these risks, many legislators and other policymakers have proposed amending or supplementing Section 230 rather than eliminating it entirely. Some of these proposed amendments are aimed not at curbing harmful speech but, instead, doing the opposite: dis-incentivizing platforms from moderating content, for instance, by limiting the types of content that interactive computer services may remove under Section 230.[85] The extent to which such proposals would actually have their intended effect is doubtful, because users do not have a legal right to disseminate speech—even lawful speech—via a social media platform. Accordingly, under current law and the platforms' terms of service, they do not now and would not have in the future a particularly robust cause of action against a platform for removing content absent some legal basis (such as racial discrimination).

Other proposals seek to incentivize platforms to remove unlawful speech by eliminating Section 230 immunity for civil claims based on violations of certain state or federal laws. Such proposals are necessarily limited to unlawful speech, because Section 230 applies only where the aggrieved party can bring a viable legal claim against the platform. Even within these limitations, these proposals could reach online harassment (including revenge porn or doxing to the extent that it meets the elements of state-law torts or might be outlawed by statute), certain forms of hate speech, and potentially tortious speech, such as defamation.[86]

A more modest approach, which is included in the Platform Accountability and Consumer Transparency Act (PACT Act) proposed by Senators Brian Schatz (D-HI) and John Thune (R-SD), is to amend Section 230 to create a notice-and-takedown process.[87] This would be similar to the notice-and-takedown regime of US copyright law, 17 U.S.C. § 512, which limits the copyright liability of an online service provider if it receives a notice that it is storing infringing material and then takes it down promptly. A measure of this kind would not put the onus on platforms to identify and remove unlawful content.[88] Instead, under that model, an internet content provider that did not otherwise know it was hosting or distributing illegal content would be entitled to immunity from suit for intermediary liability if and to the extent it acts expeditiously, on receipt of notice, to take down such content. By putting the burden on users to report unlawful content, a notice-and-takedown regime also alleviates at least some concerns about the compliance costs for platforms to monitor the enormous volume of content they store and disseminate for potentially illicit content and determine whether it is, in fact, unlawful. At the same time, this regime is vulnerable to abuse, and a meaningful counter-notice regime would need to

be instituted to ensure that content that ought not to have been taken down is restored.[89] In addition, policymakers would have to determine what would or should happen if the uploading user, in a counter-notice, asserts that the content is lawful. Finally, this approach, as noted, would not address harmful, but only unlawful, content.

Oversight by a New or Existing Government Agency

Some commentators have argued for express regulation of social media platforms by a government agency. The scope of agency authority could be akin to what is described above for SROs. Or the agency also could supervise the SRO, including by pursuing enforcement actions against platforms for persistent abusive practices—such as a failure to adhere to the content-moderation practices disclosed in their terms and conditions. This would be akin to the enforcement actions that the Federal Trade Commission ("FTC") has brought against social media companies for alleged failure to adhere to their stated privacy policies.[90] Under the First Amendment, the agency could not, however, regulate particular categories of lawful speech—such as false information or lawful hate speech—or discriminate among platforms based on their size unless such regulations were drafted to survive strict scrutiny (i.e., they are narrowly tailored to further a compelling interest).

Creating a new agency tasked specifically with regulating social media platforms would undoubtedly face political obstacles. Nonetheless, doing so would have the benefit of specialization. Social media platforms may also fall within existing agencies' authority. To the extent that they provide messaging services, they arguably might come within the FCC's authority to regulate information services under Title I of the Communications Act. But the current Communications Act would be an obstacle to most content regulation by the FCC because it is generally prohibited by law from regulating broadcast content, although there are exceptions for indecent and obscene speech, as well as other types of speech that are deceptive or promote certain types of illicit activity.[91] To remove any ambiguity as to the basis for an expanded reading of the FCC's jurisdiction to encompass platform regulation, an act of Congress probably would be warranted.

Alternatively, the FTC currently has jurisdiction over unfair trade practices and, under its existing authority, it has regulated widely in the analogous areas of personal privacy and data protection.[92] Establishing rules to mandate platforms' transparency in terms-of-use and content-moderation policies and to require adherence thereto arguably might fall within that authority.[93] For example, a platform's violation of its own policies, issued in response to the FTC's principles

or rules, could constitute an unfair trade practice in violation of Section 5 of the Federal Trade Commission Act, subjecting violators to monetary penalties or requirements to adopt new policies.

Comparing the United States and the European Union

Much as it has in the data-privacy context,[94] when it comes to social media content and operators, the European Union (EU) has developed a more comprehensive approach to regulation than has the United States. In recent years, the EU has adopted limited affirmative obligations for social media platforms. For example, 2018 amendments to the Audiovisual Media Services Directive require EU member states to enact national legislation requiring video-sharing platforms' terms and conditions to prohibit users from uploading hate speech, content related to terrorism offenses, child sex abuse, or other content harmful for minors. Covered platforms must also have mechanisms to allow individuals to alert platforms about such content.[95] In addition, in Germany, platforms with over two million users in the country must remove "manifestly unlawful" content within twenty-four hours of receiving a complaint about its presence.[96]

The EU also has adopted an approach of "co-regulation," whereby the regulatory role is shared between market participants and government, albeit with limited regulatory intervention. Most notable is the EU Code of Conduct on countering illegal hate speech online,[97] a nonbinding voluntary initiative among the European Commission, Facebook, Microsoft, Twitter, YouTube, Instagram, Snapchat, Dailymotion, Jeuxvideo.com, and TikTok. Under the Code, the participating companies have agreed to a number of public commitments, including adopting processes to review reports of illegal hate speech against their own rules and applicable laws, as well as to review the majority of valid reports in less than twenty-four hours.[98] Although there are no sanctions for failing to meet these commitments, social media companies are subject to periodic evaluations by anti–hate speech nongovernmental organizations[99] and public bodies[100] in EU member states, and the European Commission publishes statistics on notification assessment times, removal rates, and feedback to users.[101]

Near-final EU legislation—the Digital Services Act[102]—would go significantly further by imposing broad affirmative obligations on social media companies to take certain measures with respect to content on their platforms. Content that would potentially be subject to regulation under this proposal includes illegal harms, such as child sexual exploitation, terrorism, and hate crime offenses,[103] as well as legal content, including disinformation and harmful content for

vulnerable groups such as children.[104] In this respect, the Digital Services Act could create tensions with the fundamental right to freedom of expression safeguarded by Article 10 of the European Convention on Human Rights (the ECHR),[105] just as proposed social media regulations in the United States could conflict with the First Amendment. With that said, the ECHR permits somewhat broader limitations on free expression than does the First Amendment, for instance, by allowing for the criminalization of Holocaust denial.[106]

The Digital Services Act stops short of requiring platforms to actively search for and remove harmful content; instead, it would create a notice-and-takedown regime for unlawful content[107] and impose a range of transparency and reporting requirements on platforms. These include requirements to describe content-moderation policies in user terms and conditions, including algorithmic decision-making and human review;[108] publish annual reports regarding the platform's content moderation practices[109] and "significant systemic risks" of operating their services in the EU, including with respect to dissemination of illegal content and intentional manipulation of the service; and implement "reasonable, proportionate and effective measures" to address those risks.[110] The Digital Services Act is awaiting final approval from the European Parliament and EU Member States, and will apply from the later of fifteen months after its entry into force or from January 1, 2024. For very large online platforms and search engines, however, the Digital Services Act will apply from a potentially earlier date, being four months after their designation.

Once adopted, these reforms may have some influence on the US approach to regulation of social media platforms—given the cross-border nature of their services—and may have a more direct effect on US users' speech. In particular, the Digital Services Act may establish new soft norms that social media platforms will apply voluntarily to US users. After the adoption of the General Data Protection Regulation (GDPR), for example, Facebook extended many of its privacy protections to US users.[111] Similarly, Reddit allows users to request a copy of information that Reddit has about their account, as required under the GDPR and the California Consumer Privacy Act, even if the user is not in the EU or California.[112] Social media companies that are required to comply with the Digital Services Act may ultimately adopt a global approach to content moderation, such as notification of content removal, providing the reasons for removal, permitting users to challenge that removal, and providing transparency with respect to these measures. One reason for them to do so might be both legal and pragmatic: to ensure that they get the benefit of EU-based liability protection even with respect to content uploaded from outside the EU that is the subject of a complaint from an EU-based service recipient.

PART ONE

AN OVERVIEW OF THE PROBLEM

1

Social Media and First Amendment Fault Lines

DAVID A. STRAUSS

The popularity of social media has exposed fault lines in the law of the First Amendment. These fault lines are principles that are familiar, established, and foundational to First Amendment law. They present no difficulties in a wide range of cases. But, for one reason or another, they don't entirely make sense. Sometimes they rely on questionable factual premises. Sometimes their normative bases present an unresolved conflict. Sometimes they extend a reasonable claim to the point of implausibility. Sometimes their scope is chronically unclear. They are routinely stated as axiomatic, but they actually allow for exceptions, and the exceptions are not well defined.

These fault lines reflect genuine problems, not a failure of clear thinking on the part of courts or anyone else. They exist because of genuine dilemmas or tensions; any system of freedom of expression would have to deal with them and would have trouble solving the problems they reflect. And just as people can live for a long time without incident near geological fault lines, the constitutional law of free speech works reasonably well even though it has been built on top of these fault lines. But, like geological fault lines, these foundational aspects of the law can come apart when placed under stress. The widespread use of social media places them under stress. That does not mean the system will fracture. Relatively modest shifts in the law might relieve enough of the stress to allow things to continue pretty much in the way they have been. But the alternative outcome is also possible. The development of social media might require a large-scale rebuilding of some or all of these foundations.

To put the point another way, many of the problems that social media present for the First Amendment are not new problems. They were already there, and we were already dealing with them, in one way or another, although maybe not in a

completely satisfactory way. The prevalence of social media requires us to confront them more directly.

This way of approaching the relationship between social media and the First Amendment might enable us to avoid the dichotomous question whether the First Amendment is obsolete in the age of social media or is, instead, essentially up to the job.[1] Instead, we can deal with that relationship at retail—considering which specific principles need to be discarded, which should be modified, and which should be reaffirmed. But at the same time, being aware of the vulnerabilities in existing law might liberate us to make more dramatic changes than we otherwise would. We can be less worried about bringing down too much of the structure when, in fact, parts of the structure were already a little shaky and arguably should be reconsidered in any event. By focusing our attention on fault lines that already exist, the use of social media might become not just a problem but an opportunity, a reason to rethink certain aspects of the law even in familiar areas.

Here are some examples of First Amendment fault lines:

1. Government regulation directed at speech, not private action, is the principal threat to freedom of expression.
2. Statements of fact cannot be prohibited just because they are false.
3. Face-to-face encounters between individuals are an especially valuable form of expression.
4. Government speech—as opposed to government regulation of private speech—in general presents no First Amendment issues.
5. The best remedy for harmful speech is not suppression but speech that responds to the harmful speech.

Line-drawing problems are ubiquitous in the law, and clear rules will produce distinctions that seem arbitrary; those are familiar points. So it is not surprising to find principles like these that are often stated in categorical terms but that, in fact, allow for unacknowledged or vaguely defined exceptions. In particular, there is a tendency—understandable and beneficial—to try to distill First Amendment principles into clear rules, because clear rules make individuals more confident about when they can speak freely and make it easier for officials, including judges, to resist the pressure to suppress unpopular speech.

Clear rules can be fragile, though. Circumstances can change in ways that make the rules seem too arbitrary and that increase the pressure to create exceptions. That seems to be what is happening to these First Amendment principles with the growth of social media.

Government Versus Private Action

By its terms, the First Amendment, like nearly every provision of the Constitution, applies only to the government. But even if there were no First Amendment, a good argument can be made that government action should be the main concern of any system of freedom of expression. The government ordinarily has a greater capacity to suppress speech than any private entity. Government officials have an incentive to suppress the speech of their political opponents. And, to the extent the government reflects popular sentiment, the power of the government can be used by a dominant majority against nonconforming expression. So it is not surprising that government action directed at speech has been the central concern of the First Amendment.

But it is also clear that private action can effectively inhibit speech in ways that are not that different from government action. This can happen on a small scale, if, for example, private employers discriminate against employees because of their views. It can also happen on a larger scale, if private firms control large parts of the economy that are central to expression. And private individuals, without the help of the government, can effectively punish speech by acting in a hostile way, including by abusive speech, toward people who say things that they consider unacceptable. John Stuart Mill's *On Liberty*—unsurprisingly for a tract written in Victorian England—portrayed those kinds of privately imposed social sanctions as at least as great a threat to nonconforming expression and conduct as government action is.

First Amendment doctrine has not been entirely insensitive to these concerns about private entities. But it has not quite figured out what to do about them. That is why this is a fault line, and one that is increasingly tested, in various ways, by social media. Social media enable a speaker to reach a large audience easily and cheaply. That enhances the power of private speakers, other things equal; if private speech can cause harm, private speakers using social media can cause more harm than they could before.

The growth of social media also increases the power of media firms. Social media are characterized by network externalities—that is, the more users there are, the more benefit each user gains from a network. As a result, the industry will tend to be concentrated. One or a few firms will be dominant. The decisions of those firms about what speech to allow will affect both speakers and potential audiences. When those firms restrict speech, the effect is not exactly like government censorship, because there will be other (albeit presumably less effective) ways for speakers to reach their audience—by legacy media or potentially by a competing social media firm. But even government censorship is not entirely effective; it can be evaded, especially in a society that is generally free. So the

difference between government decisions and the decisions of private social media firms is not as great as is suggested by the axiom that government, not private, action is the main threat to free expression.

The straightforward way to deal with this problem would be to regulate private parties in a way that protects free expression. In the US legal system, that regulation could take one of two forms.[2] It could be done by applying the First Amendment to some private entities, on the ground that those entities exercise power in a way that makes them so similar to the government that they should be subject to the same rules. Or it could be done by allowing legislative or administrative action that, while not simply imposing First Amendment principles on private actors, prevents them from interfering too much with other individuals' speech. Both of these approaches have antecedents in the law, and both have advocates.

But both present significant problems. Among other things, allowing regulation of this kind threatens the free expression of the private actors themselves. In the name of protecting individuals' free expression, for example, social media firms might be required to communicate views that they do not want to communicate or might possibly be forbidden from communicating views that they do want to communicate. If this were done by statute or regulation, it might, under current law, itself violate the First Amendment, depending on what form the regulation took.

The government can, of course, regulate social media firms, just as it can regulate other firms, by requiring them to comply with laws unrelated to speech. Like everyone else, social media firms have to respect property rights, fulfill contractual obligations, pay their taxes, comply with antidiscrimination laws, and so on. Because those laws are unrelated to speech, they do not normally raise First Amendment issues. To some extent, the government might address excessive private power over expression by laws that are not (at least not explicitly) directed at speech. The antitrust laws are an obvious example, because they directly address social media firms' market power and therefore, depending on the market in question, the power those firms have over expression. Even then, there are issues: It is difficult to prevent the government from selectively enforcing laws on the basis of illegitimate criteria, because the courts give government officials discretion to choose the targets of their enforcement efforts, and the government can usually show plausible legitimate reasons for its decisions. So there is a risk that the government would selectively enforce the antitrust laws against firms (social media or legacy media firms) because it was unhappy with the speech they were circulating. But this is not a new problem.

The new problem is that regulations that are not directed at speech are unlikely to be enough to address the power that social media firms have over expression. Many of the concerns about private power over expression that are distinctive

to social media would require the government to entangle itself directly in the regulation of speech in ways that would be, at the very least, highly suspect under existing law. That is true, for example, of concerns that social media firms' content moderation practices exclude too much speech for illegitimate reasons or, on the other hand, should exclude more speech because it is false or abusive; that social media firms misuse customers' private data; or that the algorithms those firms use to direct material to subscribers have damaging effects. In order to address these concerns, the government would probably have to do things that are, under settled doctrine, almost always unacceptable—regulating the content of private parties'—the social media firms'—speech.[3] But the government would do that in order to protect the system of free expression from what might be even greater damage. That, again, is why this is a fault line, created by a genuine problem: The price of preventing private entities from abusing the system of free expression is to create the risk of government abuse.

This fault line has presented problems before. Probably the most famous example is the Federal Communications Commission's regulation of the content of radio and television. In *Red Lion Broadcasting Co. v. FCC*,[4] decided in 1969, the Supreme Court upheld the FCC's fairness doctrine, which required that "discussion of public issues be presented on broadcast stations, and that each side of those issues must be given fair coverage."[5] *Red Lion*, in allowing the FCC to regulate the content of broadcasters' speech in these ways, diverged from the principles that governed the regulation of other media; the Court refused to apply *Red Lion* to newspapers and cable television.[6]

The Court's justification for treating radio and television differently was that broadcasters were using a scarce medium that was of necessity licensed to them by the government. But all media are scarce, and the Court's claim that licensees owe a special obligation to the public begged the question. There are more plausible justifications, for example, that it was a good idea to allow two different models of free speech regulation to coexist[7]—in effect, a different approach on each side of the fault line—in the hope that experimentation with each would vindicate one or the other, produce a synthesis, or persist as a satisfactory bifurcated status quo. But later Supreme Court decisions drew *Red Lion* into question, and the FCC ultimately rescinded the fairness doctrine, asserting that it was unconstitutional.[8]

Rejecting the fairness doctrine, though, just amounted to grabbing one horn of the dilemma, by denying that private power over expression is an important enough problem to risk allowing government regulation. The fault line remains. In fact, there are other well-established regulations of speech that, perhaps less obviously, restrict the speech of some private entities in order to protect the free expression of others. Some jurisdictions, for example, have enacted measures that forbid private employers from expressing themselves in ways that might

coerce employees to refrain from political activity.[9] Similar issues might be presented by laws that prohibit employers from discriminating on the basis of an employee's political views, thereby limiting the employer's ability to dissociate itself from views it rejects. Historically, some communications media have been treated as common carriers, requiring them to transmit speech with which they may disagree; there is an obvious analogy to social media firms.[10] In general, to the extent that the growth of social media makes it obvious that private control over expression is a problem, this dilemma, unresolved in the *Red Lion* era, will reassert itself.

False Statements of Fact

The Supreme Court's opinions are full of statements to the effect that false statements of fact "have no constitutional value"; in fact, they are "particularly valueless" because they defeat, rather than advance, the purposes of a system of free expression.[11] They can corrupt public deliberation, impair the search for truth, and injure the individuals to, or about whom, the statements are made. There are several categories of false statements that are, without question, unprotected by the First Amendment: perjury, commercial fraud, defamation (if other requirements are satisfied), and false statements to a public official.

At the same time, though, speech cannot be prohibited just because it is false. The Court so held in *United States v. Alvarez*,[12] in which it reversed the criminal conviction of an individual who violated a federal statute, the Stolen Valor Act, by falsely saying that he had received the Congressional Medal of Honor. But even apart from that decision (and whether or not it was correct), there are important reasons to limit the government's power to forbid false statements. Arguable falsehoods are common in, among other settings, political debate, and if false statements were illegal, courts would routinely decide whether to punish people who made allegedly false statements about issues that are being publicly debated. The danger of chilling valuable speech, and the risk of abuse, are obvious.

That is the argument for First Amendment protection of false statements, though: not that those statements are valuable, but that while false statements are damaging, suppressing them would be worse. The premise of that argument, usually implicit, is that the falsehood will be exposed, not always but often enough to reduce the damage to acceptable levels. But of course, it is not clear how often, or when, that is true.

Current First Amendment law tries to draw a line between the two unacceptable polar positions—allowing unrestrained false speech or allowing officials to punish speech that they determine to be false—by permitting false statements

to be forbidden only when they fall within one of the familiar categories, like commercial fraud and defamation. But widespread falsehoods about, for example, the efficacy of a public health measure, the integrity of an election, or the justifications for military action can be more damaging, and less susceptible to correction by answering speech, than the harm caused by the types of false speech that can be forbidden. Competitors can address commercial fraud, and litigation adversaries might expose perjury, but the effects of widely circulated falsehoods about matters of great public importance will often be impossible to overcome. That is why this is a fault line: The line that defines when falsehoods can be forbidden is reasonably clear, but the justifications for it are questionable.

This fault line, like the one at issue in *Red Lion*, has manifested itself before. Take, for example, allegedly false statements by a firm that its employment practices in developing nations conform to humanitarian standards. Those statements have some of the characteristics of commercial speech—they are a kind of advertising—but also some of the characteristics of political speech. Fraudulent commercial speech can be prohibited; false speech on political issues cannot. So a claim that those statements are false falls on the fault line. When confronted with such a claim, the Supreme Court first said that it would decide the issue but then avoided it.[13] To some extent, defamation cases grappling with the question of whether a literally false statement should be regarded as hyperbole or opinion, and therefore not actionable, present the same kind of problem.[14]

Social media exacerbate this problem. On social media, anyone can make false statements, in large numbers, and many people can promulgate them quickly to a large audience. Legacy media gave fewer people that ability. In addition, on social media false statements can be directed to a specific audience, either by the speaker or by a social media firm's algorithms, to a greater extent than they could have been before. People have limited attention spans; if they are flooded with false speech, they will have a more difficult time sifting through it, and the risk of harm is greater.[15] False speech on social media, for these reasons, can cause more harm than false speech on legacy media, other things equal. The question is whether it is worth the harm, in order to avoid the risks that would inevitably come with empowering the government to forbid falsehoods. That question, which has been with us all along, becomes much more difficult.

Speech Directed at Specific Individuals

A face-to-face exchange of views seems like a core example of what a system of free expression should encourage. The Supreme Court has said that "normal conversation and leafletting on a public sidewalk" have "historically been more

closely associated with the transmission of ideas" than other forms of expression, and that " 'one-on-one communication' is 'the most effective, fundamental, and perhaps economical avenue of political discourse.' "[16] Face-to-face speech is especially valuable, the Court suggested, because it can compel an individual to confront speech he or she does not want to hear—"speech he might otherwise tune out."[17]

But face-to-face speech also presents special problems. Speech that would be run-of-the-mill advocacy if delivered to a public audience can be bothersome, harassing, disturbing, and even menacing when spoken face to face, especially when the message is unwelcome, and especially when the speaker persistently targets a specific individual.[18] Abusive speech spoken face to face has some of the same characteristics as a physical assault. As a result, the Court has allowed restrictions on face-to-face speech that would not be permitted otherwise—although generally without acknowledging that it is doing so.

For example, "fighting words," which under settled law are unprotected by the First Amendment,[19] are "personally abusive epithets" that are "directed to the person of the hearer."[20] But the same words that are unprotected fighting words if spoken face to face can be fully protected if they are expressed publicly as part of a political statement, for example, rather than directed at a specific individual.[21] Similarly, face-to-face solicitations, commercial or otherwise, can be regulated in ways that solicitations conducted in other ways cannot be.[22] In some circumstances, face-to-face speech can be restricted near a healthcare facility, but the same speech would be allowed if it were spoken at a distance or printed on a sign displayed nearby.[23]

Part of the reason face-to-face speech can be restricted has to do with physical proximity, which is not a concern with social media. But another part of the problem with face-to-face speech is that it is directed to a specific individual. Threats, for example, are not protected by the First Amendment, whether or not they are delivered in person. One way to understand fighting words is that they are a form of verbal violence—racial epithets are an example—even if physical menace is not part of the picture. Face-to-face commercial solicitation can be coercive because the recipient feels trapped in a one-on-one encounter, even in the absence of an implied physical threat. Courts have upheld laws forbidding harassing telephone calls, even if physical threat is not an element of the offense.[24] Sexual harassment, which is illegal under Title VII, need not involve face-to-face contact, and, as with fighting words, speech that would constitute sexual harassment when directed at specific co-workers might be protected expression if directed to a larger audience.[25] Recently the Supreme Court, while holding that the First Amendment forbade a high school from disciplining a student who made profane comments about the school on social media, made a

point of acknowledging the school's interest in preventing "bullying or harassment targeting particular individuals."[26]

Social media make it easier for someone to direct abusive speech at a specific individual more relentlessly than was possible before. There are two reasons for this. One is the familiar point that social media reduce the cost of speech to the speaker. Posting abusive messages on social media is easier than making repeated harassing phone calls or physically locating a target in order to engage in face-to-face harassment. Beyond that, though, social media allow a speaker to multiply the number of people who can direct abuse at a specific individual. That is because social media not only enable a speaker to communicate, at very low cost, with a large audience; social media also enable members of the audience to coordinate with each other much more easily.

In important ways, this aspect of social media—the ability of a speaker to mobilize a large, coordinated audience—is a good thing for the system of freedom of expression. The First Amendment protects freedom of association, which is not mentioned in the text of the Amendment, precisely because speech can be more effective when people coordinate with other speakers. Social media networks, both because they make speech inexpensive and because of network externalities, make it easier to exercise the freedom to associate with like-minded individuals. Social media make it easier for people to, for example, coordinate in expressing their opinions to politicians or to organize collective action of other forms, such as a boycott of a firm. Because social media are a decentralized way of coordinating speech, they can, famously, make it more difficult for authoritarian governments to suppress dissent.

But the dark side is that social media make it easier for people to engage in the virtual equivalent of mobbing—ganging up on victims in order to bully or harass them. Social media enable harassers to engage a much larger mob and to reach a much larger audience than they otherwise could. Even when the harassment just consists of speech, abusive speech, even if it is fully protected by the First Amendment, can inflict real harm on individuals. Beyond that, speech on social media that is unprotected—because it is threatening or defamatory, for example—will often be impossible to deter or punish, because speakers on social media can be difficult to identify and because there is safety in numbers.

In a sense, this is another way in which social media parallel government in their ability to inhibit speech. One way a government can threaten dissident speech is by providing a way for an intolerant group of people to coordinate its attack. Homogeneous communities could inflict social sanctions in the predigital world without involving the government, but social media enable many more such "communities" to form when people who otherwise have no connection—except, perhaps, that they are the social media followers of a prominent

person—discover that they have a common target. Without involving the government, they can make life miserable for individuals whose speech they dislike.

The targets might sometimes avoid online harassment by limiting their participation in social media, although even that might not help, if, for example, the harassment reveals private information about their families or where they live. But, in any event, participating in the communal aspects of social media can be important to people, and avoiding social media or significantly limiting participation to protect against harassment is costly. Telling people to avoid harassment by limiting their online presence might be not that different from saying that the target of sexual harassment should find a new job, or that someone who has been ostracized in a village can move elsewhere.

This remains a fault line, with free speech values on both sides of the issue. Outrage directed at people who violate social norms, even coordinated outrage, is a legitimate form of expression. In fact, it can be an especially valuable form of expression, because it can reinforce important values without creating all the dangers of government suppression. Other things equal, for example, social sanctions may be a better way of deterring hate speech than legal prohibitions. But, as Mill emphasized, coordinated private action directed at individuals can also create some of the same problems as government suppression.

And, again, social media make the problem more acute. The constitutional issues presented by speech directed at specific individuals may have been more or less manageable—in fact, they are so manageable that they have barely been acknowledged—when they consisted of face-to-face fighting words, unwanted telephone calls by a single individual, or sexual harassment by identified employees. In all of those situations, the ability of the speaker to direct speech at another individual was relatively limited. The speaker had to be in the presence of the target, work in the same place, or spend a lot of time and effort on the campaign of harassment; coordinating with others was difficult and costly if it was possible at all. These limits do not apply to anything like the same degree on social media; an individual bent on harassing another can do so with much less effort and can take advantage of the multiplier effect that social media networks enable.

Government Speech

Governments speak all the time—through their officials and employees, and sometimes through private individuals—and the government's speech takes sides on important issues. Government officials defend the government's own policies. They convey information to the public. They take positions on matters of general interest. All of that government speech is obviously legitimate—more

than legitimate; it is the government's responsibility to do those things. For that reason, the government's action in supporting a position by advocating it cannot be equated to the government's supporting a position by suppressing the opposing view. But, at the same time, government speech may have the effect—and sometimes has the purpose—of discrediting, and possibly effectively silencing, speech that takes a different view. In that respect, government speech resembles suppression.

That is what creates the fault line. The First Amendment, as currently interpreted, would ordinarily prohibit the government from criminalizing speech that vilified a religion, for example. But a government official who denounces religious bigotry in strong terms does not violate the First Amendment—even if the official is trying to shame the intolerant speakers into being silent or cause others to ridicule or ostracize them. As with other fault lines, this is a tension inherent in any system of free expression.

The current black-letter principle is that government speech simply does not present First Amendment issues.[27] But the Supreme Court has from time to time recognized that government speech can violate the First Amendment. A government blacklist, for example, is government speech, but a blacklist—which is specifically designed to discourage a potential audience from listening to a speaker—certainly seems as if it should raise First Amendment issues. The Supreme Court's treatment of blacklists has occasionally, but only occasionally, recognized this.[28]

Social media again make this problem more acute. In part that is because social media make it easier for a government official, like any other speaker, to reach a large audience and to propagate a message repeatedly. But it is also because social media remove the filter between a government official and the audience. A government official using legacy media has to depend on others to convey a message to a wide audience. Today, an official who wants to discredit a political opponent, for example—or to circulate misinformation in order to gain some personal or political advantage—may not have to worry about resistance from the editors of a newspaper or a radio or television station, or from the government communications bureaucracy. That official can take advantage of the prominence and credibility that come with being a government official and use Twitter or Facebook to convey a massive amount of speech to a large audience without any effective limit. In a sense, social media change the nature of government speech. To a greater extent than before, it might be the speech not of an institution but of an individual.

One possible solution, then, might be to distinguish between speech attributable to the government as an institution and speech that is properly attributed to a specific individual who happens to be a government official; speech in the latter category could, in principle, be restricted or sanctioned in whatever ways

private speech could be. That might conceivably allow an official who incites violence or commits fraud to be held liable for a tort or a crime. But this approach will have only limited effects. Federal government officials, for example, can be held liable for defamation, but only if the defamatory speech was not within the scope of their official duties.[29] And precisely because speaking is an important part of what government officials legitimately do, it will be very difficult for a plaintiff in a defamation action to establish that the speech was outside the scope of the official's duties.[30]

The problems created by government speech on social media are, to some degree, derivative of the problems that create other fault lines. A political leader, speaking in the name of the government, can propagate falsehoods that corrupt political debate. Even treating that speech as the speech of a private individual will not help, because falsehood alone does not justify suppression. And a falsehood propagated by a political leader can do more damage because it can be, for example, a weapon against political opposition. The same is true of speech directed at specific individuals. A person in power can use social media to attack a critic or a political opponent more effectively than legacy media would allow. These forms of government speech on social media can have effects that are nearly indistinguishable from suppressing dissent and punishing dissenters. Once again, a problem that current law treats in a way that may be roughly satisfactory—by treating government speech as if it presents no First Amendment problem—becomes much less satisfactory when social media are involved.

More Speech

One of the most rhetorically striking ideas in the history of the US system of freedom of expression—stated most famously in Justice Brandeis's opinion in *Whitney v. California*[31]—is that the way to deal with harmful speech is responsive speech, not suppression: "[T]he fitting remedy for evil counsels is good ones."[32] More specifically, "[i]f there be time to expose through discussion the falsehood and fallacies, to avert the evil by the processes of education, the remedy to be applied is more speech, not enforced silence."[33]

These statements identify a fault line. On one hand, they capture something central to any system of freedom of expression. But on the other hand, they are—if interpreted as claims about the efficacy of "good counsels"—often simply false. Obviously not every harm done by speech can be adequately answered by counterspeech. That is clear when, for example, the speech discloses information and the disclosure is itself harmful. But even beyond that, many unquestionably permissible restrictions on speech—laws forbidding defamation, incitement, fraud—would not be needed if "good counsels" could always remedy damaging

speech. And individuals—or, worse, a governing majority—can be persuaded to do terrible things that answering speech is powerless to prevent.

At the same time, though, Brandeis's formulation does describe something important about how a system of freedom of expression should operate. In a democracy, people must try to prevail by persuading their fellow citizens, not by silencing them; if the good counsels fail, and bad outcomes result, that is the price of democratic self-rule. The same principle applies outside of politics, when people might do harmful things: Suppression is a worse alternative. Sometimes damaging speech has to be prevented or punished, but in many contexts, suppression is so dangerous and harmful that, on balance, it is better, at least in certain circumstances—"if there be time to expose through discussion the falsehood and fallacies, to avert the evil by the processes of education"—to rely on counterspeech.

One implicit premise of this approach is that counterspeech will be effective enough, often enough, to justify the decision not to resort to suppression. The question raised by social media is whether that balance has to be rethought because social media make counterspeech less likely to be effective.

In some ways, the internet, including social media, could make counterspeech *more* effective. It can take much less time to expose "falsehoods and fallacies" when so many people can communicate with so many others so easily. "The processes of education" are—potentially—more effective in the internet age than they were when Brandeis wrote. In *United States v. Alvarez*, the case in which the Supreme Court reversed the conviction of a person who falsely claimed to have won the Medal of Honor, the plurality opinion described a concrete way in which the internet and social media could make counterspeech effective.[34]

But there are stronger reasons to think that social media will make counterspeech less effective. Social media allow people to cut themselves off from "counsels" different from what they want to hear; people have always been able to do that, to some extent, but social media allow people to make sure, to a greater degree than before, that their views are reinforced repeatedly and to seek out the most extreme and emotionally satisfying forms of reinforcement. Perhaps most important, social media firms' business models are based on giving people what they want. So while, in Brandeis's formulation, "the processes of education" provided by social media are in principle especially robust, the effects might be perverse. The speech may be abundant, but there is no reason to believe that it will generally be counterspeech. It will be speech that it is profitable for social media firms to direct to the individuals in question, or speech that the individuals have chosen, probably because it is agreeable rather than challenging. If so, "more speech," instead of being a "remedy to be applied," can make things worse.

* * *

The challenges presented by social media may force us to rethink the law of freedom of expression in fundamental ways; or it may be that those challenges can be comfortably accommodated without much change. But those challenges confer a collateral benefit. They force us to reexamine aspects of current law that are familiar and comfortable, because they work pretty well, but that were problematic in some ways even before social media became important. Those fault lines are where the tensions created by social media are most troubling. Whatever else might be said about the issues raised by social media, those issues may at least precipitate more self-awareness about the constitutional protection of free speech.

2

A Deliberate Leap in the Opposite Direction

The Need to Rethink Free Speech

LARRY KRAMER

Then and Now

A famous anecdote: Upon learning that the Constitutional Convention had concluded its business, a crowd gathered outside the Pennsylvania state house (today's Independence Hall) where the delegates had been meeting. As Benjamin Franklin left the building, a woman in the crowd asked him what form of government the Constitution would create. "A republic," he replied. Then, after a pause, "If you can keep it."[1]

A made-up anecdote: It's 1965. After a long day at work, you head to the local pub for a drink. A stranger sits down next to you, and you strike up a conversation. He's reserved at first, maybe even a little suspicious, but after a few drinks he opens up, confiding that he has inside knowledge of the biggest political scandal in American history. Eyes bright with excitement, he bursts out: "Lyndon Johnson and the leaders of the Democratic Party are running a pedophilia ring out of a D.C. pizza parlor!" Pleased by your astonished expression, he sits back. Then, turning serious again, he says, "I've got to get this story out."

"But how?" he asks. And now it's your turn to smile (inwardly). Because he's a crackpot, and you know there's no chance his story is going anywhere. He could self-publish, of course; he has the same free speech rights as any American. But all he could realistically accomplish on his own would be to produce an amateurish pamphlet that would reach, at most, a few hundred people in his immediate circle. He could offer the story to one of the nation's extremist political outlets, which might be willing to publish his outlandish claim. The Communist Party had a paper, *The Worker*, as did the John Birch Society, which published

American Opinion and *Review of the News.* These might have been delighted to put out a story like his. But their readerships were negligible, limited to the few people—ideological adherents—who went out of their way to subscribe to or buy journals like these. There also were supermarket tabloids—the *National Enquirer*, the *Globe*, the *Sun,* and such—but these, too, had small readerships, and all but a few of their craziest customers knew not to take them seriously.

The only way to reach a large audience in 1965 would have been through one of the major metropolitan newspapers or the three broadcast networks. But these prided themselves on practicing serious journalism. They might lean a little to the left or a little to the right—perhaps slightly more leaned left than right, at least outside talk radio—but all were by and large determinedly centrist in orientation, with high professional standards, and none would have aired this outlandish fiction. Bottom line: The demented man with the deranged idea that America's political leadership was trafficking children out of a pizza parlor would not have been able to get his story out to a mass audience.[2]

Flash forward a half century, to today, and it's a different story altogether. Today, the stranger's demented grandson would likely be boasting about how he *had* gotten his story into the mainstream. To begin, using technology like the Adobe Creative Cloud (which includes Photoshop, Illustrator, Acrobat, and PageMaker), he could cheaply produce something that looked both professional and believable—indistinguishable in appearance from conventional, respected news sources. Using the internet, he could then easily and just as cheaply get it out to millions of people. With relatively simple techniques, he could manipulate Google and other search engines to ensure that links to his story would pop up high on searches for things like "Democrats," "child trafficking," or even "D.C. pizza." Better still (from his perspective), with social media platforms like Facebook, YouTube, WhatsApp, TikTok, Instagram, Snapchat, Reddit, and Twitter, he could co-opt profit-hungry private enterprises to spread the story for him—relying on its salacious "clickbait" nature to trigger their algorithms (perhaps with an assist from some social bots he created), or simply letting gullible users spread the story for him through their platform networks. And as it spread, the very fact of its spreading would make the story newsworthy in a way the fragmented, competitive world of mainstream media could not or would not ignore.

The point is by now familiar: Changes in technology have produced radical changes in our speech and information environment. Free speech is no different than it ever was. The formal speech rights of our stranger circa 1965 were scarcely less robust than those of his grandson in 2021. What's different is how changes in technology enable speakers with those rights to reach vastly larger audiences—yielding a qualitative change in the nature and content of the "marketplace of ideas."

While the empirical point may be familiar, the moral of the story seems less so—namely, that qualitative changes of the sort we have experienced in the underlying structure of our information environment call for, indeed, necessitate, a similar qualitative rethinking of the law, rights, and norms of free speech that structure it.

Before explaining why this is so, I should add a couple of provisos. First is the obvious and important truth that the same changes that enable my disturbed stranger's unhinged grandson to poison the information environment with nonsense also enable good, smart, decent people to communicate more and more easily and to more people. The utopian claims made for the internet by its idealistic pioneers may have proved naïve (woefully so),[3] but there is absolutely no doubt that the qualitative changes it brought to our information environment have very much been for the better as well as the worse.

Second, the degrading of our speech and information environment has not been produced by the internet and new technologies, at least not in the first instance. Cultural and political conditions that make room for extremism have always been present in American society, and the particular upsurge we're living through now long predates the internet and has been evolving for decades. So, too, have institutions dedicated to spreading fake news and propaganda. Even with our new technologies, it's unlikely that a single individual could engineer a viral story like Pizzagate without help from well-funded boosters like Breitbart, Infowars, and Gateway Pundit.[4]

But outlets like these have always been around—recall the supermarket tabloids and Communist Party and John Birch Society periodicals from my opening vignette—and in the past they failed to thrive. What has changed, what is enabling the current generation of disinformation peddlers to succeed as they have, is the opportunity created by the internet and social media platforms. New technologies may not have created the problem, and they are not solely responsible for it, but they have *enabled* it—providing a necessary (indeed, indispensable) accelerant and catalyst for a change in degree so extreme as to amount to a change in kind.

None of this in any way gainsays the incomparable benefits the internet, search engines, and social media platforms have brought us by connecting people, driving economic growth, fueling innovation and creativity, and providing common access to so much of human knowledge. Which is, really, my point: These new technologies have *both* produced immense benefits *and* at the same time created fearsome risks and costs. For better or worse—or, rather, for better *and* worse—changes of the type these new technologies have wrought to the speech and information environment necessitate a fundamental rethinking of the legal and policy frameworks that structure and regulate it.

Explicit and Implicit Assumptions

In addition to the values they embody, all legal rights rest on empirical assumptions about facts and an analysis of costs and benefits, on the basis of which we determine the scope of the right and decide how it applies in different circumstances. Like any legal construct, rights have purposes, and achieving those purposes guides interpretation. So if a right's purpose is to promote some particular value or achieve some specific policy, we ask what interpretation of the right best advances that value or policy under the circumstances of a particular case (and cases like it). We also ask whether that interpretation will impose costs we should take into account, and whether there are better interpretations that might advance the value or policy as well or almost as well but with fewer costs, and so on. This kind of empirical analysis of how a rule plays out in the world is the heart of legal argument, reflecting the familiar principle that all rights have limits.

It's this kind of analysis that leads to conclusions like "one cannot falsely yell 'fire' in a crowded theater": It's speech, but not the kind of speech whose protection advances the purpose of the First Amendment, even as it can cause serious harm. All the other categories of First Amendment speech—fighting words, libel, political speech, commercial speech, art, obscenity, and so on—likewise rest on some version of this explicit empirical reasoning. As do other rights.

But rights also rest on *implicit* empirical assumptions that shape how we think about and understand their explicit costs and benefits. These implicit assumptions reflect aspects of the world we take as given, rendering them invisible to us as a practical matter. Unseen in plain sight, they influence the shape of the law, affecting the scope of our rights in ways we do not realize. Until they change. When that happens, everything about the balance we have struck in our explicit reasoning is thrown off, and we cannot continue relying on the same empirical understandings of costs and benefits. We need to rethink the whole structure of our analysis.

I can illustrate with a familiar example. In the eighteenth and nineteenth centuries, courts developed a body of law for wrongful acts—everything from harms caused by defective products, to accidents on the highway, to injuries in the workplace, and more. It comprised doctrines familiar to anyone who has taken a first-year torts class: negligence, duty, proximate causation, assumption of risk, and so forth. Anyone who has studied first-year torts is also familiar with the kinds of arguments lawyers made to shape these doctrines, which led to rules requiring privity for liability against product manufacturers, presuming that workers assumed the risks of injuries on the job, barring liability in cases of contributory negligence, and the like.

These arguments, and the rules they produced, epitomize the kind of explicit empirical analysis referred to above and constitute what we think of as ordinary legal reasoning. Yet these explicit arguments and rules were perforce nested within a world that shaped their consequences, which in turn shaped their perceived costs and benefits, which in turn shaped their formulation—but shaped their formulation in ways no one thought about or even recognized, because it was just how things were done. Markets were small and local, and most products were made locally and by hand; mass production on a widespread scale emerged only later. Travel was still largely by horse and carriage, meaning highways were not crowded, people were not moving all that fast, and collisions were relatively uncommon. No one took *these* sorts of facts explicitly into account in fashioning doctrine, because no one was thinking about them against alternative background conditions. No one had any reason to think about their role in shaping the law. It was just the way things were.

Until the second industrial revolution came along and with its new technologies changed everything. Mass production became common, consumer markets grew vastly larger, and suddenly the costs of defective products became significantly greater, even as supply chains and chains of distribution grew longer and more complex. The car was invented, and in a few short years, travel was transformed, with more people moving faster under circumstances in which collisions and injuries and deaths became markedly more common. Across the board, new technologies changed how people lived and worked. Urbanization accelerated and intensified, and work was transformed, with many more workers crowded into factories using newfangled machinery, where they began suffering injuries at previously unheard-of rates.

These fundamental transformations in how the world worked altered the *implicit* cost-benefit structure underlying tort law in ways no one could have envisaged when this law was formulated in the first place. It's one thing to require privity when most products are made to order and sold directly from producer to consumer, quite another when there is a long chain of distribution and identical products are mass produced and used by tens or hundreds of thousands of people. One cannot adapt to such dramatic change by simply massaging eighteenth- and nineteenth-century doctrines using classic reasoning by analogy, as if the changes are incremental and can be addressed using existing precedent. The shift from hand production to mass production, from horses to cars, from small shops to factories fundamentally transformed the world in which the law was situated. These were changes in kind, not just degree, and they shifted the ground on which existing legal doctrines and arguments stood. It became necessary to rethink tort law in elemental ways to conform to a new world—which is precisely what lawyers and judges and legislators did over the course of the

twentieth century, radically reallocating the risks and costs of injuries between those whose actions lead to harm and those who are harmed.

When tectonic plates collide underneath the earth's surface, and a flat plain becomes mountains and valleys, it's irrational to continue living and working as if still on the plains. New ways of living must of necessity be devised. So, too, for technological earthquakes that reshape how the world works in ways that transform the empirical ground on which our explicit legal analysis has been standing.

The Changing Context of Free Speech

Something similar has happened to free speech law, which has likewise seen the ground on which its defined categories implicitly stood altered radically by the internet and social media platforms. Consider again the not-so-hypothetical anecdote with which I began, the Pizzagate affair.[5] Prior to the advent of the internet, the technology of news distribution, combined with the curatorial practices of the major news organizations, was such that most Americans either would never have seen this story or would have seen it in a context where its preposterousness was evident (such as on the rack near the checkout line at a grocery store, or alongside the adult magazines at a newsstand). The structure of the information environment, at least as it emerged in the modern age (by which I mean post–nineteenth century), was such that the vast majority of people were exposed the vast majority of the time only to plausible news and information presented in a reasonably responsible manner.

Something extremist might occasionally break through to reach a mass audience, but this was rare and short-lived. Father Charles Coughlin was able to use radio in the early 1930s to reach a huge audience for his anti-Semitic rants and harangues in support of Hitler and Mussolini. He was effectively cut off after CBS canceled his show, however, and his audience shrank to insignificance once he was forced to finance himself. With no one else matching even his short-term success, Coughlin can be seen as the quintessential exception that proves the rule.[6]

Today, in contrast, new technologies and the information environment they have enabled make it possible for a disturbed individual or malevolent group easily and cheaply to produce authentic-looking materials and then get them distributed to a mass audience. Bear in mind, too, that the potential of these new technologies is not limited to nonsense like Pizzagate or QAnon. It is equally available—and used—to spread cleverly crafted propaganda, whether produced by a foreign adversary, homegrown fanatics, or profit-hungry troublemakers.[7]

The change in the possibilities for *production* is purely a matter of technology, which enables anyone with basic computer skills to produce a facsimile of legitimate material. The change in the possibilities for *distribution*, in contrast, results from how technology has altered the channels through which news is circulated and accessed. Prior to the internet, the major news producers were also its major distributors: The New York Times Company employed reporters who wrote stories for the newspaper, which the company then distributed for a fee to customers directly or via newsstands; CBS hired journalists who produced stories that were then aired on the network's news shows to be watched by its viewers. The rise of social media, and the increasing tendency of people to get their news online,[8] have essentially split the news production function—something still done by the *Times*, CBS, and other major mainstream media—from the news distribution function, which even for the major news producers (including Fox News) increasingly happens through internet-based search engines and platforms.

As responsible news producers, the major media companies of the twentieth century curated their content. They were willing to run stories that came from outside their organizations, often relying on wire services and freelance investigative reporters. But they chose which stories to run wearing the hat of a responsible news producer with a commitment to professional standards and ethics. This, in turn, led them to refuse to air either deranged inanities like Pizzagate or propaganda dressed up as news (by which I mean stories that present facts in distorted ways to stoke hatred or partisan outrage). It's why CBS canceled Father Coughlin's show in the 1930s. And since these major news producers were also the major news distributors, inanities and propaganda seldom reached an audience beyond the outer fringes of the information environment.

Search engines and social media platforms are different. With a commercial motive to maximize the amount of attention the content they distribute attracts, and no production function or professional journalistic commitments to incentivize control over its substance or quality, the new news distributors curate as little as possible—or, to be accurate, no more than they must. Suddenly millions of Americans who would not have been exposed to toxic forms of misinformation and disinformation are being inundated by it. Worse, inundated in forms that are difficult to distinguish from the legitimate news, because the signals that used to make this easy—appearance, source, manner of distribution, and the like—no longer work. It all looks the same and comes through the same channels. And worst of all, precisely because it has become so much easier and cheaper to produce material and distribute it widely, the sheer amount of fake and inflammatory information produced and circulated has increased exponentially.

The consequences, sadly, have been all too predictable. Millions of Americans, on both the left and the right, now believe things that are detached from reality

and often unhinged, especially on the right. And let's be clear: This is *not* because there is a shortage of good information or high-quality journalism. Traditional mainstream news services may have shrunk as a result of changes the internet forced on their industry's economic model, but there is still plenty of good journalism.[9] Plus, the reduced profession is now supplemented by semiprofessional and amateur citizen-journalists, bloggers, and newshounds generating stories that are not fake or inflammatory, because the easing of barriers to entry has encouraged the production and distribution of responsible material as well.[10] The problem is that media literacy is low and propaganda works, and good journalism—journalism reporting facts and providing rational analysis—is now mixed indiscriminately in a swamp of propaganda, clickbait drivel, and outright fake news.

Our inherited First Amendment doctrine tells us we must simply live with this, for our free speech theories and rationales all emphasize the dangers and drawbacks of regulating speech: People are autonomous individuals entitled to make up their own minds; to do so they need to hear a wide range of views, including especially opinions that are unpopular; this exposure is necessary for individual self-actualization as well as responsible democratic citizenship; protecting hateful ideas may actually increase tolerance; self-government and deliberative democracy depend on a free and robust public dialogue; and so on. Not that hate speech, disinformation, and propaganda pose no risks or inflict no harms. But in Justice Brandeis's memorable formulation, if that's the worry, "the remedy to be applied is more speech, not enforced silence."[11] I want to quote Brandeis's elucidation of the point at length, lest familiarity with the so-called counterspeech doctrine lead us to forget the genuine power of his formulation:

> Those who won our independence believed that the final end of the State was to make men free to develop their faculties; and that in its government the deliberative forces should prevail over the arbitrary. They valued liberty both as an end and as a means. They believed liberty to be the secret of happiness and courage to be the secret of liberty. They believed that freedom to think as you will and to speak as you think are means indispensable to the discovery and spread of political truth; that without free speech and assembly discussion would be futile; that with them, discussion affords ordinarily adequate protection against the dissemination of noxious doctrine; that the greatest menace to freedom is an inert people; that public discussion is a political duty; and that this should be a fundamental principle of the American government. They recognized the risks to which all human institutions are subject. But they knew that order cannot be secured merely through fear of punishment for its infraction; that it is hazardous to discourage

> thought, hope and imagination; that fear breeds repression; that repression breeds hate; that hate menaces stable government; that the path of safety lies in the opportunity to discuss freely supposed grievances and proposed remedies; and that the fitting remedy for evil counsels is good ones. Believing in the power of reason as applied through public discussion, they eschewed silence coerced by law—the argument of force in its worst form. Recognizing the occasional tyrannies of governing majorities, they amended the Constitution so that free speech and assembly should be guaranteed. . . .
>
> Those who won our independence by revolution were not cowards. They did not fear political change. They did not exalt order at the cost of liberty. To courageous, self-reliant men, with confidence in the power of free and fearless reasoning applied through the processes of popular government, no danger flowing from speech can be deemed clear and present, unless the incidence of the evil apprehended is so imminent that it may befall before there is opportunity for full discussion. If there be time to expose through discussion the falsehood and fallacies, to avert the evil by the processes of education, the remedy to be applied is more speech, not enforced silence. Only an emergency can justify repression. Such must be the rule if authority is to be reconciled with freedom.[12]

This is legitimately stirring rhetoric—but written in and for a world in which the risks posed by "falsehood and fallacies" were qualitatively different than today (though events already beginning to unfold in Germany would show the dangers to be, even then, frighteningly real). In this sense, the limited opportunities to distribute disinformation and propaganda before the internet were rather like the limited speed of transportation on horses before the car. So when Brandeis wrote these words, perhaps he could plausibly assert that "the fitting remedy for evil counsels is good ones," which today sounds about as convincing as Wayne LaPierre's remark that "the only way to stop a bad guy with a gun is with a good guy with a gun."[13] And in the mid-twentieth century, perhaps one could with a straight face say that "discussion affords ordinarily adequate protection against the dissemination of noxious doctrine." In truth, even then everyone knew how often that's not so, but we could indulge the consequences in a world in which the volume and reach of noxious doctrine was structurally limited.

Well, that's no longer true today. With the structural limits all but obliterated by new technologies, the costs of taking so much on faith have soared and can no longer be taken for granted. Technological developments have caused a tectonic shift in the implicit empirical assumptions that framed prior perceptions of the costs and benefits of our free speech rights and tolerances. Even the most

optimistic observers recognize the strains these developments have put on democratic societies. But the threat is greater than just some additional stress. Left unchecked, the presently evolving information environment must be expected to unmake our democratic constitutional systems. Perhaps not imminently, but with certainty over time.

If history teaches anything, it is that democracies are fragile things: difficult to establish and challenging to maintain. That was the purport of Franklin's cutting challenge, "If you can keep it." And while there is room to argue about the requisite practices a nation must embrace to be authentically "democratic," democratic governance is not possible as a practical matter absent at least two minimal conditions: First, the people who comprise a democratic polity must see themselves as part of a shared political community—willing to accept their inevitable differences and accommodate governance within and across them.[14] Second, given the difficult decisions a government must make and the harsh consequences these decisions can entail, there must also be either a reasonably well-informed public that shares a basic understanding of facts and is capable of understanding these decisions, or a well-informed elite to which the public is generally prepared to defer.[15]

These are not the only demands needed to constitute a democracy, at least a democracy worthy of the name. They are, rather, prerequisites for such a government, because without them popular government has no chance of enduring. Yet these are the very conditions that the new information environment is eroding: fracturing our sense of political community by turning disagreement over political and social issues into tribal differences among perceived enemies, destroying trust in elites and elite institutions, and actively and aggressively misleading and misinforming the public. We tasted the first fruits on January 6, 2021, which ought to serve as a wake-up call to the mainstreaming of extremism in forms that hitherto lingered on the fringes of American life. This is because the Capitol riot was no one-off event; belief in a laughably implausible conspiracy to steal the election persists and continues spreading, along with assorted other forms of (mostly right-wing) extremism that have found ways to flourish in the new information environment—via the internet, of course.[16]

These are consequences our system cannot and will not withstand for long. So if our intellectual and legal frameworks for free speech and the First Amendment significantly impede our ability to address them, those frameworks need to change. Note that I only say "change." Freedom of speech remains one of the requisite values and practices needed for a nation even to call itself a democracy. But with apologies to the Platonists out there, freedom of speech is not an abstraction that applies in an unchanging, universal manner at all times and in all places. In the real world of actual governance, its precepts and principles are invariably—and appropriately—shaped by context. And insofar as the

circumstances and context that shaped our prior understanding of free speech have changed—changed in ways that make this earlier understanding a mortal threat to the republic's survival—we need to rethink our stance. And, yes, of course we need to do that in a way that doesn't throw the baby out with the bathwater. But we do need to drain the bath lest the baby drown in the increasingly filthy water.

In *Terminiello v. Chicago*,[17] Justice Jackson was sufficiently apprehensive about the risks of unregulated speech to support punishing a suspended priest for an inflammatory speech. Though the vote was 6–3, neither of the other dissenters joined Jackson's dissent or shared his fears, and the general consensus has long been that Jackson's worries were overstated. But the scale of mischief Father Terminiello could cause speaking to an audience of eight hundred in an auditorium in 1949 is not comparable to what his successors can do today via the internet and social media, and Jackson's dissent seems prescient, if not downright prophetic, when considered in the very different context of our fundamentally transformed speech environment: "The choice is not between order and liberty. It is between liberty with order and anarchy without either. There is danger that, if the Court does not temper its doctrinaire logic with a little practical wisdom, it will convert the constitutional Bill of Rights into a suicide pact."[18]

Finding "a Little Practical Wisdom"

Few things send shivers down the spine of a constitutional lawyer like the suggestion that government should regulate speech. For while regulation is hardly unprecedented—think of commercial speech, defamation, obscenity, or child pornography, among other categories—our paramount speech commitments have for many decades now been fiercely libertarian. Indeed, the "marketplace of ideas" shielded by contemporary First Amendment jurisprudence is probably closer than any commercial or commodity market has ever been to a neoliberal's fantasized free market.

First Amendment values alone do not fully explain the deeply libertarian nature of our contemporary doctrine. Underlying this dread of even modest government regulation of speech are worries about the proverbial slippery slope. Let government begin censoring bad speech or coercing good speech, we are told almost without exception, and where will it end? Of course, awareness of the logical fallacy in this argument is practically as old as the argument itself: It will end where the rational distinctions used to justify taking the first step end. Nothing about regulating one form of speech because of the harms it causes necessitates allowing regulation of other forms that don't cause such harms or

that have greater benefits. But the force of a slippery slope argument doesn't lie in its logic. It lies in our worries about how the distinctions we draw will be used by other decision makers, who might have different motives or different attachments. The persuasive power of a slippery slope argument rests on fears that future decision makers will not recognize, comprehend, or defend doctrinal lines drawn by their predecessors.[19]

Whether overstated or not, such fears might offer reasonable grounds to resist taking even a first step when the costs of not taking that step are relatively low. Let Father Terminiello or Clarence Brandenburg blather their racist, anti-Semitic drivel, stepping in only if and when their rhetoric creates a palpable danger of violence in the moment. As contemptible as their speech might be, if the actual harm it threatens is limited and remote, maybe we can and should indulge it in the name of extra security for "good" speech or as a normative expression of values.

That proposition is being tested today in the debates over speech on college campuses. The controversy pits traditional First Amendment defenders and defenses against an argument that the counterspeech doctrine has failed. "More speech" has not done its part, say the advocates of limiting free expression, and we've tried it long enough to know that protecting hate speech just enables and perpetuates the violence of racism, sexism, and other like harms. Such injuries matter, and averting them warrants shutting some speech down. Not because freedom of speech isn't important but because the harms from restricting such despicable speech, even with the attendant risks, are not greater than the harms caused by protecting it.

I take no position here on who is right or wrong insofar as we are talking about traditional speech in traditional forums. Whatever the case may be on college campuses or in ordinary real-world settings, the nature and extent of harm enabled via unregulated speech on the internet is, for all the reasons explained above, qualitatively different and greater—both in enabling the spread of identity and group-based harms and in subverting the foundations of liberal democracy. In that medium, unwillingness to risk sliding down the slope of government regulation poses an equal or greater risk of sliding into harms that are no longer peripheral or insignificant on the opposite slope of unrestricted speech. In a world whose technology limited these harms "naturally," that may have been something we could chance. But the world is no longer structured that way. The structural limits are gone, and we need to rebalance our assessment of the risks, costs, and benefits of unregulated speech on the internet.

The task of rethinking First Amendment doctrine and rebalancing free speech norms online is huge, complicated, and beyond the scope of an essay like this. I do, however, want to offer a few thoughts about places to start and potential directions to take, as well as some concrete proposals.

First, if we do revisit First Amendment law and norms for the reasons and along the lines I've suggested above, we should still be thinking about taking small steps, at least initially. Opening our minds to possibilities for new approaches to freedom of expression is not the same as inviting a comprehensive, all-at-once overturning. Once new paradigms are fully formed, we come to see them as systematically different from what they replaced (though, even then, most of the key concepts remain, albeit in different configurations or with different weights). But that's only in hindsight, and typically only after an extended period of development. The shift in tort law described briefly above, for example, took decades and involved cumulative reinforcing action by courts, legislatures, and administrators. Changes in law or norms can be incremental while still being qualitatively different enough—fecund enough, if you will—to generate truly new ways to think about and address problems.

As suggested above, no one is looking, even in the long run, to entirely overthrow existing understandings of free expression in favor of something completely new and different. It would be astonishing if centuries of thinking and experience had it all wrong. No one, least of all me, believes we should abandon the core tenets of our current free speech law, which contain much that is wise and needs to be sustained. We nevertheless need to put that wisdom into a different context, built on different circumstances, and in doing so create opportunities for new approaches to policy.

The shift needed in a free speech paradigm for online speech is more like a renovation than a teardown. Our task is to preserve what is good in the old structure, while building something more suited to the needs of current occupants. It's also something we need to do room by room, so to speak, meaning that (unlike in ordinary construction) the final structure will take shape as we go along. Which is, of course, all the more reason to approach the process pragmatically: testing ideas, seeing whether and how they work, and building on those that do.

Second, there are some steps that, while not simple or uncontroversial, nevertheless seem obvious. Under our current law, the Supreme Court has recognized categories of low-value speech that may be regulated, but has defined the categories in ways that keep them very narrow. One might think that false speech would be low value, but falsity alone is not enough for legal culpability under existing law, and untrue speech—even intentional lies—can be regulated only in very limited circumstances, as in libel, fraud, perjury, and commercial speech. Most famously, the Court in *New York Times v. Sullivan* held that an injurious false statement about a public official can give rise to liability only if it was made with " 'actual malice'—that is, with knowledge that it was false or with reckless disregard of whether it was false or not."[20] In *United States v. Alvarez*, the Court not only declined to extend this rule to false speech more broadly, it ruled that

regulation of falsehoods is permissible only where there is a strong government interest in regulating a carefully defined category of speech that causes specific, material harm and is "'part of a long (if heretofore unrecognized) tradition of proscription.'"[21]

Justice Kennedy's plurality opinion offered no explanation for this austere conclusion beyond a feeble gesture at past practice: That's how we've done it in the past, so that's how we do it. Justice Breyer did better in concurrence. There are instances, Breyer noted, where a false statement may serve "useful human objectives," as where it prevents embarrassment, protects privacy, or comforts the sick. Plus, we must worry about borderline cases, where the truth or falsity of a statement is unclear, and the risk of punishment could "inhibit the speaker from making true statements, thereby 'chilling' a kind of speech that lies at the First Amendment's heart." Add to that fear of what government might do if empowered to prosecute false speech generally, and it makes sense to interpret the First Amendment to permit action against falsehoods only in narrow circumstances where specific, present harms are likely to occur and can be proved.[22]

Except that the anxieties Breyer expresses—the potential for there to be good motives or beneficial effects, the possible uncertainty about truth and falsehood, and the concomitant worry about chilling effects—are just as present, and in exactly the same ways, when it comes to the categories of false speech for which regulation *is* permitted. It was to address concerns like these that the *Sullivan* Court added the requirement of actual malice, limiting punishment to instances in which speakers know their speech is false or act with reckless disregard for its truth or falsity. There is no reason to believe this safeguard and the "breathing room" it provides would be less adequate when it comes to speech beyond libel, including speech on issues of public interest more generally. Which leaves the fear of government abuse if we permit regulation more broadly as the only persuasive justification for confining control of false speech to narrowly defined categories.

It's tempting to dismiss this predictable resort to the slippery slope argument with a Reaganesque "there you go again." But it's a reasonable position in a world in which the dangers from allowing false speech to circulate freely are small and remote. Where that is the case, it makes sense to weigh the balance in favor of permitting even this low- or no-value speech, and to permit the law to step in only when that speech causes concrete, material, present harms.

The same argument is less reasonable or sensible in a world in which the dangers of this laid-back approach are no longer small or remote: a world in which the diffuse harms that come from a regular diet of intentional disinformation are becoming manifest; in which no single speech act threatens significant injury or mischief, but the cumulation of many threatens catastrophe. In that

world, which is our world, the risk of government abuse remains, but it must now be balanced against a very real risk of even greater harms from an unregulated speech environment. Brandeis may be right that "those who won our independence were not cowards," but neither were they naïve Pollyannas. And neither should we be.

Developing doctrine here will not be simple. We need answers for some awkward situations—one colleague asked if I would punish a father for posting a video in which he tells his child that Santa Claus is real—but neither this kind of innocent or well-intentioned falsehood nor other problems we can expect to encounter are intractable. We could, for example, deal with problems like the Santa Claus example by limiting liability to speech about issues of public interest (much as *Sullivan* was limited to public figures). In any event, we must not let our comfort and familiarity with the current way of thinking—which is hardly timeless and itself emerged only in the mid-twentieth century to displace a very different free speech regime—squelch our ability to imagine new answers to difficult challenges. Working through hard cases goes with the territory whenever the law changes, and we should not let our capacity to hypothesize *reductiones ad absurdum* force us to sit helplessly by as the public discourse essential for liberal democracy is twisted into a tool for its undoing.

Third, while regulating false speech along the lines suggested above might address some dangers associated with the new information environment, much disinformation and dangerous speech involves statements that are not easily labeled "true" or "false" or that are "true" while nevertheless creating threats we need to address. Effective propaganda often relies on real facts presented in ways that are partial or cleverly distorted to trigger fear or incite anger. Hate speech, whether directed at individuals or groups, need not rely on "facts" at all. Breaches of privacy may reveal genuine information in ways or under circumstances that create significant harms. And so on. These and other problems we can imagine also require new thinking and solutions.

In tackling such issues, we can find valuable lessons in the practices of other nations. Americans have never been especially keen about seeking guidance from foreign law sources—at least not since the founding generation, which borrowed generously from foreign precedents.[23] And certainly there are reasons for reluctance about importing rules and norms from other cultures that derive from something other than a misguided belief in American exceptionalism. Yet donning Scalian blinders[24] to the actions and activities of other nations is a mistake, for we can surely learn a few things by looking to both international law and the regulatory practices of other liberal democracies.

There is, in fact, a great deal of experience upon which to draw, because the rest of the world has not approached free speech with the same all-or-nothing, in-or-out categorical approach used in the United States. In most of the world,

restrictions on speech are allowed if they advance sufficiently important state interests in ways that are not needlessly overbroad and (in the words of the Canadian Supreme Court) there is "proportionality between the effects of the measures [on speech] and that objective which has been identified as of 'sufficient importance.' "[25] Such an approach is not entirely unfamiliar to American law, which flirted with a similar balancing approach in the 1950s and 1960s.[26]

Of course, lessons from other nations at that level of generality are not likely to be particularly enlightening. What matters is how this structured proportionality approach has been used, particularly in combination with national and international texts that permit or, in some instances, even require, the regulation of speech. For example, the International Covenant on Civil and Political Rights provides in article 19(3) that the right of free expression carries with it "special duties and responsibilities" that make it "subject to certain restrictions" when provided by law "(a) For respect of the rights or reputations of others," and "(b) For the protection of national security or of public order (ordre public), or of public health or morals."[27] Article 20 goes a step further and requires the prohibition of "[a]ny advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence."[28] In like manner, article 10 of the European Convention on Human Rights, which in turn provides interpretative guidance for the EU Charter of Fundamental Rights, states that since the exercise of freedom of expression "carries with it duties and responsibilities," it

> may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society, in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary.[29]

Provisions like these, translated into statutes and decisional law in many nations, have in turn produced very different approaches from US law to some of our most difficult challenges.[30] These include different treatment, analyses, and outcomes respecting the regulation of hate speech (including group hate speech, Holocaust denial, use of hate symbols, and more),[31] privacy and data privacy,[32] elections and election-related speech,[33] "fighting words" and offensive speech,[34] speech related to religion (e.g., blasphemy, religious dissent),[35] regulation of media,[36] the use of prior restraint,[37] and the use of "human dignity" as an independent value both promoting and limiting freedom of expression.[38]

There is no need here to go through the details of how other nations treat speech in all these categories. Not only is there a great deal of variety, but my

point is not that other nations have found the "right" answers. Everyone is struggling with the same issues as the United States, though I think a decent case can be made that many of these nations are doing better than we are. Germany and France have hardly solved the disinformation problem,[39] but their efforts have plainly had at least some dampening effect compared to this country's. But only some, and clearly no one has yet figured out how to rebalance the costs and benefits of free expression in the internet age.

That's not surprising: The problem is still relatively new as these things go, and it is made doubly challenging by the speed with which the technology—and so the problem itself—continuously evolves. But other nations have a bigger toolkit with which to search for solutions, and we can and should learn from their efforts. For instance, most Western democracies criminalize inciting hatred against groups distinguished by race, gender, sexual orientation, and the like[40] or threatening, insulting, or degrading such groups.[41] In Germany, it is illegal to incite "hatred against parts of the populace" or call for "acts of violence or despotism against them," or to attack "the human dignity of others by reviling, maliciously making contemptible or slandering parts of the populace."[42] The Icelandic Penal Code includes public denigration of groups, which it defines as publicly mocking, defaming, denigrating, or threatening them.[43] In all these nations we can find precedents for using provisions like these to tamp down extremism in ways that preserve freedom of expression and promote tolerance.

Nor have other nations confined their efforts to rein in dangerous expression to hate speech. In France, a moratorium on campaigning and media coverage of elections goes into effect two days before an election, a law that likely seems incredible to most Americans—a friend once told me to calm down and breathe when I suggested such a law might make sense. Yet this provision has, in fact, been effective in mitigating forms of election disinformation that have subverted American elections without turning France into Oceania.[44]

I don't think anyone can honestly say that life in these other countries is meaningfully less "free" than life in the United States, that their people are less self-actualized, that their societies are less tolerant, or that their public debate is less robust. If anything, the opposite seems clearly true, at least with respect to tolerance and the quality of public debate. Hence, we might want to borrow and adapt some of these tools. Indeed, given the inherently global nature of today's information systems, we might want to work on the problem together. We might even *need* to do that.

Fourth, assume we find reasonably effective ways to redefine our free speech commitments, by which I mean rules and standards that appropriately balance the need for robust free expression with the need to tamp down the unique capacity of new technologies to spread hate, lies, disinformation, and propaganda. Whatever our new law looks like, we can enforce it by conventional means,

including both criminal sanctions and civil liability. But relying on litigation—whether criminal or civil, whether by the state or by injured individuals—to control the people and organizations intentionally spreading lies and hate may be less effective and is certainly less efficient than finding ways to head off mass distribution in the first instance. Which brings me to the question of platform regulation.

There has been considerable debate about whether social media platforms like Facebook, Twitter, and YouTube should be seen as "publishers" and so speakers with responsibility for the content they make available. The companies do curate some content for themselves—I am thinking here of features like Facebook's News tab or Apple News—but those services are not a source of any serious controversy. The problem that threatens our democracy and provides the reason for a book like this grows from the social media companies' willingness to allow practically anyone, with minimal effort or scrutiny, to use their platforms to circulate material to whomever they can reach—a reach the companies then extend with algorithms designed to push material out to everyone the companies' data suggest might give it some attention.

Even here, the platforms curate by enforcing community standards that ban things like explicit sexual content, depictions of graphic violence, and, more recently, certain kinds of misinformation around elections and vaccines. Not surprisingly, they also respond to episodic bursts of intense public pressure, as in Facebook's and Twitter's banning of Donald Trump after the January 6, 2021, Capitol riot. Facebook has gone so far as to create an independent Oversight Board with authority to review individual Facebook and Instagram decisions on content moderation, though it carefully cabined the board's authority to ensure company control over policy decisions with broader consequences. This reflects the general approach taken by all the platforms, which is to moderate content no more than they must, mainly for public relations, using rules that vary from platform to platform but are designed by and large to maximize their capacity to serve as a passive conduit for user-generated content.[45] If we need a metaphor to describe the role of these unique institutions in the news ecosystem, the most accurate label may be news *distributor*: The social media companies are, in effect, the largest newsstands and news delivery services in human history, able (and willing) to deliver practically anybody's "news" to the whole world in order to earn advertising revenue.

As anyone who has paid even slight attention to the issue knows, Congress decided in the Communications Decency Act of 1996 (CDA) to protect online intermediaries from legal responsibility for third-party content they distribute by providing in Section 230 that "[n]o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information

provided by another information content provider."[46] That decision makes sense when framed, as it was, as an all-or-nothing choice between liability and immunity. Given the enormous number of users and massive scale of these operations—nearly three billion people are on Facebook, YouTubers upload five hundred hours of video every minute, Twitter posts forty million tweets a day, and so on—imposing the risk of liability for anything unlawful on their platforms would saddle the companies with an impossible burden and threaten to put them out of business.

Unfortunately, an enforcement regime that ignores the platforms and limits itself to action against the direct producers of prohibited material is likely to be ineffective. The originators of unlawful online speech can be difficult to identify and hard to reach even when identified, not to mention judgment-proof or beyond service of process. Engaging social media companies in the task of sorting lawful from unlawful speech and not letting their platforms be used to spread the latter may thus be necessary—*if*, that is, we can figure out ways to do so that are reasonably effective (which includes being sufficiently protective of lawful speech) and can be operationalized in practice.

This, in turn, calls for some creative thinking and willingness to experiment. Here, I'll offer a couple of ideas of things we might consider trying. These will not address the problem in its entirety. Rather, they are offered as illustrations to challenge settled thinking and suggest directions we might take. Both may be useful even under current rules for freedom of expression, though they would be more effective if we broadened the law to reach more harmful speech along the lines suggested above.

1. One possibility is to take a different approach to intermediary liability. Congress was not wrong, when it enacted CDA § 230, to worry that companies cannot realistically keep watch over hundreds of millions of daily postings. There is disagreement about when—or even whether—artificial intelligence will have natural language abilities good enough accurately and effectively to perform that sort of task, but even techno-optimists acknowledge that it's still a long way off.[47] Yet while unlawful communications directed to or at individuals or small numbers of people are bad, our biggest concern when it comes to disinformation and propaganda is with communications that go viral and reach mass audiences. Half a loaf being better than none, we could limit the threat of intermediary liability to this subset of communications, which is readily manageable. Indeed, the large platforms already have virality tripwires they use to monitor popular content for violations of their existing community standards.

One could do this by tweaking the damages rule for intermediaries, using what we might think of as microliability. Imagine, for instance, statutorily limiting damages to one penny for each post, like, or retweet. That would mean total damages of $10,000 for an item that reached a threshold of one hundred

thousand posts, likes, or retweets, $100,000 for items that reach one million, and so on. Platform companies could then make their own decisions about how risk-averse they are and so how much monitoring they do. We could, in turn, adjust the definitions and numbers with experience to achieve an appropriate and workable level of enforcement. This would, among other things, build in a process for responding to the ever-changing ways in which technology can be misused by incentivizing the companies themselves to adapt in light of whatever threshold we set for meaningful liability.

My suggestions for what to monitor (posts, likes, tweets) and how to size liability (one cent for each) are offered to illustrate the concept. Perhaps we should focus on "views," or maybe the liability rule should be ten cents per incidence rather than a penny. How and where to set the threshold depends on what scale of sharing we think will and should incentivize monitoring by imposing a risk of significant damages—questions that are not difficult to answer but require more data, analysis, and discussion. And, yes, there are other concerns to address along the way. I'm not blind to the perverse irony of prompting companies to squelch items because they are popular, which makes it especially important to minimize the possibility that companies overreact and close off legitimate and valuable speech. To counter this, we can incorporate an actual malice requirement along with notice as either an element of the basic liability rule or an affirmative defense. We can, in fact, manage a great many concerns about over- or underenforcement using familiar procedural tools like fee shifting; evidentiary requirements; burdens of production and persuasion; and mixing public, private, and administrative enforcement.

This particular innovation obviously will not solve the whole problem. The same lie may be embodied in numerous articles, each slightly different, no one of which goes viral, each shared among smaller audiences, but collectively having the same impact. Plus, fake news may still spread by other means, though these are harder to use effectively. We have, moreover, recently seen an uptick in direct communication of end-to-end encrypted information within groups via the instant messaging service WhatsApp (also owned by Facebook), which may require special treatment like limiting the size of groups or the kinds of materials that can be shared both within and among them. There are no silver bullets, and we'll need an array of efforts to deal with the rapidly changing, many-headed disinformation beast. But the principle I'm suggesting offers a way out of the all-or-nothing liability trap and provides a novel approach to address one very significant piece of the puzzle.

2. A second approach to mitigating the disinformation threat might be to directly regulate how social media companies operate, in particular, their practice of proactively pushing out material to users. Disinformation and propaganda have been around for a very long time, but (as noted earlier) before the

internet, its peddlers had limited capacity to reach an audience beyond those who affirmatively wanted to read or see or hear what they were peddling. That's because, before the internet, people had to *do* something—to go out of their way somehow—to get news and information: They had to mail or phone in a subscription, walk to a newsstand, tune into a TV or radio station. What was required seldom entailed a great deal of effort, but since most people weren't actively looking for fringe material, even trivial demands were enough to buffer all but a very few people. That kept exposure and sales low, which in turn discouraged the creation of such material in the first place.

The internet and social media companies have removed all the friction from this system. People no longer need to seek out extremist literature: Today, it seeks them out. With relatively little sophistication, producers of disinformation can get to enough people to let the algorithms take over, while also relying on users they reach to further push their material out to "friends" and "followers" and others in their online social networks.

Once people receive this provocative stuff, viewing and reading it becomes, if not quite irresistible, at least much more likely. I would never go out of my way to look for an auto accident on the highway, but if I happen to drive past one, I'm no better than anyone else at resisting the urge to rubberneck. With the internet and social media platforms, purveyors of disinformation are able, in effect, to purposefully place accidents along our routes. At that point, the game is all but over, because if we've learned anything over the years, it's that exposure *is* the game when it comes to lies and propaganda—particularly in a system like ours, in which the mere fact that a fake story is spreading becomes an event that mainstream outlets must cover, furthering the story's circulation and, in so doing, unwittingly promoting its legitimacy.[48]

Among the most inexplicable features of our attitude toward information technology is a tendency to assume that if such technology *can* do something, that something *should* be permitted—even if it has little or no social value or can be used for malevolent purposes. No one bats an eye about making it criminal to engineer a virus that makes people sick, and rightly so. But it's still not illegal to make a computer virus (only to intentionally spread one). Technology now makes it possible to produce so-called deepfake videos, yet despite the obvious hazards, there has been little clamor to regulate, much less ban, their development.

And so it is with social media platforms: They are able to do something news and information services in the pre-internet period could not, *viz.*, gather detailed information on the potential interests of billions of people and use it to feed them material that draws their attention, whether desired or not. And because the platforms *can* do this, we seem impelled to concede that they *should* be permitted to do so.

But why concede that? Why let these for-profit entities freely operate in a fashion that is wreaking havoc on our public discourse and democracy's future? Why not limit the platforms' business models to prohibit pushing out certain kinds of information—not forbidding them to provide access, but simply disallowing the feature that lets them put unrequested information in front of unwitting users? We could, for instance, prohibit the platforms from proactively sending users materials published or produced for general audiences, whether via predictive algorithms or by allowing other users to forward active links. The companies could still make such materials available. They could still make themselves the world's largest newsstands, where nearly anything and everything can be found; could still permit users to tell others in their networks about this or that story or item. But users would need to make an effort to acquire the materials themselves, as in the pre-internet age. We might, for instance, require users to subscribe to materials but limit the subscriptions in time, requiring them to resubscribe on a monthly basis. We could, likewise, permit platforms to show users potential content while following Roger McNamee's smart suggestion to ban algorithmic amplification.[49]

The effort required for individual users to access content would not be great with changes like these; given search features, it would, in fact, still be less than was required before the internet. But restoring even a little of the friction that used to be a natural feature of our information environment would reduce the spread of disinformation without the need to censor or meddle in anyone's autonomous decisions about what they want to read and see and hear.

This would in no way restrict or interfere with free speech rights. Speakers could still write, say, and publish as much as ever; platforms could still make all of it available to users; listeners and readers could still obtain whatever they wanted. But social media channels could no longer manufacture automatic audiences, and people would (once again) have to take a little initiative (though just a little) to get the materials they want. Social media companies would lose the small portion of their advertising revenue that comes from this component of their business, but it's a relatively straightforward kind of commercial regulation. And if this is at odds with current law respecting the regulation of commercial speech—and I don't see how it could be—it's a prime example of the kind of thing we should be rethinking.

As with my first suggestion, there are complexities we need to work out here. How should we define materials published for a "general audience"? I have in mind a carveout that leaves room for algorithms to continue pushing personal items like family photos or personal thoughts intended for actual acquaintances. But crafting a workable line will be difficult, especially when it comes to public figures and celebrities, who may have thousands or millions of followers.

Equally difficult will be calibrating the amount of friction to reintroduce, which needs to be enough to stop or slow the spread of disinformation without at the same time significantly reducing the consumption of legitimate news and information. Fortunately, while a large majority of Americans now get their news on digital devices, most don't rely on social media for access to conventional sources.[50] Very few people rely exclusively on social media, and many of those who do may be more willing to make the effort needed to access mainstream information. Still, there's a lot we don't know about how this would work, and we may want to avoid a cure that also has serious side effects. My point, in any event, is that there are benefits from restoring a little friction back into the system, and we can and should experiment to find the right amount.

If You Can Keep It

The internet and social media companies did not create the destructive dynamics that today threaten the future of liberal democracy, and mitigating their influence will not solve all our problems. These new technologies have, however, enabled, intensified, and accelerated those dynamics in ways that make it impossible to solve the problems without mitigating their impact and contributing role.

I began this essay with Ben Franklin's famous quip about how the Constitutional Convention had created a republic "if you can keep it." I remember feeling pride the first time I heard this story, which was many years ago: pride in realizing how tenuous the American experiment was at its launch, and pride in thinking how distant that fragility now seemed. Well, it no longer seems distant. The world has changed, and in ways that make "business as usual" when it comes to the law of free speech dangerous and foolhardy.

Acknowledgments

I am grateful to Yochai Benkler, Kelly Born, John Cohen, Brent Harris, Peter Kaufman, Johannes Meier, David Sasaki, and Jonathan Zittrain for helpful, challenging comments. The title of this essay is from a quote attributed to Franz Kafka: "There are some things one can only achieve by a deliberate leap in the opposite direction."

3

The Disinformation Dilemma

EMILY BAZELON

In the past few years, deliberate and widespread campaigns of misinformation and disinformation around the world have sparked terrible ethnic violence, threatened democracy—including in the United States—and worsened the COVID-19 pandemic by sowing false doubt about vaccines. The pandemic in particular led public health authorities and governments to denounce an "infodemic." "Misinformation costs lives," the World Health Organization and other groups said in September 2020 about what they saw as a crisis. "In recent years, the rapidly changing information environment has made it easier for misinformation to spread at unprecedented speed and scale," the surgeon general of the United States, Dr. Vivek Murthy, stated in an advisory on the topic in the summer of 2021. "Limiting the spread of health misinformation," he added, "is a moral and civic imperative."[1]

But not a legal imperative. In the United States, the government has limited its role in combatting false information to one of exhortation. Americans have a deep and abiding suspicion of letting the state regulate speech, and justifiably so: Laws that touch on speech or public debate have a suspect history in this country and others. In the present as well as the past, they can be thinly veiled attempts at censorship.[2] The modern First Amendment sweeps broadly to prevent any such possibility. John Stuart Mill wrote a century and a half ago that "[a]ll silencing of discussion is an assumption of infallibility."[3] There is still plenty of reason to believe that moving away from the American free speech tradition could make us too quick to dismiss apparently false ideas and factual assertions that turn out to have merit—and that airing them is the only way to find out.

The infodemic about COVID, however, like the spates of ethnic violence abroad and the undermining of the 2020 US election (among other examples), underscores the cost of free speech absolutism in our era. It leaves the country vulnerable to a different kind of threat, which may be doing serious damage to the discourse about politics, news, and science. Campaigns of misinformation, which

refers to falsehoods generally, and disinformation, which refers to falsehoods aimed at achieving a political goal—propaganda, to use the twentieth-century term—encompass the mass distortion of truth and overwhelming waves of speech from extremists that smear and distract. These forms of speech are often grimly effective at muting critical voices. And yet as Tim Wu, a law professor at Columbia Law School, points out in *The Perilous Public Square*, a recent excellent collection of essays, the "use of speech as a tool to suppress speech is, by its nature, something very challenging for the First Amendment to deal with."[4]

Wu's insight leads to another that may seem unsettling to Americans: that perhaps our way of thinking about free speech is not the best way. At the very least, we should understand that it isn't the only way. Other democracies, in Europe and elsewhere, have taken a different approach. Despite more regulation of speech, these countries remain democratic; in fact, they have arguably created better conditions for their citizenry to sort what's true from what's not and to make informed decisions about what they want their societies to be.

Here in the United States, meanwhile, we're drowning in lies. The overwhelming amount of misinformation and disinformation, the polarizing anger encoded in it—these serve to create chaos and confusion and make people, even nonpartisans, exhausted, skeptical, and cynical about politics and the underlying social compact. The spewing of intentional falsehoods isn't meant to win any battle of ideas. Its goal is to prevent the actual battle from being fought, by causing us to simply give up.

These problems and tensions were evident in the summer of 2021, when the federal government tried to lay blame for misinformation about COVID. President Joe Biden lashed out in July at one giant private actor—Facebook—for "killing people" by allowing misinformation to spread. The company said Biden was wrong. He backed off, saying he meant to point the finger at a dozen "super-spreaders" of misinformation on Facebook. The company said it was already addressing the problem, and publicly, at least, the dispute petered out, an inconclusive and unedifying tit-for-tat.

What was Facebook *actually* doing about the coronavirus infodemic? "These facts are not unknowable," the tech writer Casey Newton argued, but "for the most part Facebook doesn't share them."[5] Most other social media platforms are similarly opaque. This is possible because the government neither sets standards for transparency nor even holds the company to its own professed benchmarks. Sheera Frankel of the *New York Times* reported in July 2021 that the company refused requests by a group of its own data scientists to study COVID misinformation on the platform.[6] That spring, Facebook broke up the team that supported the data analytics tool CrowdTangle, which journalists and researchers used to track the content—often featuring misinformation—that generated high engagement. Two months later, Facebook shut down a project

called Ad Observatory at New York University that collected and analyzed data about political ads and the people they target.

The larger point is that Biden's spat with Facebook revealed the extent of the government's chosen impotence. The business model for the dominant platforms depends on keeping users engaged online. Content that prompts hot emotion tends to succeed at generating clicks and shares, and that's what the platforms' algorithms tend to promote. Research shows that lies go viral more quickly than true statements.[7]

How prevalent are disinformation and hate speech on the platforms? When Facebook, Twitter, and YouTube add information labels to false and misleading content, are people less likely to share it? When the companies say they are altering their algorithms to reduce the amplification of false information, what happens? Government policy and regulation play little to no role in answering these questions.

In many ways, social media sites today function as the public square. But legally speaking, internet platforms can restrict free speech far more than the government can. They're like malls, where private owners police conduct. The platforms get to make and enforce the rules about how to moderate content and about how its algorithms amplify it. They're not bound by the First Amendment—for example, they police hate speech, to which the Supreme Court has extended constitutional protection. And the platforms also get to decide what we know about the effects of their decisions. They have little incentive to be forthcoming, since the more a platform shares about what happens on it, "the more it risks exposing uncomfortable truths that could further damage its image," the *New York Time*s writer Kevin Roose observed.[8]

Newton, whose job is to track these power dynamics, argued for changing them. "Create legal reporting requirements for social platforms that detail what violations of their own standards they are finding, and encourage them to report at least some of their findings in real time," he urged Congress and the president. "If we really want to figure out who's killing whom, somehow this country is going to have to pass some laws."[9]

A similar argument could be made about updating the rules that govern media outlets. The problem of misinformation and disinformation, after all, is by no means limited to the internet. A 2020 working paper from the Berkman Klein Center for Internet and Society at Harvard University found that effective disinformation campaigns are often an "elite-driven, mass-media led process" in which "social media played only a secondary and supportive role." During Donald Trump's presidency, Fox News and other conservative media outlets came to function "in effect as a party press," the Harvard researchers found. The right-wing press has gained great influence without observing the "reality-check dynamic" of mainstream journalism, which appears to "significantly constrain

disinformation," as Yochai Benkler, Robert Faris, and Hal Roberts of the Berkman Klein Center argue in their 2018 book *Network Propaganda*.[10]

Right-wing media and social media have a symbiotic relationship. Links from Fox News hosts and other right-wing figures aligned with Trump often dominated the top links in Facebook's News Feed for likes, comments, and shares in the United States.[11] As a result, though Fox News is far smaller than Facebook, the social media platform has helped Fox attain the highest weekly reach, offline and online combined, of any single news source in the United States, according to a 2020 report by the Reuters Institute.[12]

In light of these realities, is the American faith that more speech is better, and that the government should regulate it as little as possible, worthy of reexamination? Should we continue trusting a handful of chief executives policing spaces that have become essential parts of democratic discourse?

The dilemma is this: When it comes to the regulation of speech, we are uncomfortable with government doing it; we are uncomfortable with social media or media titans doing it. But we are also uncomfortable with nobody doing it at all. This is a hard place to be—or, perhaps, two rocks and a hard place.

* * *

In the wake of the January 6, 2021, attack on the US Capitol that temporarily halted the counting of the 2020 electoral votes, Twitter permanently took away President Trump's account on the site. The company's decision drew criticism from Chancellor Angela Merkel of Germany, who criticized it as a "problematic" breach of the right to free speech. But what troubled Merkel was not the deplatforming of Trump itself. Instead, she took issue with the fact that a private company made the call, rather than the government invoking a *law* restricting online incitement. Germany passed such a law in 2017 to prevent the dissemination of hate speech and fake news stories.[13] With some exceptions, the law (known as NetzDG) requires platforms with more than two million users in Germany to remove or block access, within twenty-four hours of notice, to content that is "manifestly unlawful" because, for example, it's defamatory, includes nonconsensual sexual photographs, constitutes "public incitement to crime," or forms "criminal or terrorist organizations."

This countertradition to the US approach is alert to the ways in which demagogic leaders or movements can use propaganda/disinformation. A crude authoritarian censors free speech. A clever one invokes it to play a trick, twisting facts to turn a mob against a subordinated group and, in the end, silence as well as endanger its members. Looking back at the rise of fascism and the Holocaust in her 1951 book *The Origins of Totalitarianism*, the political philosopher Hannah Arendt focused on the use of propaganda to "make people believe the

most fantastic statements one day, and trust that if the next day they were given irrefutable proof of their falsehood, they would take refuge in cynicism."[14]

Many nations shield themselves from such antipluralistic ideas. In Canada, it is a criminal offense to publicly incite hatred "against any identifiable group." South Africa prosecutes people for uttering certain racial slurs. A number of countries in Europe treat Nazism as a unique evil, making it a crime to deny the Holocaust.[15]

In the United States, laws like these surely wouldn't survive Supreme Court review, given the current understanding of the First Amendment—an understanding that comes out of *our* country's history and our own brushes with suppressing dissent. In the 1960s, the Supreme Court enduringly embraced the vision of the First Amendment expressed, decades earlier, in a dissent by Justice Oliver Wendell Holmes Jr.: "The ultimate good desired is better reached by free trade in ideas."[16] In *Brandenburg v. Ohio*, that meant protecting the speech of a Ku Klux Klan leader at a 1964 rally, setting a high bar for punishing inflammatory words.[17] *Brandenburg* "wildly overprotects free speech from any logical standpoint," the University of Chicago law professor Geoffrey R. Stone points out. "But the court learned from experience to guard against a worse evil: the government using its power to silence its enemies."[18]

It's important to understand, however, that the Supreme Court decisions that appear to impede even modest efforts to regulate social media or media companies have little or nothing to do with individual rights. They come from another line of precedents that entrenched the power of corporations.

In the 1960s, the heyday of foundational decisions like *Brandenburg* and *New York Times v. Sullivan* in libel law,[19] the Supreme Court was still willing to press private entities to ensure they allowed different voices to be heard. A hallmark of the era was the Court's "sensitivity to the threat that economic, social and political inequality posed" to public debate, the University of Chicago law professor Genevieve Lakier wrote in a 2020 law review article.[20] As a result, the Court sometimes permitted the government to require private property owners like TV broadcasters to grant access to speakers they wanted to keep out. "It is the right of the viewers and listeners, not the right of the broadcasters, which is paramount," the Court wrote in its 1969 decision (now sidelined) in *Red Lion Broadcasting Co. v. FCC*. "It is the purpose of the First Amendment to preserve an uninhibited marketplace of ideas in which truth will ultimately prevail, rather than to countenance monopolization of that market, whether it be by the Government itself or a private licensee."[21] The Court's language suggests "an affirmative obligation on the legislature to act," Lakier argues.[22]

But the Court later shifted, she explains, toward interpreting the First Amendment "as a grant of almost total freedom" for private owners to decide who could speak through their outlets. In 1974, it struck down a Florida law

requiring newspapers that criticized the character of political candidates to offer them space to reply. Chief Justice Warren Burger, in his opinion for the majority, recognized that barriers to entry in the newspaper market meant this placed the power to shape public opinion "in few hands." But in his view, there was little the government could do about it.[23]

In the same decade, the Court also started protecting corporate campaign spending alongside individual donations, treating corporate *spending* on speech that was related to elections much like the shouting of protesters. This was a "radical break with the history and traditions of U.S. law," the Harvard law professor John Coates wrote in a 2015 article published by the University of Minnesota Law School.[24] Over time, the shift helped to fundamentally alter the world of politics. In the 2010 *Citizens United* decision, the Court's conservative majority opened the door to allowing corporations (and unions) to spend unlimited amounts on political advocacy, as long as they donated to interest groups and political action committees rather than to campaigns.[25]

By requiring the state to treat alike *categories* of speakers—corporations and individuals—the Supreme Court began to go far beyond preventing discrimination based on viewpoint or the identity of an individual speaker. "Once a defense of the powerless, the First Amendment over the last hundred years has mainly become a weapon of the powerful," Catharine A. MacKinnon, a law professor at the University of Michigan, wrote in *The Free Speech Century*, a 2018 essay collection.[26] Instead of "radicals, artists and activists, socialists and pacifists, the excluded and the dispossessed," she wrote, the First Amendment now serves "authoritarians, racists and misogynists, Nazis and Klansmen, pornographers and corporations buying elections." In the same year, Justice Elena Kagan warned that the Court's conservative majority was "weaponizing the First Amendment" in the service of corporate interests, in a dissent to the *Janus* ruling against labor unions.[27]

Somewhere along the way, the conservative majority has lost sight of an essential point: The purpose of free speech is to further democratic participation. "The crucial function of protecting speech is to give persons the sense that the government is theirs, which we might call democratic legitimation," says the Yale law professor Robert Post. "Campbell Soup Company can't experience democratic legitimation. But a person can. If we lose one election, we can win the next one. We can continue to identify with the democratic process so long as we're given the opportunity to shape public opinion. That's why we have the First Amendment."[28]

The Court's professed interest in granting all speakers access to the public sphere becomes a mechanism for deregulation of media companies and, if the same approach applies, social media companies as well. There's a real tension here. Democratic values justify protections against government interference

with what media and social media platforms publish. But in its stronger form, this conception of corporate speech rights, especially in a legal regime lacking strong antitrust enforcement, can also be a threat to democracy. It undermines the government's ability to take the kind of action the Court permitted and even encouraged in *Red Lion*. As Lakier puts it, "The Court invoked democratic values to justify extending protection to corporations, but the result was a body of free speech law that made it hard for the legislature to ensure equality of access to the democratic public sphere."[29]

* * *

In US media law, in the not so distant past, there was another way. For most of our history, ensuring a vibrant free press made up of competing outlets was an express aim of federal policy. From the founding until the early twentieth century, Congress lowered the cost of starting and running a newspaper or magazine by setting low postage rates for mailed copies. The advent of radio raised questions about how to foster competition and public access. "Lawmakers of both parties recognized the danger that an information chokehold poses to democratic self-government," says Ellen P. Goodman, a law professor at Rutgers University. "So policymakers adopted structures to ensure diversity of ownership, local control of media and public broadcasting."[30]

In 1927, when Congress created the licensing system for exclusive rights to the broadcast spectrum, so that radio broadcasters could secure a place on the dial, lawmakers told broadcasters to act "as if people of a community should own a station." The 1934 Communications Act similarly required anyone with a broadcast license to operate in the "public interest" and allocated the spectrum to ensure that local communities had their own stations. In 1949, the Federal Communications Commission (FCC) established the fairness doctrine, which interpreted operating in the public interest to require broadcasters to cover major public policy debates and present multiple points of view. And in 1967, Congress created and funded the Corporation for Public Broadcasting, whose mission is to "promote an educated and informed civil society," and reserved broadcast spectrum for local NPR and PBS stations.[31]

During these decades, broadcasters were held to a standard of public trusteeship, in which the right to use the airwaves came with a mandate to provide for democratic discourse. Broadcasters made money—lots of it—but profit wasn't their only reason for existing. "The networks had a public-service obligation, and when they went to get their licenses renewed, the news divisions fulfilled that," says Matthew Gentzkow, an economist at Stanford University who studies trust in information.[32] The model coincided with a rare period in American history of relatively high levels of trust in media and low levels of political polarization.

But public trusteeship for broadcasting and diverse ownership began to unravel with the libertarian shift of the Reagan era. In the mid-1980s, the Reagan administration waived the FCC rule that barred a single entity from owning a TV station and a daily newspaper in the same local market to allow Rupert Murdoch to continue to own the *New York Post* and the *Boston Herald* after he bought his first broadcast TV stations in New York and Boston. Then, in 1987, the FCC repealed the fairness doctrine, which had required broadcasters to include multiple points of view. "When that went, that was the beginning of the complete triumph, in media, of the libertarian view of the First Amendment," Goodman says.[33]

With the advent of the internet, in short order, the libertarian principles that weakened media regulation allowed a few American tech companies to become the new gatekeepers. The United States gave platforms like Google, Facebook, and Twitter free rein to grow. Google bought YouTube. Facebook bought Instagram and WhatsApp.

The European contrast is illustrative. After World War II, European countries also promoted free speech, and the flow of reliable information, by making large investments in public broadcasting. Today France TV, the BBC, ARD in Germany, and similar broadcasters in the Netherlands and Scandinavia continue to score high in public trust and audience share. Researchers in Germany and France who have mapped the spread of political lies and conspiracy theories there say they have found pockets online, especially on YouTube, but nothing like the large-scale feedback loops in the United States that include major media outlets and even the president.

The difference between the political speech traditions of the United States and Europe was acutely apparent in the American and French presidential elections of 2016 and 2017. When Russian operatives hacked into the computers of the Democratic National Committee (DNC), they gave their stolen trove of DNC emails to WikiLeaks, which released the emails in batches to do maximum damage to Hillary Clinton and her party in the months before the election. The news media covered the stolen emails extensively, providing information so the public could weigh it, even if a foreign adversary had planted it.

The French press responded quite differently to a Russian hack in May 2017. Two days before a national election, the Russians posted online thousands of emails from En Marche!, the party of Emmanuel Macron, who was running for president. France, like several other democracies, has a blackout law that bars news coverage of a campaign for the twenty-four hours before an election and on Election Day. But the emails were available several hours before the blackout began. They were fair game. Yet the French media did not cover them. *Le Monde*, a major French newspaper, explained that the hack had "the obvious purpose of undermining the integrity of the ballot."[34]

Marine Le Pen, Macron's far-right opponent, accused the news media of a partisan coverup. But she had no sympathetic outlet to turn to, because there is no equivalent of Fox News or Breitbart in France. "The division in the French media isn't between left and right," said Dominique Cardon, director of the Media Lab at the university Sciences Po. "It's between top and bottom, between professional outlets and some websites linked to very small organizations, or individuals on Facebook or Twitter or YouTube who share a lot of disinformation."[35] The faint impact of the Macron hack "is a good illustration of how it's impossible to succeed at manipulation of the news just on social media," said Arnaud Mercier, a professor of information and political communication at the University Paris 2 Panthéon-Assas. "The hackers needed the sustainment of the traditional media."[36]

The challenge of informing the public accurately about the coronavirus has also played out differently in the United States and Europe. In August 2020, the global activist group Avaaz released a report showing that through at least May, conspiracies and falsehoods about the coronavirus and other health issues circulated on Facebook far more frequently than posts by authoritative sources like the World Health Organization and the Centers for Disease Control and Prevention.[37] Avaaz included web traffic from Britain, France, Germany, and Italy, along with the United States, and found that the United States accounted for 89 percent of the comments, likes, and shares of false and misleading health information. "A lot of U.S.-based entities are actually targeting other countries with misinformation in Italian or Spanish or Portuguese," said Fadi Quran, the campaign director for Avaaz. "In our sample, the U.S. is by far the worst actor."[38]

* * *

America's information crisis was not inevitable. Nor is it insoluble. Whatever the Supreme Court does, there's no legal barrier to increasing the delivery of reliable information. The government, federal or state, could invest in efforts to do exactly that. It could stop the decline of local reporting by funding nonprofit journalism. It could create new publicly funded TV or radio to create more alternatives for media that appeals across the ideological spectrum. It could create a "PBS of the Internet," as Goodman and a co-author, Sanjay Jolly, argue in a 2021 paper.[39] The only obstacles to such cures for America's disinformation ills are political.

Trump, as president, and Biden, during his campaign to challenge him, called for revoking Section 230 of the 1996 Communications Decency Act, which Congress wrote in an early stage of the internet to help it grow. Section 230 effectively makes internet platforms, unlike other publishers, immune from libel and other civil suits for much of the content they carry.

Taking away the platforms' immunity altogether, however, does not seem like the best fit for the problems at hand. The immunity that platforms currently enjoy from lawsuits directly affects only a narrow range of cases, mostly involving defamation.[40] Yet the threat of being sued for libel could encourage platforms to avoid litigation costs by preemptively taking down content once someone challenges it. Some of that content would be disinformation and targeted hate speech—when either is defamatory—but other material might be offensive but true, posing a risk of overcensorship.

But there are other ideas with plausible bipartisan support: For one, make the platforms earn their immunity from lawsuits. Ellen Goodman and others have proposed using Section 230 as leverage to push the platforms to be more transparent, for example, by disclosing how their algorithms order people's news feeds and recommendations and how much disinformation and hate speech they circulate.[41] A quid pro quo could go further, requiring the companies to change their algorithms or identify superspreaders of disinformation and slow the virality of their posts.[42] To make sure new media sites can enter the market, the government could exempt small start-ups but impose conditions on platforms with tens of millions of users.

Congress, as well as the Justice Department, can also promote competition through antitrust enforcement. In early October 2020, the House Judiciary Committee's Democratic leadership released a 449-page report, based on an extensive investigation, that said Facebook, Google (which owns YouTube), Amazon, and Apple have monopoly power in their markets like that of the "oil barons and railroad tycoons" of the early twentieth century. "Because there is not meaningful competition, dominant firms face little financial consequence when misinformation and propaganda are promoted online," the report stated[43]. When the Federal Trade Commission filed suit against Facebook for anticompetitive conduct at the end of 2020, the case was described as one of the most important such actions in decades. Then in June 2021 a district court judge threw out the case (for lack of evidence to support the government's claim that Facebook controlled 60 percent of the social media market). The government returned to court with a new filing; the upshot is that it remains to be seen how effective antitrust enforcement will be in the social media domain.[44]

For decades, tech companies mostly responded to criticism with antiregulatory, free speech absolutism. But external pressures, and their own controversial records, have had an impact. Mark Zuckerberg of Facebook and Jack Dorsey of Twitter have lately suggested they are open to government regulation that would hold platforms to external standards. This could at least solve the problem of transparency exposed by Biden's fight with Facebook over COVID misinformation. And there are other medium-sized ideas floating around Washington for steps to improve the online speech environment, like

banning microtargeted political ads, requiring disclosure of the ad buyers, and making the platforms file reports detailing when they remove content or reduce its spread.[45]

But the United States may miss the chance to lead. To fend off regulation and antitrust enforcement, the internet platforms spend millions of dollars on lobbying in Washington. They align their self-interest with a nationalist pitch, warning that curbing America's homegrown tech companies would serve the interests of Chinese competitors like TikTok.

Europe once more offers an alternative. The European Union and its members don't have a stake in the dominance of American tech companies. European lawmakers have proposed requiring platforms to show how their recommendations work and giving users more control over them, as has been done in the realm of privacy. They could also require that the platforms set and enforce rules for slowing the spread of disinformation from known offenders. Miguel Poiares Maduro of the European Digital Media Observatory has proposed treating the platforms like essential facilities, the European version of public utilities, and subjecting them to more regulation. (Senator Elizabeth Warren, the Massachusetts Democrat, has outlined a similar idea in the United States.) It would be a huge shift.[46]

There is no consensus on a path forward, but there is precedent for some intervention. When radio and television radically altered the information landscape, Congress passed laws to foster competition, local control, and public broadcasting. From the 1930s until the 1980s, anyone with a broadcast license had to operate in the "public interest"—and starting in 1949, that explicitly included exposing audiences to multiple points of view in policy debates. The Court let the elected branches balance the rights of private ownership with the collective good of pluralism.

This model coincided with relatively high levels of trust in media and low levels of political polarization. That arrangement has been rare in American history. It's hard to imagine a return to it. But it's worth remembering that radio and TV also induced fear and concern, and our democracy adapted and thrived. The First Amendment of the era aided us. The guarantee of free speech is *for* democracy; it is worth little, in the end, apart from it.

4

A Framework for Regulating Falsehoods

CASS R. SUNSTEIN

One of the most serious concerns about speech on social media focuses on the increasing amount of false information. This can have serious consequences for individuals, for our society, and for our democracy. What is the role of truth and falsehood in human life? In business? In health care? In politics? In this essay, I will try to shed some light on how we might think about these questions.

Consider three problems. All of them are hypothetical, but they are based directly on real events, happening all over the world.

- Thomas Susskind, falsely claiming to be a doctor, writes on his Facebook page that COVID-19 does not create a serious health problem. Purporting to analyze the data with care, Susskind insists that unless you're at least eighty years old, you really don't have to worry about it. Susskind knows that the statement is false; for reasons of his own, he is trying to create problems in Europe and the United States. Should Facebook take the statement down, or accompany it with some kind of correction? Should public officials order Facebook to do so?
- A political candidate named John Jones buys an advertisement on the website of a prominent television network. The advertisement falsely states that Jones's opponent, a politician named Eric Munston, sexually assaulted a female employee ten years ago. Jones knows that the statement is false. Should Munston be allowed to sue the network or Jones? Should he be allowed to force the network to take the advertisement down?
- Mary Winston runs a column in the local newspaper stating that vaccinations are responsible for autism. Winston believes that the statement is true. But it isn't. Even so, the column is convincing many parents not to vaccinate their

children and is thus creating serious health risks. Can the local authorities order the newspaper to remove the column? Can they fine Winston? Can they mandate some kind of disclaimer, at least online?

Is there a right to lie? About a pandemic? About health and safety? About a public official? About an actor or musician? About a neighbor? About science? If we are committed to freedom of speech, must we tolerate lies? What about falsehoods in general? How important is truth, and what should governments do to protect it?

Intimate relationships are defined by trust. ("I trust you," one person might declare to another, at a defining moment.) So are close friendships. Truth-telling is central to trust. Something similar can be said about relationships between employers and employees, or among people who work together (at, say, a restaurant, a hospital, or a school). None of these relationships is likely to be entirely free from lies, but in some cases, deception turns out to be shattering. In politics, truth-telling is not exactly universal. But in politics, some lies can also be shattering. They are beyond the pale. What can be done about them?

In a famous opinion, Supreme Court Justice Oliver Wendell Holmes Jr. wrote: "The most stringent protection of free speech would not protect a man in falsely shouting fire in a theatre and causing a panic."[1] Right now, a lot of people are falsely shouting fire in a crowded theater. They are causing panics. At the very least, what they are doing is pretty close to that. They are certainly shouting, and what they are shouting is false. In some cases, their lies lead to illnesses and deaths. In other cases, their lies cut at the heart of democratic self-government. Some of those lies come from foreign governments, such as Russia. Some of them are homegrown. They come from public officials and from politicians or those who support them.

Importantly, many false statements are not lies; people who make or spread them sincerely believe them to be true. Falsehoods are a broad category of which lies are a mere part. Some people say what they know to be false. Others are reckless; it should be obvious that they are spouting falsehoods, but they do not know that that is what they are doing. Still other people are simply mistaken; they had reason to say what they did, but they turned out to be wrong.

These differences matter. When we are deciding whether a falsehood can or should be punished or regulated, it might be crucial to know whether the speaker was a liar, or reckless, or merely mistaken. But even the most innocent mistakes can be damaging and dangerous. Consider the case of Mary Winston, who was not a liar. People make mistakes about health or safety, and their mistakes cost lives.

Corrections

In general, falsehoods ought not to be censored or regulated, even if they are lies. Free societies protect them. Public officials should not be allowed to act as the truth police. A key reason is that we cannot trust officials to separate truth from falsehood; their own judgments are unreliable, and their own biases get in the way. If officials are licensed to punish falsehoods, they will end up punishing dissent. As Justice Robert Jackson wrote in the greatest opinion in the long history of the US Supreme Court, "Those who begin coercive elimination of dissent soon find themselves exterminating dissenters. Compulsory unification of opinion achieves only the unanimity of the graveyard."[2]

The best response to falsehoods is usually to correct them, rather than to punish or censor them. Punishment or censorship can fuel falsehoods. In some contexts, they operate like oxygen. These are time-honored ideas, but in some ways, they are now on the defensive. We need to understand them better. We need to appreciate them more. We need to do that, above all, to protect against government overreach, but also to allow freedom to flourish on television and in magazines and newspapers, and also online and on social media platforms such as Facebook and Twitter.

Still, we need to qualify these conclusions—and to take some of them back. As William Blake wrote, commenting on lectures by Sir Joshua Reynolds, who praised generalization: "To Generalize is to be an Idiot. To Particularize is the Alone Distinction of True Merit." Blake added, "I thank God I am not like Reynolds."[3]

Under the US Constitution, government can already do a great deal to control defamation. It can already regulate false advertising. It should be allowed to do more. It should be able to restrict and punish certain kinds of lies and falsehoods that pose serious threats to public health and safety. To protect the democratic process, it should be able to regulate other kinds of lies and falsehoods, even if they are not defamatory. It should be able to regulate doctored videos, certainly when they are defamatory, and even when they are not, to ensure that people who see them know that they are doctored. In defending these conclusions, one of my main purposes is to draw attention to the sheer diversity of tools. Government need not censor or punish; it might (for example) require disclosure, or some form of choice architecture that reduces the likelihood that falsehoods will spread.

Private institutions, including television networks, magazines, and newspapers, and social media platforms (such as Facebook, YouTube, and Twitter) have considerable room to slow or stop the spread of lies and falsehoods. To their credit, some of them are doing a great deal already, and their creativity

offers a host of lessons for public officials. But they should be doing more. Real people are being badly hurt by their inaction. So are institutions, both public and private.

Fake News

If you want to ban each and every lie, or to excise lies and falsehoods from human life, you're probably not a lot of fun.

People boast; they exaggerate their achievements. Some of us flatter; we tell people things that they want to hear. People protect themselves; they lie to those who threaten them. (Do the ends justify the means? Not never; but sometimes.) Some of us joke; we tell tall tales. Journalists spread falsehoods, even when they are trying very hard to tell the truth. No one should have to live in a nation that makes it a crime not to tell the truth, or even to lie. Such a nation would crush freedom.

But some lies, and some falsehoods, are beyond the pale. Suppose that Barton Medical, a (hypothetical) company that sells medicines, markets a new product, promising, "If you take this daily, you will never get cancer!" If the product does nothing to prevent cancer, the company will almost certainly get into trouble with the authorities. This is so even in the freest of free societies. But it is not simple to come up with principles to distinguish between what is intolerable from what must be allowed. To do that, we have to explore the foundations of a system of free expression. We need to understand what such a system is for—what it is designed to do.

These issues are always important, but in the modern era, they have new urgency. One reason, of course, is the rise of modern technologies, which allow falsehoods to be spread in an instant. If you want to circulate a lie about safety or health, or about a prominent person, you can do that with ease. If you want to sell a product by lying about it, you can try, starting today. If you are a public official and you want to lie about what you are doing, you can get a large audience in essentially an instant. You might be able to do the same thing if you want to attack a public official, a former lover, a neighbor, or someone you hate or just don't like. Or consider the focus on "fake news," disseminated by both foreign and domestic agents in an effort to generate traffic, to sow social divisions, or to drive political outcomes in North America, Europe, and elsewhere in particular directions. Much news is indeed fake, and that is a major problem.

Ironically, however, charges of "fake news" are often themselves fake—making it quite destabilizing to figure out what is true. Prominent national leaders cry "Fake news!" when they are subject to criticism, even when nothing fake has been said, and when the factual claims are true. The real fake news is

the cry of fake news. The result is that with respect to many questions, people now find themselves in a state of vertigo. As St. Augustine said: "When regard for truth has been broken down or even slightly weakened, all things will remain doubtful."[4] Or consider the contestation of science and established facts by prominent figures, including national leaders in many nations. Highly influential lies by public officials and in political campaigns are nothing new, but there are certainly a lot of them these days.

Television programs, newspapers, magazines, Facebook, Twitter, YouTube, and other social media platforms should be doing more than they are now doing to control the spread of falsehoods. Of course, they should prize freedom of speech and the values that it carries with it. But they should also work to protect public health and safety, democratic processes, the reputations of individuals and institutions, and most broadly, the social norm in favor of respect for, and recognition of, what is true—a matter of uncontestable fact.

As Hannah Arendt warned:

> The chances of factual truth surviving the onslaught of power are very slim indeed; it is always in danger of being maneuvered out of the world not only for a time but, potentially, forever. Facts and events are infinitely more fragile things than axioms, discoveries, theories—even the most wildly speculative ones—produced by the human mind; they occur in the field of the ever-changing affairs of men, in whose flux there is nothing more permanent than the admittedly relative permanence of the human mind's structure.[5]

A Framework

There are countless falsehoods out there, and along many dimensions, they are different from one another. A cry of "Fire!" might be a lie, designed to cause a stampede. Or it might be an innocent mistake, coming from someone who saw some smoke from audience members who were (illegally) lighting up cigarettes. A seller of a car might lie about the vehicle's gas mileage. On a date, a man might lie about his career achievements. Someone might perjure himself, saying that he was not at the scene of the accident. Someone might make an innocent error, mistakenly identifying someone as a perpetrator of a crime.

These diverse falsehoods raise very different questions. To approach them, we need a framework. To obtain one, we need to identify four sets of issues, and we need to keep them separate. As we shall see, each of them plays a role in analysis of constitutional issues and also of the obligations of private institutions, including social media providers.

1. The first question involves the speakers' State of Mind (and hence their level of culpability). In saying something that is false, people might be (1) lying, (2) reckless, (3) negligent, or (4) reasonable but mistaken. It might greatly matter into which category speakers fall. Under US constitutional law, it often does. The difference between lying and (2), (3), and (4) should be straightforward. The differences among (2), (3), and (4) are less straightforward.
2. The second question involves the Magnitude of Harm. How much damage is caused by the falsehood? There is a continuum here, but for heuristic purposes, let us say that the damage might be (1) grave, (2) moderate, (3) minor, and (4) nonexistent. A lie could start a war; that would be grave. But many lies are harmless. In deciding whether a falsehood can be regulated, surely it matters whether it is causing a grave harm or no harm at all.
3. The third question involves the Likelihood of Harm. Here, too, we have a continuum, including (1) certain, (2) probable, (3) improbable, and (4) highly improbable. A falsehood could create essentially certain harm, as when a seller lies about a product and thus induces a consumer to buy, or when someone prominently says, online, that people under the age of eighty cannot get COVID-19. By contrast, a falsehood could create highly improbable harm, as when a student announces in a class, "John F. Kennedy was not, in fact, president of the United States; he was actually vice president."
4. The fourth and final question involves the Timing of Harm. Yet again there is a continuum, but for heuristic purposes, it might be (1) imminent, in the sense of occurring immediately, (2) imminent, in the sense of the occurring in the near future, (3) occurring not in the near future but reasonably soon, or (4) occurring in the distant future. A libel can easily be seen to create imminent harm. A claim, by one teenager to another, that smoking cigarettes is actually good for you might be seen to create long-term harm.

These various possibilities might be mixed and matched in numerous ways—256, to be precise. We could construct a matrix with that number of boxes, and aim to give an indication of how the constitutional issue would be resolved, or should be resolved, on the basis of what box is involved. In order not to get ahead of ourselves, and as an act of mercy, let us not to do that. Let us allow table 4.1 to suffice.

It should be immediately clear that for purposes of thinking about freedom of speech, the combination of boxes will almost certainly matter. Suppose that we are dealing with liars who are certain to create grave harm immediately. If so, the argument for First Amendment protection seems very weak. Suppose, by contrast, that we are dealing with speakers who made reasonable mistakes that have a low probability of creating minor harms in the distant future. If so, the argument for First Amendment protection is very strong. As we shift from the

Table 4.1. **Framework for Regulating Falsehoods**

State of Mind	Lie	Reckless	Negligent	Reasonable
Magnitude of harm	Grave	Moderate	Minor	Nonexistent
Likelihood of harm	Certain	Probable	Improbable	Highly improbable
Timing of harm	Imminent	Near future	Reasonably soon	Distant future

four sets of (4) to the four sets of (1), the argument for constitutional protection gains force. Current constitutional law roughly reflects that understanding (but note: only roughly).

What Matters

To be more disciplined, of course, we need to think about each of the four scales, and about why they matter. For State of Mind, there are two major candidates. The first involves culpability. From the moral point of view, a liar seems to be a great deal worse than a complete innocent, and if people spread falsehoods recklessly, they are worse than the merely negligent. Predictably, moral outrage weakens as we go from (1) to (2) to (3) to (4).

The second explanation points to the effects of punishment or regulation on freedom of expression. Those effects differ dramatically, depending on the state of mind of the speaker. If the state punishes a liar, the deterrent effects are far less severe than if it punishes someone who has made a reasonable mistake. If those who make reasonable mistakes are punished, a lot of people will simply shut up. But if liars are punished, liars will stop lying. How bad is that? At first glance, it is not bad at all.

It is true that the state might wrongly deem someone to be a liar. It is also true that some lies are harmless. It is true, too, that punishment of liars might turn out to deter truth-tellers. For these reasons, we might demand some showing of harm—at least (3)—to justify regulation of lies. But as we shift from lying to recklessness to negligence to reasonable mistakes, regulation imposes an increasingly serious threat to freedom of expression.

If we emphasize culpability, we might converge on a simple conclusion: Lies, as such, should not receive any protection at all. The Supreme Court has rejected that view. The Court was right on the general point. The government has imperfect tools for ferreting out lies; if government acts against lies, it might be biased, going after a particular subset of lies; human beings have a right to depart from the truth, at least in some situations. We might therefore insist that government must make some demonstration of harm if it seeks to regulate lies. I shall be

defending that conclusion here. (This conclusion bears on the responsibilities of private institutions, including social media platforms.)

For Magnitude of Harm, the central idea is simple. It is no light thing to suppress or regulate speech, which means that a significant justification is almost always required in order to do that. A small loss is not enough. As they say, *de minimis non curat lex* (the law does not concern itself with trifles). This general proposition leaves many open questions.

For Likelihood of Harm, the claim seems to be that if no one is likely to be hurt, we lack the requisite justification. Why should government regulate speech—including falsehoods—if it can show only a small chance of harm? But there is a serious problem with the claim here. Suppose that a falsehood gives rise to a small chance (say, one in ten) of a very grave harm (say, a significant loss of lives). In that event, is there a weaker justification for regulation than if a falsehood gives rise to a high probability (say, nine in ten) of a modest harm (say, a small loss of money)? What would seem to matter is the expected value of the harm, not its likelihood.

For Timing of Harm, the justification seems to have everything to do with the idea of "counterspeech," captured in the view that the proper remedy for falsehoods is more speech, not enforced silence.[6] If harm is not imminent, perhaps the legal system should rely on rebuttal or counterargument, not on censorship or regulation. But this idea is also vulnerable. Suppose that a harm is inevitable but not reasonably soon. How, exactly, will counterspeech help? Should imminent harm be necessary? Perhaps it ought not to be. But there is a qualification: If a harm is not imminent, we might not be sure that it is likely at all. Officials might think that it is, but they might be wrong. Their powers of prediction are limited, and their judgments might be distorted by some kind of bias. It might make sense to care about timing for pragmatic reasons: If the harm is a long way off, maybe we should just assume that counterspeech is the right remedy.

Tools

It is essential to see that if falsehoods threaten to create harm, government can choose among an assortment of tools, with different levels of aggressiveness and intrusiveness. It might order those who spread falsehoods to stop. It might impose a jail sentence. It might impose a criminal fine. It might impose a civil fine. It might authorize people to bring damage actions in court. It might require someone—a newspaper, a television network, a social media provider—to provide some kind of disclaimer or disclosure to those who will see the falsehood. It might require labels and warnings. It might come up with a corrective strategy

of its own, perhaps by telling the truth and making it readily available. Whenever free speech is at issue, courts might require government to choose the most speech-protective alternative—for example, disclosure rather than prohibition.

Private institutions, including social media providers, also have an assortment of tools. For example, Facebook, YouTube, and Twitter might remove certain lies. Alternatively, they could require disclosure. They could use the architecture of their platform in creative ways—perhaps by downgrading falsehoods so that few people see them, or perhaps by educating their users so that they can easily find out what is true. They might combine these approaches in creative ways, restricting the most aggressive for the most harmful, and the gentlest for the least harmful. In 2020, Twitter received a great deal of attention when it accompanied two tweets by President Donald Trump containing false claims about the high level of fraud associated with mail-in ballots with a small label: "Get the facts about mail-in ballots." (Was Twitter right to do that? I think so.)

Importantly, governments might consider building on some of those practices, perhaps requiring their best practices. There are some promising and novel possibilities here. Some social media platforms have adopted creative methods for combatting falsehoods and lies, sometimes by informing users, sometimes by educating them, sometimes by using architecture to limit the distribution and influence of misinformation. Labels and warnings might be used in lieu of removing material. We could easily imagine a law that would require general adoption and use of one of these techniques. We could also imagine a judicial holding that under the free speech principle, government is required to use the least intrusive tool. And indeed, the various tools could be mixed and matched with the four scales emphasized here—with, for example, permission to use the most aggressive tools with the (1)s, and the weakest with the (3)s.

These points offer a distinctive perspective on both old and new problems. They strongly suggest that in the United States, current constitutional law fails to strike the right balance. Public officials, actors, musicians, and athletes should be able to do far more than they are now permitted to do to respond to defamation. The same is true for ordinary citizens subject to damaging falsehoods. In addition, public officials have considerable power to regulate deepfakes and doctored videos. They are also entitled to act to protect public health and safety, certainly in the context of lies, and if falsehoods create sufficiently serious risks, to control such falsehoods as well. In all of these contexts, some of the most promising tools do not involve censorship or punishment; they involve more speech-protective approaches, such as labels and warnings.

Private institutions, including television networks, magazines, newspapers, and social media providers, should be acting more aggressively to control defamation and other falsehoods and lies. They should be doing more than they are

now doing to prevent the spread of misinformation involving health and safety and of doctored videos. They should reduce the coming spread of deepfakes.

These are specific conclusions, but they bear on some of the largest and most general questions in all of politics and law, and indeed in daily life itself. Hannah Arendt put it this way: "What is at stake here is this common and factual reality itself, and this is indeed a political problem of the first order."[7] The principle of freedom of speech should not be taken to forbid efforts to protect reality.

PART TWO

REFORMING SECTION 230

5

The Free Speech Industry

MARY ANNE FRANKS

Introduction

During a September 30, 2021, Senate Commerce Subcommittee hearing on Facebook and mental health harms, Sen. Edward Markey (D-Mass.) stated that "Facebook is just like Big Tobacco, pushing a product they know is harmful to the health of young people."[1] Comparisons between big tech and big tobacco have proliferated in recent years, as journalists, scholars, and legislators note the similarities between two massively profitable, largely unregulated industries that knowingly inflict harms on the public. One major difference between the two is that the tobacco industry was eventually subjected to extensive regulation, through both agency oversight and lawsuits. The tech industry, by contrast, remains as yet virtually untouched by either administrative rules or by tort law.

The tech industry's broad immunity from civil suits is the product of Section 230 of the Communications Decency Act. Though not usually described as such, Section 230 is an astonishingly comprehensive piece of tort reform. While a straightforward reading of the 1996 law indicates that it forbids imposing publisher liability in defamation cases, courts have interpreted Section 230 to preemptively absolve online intermediaries of nearly all tort liability (as well as state criminal liability) for wrongful content and conduct on their sites and services.

Tort reform has historically been a partisan affair. Progressives have often decried tort reform as a tool of neoconservatives to advance corporate interests over individual welfare. For example, liberals have criticized the 2005 Protection of Lawful Commerce in Arms Act (PLCAA), which broadly immunizes the firearms industry for unlawful acts committed with their products. Liberals have highlighted the constitutional significance of access to the courts to obtain redress for wrongs, especially for marginalized groups, and emphasized how "tort liability encourages industries to develop new safer technologies."[2] But liberals have been among the strongest supporters of Section 230, even as

the law has been used to close the courtroom door on some of the most vulnerable members of society, including women, racial minorities, and minors, who have suffered grave injuries stemming from the action or inaction of powerful corporations.

Until very recently, Republicans and Democrats were united in their deference to the tech industry. Today, while politicians on both sides of the aisle claim to be appalled by big tech's power and influence—though often for very different reasons—they rarely acknowledge how both parties' decades-long, enthusiastic embrace of Section 230 has helped create it. Nor has either party yet made significant strides to restore the right of individuals to their day in court for a vast array of harms facilitated by tech companies, including surveillance, harassment, threats, privacy violations, and defamation.

Why has such an incredibly powerful industry been allowed to essentially regulate itself for so long? Why have liberals, in particular, not only failed to sound the alarm on tech corporations' predatory practices, but often been the industry's biggest cheerleaders? An important part of the explanation is the truly extraordinary degree to which the tech industry has convinced the public that free speech is a product, and it has cornered the market.

From the earliest days of the commercial internet, technolibertarians asserted that cyberspace was the true home of free speech, an assertion inevitably wrapped in antiregulatory sentiment. Tech companies invoked the laissez-faire principles of the First Amendment to justify their failure to address extremism and abuse, elevating passivity into a virtue. Perhaps most seductively, companies like Google, Facebook, and Twitter appeared to provide "free speech" in a dual sense: free from censorship and free from cost.

But as the world is increasingly coming to recognize, there is nothing free about what the tech industry is offering. For tech companies, "free speech" is "free labor," a commodity created by "users" to sell to other "users." Far from being uncensored, speech online is filtered, arranged, promoted, altered, and labeled in accordance with corporate interests. In exchange for the illusion of free expression, multi-billion-dollar corporations extract troves of personal data from individuals to be used for marketing, advertising, and surveillance purposes. The relentless pursuit of "engagement" means that there is a premium placed on extreme, polarizing content, including vicious harassment, sexual exploitation, and dangerous disinformation.

The tech industry's commercialization of "free speech" has also contributed greatly to the American public's misunderstanding of the meaning and scope of the First Amendment. In particular, it has created confusion about to whom, to what, and how the First Amendment applies. The invocation of the First Amendment and free speech in the text of Section 230 further confuses the issue

for legislators and courts as well as the general public. The concept of free speech that emerges from this confusion is impoverished, commodified, and passive.

Far from unleashing the radical and revolutionary power of free speech, the free speech industry tends to tame free speech and cement the status quo. While the antiregulatory, pro-corporation, techno-utopian system enabled by Section 230 immunity generates enormous capital, both literal and symbolic, the vast majority of that capital stays firmly in the hands of those who have always had more of it than everyone else: the wealthy, the white, the male. While Section 230 does help amplify free speech, increase profits, and enable informational dominance for the powerful and the privileged, it also assists in the silencing, bankrupting, and subordination of the vulnerable.

Like other pernicious forms of tort reform, Section 230 curtails the rights of the vulnerable in favor of the privileges of the powerful. Insulating the tech industry from liability for extremism, abuse, and misinformation has threatened free expression, worsened existing inequalities of gender, race, and class, and gravely undermined democracy. Section 230 should be amended to allow those who have been injured to have their day in court. Unless the harmful content or conduct in question is clearly speech protected by the First Amendment, plaintiffs should not be barred from suing online intermediaries, and online intermediaries that demonstrate deliberate indifference to harmful content unprotected by the First Amendment should not be able to take advantage of Section 230's protections.

Section 230: Tort Reform for the Tech Industry

In 1996, Congress attempted to regulate pornographic content on the internet by passing the Communications Decency Act. While the Supreme Court soon invalidated nearly all of the Act for violating the First Amendment, one part, now popularly known as Section 230, remained. Section 230 limits how and when online intermediaries ("interactive computer services") can be held legally accountable for the actions of those who use their platforms and services. Subject to exceptions for violations of federal criminal law and intellectual property law, "providers or users of an interactive computer service," as the statute called them, are not liable for content created by other users.

The tech industry looks considerably different today than it did in 1996. As of September 2021, the five largest US companies by market capitalization were Apple, Microsoft, Alphabet, Amazon, and Facebook.[3] These five tech companies "constitute 20 percent of the stock market's total worth, a level not seen from a single industry in at least 70 years."[4] The dominance of the tech industry extends

far beyond economic markets, shaping how people work, vote, socialize, learn, shop, and communicate.

As the influence and power of "big tech" have continued to expand, calls to reform the industry have emerged with increasing frequency and urgency across the political spectrum. A 2021 Pew Research survey found that 68 percent of Americans think that "major tech companies have too much power and influence in today's economy," and that more than half of Americans think that major tech companies should be more highly regulated.[5] Executives from Facebook, Twitter, and Google have been summoned before Congress multiple times in recent years to address issues such as anticompetitive practices, misinformation, and algorithmic amplification.[6] President Joe Biden signed an executive order in July 2021 urging the Federal Trade Commission (FTC) to examine anticompetitive restrictions, unfair data collection, and surveillance practices by dominant internet platforms.[7] The FTC filed an antitrust suit against Facebook in August 2021,[8] and several bills have been introduced in Congress aimed at reducing the liability shield afforded to tech companies by Section 230 of the Communications Decency Act.[9]

Despite increasingly impassioned calls for reform, the profit, power, and influence of major tech companies only continue to grow. In August 2021, Google, Apple, and Microsoft "reported record-breaking quarterly sales and profits. Facebook doubled its profits and reported its fastest growth in five years. In the last three months alone the US's five largest tech companies made combined profits of over $68bn."[10] In 2005, only 5 percent of Americans reported using social media; today, that number is 72 percent.[11] Seven in ten Facebook users report daily use of the platform.[12]

According to its supporters, Section 230 is as necessary today as it was back in 1996 to ensure that the internet remains the most powerful medium of free expression. Section 230 enthusiasts refer to the law as the "Magna Carta of the internet," the "foundation of the internet," the "cornerstone of internet freedom," and "the First Amendment of the internet." Section 230 has three main provisions:

- 230(c)(1), which states that "[n]o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."
- 230(c)(2), which shields providers and users of an interactive computer service from civil liability with regard to any action that is "voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable" or "taken to enable or make available to information content providers or others the technical means to restrict access" to such material.

- 230(e)(3), which states, inter alia, that "no cause of action may be brought and no liability may be imposed under any State or local law that is inconsistent with this section."

In brief, Section 230(e)(3) sets out the principle of broad immunity for tech companies, and the other two sections detail the two situations in which this principle is applied: when a company leaves harmful content up—230(c)(1)—and when it takes it down or restricts it—230(c)(2). As will be discussed in more detail below, the takedown provision, Section 230(c)(2), is a fairly unremarkable provision that essentially restates rights that interactive computer service providers, as private actors, already have under the First Amendment.

Section 230(c)(1), on the other hand, has been interpreted by courts to provide wide-ranging, preemptive immunity to online intermediaries for the choices they make about what content they choose to leave up, even if they know of its unlawful nature and take no reasonable steps to address it. Courts have interpreted Section 230 to protect online classifieds sites from responsibility for advertising sex trafficking,[13] online firearms sellers from responsibility for facilitating unlawful gun sales,[14] and online marketplaces from responsibility for putting defective products into the stream of commerce.[15]

Section 230(c)(1) affirmatively strips plaintiffs who have been injured by online harms of the right to pursue claims against online intermediaries. Although it is not often characterized this way, the law is a particularly sweeping and troubling example of what is euphemistically called "tort reform," a movement primarily associated with neoconservative values. Tort reform generally refers to changing civil law to make it more difficult for injured parties to bring claims, receive compensation, or obtain jury trials. A common tactic of tort reform is to portray the civil system as being overrun by greedy, unscrupulous plaintiffs, and to cast corporations in the role of the victim. "Corporate victimhood deflects attention away from the true victims: those who suffered from defective products, negligent medicine, investor fraud or unreasonably risky financial activities," write Michael L. Rustad and Thomas H. Koenig. Similarly, neoconservatives redefined the term "reform" to mean caps and other limitations on recovery for injured plaintiffs "to improve the functioning of the American civil justice system."[16]

Section 230 is often described as providing incentives for the tech industry to self-regulate, but this description is only intelligible with regard to Section 230(c)(2), which protects online intermediaries from suit if they choose to remove content they find objectionable. By also protecting online intermediaries from suit if they choose not to regulate, Section 230(c)(1) removes the incentive for them to act. Rather than encouraging the innovation and development of measures to fight online abuse and harassment, Section 230 removes incentives

for online intermediaries to deter or address harmful practices no matter how easily they could do so.

Section 230(c)(1) functions for the tech industry in much the same way that the PLCAA does for the firearms industry: It provides a superimmunity for powerful corporations, encouraging them to pursue profit without internalizing any costs of that pursuit. Just as the PLCAA eliminates incentives for manufacturers to develop safer guns or secure storage, Section 230 eliminates incentives for tech corporations to design safer platforms or more secure products. Private individuals are left to deal with the fallout of a reckless tech industry moving fast and breaking things—including life-destroying harassment, publicized sexual exploitation, and ubiquitous surveillance—on their own, just as they are forced to absorb the costs of deaths and injuries associated with a reckless firearms industry. Section 230 preemptive immunity ensures that no duty of care ever emerges in a vast range of online scenarios and eliminates the incentives for the best-positioned party to develop responses to avoid foreseeable risks of harm.[17]

Not only does 230(c)(1) fail to incentivize safer tech products and practices, it denies members of the public access to the courts to seek redress for injuries. In *Chambers v. Baltimore & Ohio R. Co.* (1907), the Supreme Court stated that "[t]he right to sue and defend in the courts is the alternative of force. In an organized society it is the right conservative of all other rights, and lies at the foundation of orderly government. It is one of the highest and most essential privileges of citizenship."[18] Section 230 preempts plaintiffs from ever bringing suit in many cases, and ensures that the suits that are brought will rarely survive a motion to dismiss on Section 230 grounds.

It may be objected that the appropriate party to be sued in many if not all cases is the person who directly caused the injury, or that many suits against online intermediaries will ultimately fail on the merits. As to the first objection, the nature of online interaction makes it challenging, often impossible, for plaintiffs to discover the identity of the person who has most directly wronged them. More important, many harms are the result of more than one actor—the principle of collective responsibility is well recognized in the law. As to the second objection, whether a plaintiff's claim will ultimately succeed in any given case is necessarily indeterminate. The value of the right to bring the claim does not turn on whether the claim is vindicated in the end. As Douglas A. Kysar notes, "Even when a plaintiff's case fails on the merits, judicial engagement with the details of her claim helps to frame her suffering as a legible subject of public attention and governance."[19] Moreover, in many cases, the discovery process will provide significant value not just to the plaintiff in the case at hand, but to legislators, regulators, future plaintiffs, and the public, according to Rustag and Koenig:

> Prolonged discovery in cases where the ISP is classified as a distributor will enable plaintiffs to uncover more information about the nature, nexus, and extent of prior crimes and torts on websites. Plaintiffs could use the locomotive of discovery to unearth aggravating factors, such as whether the ISP profited by being too closely connected to fraudulent schemes that injured consumers. Discovery in these cases might even result in ISPs or websites being stripped of their immunity as primary publishers because of a close connection to the creators of illegal content. If ISPs were liable as distributors with knowledge, the gravamen of a case would shift to determining how much the web host or service provider knew about the dishonest scheme and when they knew it.[20]

Given that the costs of online injuries so often disproportionately fall on marginalized populations, the ability to hold online intermediaries responsible is also key to protecting "cyber civil rights"[21] (a phrase coined by Professor Danielle Keats Citron in 2009).[22] The anonymity, amplification, and aggregation possibilities offered by the internet have allowed private actors to discriminate, harass, and threaten vulnerable groups on a massive, unprecedented scale. As the internet has multiplied the possibilities of expression, it has also multiplied the possibilities of repression, facilitating a censorious backlash against women and minorities. The internet lowers the costs of abuse by providing abusers with anonymity and social validation, while providing new ways to increase the range and impact of that abuse. Abundant empirical evidence demonstrates that online abuse further chills the intimate, artistic, and professional expression of individuals whose rights were already under assault offline.[23]

Section 230's preemption of the right of injured individuals to sue for online harms has led to a dystopian state of affairs where expressive, economic, and information inequalities divide our society; where the leader of a country can use social media platforms to incite violence against his own citizens; where domestic terrorists can coordinate bloody attacks on the Capitol; where global corporations can extract astronomical profits from exploiting private data; where women and minorities are silenced by online mobs; and where massive disinformation and misinformation campaigns can microtarget populations to create public health crises, foment armed rebellions, and undermine democracy itself.

Selling Free Speech

Comparing the tech industry to other industries, and Section 230 to other forms of tort reform, invites the objection that the tech industry *should* be treated differently than any other industry because its business is speech. Indeed, the

findings section of Section 230 itself stakes out this claim: "The Internet and other interactive computer services offer a forum for a true diversity of political discourse, unique opportunities for cultural development, and myriad avenues for intellectual activity." Because the internet has "flourished, to the benefit of all Americans, with a minimum of government regulation," one of the stated goals of Section 230 is "to preserve the vibrant and competitive free market that presently exists for the Internet and other interactive computer services, unfettered by Federal or State regulation."

The close association of the internet and the First Amendment is reflected in a 2019 First Amendment Center of the Freedom Forum Institute survey that found that more than half of respondents agreed with the erroneous statement that "[s]ocial media companies violate users' First Amendment rights when they ban users based on the content of their posts."[24] The error is not limited to members of the public; it also animates multiple lawsuits filed against companies such as Facebook, Twitter, and Google for alleged free speech violations as well as an increasing number of legislative and executive actions purporting to force social media companies to carry certain speech or provide access to certain speakers. This is despite the fact that the text of the First Amendment makes abundantly clear that it restrains *government* interference with speech and case law has abundantly demonstrated the corollary that private entities like social media companies are not so restrained. Indeed, the right of social media companies to create and enforce terms of service is itself protected by the First Amendment.

One explanation for the widespread but mistaken belief that tech companies are subject to First Amendment obligations is modern society's extraordinary dependence on social media for communication, news, commerce, education, and entertainment, such that any restriction of access feels like a violation of constitutional significance.[25] The outsized influence of the internet over daily life leads users to think of online platforms and tech companies not as the premises and products of private businesses but as public forums controlled by quasi-governmental actors.

But another influential factor is how the telecommunications industry has for decades actively encouraged the public to think of online platforms as natural, essential, and unmediated outlets for free speech. The tech industry's unprecedented economic, political, and cultural dominance relies in significant measure on its successful, and successfully disguised, commodification of free speech.

Journalist Nicholas Thompson observes that "the idea of free speech has long been a central concern in Silicon Valley," from the early "hacker ethic" that "prized the free flow of information" to a Twitter executive's (in)famous 2012 characterization of the site as the "'free speech wing of the free speech party.'"[26] As John Perry Barlow, a co-founder of the influential Electronic

Frontier Foundation (EFF), wrote in his 1996 Declaration of the Independence of Cyberspace, "[w]e 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."[27] Barlow's manifesto rejects what he characterizes as governmental attempts to regulate and censor cyberspace, declaring that "all the sentiments and expressions of humanity, from the debasing to the angelic, are parts of a seamless whole, the global conversation of bits. We cannot separate the air that chokes from the air upon which wings beat."[28]

For many years, tech companies and platforms used First Amendment principles to justify their hands-off approach to abusive and harmful content on their platforms and services. Marvin Ammori observes that "the First Amendment—and American free speech doctrine—still influences top tech lawyers tremendously. . . . It does so not as law but as a way of thinking about issues and viewing the world."[29] Kate Klonick writes that reliance on First Amendment law is "a common theme" of major technology platforms:

> American lawyers trained and acculturated in American free speech norms and First Amendment law oversaw the development of company content-moderation policy. Though they might not have "directly imported First Amendment doctrine," the normative background in free speech had a direct impact on how they structured their policies.[30]

Following controversy in 2012 about a subreddit featuring surreptitious photographs of underaged girls, then-CEO of Reddit Yishan Wong wrote, "We stand for free speech. This means we are not going to ban distasteful subreddits. We will not ban legal content even if we find it odious or if we personally condemn it. Not because that's the law in the United States . . . but because we believe in that ideal independently, and that's what we want to promote on our platform."[31] When Fredrick Brennan, the creator of the site 8chan, was asked in 2014 whether sexually explicit imagery featuring minors "is an inevitable fixture in the landscape of hardcore free-speech zones like his website," he replied in the affirmative: "It is simply the cost of free speech and being the only active site to not impose more 'laws' than those that were passed in Washington, D.C."[32] Newer social media sites like Gab and Parler are even more explicit about their professed First Amendment ideals. Gab's terms of service mention the First Amendment no fewer than eight times; Parler's community guidelines state that "[w]hile the First Amendment does not apply to private companies such as Parler, our mission is to create a social platform in the spirit of the First Amendment to the United States Constitution."[33]

But, as succinctly expressed in Mark Zuckerberg's testimony before a 2018 joint hearing before the Senate Judiciary and Senate Commerce, Science, and

Transportation Committees, the business of social media platforms isn't free speech, but profit. When asked by Senator Orrin Hatch, "How do you sustain a business model in which users don't pay for your service?" Zuckerberg answered, "Senator, we run ads."[34]

The internet is dominated by a handful of multi-billion-dollar companies, including Facebook, Google, and Twitter. The point of what is euphemistically referred to by these companies as "engagement," "community," or "user-generated content" is not the promotion of free speech or the public interest; it is to harvest as much data as possible to sell to advertisers. These industry players operate "commercial enterprises designed to maximize revenue, not defend political expression, preserve our collective heritage, or facilitate creativity."[35] The commodification of free speech is an essential element of what Shoshana Zuboff calls "surveillance capitalism": "a boundary-less form that ignores older distinctions between market and society, market and world, or market and person. It is a profit-seeking form in which production is subordinated to extraction as surveillance capitalists unilaterally claim control over human, societal, and political territories extending far beyond the conventional institutional terrain of the private firm or the market."[36]

Free speech is the free labor that fuels the tech industry. Search engines and social media platforms create nothing; they amplify, sort, and sell the speech of users who increasingly cannot conceptualize a right of free speech that exists apart from the internet: "to exist is to be indexed by a search engine."[37] The tech industry masks its corporate manipulation, extraction, and exploitation of speech through an increasingly wide range of "free" services promising connection, entertainment, and convenience. "Practically anyone can set up a Facebook or Twitter or Snapchat account (or all of these together), and using Google does not even require that," writes Moran Yemini. "With all of these opportunities for speech, it is sometimes easy to forget that, whatever users wish to do and to be through the use of these platforms, their interests are always subject to the grace of the platform."[38]

Crossing a Line

Experts on online disinformation and harassment have been arguing for years that allowing extremism and abuse to flourish on the internet leads to real-world harms. But the major players in the tech industry have until very recently largely eschewed moderation or even enforcement of their own terms of service with regard to white supremacist content, coordinated misogynist campaigns, and political misinformation. In the wake of high-profile incidents and sustained public pressure by antiabuse advocates, several tech companies began to make

sporadic attempts to address nonconsensual pornography, conspiracy theories, and targeted harassment of individuals around 2015. But it was the January 6, 2021, insurrection that led to the most significant and sustained interventions by major tech companies to address harmful content.

On January 8, 2021, two days after a violent mob attacked the Capitol in an attempt to prevent Congress's certification of the 2020 presidential election, Twitter permanently banned then-president Donald Trump's personal account.[39] Twitter had first temporarily locked the @realDonaldTrump account on January 6 after Trump posted a video and a statement repeating false claims about the election and expressing his "love" for the rioters,[40] requiring Trump to delete the tweets before being able to post again. At the time of the lockout, the Twitter Safety team noted that if Trump violated Twitter's policies again his account would be banned.[41] In a blog post on January 8, the company explained that it had determined that two of the Trump tweets that followed the riots, one referencing "American Patriots" and another stating that Trump would not be attending President-Elect Biden's inauguration, were "likely to inspire others to replicate the violent acts that took place on January 6, 2021, and that there are multiple indicators that they are being received and understood as encouragement to do so."[42]

The rioters who attacked the Capitol on January 6 bludgeoned a police officer to death with a fire extinguisher; dragged another officer down several steps and beat him with an American flag; attempted to locate and assassinate Speaker of the House Nancy Pelosi; constructed a gallows on Capitol grounds and called for the hanging of Vice President Mike Pence; ransacked congressional offices; looted federal property; and forced terrified elected officials and their staff into hiding for several hours.[43] They organized their efforts on sites such as Facebook, Twitter, and Parler, where false claims about election fraud and increasingly unhinged conspiracy theories like QAnon had proliferated for months.[44]

Twitter's decision to ban Trump came after Facebook's announcement that it would be suspending Trump's account indefinitely;[45] more social media bans—not just of Trump, but of other individuals who promoted lies about the election, endorsed white supremacist rhetoric and violence, or encouraged further insurrection efforts—quickly followed.[46] On January 9, Google and Apple removed the right-wing-dominated social media site Parler from their app stores after the site refused to moderate violent content, and Amazon removed the site from its web hosting services later that same day, citing the platform's multiple violations of Amazon's terms of service.[47]

While some praised the social media crackdown, several prominent Republican figures characterized it as an attack on free speech and the First Amendment—often taking to social media to do so.[48] Secretary of State Mike Pompeo tweeted, "Silencing speech is dangerous. It's un-American. Sadly, this

isn't a new tactic of the Left. They've worked to silence opposing voices for years." Donald Trump's son, Donald Trump Jr., tweeted, "Free Speech Is Under Attack! Censorship is happening like NEVER before! Don't let them silence us." Congressman Matt Gaetz proclaimed, on Twitter, "We cannot live in a world where Twitter's terms of service are more important than the terms in our Constitution and Bill of Rights."[49]

Many conservatives also complained about how many followers they were losing as Twitter purged accounts violating their terms of service.[50] Pompeo tweeted a graphic purporting to show how many thousands of followers he and other high-profile right-wing individuals had lost. Scott Atlas, who served as a Trump advisor on COVID-19 policy, bemoaned on January 11, "I have lost 12k followers in the past few days." Sarah Huckabee Sanders, the former White House press secretary, tweeted on January 9, "I've lost 50k+ followers this week. The radical left and their big tech allies cannot marginalize, censor, or silence the American people. This is not China, this is United States of America, and we are a free country."

But it was not only conservatives who raised concerns about social media platforms' banning Trump and cracking down on election disinformation and violent propaganda. Following Facebook's indefinite suspension of Trump's account, NSA whistleblower Edward Snowden tweeted, "Facebook officially silences the President of the United States. For better or worse, this will be remembered as a turning point in the battle for control over digital speech."[51] The Electronic Frontier Foundation (EFF) somberly observed, "We are always concerned when platforms take on the role of censors."[52] A senior legislative counsel for the American Civil Liberties Union (ACLU) wrote, "It should concern everyone when companies like Facebook and Twitter wield the unchecked power to remove people from platforms that have become indispensable for the speech of billions."[53] Ben Wizner, an attorney for the ACLU, criticized Amazon's decision to cut off Parler, telling the *New York Times* that "there will be times when large majorities of people want to repel speech that is genuinely important. . . . I think we should encourage, in a broad sense, companies like Amazon to embrace neutrality principles."[54]

The swift social media crackdown on harmful online content following the events of January 6 demonstrated that technology companies have long had the capacity to address online extremism and abuse—they have only lacked the will. And the cries of censorship that these belated and modest moves have triggered from influential figures across the political spectrum is a testament to how successfully the tech industry has colonized the free speech imagination of the American public. Speech is only real if it appears online; free speech is not merely a negative freedom from government interference, but an

affirmative right to an audience and to amplification; any moderation, editorial control, or enforcement of quality standards by social media companies is a form of censorship.

The consequences of this confusion have been thrown into sharp focus in recent years as tech companies have made modest, belated attempts to address severe online harms such as misinformation and violent rhetoric. Platforms have begun deleting, flagging, or providing context to false information, as well as suspending or permanently banning users who violate their terms of service. As private entities, they have the First Amendment right to undertake these actions; a different part of Section 230, 230(c)(2), explicitly provides them with additional procedural protections in doing so. And yet these actions have led to a barrage of criticism, especially from the right, including claims that the companies are engaging in "censorship" forbidden by the First Amendment. A number of high-profile Republicans have initiated lawsuits against tech companies and sought to gut Section 230(c)(2) in the name of defending free speech. Republican legislators have introduced laws attempting to force platforms to carry certain speech or grant access to certain speakers: exactly the kind of government action that the First Amendment explicitly prohibits.

These Orwellian attacks on tech companies and Section 230 obscure and undermine legitimate, urgently needed efforts to limit the tech industry's terrifying power over the boundaries of public discourse. The most serious problem with Section 230 is not that it reinforces tech companies' First Amendment rights to reduce, remove, or ban speech as they see fit. Rather, it is that it incentivizes tech companies to do the opposite: to leave up, boost, and even solicit content that causes foreseeable, measurable harm because they will not bear the costs for it. When it comes to private actors, passivity toward or complicity with harmful speech is neither an obligation nor a virtue. The tech industry should not be rewarded for turning a blind eye to conduct that incites violence, invades privacy, or threatens physical health, and certainly not for profiting from it. Such conduct is not only adverse to public welfare; it is adverse to free speech.

Most people know, on some level, that the First Amendment does not generally apply to private actors. They know that declining to let a solicitor into their private home isn't censorship, that restaurants displaying signs that they have the right to refuse service to anyone isn't a violation of free speech, that a newspaper doesn't violate their constitutional rights when it fails to publish their letter to the editor. But the internet, and social media platforms in particular, feel different. That is due in no small part to how aggressively tech companies have cultivated the image of their products and services as being open to everyone to say anything: a "modern day public square," as the Supreme Court put it in *Packingham*

v. North Carolina.[55] Congress has reinforced this association by asserting that the internet provides "a forum for a true diversity of political discourse."[56]

These characterizations make it tempting to view the internet as a public forum. Public fora occupy a " 'special position in terms of First Amendment protection' because of their historic role as sites for discussion and debate."[57] If the internet is a public forum, then social media companies, despite being private companies, should be treated like government actors with obligations to allow virtually all comers and all content.

But whatever hopes *Packingham* may have raised in terms of the Supreme Court's expanding or obliterating the state-action doctrine for private fora that host speech, they were dashed in the 2019 case *Manhattan Community Access Corp. v. Halleck*. The Court held that "merely hosting speech by others is not a traditional, exclusive public function and does not alone transform private entities into state actors subject to First Amendment constraints."[58] In the majority opinion, Justice Kavanaugh noted that to hold otherwise would be to intrude upon a "robust sphere of individual liberty."[59] This would be "especially problematic in the speech context, because it could eviscerate certain private entities' rights to exercise editorial control over speech and speakers on their properties or platforms."[60] In other words, protecting free speech in a private forum requires the exact opposite of what it takes to protect free speech in a public forum: Private actors must be allowed to exercise their free speech rights to counter, ignore, or exclude speech as they see fit, even where state actors would be restrained from doing so.

Though *Halleck* involved a cable channel, not a social media platform, the case had clear implications for the application of the state action doctrine to the internet. As private entities, Twitter and Facebook can employ fact-check labels to exercise their own free speech rights, and they are under no First Amendment obligation to allow any government official to use its platform as a propaganda channel. They are also free to remove or reduce the visibility of the speech of some individuals so that speech by other individuals might flourish.

The problem is, except for rare situations such as the bans and fact-checks that occurred in the wake of the insurrection, the tech industry generally declines to take any responsibility for harmful content or conduct facilitated through online platforms. As long as this is the case, free speech will remain the privilege of those with the most power. And while power imbalances and gatekeeping are a problem in traditional media as well, traditional media do not enjoy the same extensive tort reform as online media have been granted through Section 230. The tech industry is incentivized to serve its own bottom line due to the extraordinary privileges granted to it by the current interpretation of Section 230.

Dismantling the Free Speech Industry

To hold the free speech industry accountable requires reforming Section 230. Specifically, Congress should amend Section 230 to allow people who been injured by online harms to have their day in court. Unless the harmful content or conduct in question is clearly speech protected by the First Amendment, plaintiffs should not be barred from suing online intermediaries, and online intermediaries that demonstrate deliberate indifference to harmful content unprotected by the First Amendment should not be able to take advantage of Section 230's protections.

Limit Section 230's Protections to Speech Protected by the First Amendment

Both critics and defenders of Section 230 agree that the statute provides online intermediaries broad immunity from liability for a wide range of internet activity. While critics of Section 230 point to the extensive range of harmful activity that the law's deregulatory stance effectively allows to flourish, Section 230 defenders argue that an unfettered internet is vital to a robust online marketplace of ideas. The marketplace of ideas is a familiar and powerful concept in First Amendment doctrine, serving as a justification for a laissez-faire approach to speech. Its central claim is that the best approach to bad or harmful speech is to let it circulate freely, because letting ideas compete in the market is the best way to sort truth from falsity and good speech from bad speech, and because government cannot be trusted to make such decisions wisely or fairly.

The internet-as-marketplace-of-ideas presumes, first of all, that the internet is primarily, if not exclusively, a medium of speech. The text of Section 230 reinforces this characterization through the use of the terms "publish," "publishers," "speech," and "speakers" in 230(c), as well as the finding that the "Internet and other interactive computer services offer a forum for a true diversity of political discourse, unique opportunities for cultural development, and myriad avenues for intellectual activity."

When Section 230 was passed, it may have made sense to think of the internet as a speech machine. In 1996, the internet was text-based and predominantly noncommercial. Only twenty million American adults had internet access, and these users spent less than half an hour a month online. But by 2019, 293 million Americans were using the internet, and they were using it not only to communicate, but also to buy and sell merchandise, find dates, make restaurant reservations, watch television, read books, stream music, and look for jobs. According to one Section 230 enthusiast,

> the entire suite of products we think of as the internet—search engines, social media, online publications with comments sections, Wikis, private message boards, matchmaking apps, job search sites, consumer review tools, digital marketplaces, Airbnb, cloud storage companies, podcast distributors, app stores, GIF clearinghouses, crowdsourced funding platforms, chat tools, email newsletters, online classifieds, video sharing venues, and the vast majority of what makes up our day-to-day digital experience—have benefited from the protections offered by Section 230.[61]

But many of these "products" have very little to do with speech and, indeed, many of their offline cognates would not be considered speech for First Amendment purposes. If, as many defenders of Section 230 as currently written would have it, the broad immunity afforded online intermediaries is justified on First Amendment principles, then it should apply only with regard to online activity that can plausibly be characterized as speech protected by the First Amendment. What is more, it should only apply to third-party protected speech for which platforms serve as true intermediaries, not speech that the platform itself creates, controls, or profits from.

To accomplish this, the word "information" in Section 230(c)(1) should be replaced with the word "speech protected by the First Amendment." This revision would put all parties in a Section 230 case on notice that the classification of content as protected speech is not a given, but a fact to be demonstrated. If a platform cannot make a showing that the content or information at issue is speech, then it should not be able to take advantage of Section 230 immunity.

Honor the Long-Standing Principle of Collective Responsibility

Many harmful acts are only possible with the participation of multiple actors with various motivations. The doctrines of aiding and abetting, complicity, and conspiracy all reflect the insight that third parties who assist, encourage, ignore, or contribute to the illegal actions of another person can and should be held responsible for their contributions to the harms that result, particularly if those third parties benefited in some material way from that contribution. While US law, unlike the law of some countries, does not impose a general duty to aid, it does recognize the concept of collective responsibility. Third parties can be held both criminally and civilly liable for the actions of other people for harmful acts they did not cause but did not do enough to prevent.

Among the justifications for third-party liability in criminal and civil law is that this liability incentivizes responsible behavior. Bartenders who serve alcohol to obviously inebriated patrons can be sued if those patrons go on to cause car accidents; grocery stores can be held accountable for failing to clean up spills that lead to slip and falls; employers can be liable for failing to respond to reports of sexual harassment. Such entities are often said to have breached a "duty of care," and imposing liability is intended to give them an incentive to be more careful in the future. It is a central tenet of tort law that the possibility of such liability incentivizes individuals and industries to act responsibly and reasonably.

Conversely, grants of immunity from such liability risk encouraging negligent and reckless behavior. The immunity granted by Section 230 does just that, despite the evocative title of its operative clause, "Protection for 'Good Samaritan' blocking and screening of offensive material." This title suggests that Section 230 is meant to provide Good Samaritan immunity in much the same sense as Good Samaritan laws in physical space. Such laws do not create a duty to aid, but instead provide immunity to those who attempt in good faith and without legal obligation to aid others in distress. While Good Samaritan laws generally do not require people to offer assistance, they encourage people to assist others in need by removing the threat of liability for doing so.

Subsection (c)(2) of Section 230 is a Good Samaritan law in a straightforward sense: It assures providers and users of interactive computer services that they will not be held liable with regard to any action "voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable" or "taken to enable or make available to information content providers or others the technical means to restrict access" to such material. Importantly, because most interactive computer service providers are private entities, their right to choose whether to carry, promote, or associate themselves with speech is not created by Section 230, but by the First Amendment. Subsection (c)(2) merely reinforces this right by making it procedurally easier to avoid specious lawsuits.

On the other hand, Section 230(c)(1)'s broad statement that "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider" has been interpreted in ways directly at odds with Good Samaritan laws, as well as with a host of other legal principles and settled law. Where (c)(2) offers immunity to interactive computer service providers in exchange for intervening in situations where they have no duty of care, (c)(1) has been read to provide the same immunity to providers who do nothing at all to stop harmful conduct—and, even more perversely, extends that same immunity to providers who actively profit from or

solicit harmful conduct. For example, Section 230(c)(1) has been invoked to protect message boards like 8chan (now 8kun),[62] which provide a platform for mass shooters to spread terrorist propaganda; online firearms marketplaces such as Armslist, which facilitate the illegal sale of weapons used to murder domestic violence victims;[63] and classifieds sites like Backpage (now defunct), which was routinely used by sex traffickers to advertise underage girls for sex.[64]

In subsidizing platforms that directly benefit from illegal and harmful conduct, Section 230(c)(1) creates a classic "moral hazard," ensuring that the multi-billion-dollar corporations that exert near-monopoly control of the internet are protected from the costs of their risky ventures even as they reap the benefits.[65] Given that the dominant business model of websites and social media services is based on advertising revenue, they have no natural incentive to discourage abusive or harmful conduct. As Kalev Leetaru notes, "abusive posts still bring in considerable ad revenue . . . the more content that is posted, good or bad, the more ad money goes into their coffers."[66]

Online intermediaries who do not voluntarily intervene to prevent or alleviate harm inflicted by another person are in no sense Good Samaritans. They are at best passive bystanders who do nothing to intervene against harm, and at worst, they are accomplices who encourage and profit from harm. Providing them with immunity flies in the face of the long-standing legal principle of collective responsibility that governs conduct in the physical world. In physical spaces, individuals or businesses that fail to "take care" that their products, services, or premises are not used to commit wrongdoing can be held accountable for that failure. There is no justification for abandoning this principle simply because the conduct occurs online. In fact, there are more compelling reasons for recognizing collective responsibility online, because online interaction provides so many opportunities for direct tortfeasors to escape detection or identification.

Creating a two-track system of liability for offline and online conduct not only encourages illegality to move online, but also erodes the rule of law offline.[67] Offline entities can plausibly complain that the differential treatment afforded by broad interpretations of Section 230 violates principles of fairness and equal protection, or to put it more bluntly: If they can do it, why can't we? There is a real risk that Section 230's abandonment of the concept of collective responsibility will become the law offline as well as on.

To undo this, Section 230 (c)(1) should be further amended to clarify that providers or users of interactive computer services cannot be treated as the publisher or speaker of protected speech *wholly provided by* another information content provider, *unless such provider or user intentionally encourages, solicits, or generates revenue from this speech.* In addition, a new subsection should be added to Section 230 to explicitly exclude from immunity intermediaries who exhibit deliberate indifference to unlawful content or conduct.

The revised version of Section 230(c) would read:

(1) Treatment of publisher or speaker
No provider or user of an interactive computer service shall be treated as the publisher or speaker of any ~~information~~ speech protected by the First Amendment wholly provided by another information content provider, unless such provider or user intentionally encourages, solicits, or generates revenue from this speech.

(2) Civil liability
No provider or user of an interactive computer service shall be held liable on account of—

(A) any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected; or

(B) any action taken to enable or make available to information content providers or others the technical means to restrict access to material described in paragraph (1);

(3) Limitations. The protections of this section shall not be available to a provider or user who manifests deliberate indifference to unlawful material or conduct.

Objections

Those skeptical of Section 230 reform worry that narrowing the immunity for online platforms will encourage these platforms to be overly cautious in a way that would negatively impact valuable speech now available on the internet. That is, the specter of potential liability might lead platforms to remove content that seems even remotely controversial, including speech by women, minorities, and other historically marginalized groups. This is a serious concern, but not one that overrides the need for reform.

It must first be noted that whether an online platform leaves content up or takes it down, it makes a choice—there is no neutral position. Platforms *already* make choices about whether to leave content up or take it down, and they make that choice primarily based on their corporate bottom line. Right now, the tech industry's calculus for making choices about potentially harmful content is very simple: There is virtually no downside to leaving up even egregiously harmful content. Such content generates engagement, which in turn generates profit. It may also draw some bad public relations, but even that isn't necessarily a negative

consequence from a financial perspective. Without any credible fear of liability, there is simply no real incentive for social media companies to do anything other than leave content up unless it drives down user engagement.

If Section 230 immunity were limited in the way described above, it would complicate this calculus. Would it mean that valuable content would disappear? Possibly, although this is by no means inevitable. Imposing a modest restriction on the extravagant immunity currently enjoyed by the tech industry does not mean that social media companies would suddenly become automatically liable for user content, no matter how controversial. Plaintiffs would still at a minimum need a theory of liability and need to demonstrate the basic element of causation—high bars to clear with regard to speech. And whatever risk of liability for not taking content down would still need to be weighed against the potential benefit of keeping it up. Controversial speech would still generally mean more engagement, and engagement would still generally mean more profit.

Virtually every industry other than the tech industry (with the possible exception of the firearms industry) has to contend with the potential for liability when its products cause harm. Auto manufacturers can be sued when design flaws cause injury and death. Tobacco companies can be sued for deceptive advertising practices. Universities can be sued for failing to address sexual harassment. And yet these industries survive.

Finally, it is worth noting that under the status quo, certain forms of valuable content are never able to appear. When platforms are overrun by death threats, rape threats, harassment campaigns, and exposure of private information, many people—especially women, minorities, and other marginalized and vulnerable groups—go silent. They exit. Their valuable speech goes missing from those platforms.

Some also argue that any reform of Section 230 jeopardizes free speech in a larger sense, even if not strictly in the sense of violating the First Amendment. Of course, free speech is a cultural as well as a constitutional matter. It is shaped by nonlegal as well as legal norms, and tech companies play an outsized role in establishing those norms. There is indeed good reason to be concerned about the influence of tech companies and other powerful private actors over the ability of individuals to express themselves. This is an observation scholars and advocates who work on online abuse issues have been making for years—that some of the most serious threats to free speech come not from the government, but from nonstate actors. Marginalized groups in particular, including women and racial minorities, have long battled with private censorial forces as well as governmental ones.

But the unregulated internet—or rather, the selectively regulated internet—is exacerbating, not ameliorating, this problem. The current model, which shields platforms from liability, may ensure free speech for the privileged few, but protecting free speech for all will require legal reform.

Conclusion

In place of the free speech industry, we might envision the flourishing of multiple spaces—online and off, public and private—that provide the conditions necessary for free expression and democratic deliberation for different groups with different needs. This vision entails crafting law and policy to ensure that no one host or forum, or even one medium, dominates the shaping of public opinion or the boundaries of free speech. In addition to Section 230 reform, we should contemplate meaningful investments in traditional media, public education, universities, community centers, and small businesses to return to or become alternate sites of free expression and informed debate. A democratic society should demand the myriad possibilities of a free speech culture, not settle for the constraints of a free speech industry.

6

The Golden Era of Free Speech

ERWIN CHEMERINSKY AND ALEX CHEMERINSKY

It is not hyperbole to say that the internet and social media are the most important developments with regard to expression since the creation of the printing press. Throughout history a central problem with speech was scarcity of opportunities for expression. Even when newspapers were far more prevalent, there were a limited number, and a person had to be fairly rich to own one. When broadcasting developed, first radio and then television, there was the scarcity created by the limited number of frequencies on the spectrum. Indeed, in *Red Lion Broadcasting Co. v. Federal Communications Commission*, in 1969, the Supreme Court upheld the fairness doctrine for television and radio, including a right to reply law, based on the inherent scarcity of the broadcast spectrum.[1]

But with the internet and social media, no longer is scarcity an issue. Anyone can speak and potentially reach a mass audience instantly. Everyone has access to virtually unlimited information. A half-century ago, if free speech advocates had engaged in science fiction and tried to devise media to maximize expression, the Web and Twitter and Facebook and YouTube likely still would have been beyond their imaginations.

But every advance has costs as well as benefits. Many now point out that because it is easy to speak and reach a mass audience, there is more false speech and that it risks undermining democracy.[2] Also, there is no doubt that these developments have had a detrimental effect on traditional media, a tremendous loss of newspapers, and of the benefits they provide, such as investigative journalism.[3] The huge proliferation of sources of information means that it is easy for people to hear only what reinforces their beliefs. This contributes to the stark political polarization of American society.[4]

Ironically, while in *Red Lion* the justification for regulation was scarcity, now the call for regulation is based on there being too much speech. Although there are problems that require solutions, overall the internet and social media should

be heralded for their huge contributions to free speech, and great care should be taken to not enact regulations that could undermine what has been gained.

Simply put, the assumption of the First Amendment is that generally more speech is better. To be sure, that is not always so; if the speech is child pornography or false advertising, more is not better. But overall, the premise in analyzing speech should be to increase expression and in that way the internet and social media have brought us a golden age of free speech.

In the first section, we explain why the internet and social media should be heralded as radically changing free speech for the better. In the second section, we focus on Section 230 of the Communications Decency Act, which has become the focus of much of the debate over speech over the internet and social media. Although it has attracted significant criticisms from both the left and the right, we believe that Section 230 is desirable and only minor revisions would be desirable. Finally, in the third section we briefly identify some of the crucial problems with the internet and social media—false speech and foreign speech—and point out that there are no easy solutions under current First Amendment doctrine.

In Praise of the Internet and Social Media

A quarter of a century ago, Professor Eugene Volokh wrote a prescient article, *Cheap Speech and What It Will Do*, predicting how online communication would promote democracy by making publishing and receiving a vast amount and variety of information easily available to a wide range of individuals.[5] Professor Volokh wrote that before there was Facebook or Snapchat or Twitter. What he described has come to fruition in an exponential way.

In *Packingham v. North Carolina*, decided in June 2017, the Supreme Court spoke forcefully about the importance of the internet and social media as places for speech.[6] The Court declared unconstitutional a North Carolina law that prohibited registered sex offenders from using interactive social media where minors might be present.[7] The Court explained that cyberspace, and social media in particular, are vitally important places for speech.[8] Justice Kennedy, writing for the majority, explained:

> Seven in ten American adults use at least one Internet social networking service. . . . According to sources cited to the Court in this case, Facebook has 1.79 billion active users. This is about three times the population of North America.
>
> Social media offers "relatively unlimited, low-cost capacity for communication of all kinds." On Facebook, for example, users can debate

> religion and politics with their friends and neighbors or share vacation photos. On LinkedIn, users can look for work, advertise for employees, or review tips on entrepreneurship. And on Twitter, users can petition their elected representatives and otherwise engage with them in a direct manner. Indeed, Governors in all 50 States and almost every Member of Congress have set up accounts for this purpose. In short, social media users employ these websites to engage in a wide array of protected First Amendment activity on topics "as diverse as human thought."[9]

Three characteristics of the internet are particularly important for free speech. First, the internet has democratized the ability to reach a mass audience. It used to be that to reach a large audience, a person had to be rich enough to own a newspaper or to get a broadcast license. Now, though, anyone with a smartphone—or even just access to a library where there is a modem—can reach a huge audience instantaneously. No longer are people dependent on a relatively small number of sources for news.

It is difficult to overstate how much this changes the most fundamental nature of speech. As mentioned in the beginning of this essay, throughout world history, there always was a scarcity of most media for communication. Only so many books or newspapers or radio or television programs could exist. A half-century ago, the Court unanimously held that the federal government could regulate the broadcast media because of the inherent scarcity of spectrum space.[10] No longer is there such scarcity.

To be sure, this also means that false information can be quickly spread by an almost infinite number of sources. True information that is private can be quickly disseminated.[11] There is even a name for it: "doxing," or publishing private information about a person on the internet, often with the malicious intent to harm the individual.[12] The internet and social media can be used to harass.[13] A study by the Pew Research Center "found 40 percent of adult Internet users have experienced harassment online, with young women enduring particularly severe forms of it."[14]

Second, the internet has dramatically increased the dissemination and permanence of information, or to phrase this differently, it has enormously increased the ability to access information. The internet has the benefit of providing us all great access to information. As lawyers and law students, we can access Westlaw and all of the cases and secondary sources that would have required a trip to the law library. We can visit the great museums of the world online. We have access to virtually unlimited information from a myriad of sources.

But there is a downside to this easy access and permanence. Take defamation as an example. Imagine before the internet that a local newspaper published false information about a person that harmed his or her reputation. The falsity would

be known by readers of the paper and could be circulated by word of mouth. There could be great harm to the person's reputation. But the newspaper itself would largely disappear except to those wanting to search for it on microfilm or microfiche.

Now, though, the defamatory story can be quickly spread across the internet and likely will be there to be found forever. It is enormously difficult, if not impossible, to erase something from the internet.

Finally, the internet does not respect national boundaries. Again, there are great benefits to this. Totalitarian governments cannot easily cut off information to their citizens. When the revolution began in Egypt, the government tried to stop access to the internet, but people with satellite phones could maintain access and, consequently, disseminate what they learned.[15] The Supreme Court has estimated that 40 percent of pornography on the internet comes from foreign countries, making any attempt to control it within a country impossible.[16] Of course, as we saw in the 2016 presidential election and evidenced by Special Counsel Robert Mueller's report, this also allows foreign countries and foreign actors a vehicle for trying to influence the outcome of US elections.[17]

This, of course, is just a brief sketch of how the internet has changed free speech. But our point, like the Court's in *Packingham*, is that the internet is in-kind different from other media that exist for speech. The benefits are great, but so too are the potential costs.

Section 230

This golden age of free speech is possible thanks to the powerful shield against liability for online platforms created by Congress in 1996: 47 U.S.C. § 230. Under Section 230, "interactive computer services"—a capacious term that includes websites such as Facebook, Twitter, and TikTok—are immune from liability for most content posted by users. Platforms are permitted to solicit[18] and host user-generated content,[19] and to make editorial decisions regarding that content,[20] without fear of being sued. But Section 230 does not immunize platforms for content they have a hand in generating.

In the first major case analyzing Section 230, in 1997, the Fourth Circuit observed that

> [t]he purpose of this statutory immunity is not difficult to discern. Congress recognized the threat that tort-based lawsuits pose to freedom of speech in the new and burgeoning Internet medium. The imposition of tort liability on service providers for the communications of others represented, for Congress, simply another form of intrusive

> government regulation of speech. Section 230 was enacted, in part, to maintain the robust nature of Internet communication and, accordingly, to keep government interference in the medium to a minimum.[21]

Section 230 is essential to the functioning of the modern internet. It fosters the maximum amount of online speech by immunizing platforms for hosting content, and at the same time encourages a tolerable internet environment by permitting companies to freely edit or remove speech on their platforms. And it does it with remarkably low social costs because it protects online platforms from lawsuits they would almost always win anyway. Professor Jeff Kosseff has suggested that Section 230 contains "the twenty-six words that created the internet."[22]

In recent years, Section 230 has endured attack from both progressives and conservatives.[23] Progressives complain that online speech platforms create breeding grounds for bigotry and misinformation. Conservatives bemoan content moderation practices by tech giants such as Facebook and Twitter, which they claim favor left-leaning views at the expense of conservative speech.[24] Calls to amend Section 230 abound from both sides of the political aisle.[25] In December 2020, former president Donald Trump vetoed a major defense appropriation bill in protest of Congress's refusal to amend Section 230.[26] More than one hundred bills have been introduced in legislatures nationwide in the last several years in attempts to influence the content moderation decisions of social media platforms, either by encouraging more moderation or by requiring less moderation.[27] Although we are sympathetic to the criticisms of the current online speech environment—including online hate speech, violence, misinformation, revenge porn,[28] deepfakes,[29] and unequal access to critical informational technologies—we nevertheless believe that the issue here is not Section 230. Amending Section 230 significantly would be difficult to accomplish, likely unconstitutional, and more likely to cause harm than good. Indeed, we believe that Section 230 has been unfairly blamed for many of the problems that are inherent to the internet and social media.

Many proposals to modify Section 230's protection violate the First Amendment. In June 2021, a federal court preliminarily enjoined a Florida statute that "prohibits a social media platform from taking action to 'censor, deplatform, or shadow ban' a 'journalistic enterprise' based on the content of its publication or broadcast."[30] The court held that the statute was a content-based restriction of the speech of online platforms subject to strict scrutiny, and that the plaintiffs' First Amendment challenge was likely to prevail.[31] In a persuasive thirty-one-page opinion, the court identified the problem with most proposed Section 230 reforms: Online platforms have the right, protected by the First Amendment, to make editorial decisions regarding the content they host. Any

attempt to limit that right by restricting editorial power or mandating certain editorial decisions will be deemed a restriction or compulsion of speech that is subject to heightened scrutiny.

Several proposals put forward in Congress recently, such as the Protecting Constitutional Rights from Online Censorship Act, which would make it unlawful for internet platforms to restrict access to content, would similarly be unconstitutional on First Amendment grounds.[32] They would require online platforms to host messages the platforms would prefer to remove and would thus trigger strict scrutiny under the Supreme Court's compelled speech jurisprudence.

Proposals that would *require* platforms to moderate content also have serious constitutional problems. The government cannot directly require platforms to moderate objectionable speech if that speech is protected by the First Amendment. This means that Congress cannot directly impose requirements that platforms prohibit hate speech or misinformation, both of which are at least sometimes protected speech.[33] Any statutory attempt to restrict hate speech online, even if desirable, would need to meet strict scrutiny under the Supreme Court's rules for judicial analysis of content-based restrictions on speech.

But any reasonable observer can agree that some content moderation is required for the internet to be a tolerable place. If platforms could not remove First Amendment–protected terrorism accounts, or misinformation, or harassment, or posts espousing bigotry, the internet would be unusable. Section 230 gets around the basic First Amendment problem in content moderation by taking the government out of the equation and leaving editorial decisions up to the platforms and their users. It also is predicated on the very basic notion in the American free speech tradition that, if a person doesn't like someone's speech, he or she can just ignore it or respond with more speech. Section 230 encourages private companies to do what the government cannot do itself—moderate "cheap speech" that is harmful—while leaving open a channel for online speech available to everyone. In doing so, it permits a balance between free speech and decent speech that would be impossible to achieve through state action.[34]

Even where proposals to increase moderation do not pose constitutional difficulties, they are frequently undesirable as a matter of policy because they would significantly reduce liberty of speech online. One proposal, the PACT Act, co-sponsored by Sen. Brian Schatz (D-HI) and Sen. John Thune (R-SD), would make Section 230 immunity contingent on online platforms' complying with elaborate notice-and-takedown procedures for illegal or policy-violating content, publishing quarterly transparency reports, and even setting up toll-free call centers.[35] The point of the PACT Act is to ensure that companies rapidly respond to complaints of illegal content, and to encourage them to enforce their own acceptable use policies. The PACT Act cleverly attempts to quell lawful

but awful online speech by exhorting platforms to promulgate acceptable use policies and to enforce them. All platforms, including ones that exist for the express purpose of offering a safe haven from censorship, prohibit uses that would be constitutionally protected if restricted by the government.[36] The PACT Act encourages moderation of that speech, without directly requiring it, to restrict speech without running into a looming state action problem. But it may lead to a decrease in freedom of speech, as platforms would overregulate to avoid the risk of tort or criminal liability.[37]

There is considerable evidence in the context of the Digital Millennium Copyright Act (DMCA)[38] (which uses a notice-and-takedown system) that notice-and-takedown procedures can chill online speech by giving anyone a way to veto speech they do not like. In a persuasive article, Professor Wendy Seltzer explains that the DMCA serves as a "prior restraint by proxy," encouraging platforms to overmoderate online speech without the presence of state action, by requiring them to set up private adjudication schemes for online speech that may constitute copyright infringement.[39]

The PACT Act would surely compound the potential free speech problems caused by the DMCA. Not only would it impose a similar notice-and-takedown requirement, but it would do it in a context in which adjudication is much more complicated. It is much easier for a platform to look at two photos, or listen to musical compositions, and then make a quick determination as to whether one infringes the copyright of the other, than it would be in the context of speech torts. For example, if a platform receives a notice that a post is defamatory, it would be required to rapidly determine whether the post is true. That would require significant fact-finding resources that platforms are not prepared to expend. Rather than investigating the truthfulness of every claim of defamation on the internet (of which there would be many), platforms would lean toward the side of overmoderation.

This approach to regulating truthfulness has been visible in the context of the COVID-19 pandemic. Facebook has struggled mightily to reduce antivaccine misinformation on its platform. Although Facebook bans antivaccine advertisements,[40] it admits that the most popular Facebook post from January to March 2021 was one that promoted antivaccine doubts.[41] This occurred even though its algorithm frequently moderates posts that are well meaning. For example, the *Washington Post* reported that although Facebook has put its algorithms and fifteen thousand human moderators to the task of detecting vaccine misinformation—it has taken down more than twenty million posts since the start of the pandemic—"it routinely misses new memes and conspiracy theories, while at times scrubbing legitimate information."[42] If it is too difficult for Facebook to moderate the truthfulness of posts about commonly accepted scientific information, we shudder to think how it might respond to

private claims of defamation, or how it might respond when pressed with factual understandings that shift over time.

Content moderation occurs at too massive a scale for companies to risk liability for each item of user-generated content they host. In a white paper published in 2013, Facebook revealed that at that time—and the number is likely much larger now—more than 350 million photos were uploaded to the platform (not inclusive of Instagram) *per day*.[43] According to the paper, 4.75 billion items are shared on Facebook every day.[44] In the first quarter of 2021, Facebook took action against five million pieces of content that violated its child nudity and sexual exploitation policies alone.[45] From 2015 to 2018, Twitter removed more than 1.2 million Twitter accounts promoting terrorism.[46] At this scale, content moderation acts on a system of probability and proportionality, not individual fairness.[47] If companies face increased risk, they will only respond by hedging their bets and deleting lawful user speech to avoid expensive lawsuits, even if they would win those suits easily on First Amendment grounds.

The overmoderation problem was visible following the only major amendment that there has been to Section 230. In 2018, Congress enacted the Allow States and Victims to Fight Online Sex Trafficking Act (FOSTA), which creates exceptions to Section 230 for claims under sex-trafficking laws.[48] Shortly after the passage of FOSTA, Craigslist removed its entire personal ads section.[49] It explained that it could not continue hosting users' personal advertisements "without jeopardizing all our other services," because it would face potential civil and criminal liability if it hosted content that violated certain sex-trafficking laws.[50] Any future attempts to add carveouts to Section 230 should contend with the fact that restriction of illegal or objectionable speech online, even if constitutional, will inevitably lead to restriction of desirable speech as well. Given the presumption that more speech is better underlying the American First Amendment tradition, we argue that such carveouts should be made sparingly and narrowly.

Moreover, carving categories of illegal speech out of Section 230's protection will not resolve many of the problems on the internet. Hate speech and bigotry will still thrive because it is not illegal to be a bigot. Misinformation will still be present because of the enormous challenge of sorting fact from fiction,[51] and because much false speech is protected by the First Amendment.[52] Users will still be able to discuss and organize violent activity without rising to the level of incitement. Despite inevitable overregulation, illegal and objectionable speech will still survive simply because content moderation at scale is impossible to perform.

Pressed with the practical and constitutional limitations of statutorily required content moderation, we think that the best way to approach online

speech is to trust in the First Amendment rights of private entities. Section 230 is an elegant and simple method of attaining a balance between free speech and online decency. It encourages online platforms to permit their users to speak freely, to comment on each other's posts and on articles, and to upload anything they like, and empowers those platforms to moderate that speech in a way that the government cannot. We recognize that a great deal of online speech is abhorrent, and that content moderation is not always perfect or fair, but we ultimately conclude that the current regime is the best possible way to ensure the endurance of the golden age of speech.

This does not mean, however, that online speech regulation should stop only at an immunity. Social pressures motivated by thorough reporting and terrific scholarship are already causing platforms to begin improving moderation practices. Those efforts can and should be improved by efforts to improve the transparency of content moderation, perhaps through the establishment of a commission or through transparency reports, such as those that would be required by the PACT Act. It will be easier for everyone to have a conversation about online speech if we have a better sense of the facts. And we can have that conversation without further carving away at Section 230's core protection.

The Problems of the Internet

Although we believe that the internet and social media have created a golden age of speech, we do not deny that the ease of speech that reaches a mass audience creates serious problems. Some already have drawn legislative fixes. California, for example, has adopted laws to address revenge porn (posting sexually explicit photos of a person without consent). We believe that such regulations are constitutional. As for revenge porn, the most basic notions of privacy support preventing sexually explicit photos of a person being publicly disseminated without his or her consent.

But the internet and social media also expose some serious issues where there are not easy answers and current First Amendment jurisprudence makes solutions difficult. We identify two: false speech and foreign speech.

False Speech

False information can cause great harms. In the political realm, it can change the outcome of elections. We have seen, in the area of public health, that it can exacerbate the spread of a communicable disease and undermine efforts to control it through masks and vaccinations.

There is no consistent answer as to whether false speech is protected by the First Amendment. In some areas, the Court has found constitutional protection for false expression, but in other instances it has upheld the ability of the government to punish false speech. This seems inevitable because analysis must be contextual and must be the result of balancing of competing interests, which will prevent a consistent approach to false speech. That is, the Court never will be able to say that all false speech is outside of First Amendment protection or that all false speech is constitutionally safeguarded.

In some instances, the Court has emphatically declared the importance of protecting false speech. The most important case in this regard—and one of the most important free speech decisions of all time—is *New York Times Co. v. Sullivan*.[53] One L. B. Sullivan, an elected commissioner of Montgomery, Alabama, successfully sued the *New York Times* and four African American clergymen for defamation for an advertisement that had been published in the newspaper on March 29, 1960. The ad criticized the way in which police in Montgomery had mistreated civil rights demonstrators. There is no dispute that the ad contained minor false statements.

The Supreme Court held that the tort liability violated the First Amendment. Justice Brennan, writing for the Court, explained that criticism of government and government officials was at the core of speech protected by the First Amendment.[54] The Court said that the fact that some of the statements were false was not sufficient to deny the speech First Amendment protection.[55] The Court explained that false "statement is inevitable in free debate and [it] must be protected if the freedoms of expression are to have the 'breathing space' that they 'need . . . to survive.' "[56]

Subsequently, in a very different context, in *United States v. Alvarez*, the Court again recognized the importance of judicial protection of false speech.[57] *Alvarez* involved the constitutionality of a federal law that made it a crime for a person to falsely claim to have received military honors or decorations.[58] Justice Kennedy wrote for a plurality of four and concluded that the law imposed a content-based restriction on speech and thus had to meet the most "exacting scrutiny."[59] He explained that the government failed this test because it did not prove any harm from false claims of military honors and because the government could achieve its goals through less restrictive alternatives.[60]

Most importantly, Justice Kennedy expressly rejected the government's argument that false speech is inherently outside the scope of the First Amendment. Justice Kennedy declared:

> Absent from those few categories where the law allows content-based regulation of speech is any general exception to the First Amendment for false statements. This comports with the common understanding

> that some false statements are inevitable if there is to be an open and vigorous expression of views in public and private conversation, expression the First Amendment seeks to guarantee.[61]

Justice Kennedy further explained: "Even when considering some instances of defamation and fraud, moreover, the Court has been careful to instruct that falsity alone may not suffice to bring the speech outside the First Amendment."[62]

Yet there are other contexts in which the Supreme Court has refused to provide protection for false speech. For example, it is clearly established that false and deceptive advertisements are unprotected by the First Amendment.[63] The government, of course, can constitutionally prohibit making false statements under oath (perjury) or to law enforcement officials.[64] As for the former, the Court generally has treated commercial speech as being of lower value than political speech and that may make it easier to say that false advertising is entitled to no constitutional protection at all. As for perjury, there is a requirement for proof of a knowing and intentional falsehood.

More generally, the Court has declared that "[f]alse statements of fact are particularly valueless [because] they interfere with the truth-seeking function of the marketplace of ideas,"[65] and that false statements "are not protected by the First Amendment in the same manner as truthful statements."[66] Indeed, the Court has declared that "the knowingly false statement and the false statement made with reckless disregard of the truth, do not enjoy constitutional protection."[67] Whether these statements remain good law after *Alvarez* is an open question, but whether *Alvarez* would come out the same way today is also uncertain. It was a 6–3 decision with Justices Kennedy and Ginsburg in the majority, and Justices Scalia, Thomas, and Alito dissenting. It is not clear how the three newest Justices—Gorsuch, Kavanaugh, and Barrett—would have voted in the case. But ideologically they are much more like the dissenters in *Alvarez* than those in the majority.

The Court's inconsistent statements about false speech can be understood as reflecting the competing interests inherent in First Amendment analysis. On the one hand, false speech can create harms, even great harms. Speech is protected especially because of its importance for the democratic process, but false speech can distort that process. Speech is safeguarded, too, because of the belief that the marketplace of ideas is the best way for truth to emerge. But false speech can infect that marketplace and there is no reason to believe that truth will triumph. False speech can hurt reputation, and it is fanciful to think that more speech necessarily can undo the harms.

But at the same time, there is great concern about allowing the government to prohibit and punish false speech. *New York Times Co. v. Sullivan* was unquestionably correct when it said that false "statement is inevitable in free debate, and . . . it

must be protected if the freedoms of expression are to have the 'breathing space' that they 'need . . . to survive.' "[68] Also, allowing the government to prohibit false speech places it in the role of being the arbiter of truth. Justice Kennedy captured the dangers of this in *Alvarez*:

> Permitting the government to decree this speech to be a criminal offense, whether shouted from the rooftops or made in a barely audible whisper, would endorse government authority to compile a list of subjects about which false statements are punishable. That governmental power has no clear limiting principle. Our constitutional tradition stands against the idea that we need Oceania's Ministry of Truth.[69]

The result is that it always will be impossible to say either that false speech is always protected by the First Amendment or that it never is protected by the First Amendment. Inescapably, the Court will need to balance the benefits of protecting the false speech against the costs of doing so. Such balancing is inherently contextual and will yield no general answer as to the Constitution's protection of false speech. This, then, creates an enormous challenge for dealing with false speech over the internet and social media. Without a doubt, the internet and social media make the problem of false speech much greater: False speech is easier to disseminate to a large audience and has much greater permanent availability. But we do not think that there can be a general rule about how to treat false speech over these media without either doing great harm to free speech or providing protection in instances where restrictions should be constitutional.

Foreign Speech

There is now incontrovertible evidence that Russia engaged in a concerted effort to use speech, including false speech, to influence the outcome of the 2016 presidential election.[70] American intelligence agencies recognized this soon after the election.[71] In February 2018, Special Counsel Robert Mueller issued a thirty-seven page indictment charging thirteen Russians and three companies with executing a scheme to subvert the 2016 election and help to elect Donald Trump as president.[72] Mueller's indictment details "how the Russians repeatedly turned to Facebook and Instagram, often using stolen identities to pose as Americans, to sow discord among the electorate by creating Facebook groups, distributing divisive ads and posting inflammatory images."[73] Russia's efforts to influence the election primarily were through the internet and social media.

There is understandable widespread outrage at the idea of Russia's engaging in a concerted effort to influence the outcome of the 2016 presidential election. Yet, it must be remembered that the United States long has been doing exactly

this, using speech—including false speech—to try to influence the outcome of elections in foreign countries. Dov Levin, a professor at Carnegie Mellon University, identified eighty-one instances between 1946 and 2000 in which the United States did this.[74] As one report explained,

> Bags of cash delivered to a Rome hotel for favored Italian candidates. Scandalous stories leaked to foreign newspapers to swing an election in Nicaragua. Millions of pamphlets, posters and stickers printed to defeat an incumbent in Serbia. The long arm of Vladimir Putin? No, just a small sample of the United States' history of intervention in foreign elections.[75]

Although condemnation of Russian meddling in the American election is easy, the underlying First Amendment issue is difficult. Obviously illegal conduct, such as hacking into the Democratic National Committee headquarters and the subsequent dissemination by the hackers of unlawfully gained information,[76] is not constitutionally protected. But what about foreign speech that is legal and that expresses an opinion—even false speech?

The Supreme Court repeatedly has said that the source of information does not matter for First Amendment purposes. In *First National Bank of Boston v. Bellotti*, the Supreme Court declared unconstitutional a Massachusetts law that prohibited banks or businesses from making contributions or expenditures in connection with ballot initiatives and referenda.[77] Justice Powell, writing for the Court, concluded that the value of speech is in informing the audience. Any restriction on speech, regardless of its source, therefore undermines the First Amendment. Justice Powell explained: "The inherent worth of the speech in terms of its capacity for informing the public does not depend upon the identity of its source, whether corporation, association, union, or individual."[78]

The Court relied on this in *Citizens United v. Federal Election Commission* to hold that corporations have the constitutional right to spend unlimited amounts of money directly from their treasuries to elect or defeat candidates for political office.[79] The Court stressed that the value of the speech does not depend on the identity of the speaker and held that corporate speech is protected not because of the inherent rights of corporations, but because all expression contributes to the marketplace of ideas. The Court wrote: "The basic premise underlying the Court's ruling is its iteration, and constant reiteration, of the proposition that the First Amendment bars regulatory distinctions based on a speaker's identity, including its 'identity' as a corporation."[80] On other occasions, too, the Court has declared that "[t]he identity of the speaker is not decisive in determining whether speech is protected."[81]

But if this is so, why should it matter whether the speaker is a foreign government or foreign individual? Federal law prohibits foreign governments, individuals, and corporations from contributing money to candidates for federal office.[82] A federal court upheld this restriction on foreign speech, declaring: "It is fundamental to the definition of our national political community that foreign citizens do not have a constitutional right to participate in, and thus may be excluded from, activities of democratic self-government. It follows, therefore, that the United States has a compelling interest for purposes of First Amendment analysis in limiting the participation of foreign citizens in activities of American democratic self-government, and in thereby preventing foreign influence over the US political process."[83] But can this be reconciled with the Supreme Court's declaration that the identity of the speaker should not matter in First Amendment analysis? Although it is not a comfortable answer, we do not see a way to exclude foreign speakers consistent with the Court's premise that the identity of speaker cannot be the basis for regulation. The assumption of the First Amendment is that more speech is better and that is true whether the speaker is foreign or domestic.

At the very least, it would be desirable to have disclosure of the identity of speakers so that people can know when the speech is coming from a foreign government or other foreign source. But this, too, raises First Amendment issues, as the Supreme Court has held that there is a First Amendment right to speak anonymously. In *Mclntyre v. Ohio Elections Commission*, the Court declared unconstitutional a law that prohibited the distribution of anonymous campaign literature.[84] Justice Stevens, writing for the Court, stated: "An author's decision to remain anonymous, like other decisions concerning omissions or additions to the content of a publication, is an aspect of the freedom of speech protected by the First Amendment."[85] Moreover, Justice Stevens said that anonymity also provides a way for a speaker "who may be personally unpopular to ensure that readers will not prejudge her message simply because they do not like its proponent."[86] Wouldn't that be especially true of a foreign government or foreign individuals who were trying to influence an American election?

We would draw a distinction between a right to speak anonymously and a right for a speaker to mask its identity in order to deceive and manipulate voters. Whether the speaker is domestic or foreign, that should be regarded as a form of fraud that is unprotected by the First Amendment.

The transnational nature of the internet makes controls elusive even if they are constitutional. As the 2016 presidential election shows, foreign governments can use the internet and social media to influence elections without their officials' and agents' ever entering the United States. It is unclear how the law can be applied to them. The internet gives them the ability to engage in false speech (and all other kinds of expression) with relatively little fear of legal sanctions. The United

States could attempt to impose international sanctions on nations that interfere with our elections. Under the Court's recent decision in *Agency for International Development v. Alliance for Open Society International*, the First Amendment does not apply to the government's extraterritorial actions.[87] The United States' imposing sanctions on foreign entities that engage in such behavior is the appropriate response, not the government amending Section 230 to require online platforms to engage in moderation of foreign speakers. But whether such international sanctions would make any difference is at best uncertain.

Conclusions

Our basic point is that the internet and social media provide enormous benefits in terms of speech and also unique challenges. Great care should be taken to make sure that any regulation actually solves the problems without sacrificing the great gains in expression from these relatively new media.

7

Section 230 Reforms

SHELDON WHITEHOUSE

Introduction

For better or worse, a few social media companies dominate today's internet. In recent years, with the spread of disinformation and other dangers online, the worse now overshadows the better.

Social media platforms—companies that facilitate information sharing through virtual networks—have shielded themselves more than any other media from responsibility for destructive content that they house and propagate.[1] They claim that their algorithms simply promote whatever is selected by the collective wisdom of the public,[2] and that they lack the resources or expertise to identify and remove unlawful or untruthful content. But the truth is they are not neutral or incapable observers. Social media companies spread disinformation,[3] exacerbate preexisting biases,[4] and disseminate unlawful content[5] because of deliberate, profit-seeking choices. These platforms choose how to structure their services; what content to allow or disallow; what content to promote; what ads to sell, and to whom; and how they connect advertising to the content users consume or create.

These deliberate choices create real-world harm. Although Facebook Chief Operating Officer Sheryl Sandberg initially tried to deflect blame to other platforms that she claimed "don't have our abilities to stop hate, don't have our standards and don't have our transparency," a leaked internal Facebook report acknowledged that the company chose not to act against January 6, 2021, insurrection plotting on its platform.[6] A *New York Times* study of the 2020 Senate run-off election in Georgia found that the "top 20 Facebook and Instagram accounts spreading false claims aimed at swaying voters in Georgia accounted for more interactions than mainstream media outlets."[7] A 2021 study by University of Southern California researchers found Facebook job advertisements were infected with gender bias, disproportionately targeting men for male-dominated

industries and women for female-dominated industries.[8] These incidents signal problems that demand our attention.

Large social media companies play a dominant role in how we share and consume information. For the purposes of this essay, "[s]ocial media companies" are platforms that facilitate online communities among users who create and disseminate user-generated content. Social media companies also collect and sell that user data to third parties. While other companies can facilitate the spread of illegal content or misinformation, the scale and anonymity offered by major social media companies makes them a particular hazard.

The first step to stopping the spread of illicit content and misinformation online is to address the shortcomings of Section 230 of the Communications Decency Act of 1996. Section 230 was originally meant to help platforms create a safe and inclusive internet by granting legal immunity, unheard of in other industries.[9] The law was predicated on a vision of the internet as a community notice board rather than its current role as an influential and manipulated information network. Since Section 230's enactment, the internet has grown more than one hundredfold, from 40 million users in 1996[10] to 4.7 billion users in 2021.[11] Profit-seeking manipulation of the supposed "community notice board" has become the operating practice of mighty social media giants whose platforms dominate the flow of online information. Section 230 has not kept pace. Indeed, its unprecedented immunity protection has grown, as courts allowed Section 230 to shield a range of illicit corporate behavior unimaginable at the time of its passage.

Simply put, Section 230 is outdated, and we need to be rid of it. The Senate has the broad bipartisan consensus that we need to be rid of it. However, once you try to get rid of it, a dispute over alternatives emerges. The broad bipartisan agreement in the Senate that Section 230 must go breaks down over the alternatives.[12] So we must grapple with how a replacement should clarify the unique responsibilities of social media platforms, regulate their moderation and advertising policies, and give redress to parties whom they injure.

I should say at the beginning that I would support simply repealing Section 230 and letting courts sort it out. This has the advantage of legislative simplicity and speed. It also minimizes the hand of Congress in an area that relates to speech, where our own political motives—whether of incumbency or party—create their own hazards. Better to minimize Congress's hand in this.

Most of the questions that would come up in court post-repeal would find ready answers in existing legal doctrines, with familiar structures and duties. Repeal is not a ticket to an alien legal environment; it's actually a return to established legal norms.

At the same time, because Section 230 has effectively prevented our existing legal norms from adapting to the way social media work, repeal may not fully address the unique problems social media have created. While the existing legal

regimes may be sufficient to stop the spread of illicit content, the volume of content online presents practical challenges for social media platforms seeking to root out illicit content. Repealing Section 230 will also not entirely solve the pernicious problem of disinformation on social media platforms. Where disinformation targets an injured individual, liability law will usually clean up the mess. But where disinformation is general, dispensed like pollution into the information environment, it will be hard for a litigant to show standing to sue for particularized harm. Indeed, the harm may be too widespread for any individual litigant to take an interest.

Moreover, litigation (or the threat of it) could be used by well-resourced interests to influence platforms' behavior, if not the law; and well-resourced platforms could use litigation (or the threat of it) to protect their incumbency. We need to think in advance about how to sort through that tangle.

To craft a replacement for Section 230, Congress must first recognize that illicit content and disinformation are distinct problems that should be treated differently. While we can ban illicit content, we must fight disinformation through transparency and accountability measures that stop the spread of false facts and reveal their true sources. Consumers deserve to know who is behind content that is presented to them.

Second, Congress should account for the ways in which social media operate differently than legacy media, and the unique challenges posed by the volume of content online. Congress should create a notice-and-takedown system for illegal content or conduct, mandate stronger transparency requirements, ensure that platforms are liable for their own programs or practices, and require that platforms share their algorithms with third parties so that we can study how they facilitate disinformation.

On their own, these changes would not be enough to end the proliferation of disinformation online, but they would bring success within reach. The First Amendment limits how far the government can go in regulating speech. We should feel similar discomfort with empowering private corporations to decide what is true; as we have already seen, they may do so in ways that benefit their own interests. Transparency is the surest course. These measures, combined with an engaged citizenry to act on that information, can pressure social media companies to take their role as stewards of the public sphere seriously.

Section 230 Is Outdated and Stretched Beyond Its Original Bounds

When Section 230 of the Communications Decency Act was enacted in 1996, the internet was very different than it is today. In October 1995, the Pew Research Center observed: "Few see online activities as essential to them, and

no single online feature, with the exception of E-Mail, is used with any regularity. Consumers have yet to begin purchasing goods and services online, and there is little indication that online news features are changing traditional news consumption patterns."[13] Those were the days. Only forty million people worldwide had any internet access—1 percent of users now.[14] When Congress debated the Communications Decency Act, lobbyist and media attention focused on its regulation of long-distance carriers, phone companies, and cable providers, not on Section 230.[15]

The internet's "Big Three information services" then were AOL, CompuServe, and Prodigy.[16] These three providers were the original online portals for news, websites, forums, chats, and file-sharing services.[17] Unlike modern social media platforms, the "Big Three" charged monthly rates to use their services.[18] Subscriptions were their primary source of revenue. Internet advertising was still in its infancy.[19]

Under the legal regime that existed at that time, entities that affirmatively spoke (publishers), or entities that published or disseminated others' speech (distributors), could be subject to liability.[20] The internet did not fit neatly into those definitions. In 1991, a columnist for a website within CompuServe's umbrella posted negative comments about a competitor, and that competitor sued.[21] The court found that CompuServe could not be held liable for libel: Because CompuServe did not review any of the content posted on its forums and websites, it could not be considered a publisher.[22] Four years later, the brokerage firm Stratton Oakmont sued Prodigy for user-generated comments on one of its bulletin boards that called into question Stratton Oakmont's business practices.[23] The court found that, because Prodigy moderated some content for "offensiveness and bad taste" and established online community policies, it was akin to a publisher and could be sued for failing to delete posts that allegedly defamed Stratton Oakmont.[24] In effect, the courts ruled that if platforms moderated anything, they were liable for everything. The *Stratton Oakmont* court awarded $200 million in damages—roughly equivalent to Prodigy's estimated annual revenue for 1994.[25]

Many believed the *Stratton Oakmont* decision would incentivize all platforms to abandon all moderation efforts, as only a totally hands-off approach would protect platforms from liability.[26] In response, Rep. Christopher Cox (R-CA) and Rep. Ron Wyden (D-OR) (now Senator Wyden) proposed what would become Section 230.[27] The goal, according to Rep. Cox, was "to make sure that everyone in America has an open invitation and feels welcome to participate in the internet. . . . I want to make sure that my children have access to this future and that I do not have to worry about what they might be running into online."[28] The intent of Cox and Wyden's original Section 230 was to afford platforms enough immunity to let them moderate content, while keeping liability for intellectual property abuses and federal criminal law violations.

But soon content creation flew past platforms' ability to monitor the content being created. No one anticipated the technological changes that shortly made it feasible to offer masses of internet users access to "user-centric spaces they could populate with user-generated content."[29] As of January 2021, four of the world's six largest websites by monthly search visitors (YouTube, Facebook, Twitter, and Instagram) occupied this new space.[30] More than 50 percent of Americans now receive at least some of their news from social media.[31] The volume of content soared. In 1995, all of Prodigy's websites and message boards received a combined sixty thousand new comments per day.[32] In 2021, Facebook processed more than eight times as many comments *per minute*—in addition to innumerable status updates, likes, and photographs.[33]

Judicial decisions have since expanded Section 230 far beyond the bill's original purpose. As the internet changed, courts construed Section 230 to shield websites that failed to effectively moderate even known objectionable content.[34] In 1997, Kenneth M. Zeran sued AOL after his name was improperly attached to messages celebrating the Oklahoma City bombing,[35] arguing that the company had a duty to quickly process and take down repeated instances of these comments.[36] The Fourth Circuit held that Section 230 immunizes even a service provider that receives adequate notice about illegal material on its system and nevertheless fails to act.[37] Similarly, in 2014, a court found that Section 230 immunized the web hosting platform GoDaddy for its failure to take down Texxxan.com, a revenge porn website that utilized GoDaddy's hosting services.[38] As a result, Section 230 now "immunizes platforms even when they negligently, recklessly, or knowingly fail to curb abuse of their services."[39] Courts have created a safe harbor for the very evil that Section 230 was originally intended to counter: willful blindness by platforms to illicit and illegal content on their sites.

No other business enjoys the blanket immunity that courts, interpreting Section 230, have granted to internet service providers. Congress should update Section 230 and restore traditional lines of accountability.

Without an Effective Regulatory Regime, Illegal Content and Misinformation Proliferate

Platforms know illegal and harmful content spreads widely across their sites, but fail to stop it. Some blame for this inertia falls on the platforms themselves, which have political and business incentives to permit harmful content to propagate. Blame also falls on outside forces, including America's geopolitical adversaries, who exploit our system's vulnerabilities to their own nefarious ends. And blame must fall on malign propagators, who take easy advantage of the platforms' failures. To address the shortfalls of Section 230, we must remedy the incentives for platforms and users that accelerate harmful content's spread.

A caution light should go on whenever a politician talks about regulating "harmful content." I get that. But harmful content is real, and has been dealt with by the law for centuries. The closer we hew to well-established legal traditions, the less the opportunity for new mischief, and the more we will benefit from the wisdom of the past that was suppressed by Section 230.

The Rampant Spread of Misinformation and Illegal Content

Without incentives to cleanse platforms of illegal content that managers know is present, this toxic material has proliferated.[40] In 2019, a team of researchers discovered that YouTube's automated recommendation system was encouraging consumption of videos of prepubescent, partially clothed children.[41] Although YouTube had learned of the issue months earlier, it failed to act.[42] Drug dealers use Facebook to post images of illicit drugs (or images that are a code for drugs) with contact information that can be used to buy the drugs. Facebook knows this, yet the illegal content is not removed, the user is not banned, and the illicit sale is permitted to proceed. It has also failed to respond effectively to drug cartels and human traffickers, who use its platform to facilitate illegal activities.[43] Similarly, Airbnb claims that Section 230 shields it from liability for noncompliant listings on its site, including listings for residences that violate state and municipal safety codes.[44]

A growing body of evidence shows that disinformation and misinformation spread like wildfire on social media. A 2019 MIT study found that "falsehoods are 70% more likely to be retweeted on Twitter than the truth, and reach their first 1,500 people six times faster."[45] By contrast, true stories are rarely retweeted by more than a thousand people.[46] Making matters worse, the spread of falsehoods tends to accelerate at critical moments when salience is high and truth counts the most, such as presidential elections.[47] To understand this phenomenon, consider: The Georgia Star News, a fake news outlet devoted to boosting former president Trump, published a story claiming that Dominion voting machines flipped votes from Donald Trump to Joe Biden during the 2020 election[48]; Facebook users shared the story more than two thousand times, eventually reaching up to 650,000 people.[49]

The pattern holds across different types of damaging disinformation. The Center to Counter Digital Hate studied 409 accounts from Facebook, Instagram, YouTube, and Twitter, and found that the largest accounts associated with antivaccination content added close to eight million followers since 2019.[50] Clusters of users opposing or questioning the use of vaccines are more numerous, and are more effective at reaching people who are undecided about

vaccines, than are clusters who support vaccines.[51] As a result, Facebook users who receive all their COVID-19 news from Facebook are far likelier to refuse vaccination, a potentially fatal decision.[52]

The spread of climate denial and other scientific misinformation is a major problem for platforms. *E&E News* recently revealed how *The Epoch Times*, a top mouthpiece for polluter-funded climate denial groups, rose to the top-ten most popular purveyors of content on Facebook, with more than forty-four million views between April and June 2021 alone.[53] This popular content included false or wildly misleading stories and opinion pieces casting doubt on the science of climate change. For example, *The Epoch Times* ran articles suggesting the sun, not fossil fuels, is driving climate change, and that renewable energy threatens the electricity grid.[54]

Social Media Companies Fail to Stop the Spread of Misinformation

While social media companies would undoubtedly deny that they intend to spread disinformation, and may even have policies aimed at minimizing its impact, they have not responded adequately to stop the spread of known misinformation on their platforms.

On several notable occasions, Facebook has failed to intervene to stop the spread of misinformation:

- *VICE News* examined multiple Facebook groups devoted to sharing unverified or discredited information about using a livestock dewormer called ivermectin to treat COVID-19. The Food and Drug Administration has pleaded with the public to avoid using the drug,[55] including broadcasting warnings directly on social media platforms.[56] The platform's explicit policies require it to remove any "false claims about how to cure or prevent COVID-19."[57] Nonetheless, Facebook allowed the groups—some with many thousands of users each—to promote ivermectin widely, including sharing strategies for obtaining it without a prescription from overseas.
- In July 2020, Senators Elizabeth Warren (D-Mass.), Thomas Carper (D-Del.), Brian Schatz (D-Haw.), and I (D-R.I.) sent a letter to Mark Zuckerberg regarding reports that Facebook created fact-checking exemptions for people and organizations who spread disinformation about climate change.[58] This letter responded to a *Washington Examiner* climate denial piece which five climate scientists declared to be "highly misleading." The piece was allowed to stay on Facebook, despite being flagged by Facebook's own fact-checking moderation partners, because Facebook claimed it was an "op-ed" and hence apparently immune to fact-checking.[59]

- In March 2021, thirteen environmental groups sent a letter to Zuckerberg urging Facebook to take a more proactive role in combatting climate change disinformation.[60] Despite being given a list of climate-change-denying publications, Facebook failed to act against posts from these sites.[61]

Facebook benefited from climate disinformation, as climate disinformation ads on Facebook receive tens of millions of impressions.[62]

Misinformation Is Spread by Bad Actors and the Companies' Own Algorithms

While social media companies may argue that the spread is driven by the innate appeal of certain falsehoods, certain falsehoods are strategically accelerated on social media by deliberate propagators.

Almost all of what we know about how these companies operate comes from outside researchers. The vast majority of social media platforms do not allow outside researchers access to their algorithms or data.[63] With some services, such as closed messaging products like WhatsApp, it's impossible to know what happens online "without labor-intensive ethnographic techniques like sitting next to [users] or interviewing them."[64] The opaque nature of the platforms obscures what they're really up to. "If the public, or even a restricted oversight body, had access to the Twitter and Facebook data . . . it would be harder for the companies to claim they are neutral platforms who merely show people what they want to see."[65]

Yet this research indicates that the platforms' own algorithms play a significant role in spreading misinformation. A recent NYU study found that users of Facebook are more engaged with misinformation: "[F]rom August 2020 to January 2021, news publishers known for putting out misinformation got six times the amount of likes, shares, and interactions on the platform as did trustworthy news sources, such as CNN or the World Health Organization."[66] Similarly, a 2018 MIT study found that "misleading stories on Twitter . . . performed better than factual stories."[67] *Guardian* journalist Julia Carrie Wong has reported how Facebook's algorithm repeatedly encouraged her to join QAnon groups. Wong's retelling of this process details how common an occurrence it is for Facebook to recommend extremist content, with Facebook's "own internal research in 2016 [finding]" that "64% of all extremist group joins are due to our recommendation tools."[68]

The information environment, in turn, responds to the incentives created by the algorithms. According to internal Facebook documents obtained by the *Wall Street Journal*, changes in the company's algorithm led publishers and political

parties to reorient their posts toward outrage and sensationalism because "[t]hat tactic produced high levels of comments and reactions that translated into success on Facebook."[69] Facebook's data scientists acknowledged, "Our approach has had unhealthy side effects on important slices of public content, such as politics and news."[70] Episodic discussions like this give a mere anecdotal glimpse into social media platforms' behavior, but it's a disturbing glimpse.

Notably, it is in social media companies' financial interest for misinformation and disinformation to spread. This content drives user engagement; platforms' algorithms amplify content that drives engagement because engagement drives ad dollars. Internal Facebook documents indicate the company resisted changing algorithms that reduced the spread of misinformation and divisive content at the expense of user engagement.[71]

Social media platforms' algorithms can drive radicalization. Far-right content consistently received the highest engagement per follower of any partisan group, with far-right misinformation purveyors having "on average 65% more engagement per follower than other far right pages."[72] In 2018, the investigative journalism site Bellingcat obtained a leaked Discord server for the neo-Nazi organization Atomwaffen. One thread from the server featured seventy-five members discussing the process by which they were radicalized. Thirty-nine credited the internet with their radicalization, with fifteen specifically mentioning YouTube.[73] Similarly, in 2018 two researchers created new YouTube accounts and viewed content from right-leaning channels, left-leaning channels, and mainstream channels, and then let the algorithm go wherever it wanted. They found that "being a conservative on YouTube means that you're only one or two clicks away from extreme far-right channels, conspiracy theories, and radicalizing content."[74] According to a 2018 study by the National Consortium for the Study of Terrorism and Responses to Terrorism, social media—plus personal blogs and forums—played a role in radicalizing and mobilizing roughly 90 percent of lone extremist actors.[75]

Compounding the problem, America's geopolitical rivals exploit social media companies' hesitance to police content to conduct malicious disinformation campaigns. Russian intelligence operatives have been deeply engaged in a coordinated campaign to undermine American faith in COVID-19 vaccines.[76] Even before COVID, a 2018 study found that "Russian trolls and sophisticated bots promote both pro- and antivaccination narratives."[77] This makes sense: Russia's goal is not the spread of antiscience propaganda so much as the promotion of "political discord,"[78] with the ultimate aim of reducing America's effectiveness as a global leader. It pursued a disinformation campaign with similar aims that tried to plant the idea that the United States conducted a biological weapons research project at Fort Detrick, Maryland.[79]

Social media platforms' willful blindness facilitates our adversaries' misinformation campaigns. Senator Al Franken memorably challenged Facebook founder and CEO Mark Zuckerberg over accepting political ad payments denominated in rubles. Even now, Facebook won't look behind the nominal ad buyer to assure that American shell corporations are not being used as screens for foreign influence. An October 2020 analysis found that many Facebook ads did not even disclose their source, despite Facebook's own nominal policies to the contrary.[80]

Looking Toward Model Solutions

Things don't need to be this way. Indeed, social media companies have taken some steps to combat illicit content and disinformation and misinformation. Platforms have worked to identify child sex abuse material; banned some communities used to promote hate speech and radicalization;[81] worked with third-party fact-checkers to identify, review, and label some disinformation;[82] deleted some White supremacist content and redirected White supremacist–related search terms to Life After Hate;[83] and fought back against some misinformation by posting descriptions from Wikipedia and *Encyclopedia Britannica* next to videos on topics that encourage conspiracy theories.[84] These efforts are steps in the right direction that Congress should incentivize in its replacement for Section 230.

Disinformation may be one area where we need to go beyond restoring traditional legal definitions of liability. Deliberate schemes of general disinformation are different from a deliberate lie about an individual. Platforms have been particularly vulnerable to general disinformation campaigns. Usually, the worst campaigns have something or someone behind them manipulating the disinformation. In these cases, transparency as a remedy has the particular civic value that motives and conflicts of interests can be disclosed and discussed. Citizens deserve that information if they are to be informed citizens.

Replacement, Not Just Repeal, of Section 230

Despite Section 230's present dangers, repealing it without a replacement leaves potential gaps in our defense against the spread of illegal content or misinformation.

In the absence of new regulatory guidance, a period of uncertainty could ensue. While existing federal and state laws would encourage platforms to police themselves, a coherent body of law may take time to develop in courts, leaving much uncertainty in the meantime. While legal liability may ultimately prod companies to take down some illegal content, it would not adequately address deliberate patterns of misinformation that don't have a harmed target with standing to sue.

Opening social media platforms to liability could also cause them to overcorrect until the law is settled.[85] Powerful interests able to finance litigation could carve favored paths for themselves. Small social media companies would face barriers that could make it harder for them to compete against big incumbents.

Eliminating Section 230 May Not Solve the Spread of Misinformation

Libel and defamation suits can deter and stop the spread of false or misleading information that is legally actionable.[86] When the voting machine company Dominion sued several news organizations for defamation, the wave of lies about the company slowed to a trickle, and news anchors started reading disclaimers that Dominion's machines had been repeatedly found to be safe and secure.[87] Similarly, in *Competitive Enterprise Institute v. Mann*,[88] a climate scientist sued for defamation bloggers who criticized his work and questioned the validity of investigations that cleared him of scientific misconduct.

These cases parallel successful litigation brought by Sandy Hook parents and the parents of a murdered Democratic National Committee employee for lies propagated in traditional right-wing media about their murdered children. Other forms of online speech could violate state laws against harassment, fraud, election interference, cyberbullying, cyberstalking, or revenge porn.[89]

Things get more difficult when the effect of the scheme is to pollute the general information environment with lies, rather than to harass or slander an identified individual or company. How do we address a pattern of lies about "Italygate" or vaccine efficacy? Who would be a proper plaintiff? How would you establish causation between the misinformation and the harm?

Congress can't readily solve these problems by creating new causes of action. Causes of action based on the content of speech—for example, a new cause of action for knowingly publishing misinformation online—will be subject to strict scrutiny in court. Many statutes seeking to criminalize cyberbullying or other online speech have been struck down on vagueness grounds.[90] Transparency rather than liability may be the best policy.

Unlimited Liability Could Privilege Wealthy Special Interests

We have an unbalanced information ecosystem populated by predators who conduct persistent, stealthy campaigns of special-interest propaganda.[91] The fossil fuel industry's long campaign of climate denial through fake science and phony front groups is a particularly poisonous example. If Section 230 is

repealed without additional guidelines, an already unbalanced information ecosystem could be unbalanced further as platforms yield to legal pressure from big, deliberate manipulators of information. Powerful special interests can bring lawsuits they are unlikely to win in order to scare off social media companies in terms of how they police certain content.[92]

Users do not have a legal right to disseminate speech—even lawful speech—via a social media platform. The First Amendment "has plenty to say about government regulation of the internet. . . . It has nothing at all to say about any supposed constitutional right to use a private platform. . . . The First Amendment . . . addresses only government action, not the action of private property owners."[93] Thus, under current law and the platforms' terms of service, users would have no First Amendment cause of action against a platform for removing content.

The ultimate success of a lawsuit, however, may not matter to well-funded interests with the means to threaten nuisance suits, and there are other doctrines of tort law that could be used to frame a dispute or a threatened dispute.[94] Consider former president Trump's July 2021 announcement that he would sue Facebook, Twitter, and Google for removing him from their platforms following the January 6 insurrection.[95] Within minutes, the Republican National Committee used Trump's lawsuit as a fundraising opportunity and sent out emails and text messages soliciting donations.[96] Legal experts roundly criticized the suit as baseless.[97] Nonetheless, the defendant companies must bear the costs of defending themselves against the suit, and face the reputational damage such a claim could inflict in the eyes of Trump's numerous supporters.

Repealing Section 230 Could Benefit the Biggest Companies

Trillion-dollar social media companies could be beneficiaries as well as victims of nuisance litigation. With an abundance of resources at their disposal, Google and Facebook can easily afford to litigate. This gives them an incumbency advantage: New social media startups cannot afford to spend millions of dollars on litigation. Startups also can't afford to spend millions of dollars developing automoderation mechanisms. Section 230 reform needs to recognize the power discrepancies among the contestants in this market, and avoid turning repeal of Section 230 into an incumbency protection scheme.

In summary: This is complicated, and sorting responsibly through these questions can spare us from dangerous pitfalls. Even though it's complicated, it's necessary, as the worst-case scenario is to leave Section 230 undisturbed. Congress would do well to listen to the various interests and concerns and offer clear guidelines to platforms and users.

A Framework for Section 230 Reform

Congress must replace Section 230 to account for the colossal changes to the internet that have transpired since the law's enactment. A replacement should ensure greater oversight of the policies of major social media platforms that most directly contribute to the spread of hate speech, disinformation, radicalization, and white supremacism. It should hold sites accountable for their own actions, advertisements, and algorithms. It should ensure transparency where necessary to expose the actors behind schemes of deliberate disinformation. It should also align incentives for social media companies with the interests of the public, with companies earning liability protections by reducing the illicit content and misinformation on their platforms.

Democrats and Republicans agree Section 230 needs serious reform, but approach the issue from different angles. Republicans accuse the platforms of anticonservative bias,[98] while Democrats accuse the platforms of propagating disinformation, hate speech, and right-wing extremism.[99] Despite the vastly different conclusions of the two parties, there is a path forward that should earn support from both sides.

Creating a Notice-and-Takedown System for Illicit Content

As discussed above, social media platforms are not liable even when they fail to take down illegal content they're aware of. The best solution would be for Congress to require a "notice-and-takedown" system removing Section 230 protections when a company willfully refuses to remove unlawful content. As part of this system, major social media platforms should maintain an "acceptable use" policy, explain how the platform enforces its content moderation policies, and describe the methods for reporting content or speech that violates these policies or other laws.[100] They should notify users when their content is taken down, and give users a forum for appeal if they think they've been wrongly removed or if the company has failed to act. (When users who have been removed don't come forward, it can be a pretty good sign that you've hit on fraudulent actors who don't want to show their faces.) All of these steps can be—indeed, as a practical matter, must be—systematized. Congress's focus should be not on individual cases, but on adequate systems.

We would be in good company in proposing a notice-and-takedown system for illegal conduct on social media platforms. In December 2020, the European Commission proposed the Digital Services Act, which, among other things, would require platforms, after adequate notice of illegal material, to remove that material to retain the act's protections.[101] The act would also require platforms to

establish redress mechanisms for removed user content.[102] Requiring a notice-and-takedown system would also bring social media regulation more in line with legal theory. Notice of risk is a common ingredient of tort liability.

Imposing Stronger Transparency Requirements

Social media companies can make a lot of money from inadequate transparency. In order to stop the spread of disinformation, we need more insight into the role that bad actors and the platforms' own algorithms play in propagating it.

In our modern disinformation environment, platforms must help identify deliberate disinformation campaigns and flag systematized disinformation. Much of this disinformation is propagated by domestic interests with economic incentives to seek out avenues of malign influence, the capability to deploy large-scale influence campaigns, and the resources to bring legal, economic, and political pressure to bear on the platforms.[103] If the platforms don't have systems and safeguards that help them resist this pressure, their incentive to yield to the disinformation campaigns' sponsors will be powerful.

In order to limit the use of paid content to spread disinformation, we ought to require platforms to disclose what is paid content and who is behind it, and to separate paid content from other content. Users should also be required by platforms' terms of service to disclose when they are paid to disseminate particular messages. In some circumstances, where dissenting voices are legitimate and danger is real, anonymity can provide a valuable protection. But usually, anonymity degrades public debate and discourages active citizenship by allowing special interests to pollute public discourse from behind false fronts.[104] Facebook might not have accepted payment in Russian rubles for US election political advertising if it had known that its receipt of rubles for those ads would be disclosed to the public.[105] Disclosure ensures that people know who pays to influence their views and enables them as citizens to judge that sponsored content accordingly.

Platforms should stop allowing shell companies or other anonymizing intermediaries to buy political advertisements.[106] No one is better placed than the platform to ensure that the voices they propagate are real voices, not bots or foreign intelligence operations or corporate disinformation campaigns. We know how to press through misleading shells to determine the real "beneficial owner" of an entity,[107] and we know how to craft policies to pierce corporate veils.[108] Citizens are owed that information in a democracy. Indeed, they are disabled in their citizenship responsibilities without it.

Finally, social media companies should open their algorithms up to scrutiny. In cybersecurity, firms deploy "white hat" researchers to probe their systems for glitches and weaknesses. Mandating some form of access to social media platforms' algorithms would allow researchers to determine whether and

how those algorithms contribute to illegal activity, and also provide a way to counteract any illegal actions found. Giving this information to outside "white hat" researchers would oblige platforms to be more serious about identifying and counteracting problems in their algorithms without government directly probing into their affairs. The law countenances "private attorneys general"; one could adapt "private inspectors general" to help keep social media platforms alert and honest.

Reaffirming That Platforms Are Immunized for Others' Content—Not Their Own Algorithms and Practices

Congress should make clear that social media companies are only immunized for user-generated content, not for their own profit-making decisions. Courts already assign liability to platforms that cross the line from hosting another's content to co-creating it. For instance, some courts have held that a service provider may be subject to liability if it "materially contribute[d] to the illegality" of the disputed content.[109] Others have said that service providers may be liable if they "specifically encourage[d] development of what is offensive about the content."[110]

Platforms should be liable for their own decision-making, for harms resulting from how they use the data they collect on their users, and for the algorithms they rely on to boost and amplify content. If an algorithm itself violates federal civil rights laws—for example, by only serving certain kinds of employment opportunities to certain users[111]—platforms should be held responsible. This would require platforms to internalize the costs imposed by their algorithms, instead of solely reaping the profits. The threat of legal liability, for example, could make Facebook and other companies more likely to adopt measures that stop the spread of misinformation even if they also reduce user engagement, as failing to act would carry its own financial risks.

Platforms should also be subject to liability for paid advertisements and sponsored content. When a platform is paid, the notion that it is just providing a public forum vanishes, and the same laws should apply to the platform that govern paid advertising in other media. Since this would bring social media liability in line with legacy media, it would likely survive legal challenges.[112]

Conclusion

Social media platforms generate immense value, enjoyment, and convenience. They have given rise to new industries, and transformed existing ones. They have forever changed our social lives. We must recognize their benefits, as well as the clear fact that social media are now woven into the fabric of society.

But we cannot ignore social media's many dangers, primary among them deliberate disinformation on a massive scale, whose targeting of consumers the platforms facilitate.

In reforming Section 230 to hold major social media platforms accountable, we can more closely align Section 230 to its original purpose, more closely align this area with long-standing principles of law, and ensure that platforms are duly diligent to stop the spread of fraud, white supremacism, climate denial, disinformation, and illegal conduct online. In the process, we will create a healthier and more accessible internet for all, and improve the functioning of our democracy.

It is this latter point that is most important. Our experiment in American democracy is facing grave challenges. Lies and disinformation designed to whip up public animus and dissatisfaction are one grave challenge. Schemes of disinformation designed to mislead the public to advance corporate and political interests are another. Both are hazards to the experiment in governance we treasure. "Truth will ultimately prevail where there is pains to bring it to light," said George Washington; "An educated citizenry is a vital requisite for our survival as a free people," said Thomas Jefferson. In its present state, and thanks to Section 230, the social media arena is corrosive to our democracy, and that must be—and can be—repaired.

PART THREE

CONTENT MODERATION AND THE PROBLEM OF ALGORITHMS

8

Algorithms, Affordances, and Agency

RENÉE DIRESTA

In 2003 and 2021, two momentous events occurred in US politics, in very different information environments. In 2003, Americans watched on their television screens, listened to talk radio, and read the morning paper to keep up with the latest news about the invasion of Iraq. In 2021, they watched, in real time, the live-streamed breach of the US Capitol by an angry mob as it unfolded; many amplified, on social media, specific posts about the moment that most resonated with their politics.

Although these events were separated by fewer than two decades, and a mere two and a half presidencies, the ways the American public contributed to, talked about, and understood them were wildly different. The key drivers behind those differences have had a tremendous bearing on society: An unprecedented agency to shape the media environment, available to anyone with a keyboard, internet connection, and grievance. New affordances: the tools and features of centralized social media platforms that enable individuals to livestream, like, retweet, and share. And influential algorithms, which amplify, curate, and proactively recommend content with speed and precision. Put more simply: The way people encounter, engage with, and amplify speech has changed.

In 2003, the rationale for the Iraq War was developed in a decidedly top-down fashion. Following 9/11, the Bush administration outlined reasons[1] to invade Iraq and depose its president, Saddam Hussein: weapons of mass destruction (WMD); the aiding and abetting of Al Qaeda, which was responsible for the 2001 September 11 attacks; and the necessity of bringing democracy to a brutal authoritarian state. With these justifications, the executive branch—and also the legislative, via Congress's Iraq Resolution[2]—communicated to the public the necessity of invading Iraq.

Some were more cynical about the motives for the impending war, of course. Left-leaning media suggested a different set of justifications: a president's son, now the leader of the free world himself, finishing his father's unfinished

business; control of oil; and purported divine inspiration (which was roundly mocked). But for the most part, the Bush administration and Congress found an ally in one of the most powerful institutions of the United States: the media. In the months leading up to the war, the nation's top newspapers and television networks repeated and endorsed the administration's narratives for the American people. Fox News perennially displayed American flags onscreen during their war coverage. Even many left-of-center reporters and publications—like Judith Miller at the *New York Times*[3] and David Remnick at the *New Yorker*[4]—loudly repeated claims about WMDs and Saddam's brutality. The few mainstream journalists who rejected the justifications found themselves unpopular: Iraq War critic and MSNBC host Phil Donahue saw his show canceled, for example.[5] In a study by Fairness & Accuracy in Reporting (FAIR), researchers revealed that critics of the war were vastly misrepresented on network television broadcasts.[6] And so, when the war officially began in March 2003, the public consensus was clear: Seventy-two percent felt that America's invasion of Iraq was justified.[7]

Throughout all this, ordinary Americans had few affordances with which to meaningfully engage with the dominant narrative. They could attend a protest. They could change the channel, cancel a subscription, or write a letter to the editor—though those options provided minimal agency. By and large, Americans were largely passive consumers of a top-down information machine.

Eighteen years later, on January 6, 2021, the same halls where the Iraq War blueprints were crafted were overrun with rioters.[8] The mob had come to hold the (mistaken) belief that the 2020 election had been stolen and that Donald J. Trump remained the rightful president of the United States, and they were there to protest the ratification of election results showing that Joe Biden had won. But this time, the narratives that preceded the infamous riot, that shaped a false reality that many came to believe, were *not* solely coming from the top down. That old dynamic was present: The president's acolyte media properties, such as Newsmax and OANN, filled their segments with election fraud claims. But such claims were also, in fact, frequently bottom-up, formed on social media by way of a process that played out repeatedly for months leading up to election day, in which unsubstantiated allegations of fraud were made to go viral by way of influencers and online crowds, which created and spread content within online echo chambers full of people likely to believe them.[9] Although populist politicians and shifty polemicists on Fox News certainly participated in shaping the overall narrative that there had been a "steal," ordinary Americans were no longer passive recipients of a top-down narrative—they themselves played a very significant role in facilitating that consensus and parlaying it into in-the-street action.

In the days following the 2020 election, a collection of communities ranging from conspiracy theorists and trolls, to militia groups, to partisan ideologues

and the otherwise disaffected gathered in Facebook groups and Parler threads to seethe and plan protests—and, for some, to plot violence and a breach of the Capitol.[10] Social media are remarkably effective at bringing the like-minded together. These angry and disappointed communities reinforced members' beliefs that ballots were being destroyed, that voters were being intimidated, and that results had been distorted. These communities of ordinary people, alongside the "blue checks" with millions of followers who served as both content creators and key amplifiers themselves, facilitated a parallel system of mass influence on a par with the propagandists of old.[11] While the debate about whether media or social media are more impactful continues,[12] the seeming impact of pervasive misinformation is startling: A few days after the election, seven in ten Republicans believed it had been stolen.[13] Three months after the riots of January 6, 55 percent of Republicans believed that the rioters themselves had been left-wing agitators[14]—not Trump supporters at all—despite an overwhelming abundance of evidence to the contrary.[15]

Today, information moves by way of a system of algorithms and affordances and a historically unprecedented degree of agency, wholly unavailable to those Americans of 2003. Ordinary people use the tools of social media to share memes and retweet calls to action—like #StopTheSteal—in a high-velocity, high-virality ecosystem with the potential to reach millions in a moment. These tools are situated alongside a more opaque force in communications technology: the algorithms that suggest who we follow and determine what is recommended or amplified. These automated curators have facilitated the proliferation of echo chambers, bespoke realities within which highly activist online factions coordinate. These are mutually shaping complex systems; MIT researcher Sinan Aral, for example, describes the relationship between humans and algorithms as a "hype loop" in which users provide signals to machines via their actions and content consumption, the algorithms make sense of it and reflect it back in their suggestions, and then the humans react to the new provocations.[16] There is not one sole driving force at fault, but rather the *interplay* between algorithms, affordances, and agency that has transformed speech and community.

In this essay, I explore this interplay, and the resultant new infrastructure of human connection, contextualizing it within past communication ecosystems, examining its unintended consequences, and offering suggestions for a way forward.

The Infrastructure of Influence

Throughout history, punctuated advances in communications technology have had profound impacts on how information reaches the public and how

communities come to consensus. The printing press, the radio, the television—each of these technological advances changed how we communicate, and precipitated shifts in social behavior, politics, and power dynamics. The invention of the printing press ushered in the era of mass media, battering the monopoly that the Church and wealthy nobles had on information. The era of television made it possible for the public to see what was actually happening in far-off places and to contrast what was on their screen with how newspaper journalism and politicians' commentary spun certain stories. The most recent transformation—the introduction of social media platforms, which followed soon after mainstream adoption of the internet—rearranged who had access to the tools of content creation and dissemination, and by extension, who had the ability not only to speak, but to reach millions. Even among these punctuated technological leaps, the impact of social media platforms stands out because of the extent to which it turned the audience from passive recipient to active participant, upending the power dynamics of mass media. Social networks remade not only the way information moves through society, but the way society itself is organized—the topology of human connection—and in doing so disrupted long-established mechanisms for coming to consensus.[17]

For centuries, information was relayed from the top down: Elite institutions and media communicated to the public through a particular lens, framing the narrative, shaping what society thought and felt. People were largely passive recipients of information. Affordances were few; computationally supported algorithms were, of course, nonexistent. This molding of public opinion through the strategic presentation of information—propaganda—was, in the early to mid-1900s, viewed as a necessary component of democratic society. The purpose of the elite-manufactured narratives, those who produced them believed, was to help the public understand issues about which they had no firsthand expertise.

The merit of this top-down power structure, particularly within a democracy, has long been scrutinized and critiqued by prominent communication and media theorists such as Walter Lippman, who coined the phrase "the manufacture of consent" to describe the process of elites and media shaping public opinion in 1922,[18] to Edward Bernays, who argued that it was in fact the responsibility of democratic government to influence the public in this way because the complexities of modern policy meant that ordinary people were rarely truly informed citizens.[19] Later, that position came to be seen as decidedly antidemocratic; the work of Marshall McLuhan,[20] Jacques Ellul,[21] Daniel Boorstin,[22] and Noam Chomsky[23] chronicled the extent to which the incentives of media and government often led to messaging that was, in fact, decidedly manipulative. The top-down model was the communication dynamic that shaped the disastrously incorrect 2003 consensus around the Iraq War.

The specific elite institutions driving the manufacture of consent changed over the centuries: the Catholic Church, European empires, and the US government. The communication technology available for messaging has also varied: newspapers, radio broadcasts, and network television. But the fundamental principle—entrenched elites crystallizing the opinions of the many—remained largely the same until the internet arrived.

This shift began in the late 1990s with the emergence of the World Wide Web. In its first iteration, the web offered an affordance not previously available to the public: zero-cost publishing. This Web 1.0 colloquialism referred to the free, readily available turnkey sites that anyone could sign up for. Any individual with an internet connection, computer, and basic grasp of HTML could create and disseminate speech, with minimal financial or temporal hurdles. Soon after, blogging engines appeared, and blogs quickly became the primary mode of publishing one's thoughts online. However, although this new infrastructure circumvented some of the old gates and gatekeepers—like access to a printing press or the favor of an editor—the decentralized environment made it hard to reach large quantities of people or amass persistent audiences. Anyone could write a blog—but how many people would find it and read it? The overwhelming majority of online speech, whether offensive or anodyne, had virtually no readership.[24] Many aspiring citizen journalists and writers, originally optimistic about the web, were discouraged by the challenge of distribution.[25] "Why do blogs have a higher failure rate than restaurants?" asked the *New York Times* in 2009, noting that 95 percent of bloggers gave up.

But over the next decade, people gradually coalesced onto a small handful of centralized "social networks" that offered them easy ways to share their thoughts with friends: Facebook, Twitter, Reddit. Compared to blogs, even in the early days the social platforms rewarded brevity. Communication norms shifted as platform interface design nudged users to create, and behave, in particular ways. Text-heavy blog posts were usurped by far more economic modes of speech: a gif, a meme, a 140-character bon mot. These new norms emerged as a result of the affordances offered to users and the incentives set by platforms—which encouraged content that would lend itself to virality and capture the attention of other users.

Virality was encouraged through other affordances as well. Perhaps most critically, the Share button and the Retweet feature were material steps toward solving the distribution challenges of the old decentralized blogosphere; ordinary people became arbiters of what should spread, and had the power to amplify content that they liked or agreed with. Suddenly, everyone had access to tools for creation *and* distribution—affordances that enabled not only speech but the potential for *reach*.

As users flocked to social media platforms en masse, writing their status updates and posting photos of their lunch, the volume of content exploded, creating something of a glut.[26] There was simply no way for people to read every tweet, peruse every comment, and click every clickbait headline that filled their feeds. The web's original curation mechanism, reverse-chronological order, became woefully inadequate—and platforms realized they needed to not just host speech, but curate it, at a scale far beyond what human editors had ever attempted. Drawing on human-generated signals about what speech was most engaging—What were people clicking on? Hovering over? Scrolling through?—platforms debuted AI-powered feed rankings and recommendations. Now powerful algorithms existed alongside the affordances. Both are, of course, simply code—but in the colloquial conversation about social media, the somewhat imprecise phrase "the algorithms" is primarily used to reference functions that act *on* the user without the user's taking a direct action in that moment. In the case of an affordance—a Like or Share button, a livestreaming feature—the user is proactively and consciously choosing to click a button and take an action; in the case of "the algorithms," such as curated search results, topics sorted by a trending algorithm, or groups suggested by a recommendation engine, the code is shaping the user's experience, determining what will hit their field of view, without any direct user input in that moment.

Unlike their human counterparts—network newscasters, newspaper editors, librarians—who prioritize giving their audiences factually accurate information from reputable sources, the algorithmic curators and recommenders generally prioritize one key metric: engagement. As users click on, share, or "like" a piece of content, they signal to the algorithms that something is worth paying attention to. If many people click en masse, or a cluster of a particular type of content suddenly attracts attention from many users, that is a signal that platforms take into account for another type of algorithmic action: amplification, further boosting of the topic or piece of content to even more users. And in the early days of the emergence of these curation and amplification algorithms, trending algorithms on Facebook and Twitter automatically curated the top stories of the day, pushing popular hashtags and URLs into the feeds of millions of users.

Algorithmic curation and amplification didn't just spotlight certain topics and stories—it also bred whole new communities *around* them. Recommendation engines steered people into online groups with niche focuses, independent of geography: a mutual love of corgis, a mutual loathing of Democrats (or Republicans), a mutual skepticism of vaccines. While this was, in most cases, a positive means of facilitating human connection, the recommendation engines were wholly unaware of what, exactly, they were recommending. Sometimes they inadvertently pushed people into conspiracy-theory-focused communities deeply distrustful of outsiders, and highly activist—which, to

the recommendation engine, appeared to simply be high engagement, and thus worth promoting.

By the early 2010s, the ecosystem of social media platforms had enabled unprecedented levels of virality, velocity, and scale. The social space and information space merged; increasingly large percentages of people began getting their news from social media.[27] A critical mass of people, effortless creation and distribution of content, mechanisms to facilitate virality, and the phenomenal *reach* of information—by the 2012 upheavals of the Arab Spring, it had become clear that social networks were a communication infrastructure with great potential for amassing and disrupting power.

As this new ecosystem emerged, the old began to erode. Mainstream and broadcast media did work to leverage social media as an additional channel, setting up Facebook pages and Twitter accounts and growing large audiences. But a new class of communicators had emerged: A growing number of influencers and a hyperpartisan "demi-media" flourished, enabled by affordances that let anyone become a broadcaster. These new entrants targeted their content toward the distinct communities, factions, and bespoke realities that social media had engendered. They often aggressively competed with—and, among some audiences, muscled out—the old media guard by challenging their truthfulness and legitimacy, and developed strong relationships with the fan communities that amplified them.

By the end of the 2010s, it had become clear that the public had increasingly become active participants in the process of competing to shape narratives, to amplify their preferred version of reality. Today, a handful of antivaccine zealots can make a video alleging that a global pandemic is a deliberate "plandemic"[28] and get millions of views on Facebook and YouTube. A community of conspiracy theorists can make wild allegations—that a major furniture company is a front for child trafficking, for example—go viral on Twitter.[29] From an office building in St. Petersburg, Russia, trolls can masquerade as both Texas secessionists and proud American Muslims, triggering in-the-streets conflict between the two groups in downtown Houston.[30] And, perhaps most important, ordinary people with strong political convictions become digital warriors speaking out (and coordinating) to advance their beliefs at scale.[31]

Narratives now emerge from the bottom up as well as from the top down; propaganda is increasingly democratized. Many thousands of people chose to become digital warriors in the presidential elections of 2016 and 2020, equipped with affordances and bolstered by algorithms. A successful crop of political candidates across the world wholly embraced the new dynamic: Populist leaders such as Rodrigo Duterte, Donald Trump, Narendra Modi, and Jair Bolsonaro harnessed the energy of these online factions. Some niche communities, such as QAnon, evolved into online cults;[32] recommender systems kept promoting and

amplifying these even as researchers and watchdog organizations grew more vocally concerned about their impact.

The expansion of the notion of, and participation in, "media" is in many ways a positive development; worldwide, diverse communities and powerful movements have formed online, new voices have flourished absent the traditional gatekeepers, and groups that have historically been silenced have increased their visibility. The old system of facilitating consensus by way of top-down narrative control was demonstrably flawed. And yet, today we find ourselves facing a distinctly different challenge: Although what society thinks is no longer neatly manufactured by the top-down propaganda of elites, it is increasingly shaped by whoever manages to wield the affordances of social networks most adeptly to solidify online factions and command public attention.

What we are contending with now is not the manufacture of consent but the *manipulation of consensus*: participatory propaganda[33] campaigns that often originate, bottom-up, as a meme, passed to influencers and media who reshape it for their own ends—whether that be birthing or toppling political movements, sparking mass protests, upending regimes, or something else entirely. Communities do not trust the same sources. They do not see the same content. They get agitated about specific grievances prevalent within their online reality but almost entirely invisible to anyone operating outside it. The result is perpetual dissensus and acrimony.

As this dynamic has continued, as the participants have grown in number and become better organized, the implications have become more profound: in 2020, a global pandemic began in which twenty-five percent of the American public believes, to some extent, that the outbreak was planned;[34] a vaccination campaign to address the pandemic in which the likelihood of taking a vaccine is based on the extent to which communities trust particular politicians;[35] and a free and fair election that nonetheless resulted in the storming of the US Capitol by people who not only heard *but proactively created and spread* allegations of mass voter fraud. And yet, this divisive environment, this inability to find shared reality, to come to consensus, is driven by very real people exercising their right to express their political opinions. That makes responding to this challenge—moderating, recommending, and curating responsibly; balancing the freedom of expression with minimization of harmful content—a far more significant challenge.

We presently have an underdeveloped set of norms delineating persuasion from manipulation, and a world in which the group most adept at executing social media marketing tactics determines what captures the public's attention. Foundational processes of deliberation, of coming to consensus in democracies, no longer hold;[36] the "marketplace of ideas" does not span the bespoke realities. And so, we are at a moment in time in which it is imperative for society to wrestle

with the implications of this new infrastructure. Can democracy function in this environment? Is it possible for humanity to reach consensus on existential issues such as the climate crisis or future pandemics?

A Path Forward

The status quo is untenable. Americans, immersed in this fragmented and polarized information environment, are increasingly unable to agree on critical facts. While the 2020 election itself was free and fair, around 30 percent of the public remained unconvinced that Joe Biden is the rightful president in May 2021.[37] Beyond electoral politics, the most pressing global issues—vaccinating the population during a pandemic, addressing the challenges of a changing climate—become unworkable if consensus has broken down. Historians, election officials, and legal scholars alike wonder aloud about the extent to which American democracy can function when the losing side distrusts the outcome of elections or when large communities of people are immersed in bespoke realities with no shared epistemology between them.[38]

The question of who bears responsibility for fixing this state of affairs remains a matter of debate. For years, platforms were loath to reckon with these challenges because they did not want to be the arbiters of truth or determinants of the line delineating social and antisocial behavior. That began to change in 2017, when, in response to public pressure surrounding concerns about "fake news" (a term that briefly referred to demonstrably false and misleading news, before it was appropriated by partisans), platforms began to incorporate fact-checks on disputed content. This was necessarily a reactive process; fact-checking links were often appended days after the content had gone viral and audiences had long ceased paying attention. In 2019, the platforms additionally began to remove certain types of demonstrably harmful communities—such as antivaccine groups and conspiracy theory communities prone to real-world violence—from curation algorithms that had previously suggested them. In 2020, cult-like communities prone to violence, such as QAnon, were declared persona non grata on several social platforms. These policy iterations, however, evolved reactively, often in response to situations like major measles outbreaks or the January 6, 2021, events at the Capitol. However, a series of leaks of documents from within tech platforms suggest that they could have been dealt with far more proactively: A Facebook internal presentation leaked to the *Wall Street Journal* in 2020, for example, revealed that senior leadership knew in 2018 that recommendation algorithms were having unintended, and at times dangerous, consequences.[39] The presentation warned, "Our algorithms exploit the human brain's attraction to divisiveness" and promote "more and more divisive

content in an effort to gain user attention and increase time on the platform." This type of empirical understanding of the dynamics in play is presently almost wholly lacking among outside researchers, government, and the public.

While platforms bear a significant degree of responsibility for mitigating the unintended consequences of their business models, incentive structures, and design decisions, social networks *reflect* social behaviors and societal issues even as they shape them. A loss of trust in media, and loss of confidence in government—a decline that began long prior to the advent of social media—coupled with the unpreparedness of institutions to participate in the new media ecosystem created vacuums, and an environment in which the speech of grifters flourished.

The current discussion of how to reduce polarization, and Americans' inability to come to consensus, tends to focus on social media. More specifically, it focuses on content moderation, which is inherently reactive, and centers the debate on specific types of speech to take down or deamplify. Sometimes it is pieces of content, other times it's the accounts posting the material. But this approach has led to an unending game of highly politicized whack-a-mole, and it does nothing to address the problem of receptivity. We can reduce the supply of harmful content, but doing so does not mitigate the demand.

In a mutually shaping system, we need a systemic approach to solutions. There is a path forward toward a healthier information ecosystem that balances freedom of expression and democratic ideals while minimizing harms and unintended consequences—but it involves re-envisioning the roles that algorithms, affordances, and agency play in our information landscape. There are levers at our disposal for this reformation: policy, education, and design. There are also multiple stakeholders with the power to drive and implement reforms: the tech platforms, governments and regulators, civil society, media, and even the empowered public. In the next section I discuss their respective roles and envision a whole-of-society effort for transforming the communication ecosystem.

Addressing the Algorithms

Algorithms, perhaps, are the easiest of the three factors to transform. By prioritizing engagement when choosing what to surface, what to rank, and what to draw attention to, they inadvertently incentivize sensationalism and acrimony; content creators using platform tools can attract attention by perpetuating emotionally engaging grievances that online factions subsequently act upon. Journalists, researchers, and the public have called attention to the unintended consequences of misguided curation—the impact of proactively nudging people into extreme groups or filling feeds with demonstrably false content. However, because of limitations of platform data sharing (sometimes attributed

to concerns about privacy), outside investigators have incomplete pictures of what is happening; only the social media companies have full visibility into what's happening on their platforms. Here, therefore, are five suggestions for algorithmic reform, implementable by social media companies:

1. Reduce dependency on engagement when devising algorithms to curate and recommend content.
2. Add friction, such as circuit breakers and internal-oversight models, to incorporate humans into the loop of curation and recommendations.
3. Provide researchers with data access that affords visibility into the dynamics of algorithmic creation, recommendation, and amplification.
4. Create a system of regulatory oversight.
5. Offer the public greater visibility into—and perhaps control over—curation and recommendation experiences.

Although many readers no doubt feel distrustful of social media platforms—polls suggest a dramatic decline in positive sentiment since 2017—when we consider what entities should be responsible for mitigating the myriad issues with "the algorithms," the platforms themselves are the only stakeholder with the direct ability to address emerging issues. They can shift the system through design. They've done it in the past: When Twitter observed that trending algorithms were being gamed by automated accounts, they chose to reweight the factors that determined that something was trending. They created a designation for "low-quality accounts"—many of which were automated—to reduce the ease with which someone with a botnet could create the perception that thousands of people were tweeting about a particular topic.

Another type of design intervention for mitigating algorithmic harms might be the introduction of friction[40]—a limit or constraint placed on behavior or content to reduce its use or spread. Friction in speech has historically enabled opportunities to verify information, to validate facts. In prior information environments, traditional journalists vetted sources, conducted fact-checks, and reviewed quotes before publishing their story. Algorithms do none of this. But that could change: If, for example, platform algorithms detect significant engagement on a new piece of content, rather than treating that as a signal that it should be additionally amplified into the feeds of millions of other users, they might instead temporarily throttle it while someone investigates the dynamics of the spread to ensure that it's not coordinated manipulation. This kind of "circuit breaker"[41] could be narrowly tailored to start, perhaps to temporarily throttle the dissemination of viral content on breaking news topics (particularly those in highly contested areas such as politics), or if the content appears to be manipulated (such as sensational videos that might be edited or AI-generated).

There is precedent from other industries: In finance, "trading curb" mechanisms sometimes kick in to temporarily halt trading if a stock is behaving erratically or if news breaks about the company.[42] That same principle can be applied to news feeds and recommendation engines. A viral hashtag about alleged election interference, or supposed vaccine side effects, could be decelerated until properly fact-checked.

The platforms currently use a policy rubric known as "remove, reduce, inform" to moderate their content—"removing" involves taking content down, the most severe intervention and the one most likely to trigger allegations of platform censorship. "Reducing" distribution—throttling content—and "informing" interventions, such as adding labels and interstitials to convey context around content, are less divisive and possibly even more impactful. Adding friction would enable platforms to temporarily reduce the spread of viral content, facilitating an increased ability to help inform the public. They also could proactively facilitate more trustworthy content at key times, such as with banners or public service announcements from election officials on Election Day.

But platforms' self-regulation of algorithmic design and curation policies won't be enough. Many of the design changes and policy shifts that platforms have implemented have come about because researchers or media have uncovered a significant harm. Public outcry has driven certain interventions, but regulation is needed to create real oversight.

Presently, there isn't a strong regulatory foundation to build on. That's not to say that US lawmakers haven't tried—they've just been ineffective and often too slow. Consider the Twitter bot manipulation situation discussed above; first a novelty on social media platforms, then a scourge that excelled at manipulating trending topics, bots by 2018 had largely become toothless in the information war because Twitter designated them as "low-quality accounts." Despite the platform's already having taken action to render bots largely useless for mass manipulation, California policymakers still enacted a law in 2019 to make it illegal for bot accounts to masquerade as humans. By the time lawmakers got their arms around the problem and devised even a weak solution, information combatants had moved away from automation in response to Twitter's design changes, and had instead begun to focus on infiltrating and galvanizing communities of real people—a harder problem for Twitter to take action on with simple weighting tweaks.

The toothless bot law and the timescale on which it was passed are emblematic of a broader problem within the policy space. Laws that attempt to go after problematic algorithms are the equivalent of building a digital Maginot Line—quickly rendered useless by determined adversaries who simply go around them. By the time lawmakers pass legislation to address one problematic algorithm or affordance, combatants will have moved on to the next. What we need from

government regulation of the tech industry is the establishment of comprehensive oversight of social media companies, a multitiered system of regulation similar to that which governs the financial industry: a combination of regulatory and self-regulatory policy, in which top-level government regulatory bodies can facilitate oversight while self-regulatory bodies can respond quickly to industry-wide dynamics. The Election Integrity Partnership,[43] an inter-institution partnership that spent several months monitoring voting-related misinformation during the 2020 election, offered several suggestions for regulatory oversight: first, the passage of laws for algorithmic transparency, which would mandate that platforms share reports and granular data with academia or civil society, and second, at the federal level, interagency prioritization and cooperation in identifying and countering foreign mis- and disinformation. Senator Mark Warner (D-Va.) has also laid out proposals for regulatory oversight;[44] one key blocker, however, is that partisan polarization within Congress itself is stymying consensus around what, and how, to regulate.

Finally, there is a real need for educational programs to help users understand how the trends and recommendations that they see are selected for them, to facilitate better understanding of the dynamics of how engagement begets more engagement, or how accuracy of content is not presently a primary determinant when surfacing it. "Algorithmic literacy"[45] research is in its nascent stages, but scholars such as Joëlle Swart have begun to examine ways in which young people process the recommendations that are pushing content into their feeds on platforms such as YouTube, TikTok, and Instagram.[46] Opinion is split: Some users who understand algorithmic curation find it serendipitous and time-saving; others find it creepy, rooted as it is in companies tracking user behavior. Beyond simple research, some efforts, such as MIT's Gobo project, led by Ethan Zuckerman, envisioned putting control of curation far more directly into the hands of users by giving them their own filters and sliders, increasing their agency to determine what hits their feed.[47] While Gobo was an experiment, it offers a path forward toward potentially leveraging other third-party tools to more directly craft a user experience. As a caveat, however, while this may enable users to prioritize other criteria beyond engagement in curating their feeds, it is unclear whether or not it would result in users' simply curating themselves into filter bubbles.

Addressing Affordances

In the mutually shaping system, algorithms are still ultimately serving up and amplifying content, drawing on signal from users as they create and disseminate content. And so, we additionally need to pay attention to *affordances*, to how communication tools are used and misused.

Design has a significant role to play here as well, just as it does with the algorithm problem; platforms can also add an element of friction to the affordances that facilitate virality, or enable the amplification of speech.[48] These need not be onerous; a simple interstitial screen asking users if they've read beyond the headline when the Share button is clicked can prevent people from unwittingly propelling falsehoods (and has added downstream benefit on the algorithms, which derive signal from human actions). On platforms like WhatsApp—which has no algorithms yet remains prone to people being duped into sharing viral falsehoods—designers have limited the ability to forward messages to other groups, and limited group size, to reduce the number and reach of viral misinformation incidents.

There are many types of design interventions to try, and, as platforms evolve and new features are made available to the public, more potential interventions will emerge. A consortium of design researchers committed to drawing on behavioral science has emerged: the Prosocial Design Network, with the mission statement promoting "prosocial design: evidence-based design practices that bring out the best in human nature online."[49] The group's website highlights design intervention possibilities—the creation of "Thank You" buttons, delay-by-default posting—that might create a less polarized, more "prosocial" social media experience. As the interventions are examined in peer-reviewed literature, the Prosocial Design Network highlights the research.

Another potential way to mitigate the harms of affordances before they arise is to have internal or external "red teams" tasked with envisioning how a feature is likely to be misused *before* it is rolled out. This is how cybersecurity researchers discover, and how companies proactively mitigate, vulnerabilities in software. Thinking like a troll or foreign agitator in advance could surface challenges or unintended side effects of a particular feature before it is actively misused to cause harm.

Policy has some role to play in mitigating the worst harms of social media affordances as well: On a self-regulatory front, terms of service and other policies presently govern and at times limit affordances granted to users. For example, Twitter's features enable users to tweet, but its policies state that these features may not be misused to tweet *misleading content related to COVID-19*;[50] if users do so, platform policy notes that they may receive a "strike," may be asked to delete the tweets, or may have their tweets labeled as misleading. Labeled tweets, in some cases, may no longer be able to be retweeted, stopping the further dissemination of false and misleading information even if it does not rise to the level of meriting a takedown. Additionally, after five strikes, the platform may ban the account holder. Many platforms already have existing policies that prohibit things like amplifying disinformation and violent content. These rules matter. Policy shapes propagation—it defines how users can wield affordances for the distribution of speech.

It is also worth noting that much of this essay has focused on the existing social media ecosystem; while it is easy to fall into the trap of assuming that the big-tech platforms that exist today will exist tomorrow, the emergence of new prosocial-first platforms, designed from the ground up, may be the way forward. There may be significant hurdles to the mass adoption of such platforms, but adjacent regulatory efforts—the rising popularity of antitrust action, and data portability, among regulators in the Biden administration, for example—may create an opportunity for new entrants.

"We shape our tools, and then our tools shape us," as Marshall McLuhan's friend and fellow media scholar Father John Culkin once said, describing McLuhan's work. And that mutual shaping leads to the third area for intervention: *agency*.

Addressing User Agency

No matter how carefully platforms, researchers, and regulators proactively address social media algorithms and affordances, any tool has the potential to become a weapon for manipulating the discourse. And that is because people *choose* to use them that way.

While design can nudge, and policy can shape and constrain, the fact remains that people increasingly feel compelled to participate in hyperpartisan, highly factional online behaviors. During the 2016 presidential campaign, cadres of people who put frog emojis in their Twitter bios—an allusion to the popular alt-right meme Pepe the Frog—elected to become digital warriors supporting the candidacy of Donald Trump. By the time the 2020 presidential primary got under way, a litany of political factions with emojis-in-bio had proliferated.[51] It became clear to social media researchers looking at emerging false and misleading narratives that ordinary people had become increasingly active participants in the process of shaping narratives and amplifying their version of reality. While there have been online political discussion groups since the earliest days of the internet, these highly activist online factions were vocally united in their hatred of the other side.

Harassment is not new. Spreading disinformation is not new. Both predate the internet by centuries. The dynamics that are playing out online are a function of human behavior. Many people are drawn to the camaraderie of participation in an online crowd, to amplify their preferred candidates or policies. As polarization scholar Chris Bail writes, "the root source of political tribalism on social media lies deep inside ourselves."[52] Some people feel inclined to share because members of their political "team," or online factional friend group, are sharing. Here, too, scholars such as Jonathan Haidt, Tobias Rose-Stockwell, and others have suggested design interventions such as removing visible indicators

of engagement, such as likes and retweets, to reduce the inclination toward "follow-on" sharing (in which users simply retweet the tweets of their peer group without reading the material).[53]

Another important facet of agency is informedness. Some people are unaware of how their shares and behaviors influence algorithmic feeds and curation. They are unaware of how they themselves have been slotted into echo chambers with a narrow field of view. And so media literacy education, which attempts to help people understand the online environment and the way algorithms, affordances, and agency come together, offers a way forward. Awareness of the impact of one's actions has the potential to shift norms and to make people give agency they wield when creating or sharing content on social platforms greater weight. Education has the potential to break down the walls of bespoke realities and to rescue our ability to reach consensus by helping people recognize the signs of manipulative content or behavior online.

The News Literacy Project offers one example of a notable effort to improve online discourse by raising awareness of social media dynamics.[54] A nonpartisan nonprofit, the initiative helps educators and the general public to better navigate the digital information landscape. Simple skills can go a long way in avoiding manipulation: the ability to differentiate between fact and opinion, the ability to identify trustworthy news sites, the ability to identify a bogus tweet or Twitter account. Mike Caulfield, the director of Blended and Networked Learning at Washington State University in Vancouver, has also done important work on literacy. Caulfield champions a simple but effective approach to consuming speech online: Before engaging deeply with an article or tweet thread, take a minute to learn about the source.

"Think before you share" is a simple rule that could dramatically transform the information environment today; however, it's not only individual consumers who need to learn that lesson. Media, too, amplify the most sensational trends on social media, often covering absolute nonsense pushed by a relatively small handful of people: "Some people on the internet are saying . . ." Media coverage of a small, sensational controversy can amplify its reach significantly. Researchers like FirstDraft work on educating reporters to be aware of manipulation tactics; those at Data & Society have urged reporters to practice "strategic silence"—that is, choosing not to cover (and thus amplify) speech known to be false.[55] Many reporters already do this in relation to sensitive issues like suicide—perhaps it's time to update that playbook to include other forms of speech that, while nonviolative, are simply designed to foment outrage and generate clicks.

Platforms, governments, media, and civil society each have a role to play in this broader social restoration. Platforms can expand the reach of media literacy efforts in a variety of ways, such as by incorporating lessons via pre-roll on video content or in their "informing" labeling efforts. Governments can take

action also, particularly at the state and local levels, by offering formal education programs as well as investing in developing communications channels to grow trust with their citizens. Finally, civil society organizations, which often still enjoy a high degree of trust in their communities, can inflect lessons to account for the specific concerns and needs of the people they serve.

But, ultimately, the root cause of much of the discord online is a loss of trust and social rifts offline. As Chris Bail notes, closing the perception gap—the beliefs we hold about the "other side"—is a top priority for reducing polarization.[56] Bolstering the offline work of efforts such as America in One Room[57] and other community-led efforts to bring people together, to break down misconceptions and barriers between them, is as necessary as any social media reformation project. These community-driven efforts to counter affective polarization, *in conjunction with* the rethinking of algorithms and affordances, may ultimately turn out to be the best way to reduce the vitriol, the viral spread of false and misleading information, and the tribal appeal of online bespoke realities.

The present era of participatory propaganda and pervasive manipulated consensus is unique. But there have been many historical examples of periods of social upheaval following the invention of a new technology. The introduction of the printing press in the fifteenth century democratized speech in an entirely new way—and introduced a period of societal dissensus. The ability to mass-produce speech led to the spread of radical texts like Martin Luther's Ninety-five Theses, triggering religious fragmentation and the Thirty Years' War, one of the most destructive wars in European history. But eventually, with the Peace of Westphalia nearly a century later, equilibrium was achieved. Pamphleteering continued—especially in the nascent United States—but eventually consolidated into newspapers with mastheads and codes of ethics. Consensus was hard-won, but ultimately possible.

Social media still hold immense promise: human connection uninhibited by geography and instant access to knowledge. But its current iteration, with amoral algorithms and easily weaponized affordances, is not fulfilling that promise. Creating a future where powerful communication technology enables freedom of speech and expression without also facilitating mass manipulation, where social networks facilitate consensus across borders and bolster the foundational values of Western democracy—this is the challenge that faces us in this moment.

9

The Siren Call of Content Moderation Formalism

EVELYN DOUEK

Introduction

On January 8, 2021, after years of resisting calls to do so, Twitter suspended the account of the then-sitting president of the United States, @realDonaldTrump.[1] In an unsigned blog post that looked more like a judicial opinion than a corporate press release, the company laid out its reasoning in detail, reciting Twitter's rules and ostensibly applying them to the facts at hand.[2] The post explained why, despite the company's relatively long-standing rule that Twitter would generally not remove tweets or accounts of world leaders given the public's interest in knowing what their representatives think,[3] two recent tweets from President Trump fell within its exception for glorification of violence.

Was this an even-handed application of Twitter's prior rules? Twitter *had* always said world leaders would not be above its policies entirely and would remove any account for threats of violence. But the company had also previously stated that "[w]e focus on the language of reported Tweets and do not attempt to determine all potential interpretations of the content or its intent."[4] By contrast, in defending its decision to permanently suspend @realDonaldTrump, Twitter referred not only to the wording of two relatively anodyne tweets, but also to how those tweets should be read in the context of the broader events in America (especially the violent storming of the Capitol two days earlier) and how the tweets were being interpreted both on and off Twitter as encouraging further violence.[5]

Facebook had indefinitely suspended Trump's accounts on its platforms the day before.[6] Facebook CEO Mark Zuckerberg's post announcing the decision described his reasoning in somewhat less detail than Twitter's post. But two weeks later, Facebook referred the decision to its Oversight Board,[7] a body with

quasi-judicial characteristics that reviewed Facebook's decision and issued a public, reasoned opinion upholding this decision that apes that of a court.[8]

When it comes to the content they host, these private companies can largely do whatever they like: Under American law, at least, they are not required to host any user on their platforms, nor are they obliged to afford any kind of due process or explanation for their decisions.[9] They could determine the future of former president Trump's account by coin flip and no court would uphold any challenge. So why did they instead elect to write long, tortured blog posts or invoke elaborate procedures in trying to rationalize their decisions?

The answer is both intuitive and seemingly irrational for profit-driven companies: Legitimacy matters. These companies' decisions to deplatform the leader of the free world were controversial. While many celebrated the step as long overdue, others—including politicians around the world[10] and other prominent public figures[11] not usually thought of as allies of Donald Trump—decried it as censorship and an infringement on free speech. Caught between a rock and a hard place, where there was no "right" or uncontroversial answer, Twitter acknowledged the need to work on "inconsistencies . . . [and] transparency" in its policies and enforcement,[12] while Facebook pointed to its willingness to submit the decision to independent review so the company wasn't making such a significant decision on its own.[13]

This was a particularly high-profile example of what has been a general trend, especially apparent in recent years, toward greater transparency and due process in content moderation as a means of establishing their consistency and commitment to rule-bound decision-making. This turn to a more formalistic decision-making model is a common path for institutions that exercise extensive power over public interests and are trying to assuage distrust of their competence or credentials to do so. And it has been embraced by scholars and policymakers as the solution to many of the problems that plague social media platforms.

But I argue in this essay that the quest to make content moderation systems ever more formalistic cannot fix public and regulatory concerns about the legitimacy and accountability in the way platforms moderate content on their services. Such a goal is illusory. The more formalistic way the companies handled the deplatforming of Donald Trump clearly provides no model for the vast, vast majority of content moderation decisions. These companies decide to act or not act on millions of pieces of content uploaded to their platforms every day. The variety of decisions they make, and the way in which they make them, far exceeds the narrow kind of decision involved in choosing whether to leave a particular piece of content up or take it down. The largest platforms operate massive unelected, unaccountable, and increasingly complex bureaucracies deciding how people can use these central squares of the internet at unimaginable scale.

A formalistic model, invoking judicial-style norms of reasoning and precedent, is doomed to fail at this scale and level of complexity.

This essay first traces the origins and path toward content moderation formalism and juridification in the section "The False Promise of Juridification." Never a fully thought-out plan—indeed, largely an afterthought—platforms have been building systems of governance built around elaborate rule sets enforced by large task forces and automated tools, ostensibly trying but always failing to achieve consistency and determinacy. The section of this essay titled "Toward Content Moderation Realism" describes why this is a fool's errand. The practical realities of content moderation mean formalism is even more out of reach in this context than in others. Even if the offline world to which content moderation governance is frequently compared could achieve it, the scale and messiness of online speech governance make simple translation of offline tools inapposite. But, of course, perfect formalism is not a realizable goal, online or off. Part of the dissatisfaction with content moderation is caused by this unrealistic goal. The "Conclusion" then charts a course toward content moderation realism. The goal of a better and more legitimate system of content moderation governance should not be forsaken just because the formalistic model that stakeholders currently aspire to is impossible, but being realistic about what to aim for is a necessary part of building that system. This requires accepting what can and cannot be constrained, what mistakes can and cannot be remedied, and the trade-offs inherent in every choice. Some degree of consistency and determinacy needs to be sacrificed for greater effectiveness overall.

When this essay talks about "effectiveness" it does not mean to imply that I think content moderation, done effectively, can bring about an ideal speech situation. There is not and will never be agreement on the ideal substantive content moderation rules. But content moderation of some form *is* inevitable: Content moderation is *the* commodity platforms offer.[14] It is both a First Amendment right and in their business interests for platforms to interfere with their users' speech more than a government could ever mandate.[15] And while platforms' interests will not align with the public's, there is room (albeit limited) for stakeholders like government, civil society, advertisers, and users to push platforms to reform their rules to respond to their preferences. "Effectiveness," then, has a content-neutral meaning here: It means the efficient enforcement of platform rules in a way that increases public accountability and transparency, which creates a mechanism for these stakeholders to influence platforms' operations or address the problems created by them.

Finding a way to pair a more realistic model of content moderation governance with ways to impose constraints on platform decision makers is perhaps the greatest challenge for the future of freedom of expression online. The answer

will require new tools, and not simply attempts to bring offline speech governance models to the online world.

The Juridification of Content Moderation

The story of how platforms rapidly constructed legalistic content moderation bureaucracies is well-told by now.[16] The task of "maintaining a system of freedom of expression in a society is one of the most complex any society has to face,"[17] but the governance systems of the most important speech forums in existence today were created by people who somewhat accidentally found themselves in the position of having to do so. In the early days of platforms' lives, "very few lawyers were focusing on the responsibilities that commercial online companies and platforms might have toward moderating speech."[18] But, as these platforms became ever more important and scrutiny increased accordingly, they soon found themselves hurriedly building initially ad hoc systems to manage their users' speech.[19]

In the early days these systems were fairly rudimentary. Twitter's first set of rules was 568 words long,[20] and YouTube's was a mere one-page list of instructions.[21] Internet governance was fly-by-night ("Move Fast and Break Things!")[22] and almost completely opaque.[23] Decisions descended from Silicon Valley overlords as if from on high, and they often seemed arbitrary—indeed, they often *were* arbitrary.

But as user and societal expectations of platforms have increased, especially with the techlash of the past half-decade, platforms' rulebooks have quickly expanded, becoming ever more comprehensive. They have come to resemble the "prolixity of a legal code."[24] They have moved from a standards-based approach to a rule-based one, constructing a system aiming for "consistency and uniformity: to get the same judgment on a piece of content, regardless of who was moderating it."[25] That is, there was a turn toward "formalism," or "decisionmaking according to *rule*," intended to screen off from a decision maker "factors that a sensitive decisionmaker would otherwise take into account."[26]

American law has seen this story before. When bureaucracies experience a trust deficit, as is the persistent condition of the administrative state, for example, a common strategy for attempting to rebuild trust and legitimacy is recourse to formalism.[27] American law has a deeply ingrained intuition that proceduralism and legalism can mitigate concerns about the exercise of otherwise unaccountable power.[28]

That this path is also reflected in content moderation history is not overly surprising given that it was American *lawyers* who oversaw the development of

content moderation policy at the major platforms.[29] After all, "if all you've got is a lawyer, everything looks like a procedural problem."[30] And so there has been a steady march toward a legalistic, formalistic paradigm of content moderation that is rule-based and provides some semblance of procedural justice in an attempt to ease anxieties about the enormous and unconstrained power platforms have come to exercise over modern discourse.[31]

Lawmakers, too, are enamored with formalistic content moderation. Hamstrung by substantive and practical limits on their ability to take content moderation decisions away from private platforms completely,[32] governments have increasingly turned toward measures designed to increase platform transparency and bind their hands by forcing them to stick to their own rules.[33] This follows years of civil society organizations' demanding similar transparency and due process protections in content moderation.[34]

Facebook's Oversight Board experiment is the apotheosis of this approach to content moderation.[35] Clearly invoking a court-centric conception of governance, the board decides individual cases where users have specific grievances and standing.[36] Its decisions are carefully reasoned conclusions based on Facebook's rules and values and human rights norms.[37] These reasons "have precedential value and should be viewed as highly persuasive when the facts, applicable policies, or other factors are substantially similar."[38]

The Oversight Board is not the only example of this approach to content moderation; the trend is apparent across the industry, as the example of Twitter's reasoning in the Trump example described above demonstrates. It's not just the major platforms—Twitch's blog post explaining the nuances of how its Nudity & Attire and Sexually Suggestive Content policies apply to the surprisingly prevalent genre of Hot Tub streams (which are exactly what they sound like) was more than eleven hundred words long.[39] Knitting forum Ravelry published a blog post describing its decision to ban support of Donald Trump or his administration across the site that included its new policy language and a list of "policy notes" that essentially amounted to interpretative guidance.[40] The practices of producing periodic transparency reports about their rule enforcement and offering users rights to appeal decisions taken against them have proliferated across the social media industry.

Formalism has its allure. It serves the rule-of-law ideals of predictability and consistency between cases.[41] This in turn helps guide users' behavior toward compliance, because they know what the rules are in advance.[42] Reasoned decision-making can increase users' acceptance of decisions taken against them because they have greater understanding of the reasons why.[43] Theoretically, the development of standardized norms and the commitment to treat like cases alike constrain our tech overlords in their otherwise absolute discretion to do whatever they want. The attractiveness of procedural formalism in the realm

of content moderation may be especially great because the idea of universal agreement on substantive speech rules is fanciful. Commitment to regularized processes for channeling that disagreement may be more achievable.

As this trend of formalization has developed, however, it has required ever finer distinctions to accommodate and categorize the infinite variety of online content. Again, this is not surprising: As any set of rule-based jurisprudence develops, rules converge toward standards over time as in hard cases decision makers use rule-avoiding strategies such that proliferating rules begin to "operate as standards, far more than may be apparent from the face of the rule itself."[44] This has been the course of elements of First Amendment doctrine, which "has become only more intricate, as categories have multiplied, distinctions grown increasingly fine, and exceptions flourished and become categories of their own."[45] Again, this is reflected in the path of content moderation rules, too, but on hyperdrive. The sheer volume of speech decisions and the fact that the outcome of these decisions remains permanently inscribed on the internet, unlike when someone says something on the street corner, make attempts to achieve consistency even harder. Facebook alone makes more speech decisions every day, perhaps even every hour, than the Supreme Court ever has in its entire history. Although datasets of these decisions will always comprise only a subset of those decisions,[46] even these partial archives of speech decisions are the largest in history, exposing platforms to relentless criticism. Attempting to impose order on this chaos has been . . . challenging.

The results are becoming somewhat ridiculous. Facebook's once simple and famously uncompromising ban on adult nudity[47] has given way to a lengthy policy that covers its positions on "implied stimulation," "by-products of sexual activity,"[48] and the nuances of what exactly it thinks constitutes "breast squeezing."[49] Twitter has a complex three-by-three matrix for assessing "coordinated harmful activity."[50] YouTube is consistently vaguer in its public-facing rules, but it, too, has unveiled policies that draw fine distinctions in an attempt to bring order to the enormous diversity of things humans do and say online. Its policy on eating live animals, for example, allows eating living *invertebrates*, apart from fish and frogs, but not eating living *vertebrates*.[51]

These rapidly expanding rule sets call to mind the cartographers in Borges's story *On Exactitude in Science* who, in their quest to create a perfect map of an empire, created a map the size of that empire. In their attempt to make the governance of the messiness of human communication on the internet completely predictable and consistent, platforms' rule sets seem to be trying to create a rule for how every post will be treated. Like Borges's story suggests we should be, I am skeptical of whether this is a feasible or useful approach to the project of bringing meaningful oversight to the vast terrain of online speech forums.

The False Promise of Juridification

This is not to say that the general trend of formalization is without value. From a baseline of almost complete opacity and arbitrariness, publishing rulebooks and attempting to apply them fairly and consistently is progress. Some degree of formalism, and even juridification (by which I mean "the spread of legal discourse and procedures into social and political spheres where it was previously excluded")[52] is inevitable and salutary. But more is not always better. Determining what level of formality and juridification in content moderation will result in accountability and legitimacy is an optimization problem.[53] Given those trade-offs, making formalism the lodestar of what a content moderation system should achieve—trying to create a system in which decision makers profess to do little more than call balls and strikes—is a mistake.

The aspiration to a highly formalistic system of content moderation is a false promise for three reasons: First, it is false in the sense that it is impossible to attain in practice (as offline adjudicatory systems have long realized, and as characteristics of the online world exacerbate); second, it is false in the literal sense that platforms are not truly committed to that promise and do not intend to be completely bound by precedent; and, third, it is false in the sense that even if it were attainable and platforms were truly invested in achieving it, it would not cure the distrust and legitimacy deficits that currently plague content moderation. The remainder of this part addresses each of these false promises in turn.

Content Moderation Formalism Is Impossible

The scale and scope of online speech systems make formalistic content moderation a practical impossibility. The disappointment created by platforms' resulting failure to achieve this ideal is made worse because content moderation is judged against a comparator of a utopian offline justice system that does not exist.

The quest to garner legitimacy by creating comprehensive, specific, and determinate rule sets is doomed to fail. Indeed, beyond a certain equilibrium, chasing this goal will become self-defeating because the bar of perfectly consistent and stable content moderation is not only impossible to clear but also creates complexity that confounds efforts to do so. A map the size of the world is not a useful guide, and content moderators cannot be expected to remember every page of the handbook in the few seconds they have to review every post.[54] Even if these comprehensive rule sets could be fully internalized by content moderators, the tens of thousands of them distributed around the globe will never agree on their application, like whether a particular picture meets the specific definition of breast-squeezing, to take one example. Online speech norms

develop at a pace that makes unduly specific rules obsolete at an unsustainable pace. The meaning of a hashtag or meme can, and often does, change literally overnight.[55] The context-specific nature of speech means the exact same post can have very different meanings in different cultures, societies, or situations, creating a dilemma for decision makers trying to treat like cases alike: The cases are both alike in form but not alike at all in meaning.

Then there is the enforcement problem. The scale of online speech means perfect rule enforcement is impossible.[56] Mistakes are inevitable, and they will always be discoverable. It is "exceedingly easy (and often misleading) to find cursory evidence of anything on social media because there is so much of it and . . . it is almost by definition optimized for search."[57] Content moderation at scale means content moderation errors at scale, and those errors often remain just sitting online, waiting to be found. A content moderation system that attempts to find legitimacy in perfectly predictable formalism will find itself constantly confronted with and shamed by its inevitable mistakes and inconsistencies.

The flood of speech that the internet has enabled and that content moderation must govern is truly staggering. This difference in scale is a shift in kind, not merely degree. Because a "correct" decision in each case (assuming, for the sake of argument, there is such a thing) is practically impossible at this volume, individual failures cannot be considered system failures. Content moderation is a system that relies on a distributed and underresourced workforce of tens of thousands of frontline workers[58] and artificial intelligence tools that are blunt and stupid but essential.[59] This system will get decisions wrong all the time. Effective content moderation means getting less wrong and being wrong in the right way. To be wrong in the right way involves choosing which side of the line to err on—whether that means more false positives or negatives.[60]

Critics do not expect perfection in every initial decision, but the dominant view is that mistakes should and will be corrected through the availability of an appeals mechanism that allows every case to be "reheard" by a human moderator.[61] But even the smaller fraction of content moderation decisions that are appealed would still overload anything but an impractically large workforce. And it is not clear that even then consistency would be possible; human error is irradicable, as is human disagreement on the application of rules.

Of course, my argument can only be pushed so far and shouldn't be mistaken for an argument in favor of the status quo. There are still far too many examples of flagrant errors on which there can be no disagreement. This is not an argument that content moderation systems have no need to improve because they can never be perfect. Clearly, platforms need to be more committed to abiding by their formal rules and explain how they have applied them. But it is also important to be clear-eyed about reasonable goals.

The ideal of every content moderation victim having their day "in court," being heard and getting reasons, is an unrealistically court-centric conception of justice. Many rights determinations in the analogue world are made without the possibility of meaningful appeal or with procedural barriers making access to such appeals illusive. Many of these decisions have much more significant impacts on the subject of the decision than a post being removed from a social media platform. These offline failures are often ignored in discussions of content moderation, which imagines a utopian system of governance as the standard that content moderation must meet.

Platforms' Commitment to Content Moderation Formalism Is Fickle

The promise of content moderation formalism is false in another, more literal sense too: Platforms do not mean it. They talk in the language of rights and lofty ideals,[62] but it can never be forgotten that they are profit-maximizing businesses. Content moderation is *the* commodity platforms offer.[63] There is no reason to assume that their interests align with societal interests, and every reason to think that often the case is otherwise. Certainly, a degree of consistency and predictability in their rules will lead to greater user satisfaction and may help stave off certain forms of draconian governmental regulation. But there are constant stories of platforms departing from their own rules when it serves their political or commercial interests.[64]

Let's return to the example that this essay opened with. As described above, Facebook and Twitter released extensive, and tortured, reasons justifying their decisions to suspend Trump's account as a simple application of prior rules to the facts at hand. Facing what would be one of the most high-profile and contentious content moderation decisions in their history, these platforms tried to wrap themselves in formalistic and legalistic reasoning, downplaying the extent of discretion involved and disavowing political considerations.[65] A convenient narrative, if you can believe it. But, of course, there is another explanation to the question of "Why now?" Platforms had been under pressure to ban Trump well before January 6, 2021. But there was one notable difference now: Donald Trump had been deplatformed from the presidency a few months before, and the Senate had just changed hands from Republicans to Democrats. Platforms' regulators would now be Democrats, not Republicans, and Trump's star was entering retrograde. Years of controversy had taken their toll, and platforms' continued attempts to explain why political leaders should be treated differently under their rules had never succeeded at quelling the very loud and constant criticism.

It is also true, of course, that the events at Capitol Hill on January 6, 2021, were exceptional and likely shook platform executives as people and citizens. But this does not undermine the realist take on their decisions in the wake of those events: There is no set of rules that will ever be able to account for every eventuality, and there will always be an irreducible zone of discretion needed to respond to the exigencies of the moment. Their political and commercial interests aligned with the decision they took—whether or not they were platforms' only considerations, they certainly weren't in conflict.

On a broader scale, platforms are trying to have their cake and eat it too by being two-faced about their use of formalistic reasoning. They use it both as a distraction from underlying systemic issues about how their platforms operate, while also insisting that individual errors are not necessarily indicative of their broader operations.

Focusing on individual cases in a formalistic and juristocratic manner keeps the level of analysis at the micro, preventing proper evaluation of the systemic, macro forces that shape the way their services operate and are far more consequential for society than any single post standing alone. What kinds of content platforms optimize for, the affordances they offer users to both create their own networks and empower them to deal with abuse, the nudges they provide toward healthier communications, and many other structural choices, create and tilt the playing field on which content moderation occurs. Focusing on individual cases through an individualistic lens is focusing on line calls and ignores how platforms create the environment in which the game is played. But there is nothing inevitable or natural about the way platforms are constructed or the shape of the playing field: Everything should be up for grabs. Keeping the focus on individual rules and their application to singular cases provides a distraction from this bigger picture.

But platforms also try to disavow the importance of individual cases, even as they use them as a decoy. They point to their massive scale and the inevitability of error to shrug off individual mistakes as the price of the way the internet empowers everyone to have a voice. Mark Zuckerberg has taken to likening content moderation to policing a city, arguing that cities will never be able to get rid of all crime.[66] He's not wrong: This is the very same argument I made above about how fully deterministic rule sets are impossible. But the approach of dismissing individual errors as mere "operational mistakes"[67] means that the systems that led to those errors are not interrogated.

The problem, then, is that content moderation formalism is half-baked. It is both what platforms and, increasingly, outside stakeholders hold up as the ideal content moderation should aspire to, while also—when it suits—arguing that any failures to attain it are made simply because it is impossible to achieve in practice.

Content Moderation Formalism Cannot Deliver Legitimacy

Finally, even assuming content moderation formalism was theoretically possible and platforms were truly committed to realizing it, it cannot cure the legitimacy deficits from which content moderation systems currently suffer.

On the one hand, where distrust in a system is "deserved" in the sense that it has "accrued a track record of ineptitude, bias, or unresponsiveness—then the move towards greater legal specificity, adversarial checks, and close fidelity to rules may well result in an improvement in . . . justice."[68] Many institutions have tried to garner trust in this way. Daphna Renan has, for example, described how President Carter tried to use the Office of Legal Counsel (OLC) in this way in the aftermath of Watergate, trying to "instantiate a type of legalistic credibility in response to political pressures from a wary public." This was a brief heyday for the *formalist structure* of OLC. Under this conception, formal legal decisions reached through a relatively apolitical process might cabin presidential discretion in any one-off case. But they would empower the president by helping to rebuild credibility.[69]

But legitimacy is a complicated concept, and a formalistic conception of legitimacy is only one aspect of it. Authority also needs to be sociologically and morally legitimate to be accepted, and legalistic legitimacy alone is not enough to garner social and moral respect.[70] Indeed, "the people with the deepest, most corrosive cynicism about law and legitimacy . . . are often those who began with unrealistic expectations that . . . decision making could be wholly apolitical or untouched by ideological influence."[71] Formalism will not appease these critics and can be damaging to a broader conception of legitimacy. An overly legalistic approach can be "incompatible with the need for timely delivery of policy, and with . . . expectations [that an institution] should deliver on substance rather than concentrating on procedures."[72] When focusing on a legalistic conception of legitimacy impairs effectiveness, the overall effect on legitimacy is ambiguous.[73] In a culture of deep distrust of content moderation in general, and waning faith in formalism more generally, a formalist model is less capable of signaling the very credibility it is intended to show.[74] When the pursuit of formalism stands in the way of achieving other governance goals, like speed of decision-making or responsiveness to prevailing social conditions, it will harm rather than enhance legitimacy and perceptions of accountability and effectiveness.

These are all lessons of offline American legal realism. The online world does not need to relearn them. The story that highly formalistic content moderation could exist or would be unquestionably beneficial is a myth: a story that platforms are trying to tell and in which many outsiders also place their faith, but will never be reality.

Toward Content Moderation Realism

None of this is to say we should give up on the goal of providing greater legitimacy to content moderation. Content moderation regulates the most important spaces for speech in the modern era. The rules for online speech affect almost every facet of modern life, from politics, both domestic and international, to public health, national security, and culture. There are good reasons to want a better system than leaving these decisions in the hands of completely unconstrained Silicon Valley executives.

To say formalism will not be an effective form of such constraint is not to admit defeat entirely. It is simply an acknowledgment that there is more than one way of achieving legitimacy. Content moderation needs a more fluid, experimental, and responsive form of governance than formalism provides. Achieving such governance will require institutional experimentalism. This is an entirely new problem set which the world has only started truly confronting in the last decade. Offline speech governance systems have had far longer to develop and are still plagued by inconsistencies and gaps. Effective content moderation governance will not arrive overnight and will involve trial and error. But this part sets out some general principles that should guide content moderation institutional experimentalism before offering some tentative thoughts about the paths and obstacles to implementing them. Being realistic about what content moderation is and can be means accepting that these rules do not operate in a vacuum but must account for their institutional, political, and social context, will never be able to eradicate discretion, and will always involve errors in their application.

Realistic Principles

Principle 1—Empower Decision Makers Other Than Lawyers. Content moderation formalism is an ideal created by lawyers for lawyers. Systems of governance that hinge on a culture of legalism empower lawyers and enlarge their influence.[75] A lawyer-dominated organization will tend toward ex post enforcement and case-specific policy elaboration, but experts from other disciplines will see and advocate for the benefits of more ex ante rulemaking and approach problems in terms that lawyers don't.[76]

Content moderation is as much a technical problem as a political and policy one. At every stage of the content moderation life cycle, the question of what is technically possible must inform policy. A perfect rule on paper is not a rule at all if it cannot be enforced in practice. Reducing the delta between platforms' written rules and their actual enforcement remains one of the most important challenges for content moderation. This requires technologists who can create,

tinker with, and describe to platform policy makers the automated tools that do the bulk of content moderation. It also requires quality assurance specialists to assess how effectively policies are working, and people to communicate policies and changes in rules to the human moderators that enforce them.

Such thinking about content moderation enforceability must start *before* any rule is written. There are a number of choices platforms make that are not typically thought of as "content moderation" choices, but should be because they fundamentally affect how rules are enforced: the way a platform is constructed, the ease of reporting mechanisms, the nudges it gives users in what to engage with or amplify ("Do you want to read this before retweeting it?"[77]), the extent to which it empowers victims of abusive behavior to take control of their own experience. These decisions, too, should not be made primarily by lawyers. They require technical expertise, product manager buy-in, and the input of those who bear the brunt of the abusive behavior that content moderation is intended to protect against. No content moderation system can compensate for a product or platform designed around principles that will thwart content moderation's goals.

Many other professions have relevant expertise. Some have called for increased roles for librarians in content moderation, given platforms' role in knowledge curation.[78] As platforms increasingly overrule media organizations' editorial decisions with their own,[79] there's a good case to be made that they should give more power to journalists and editors who are familiar with editorial norms, media markets, and so on and so forth. These examples are just to demonstrate that the legalistic frame is unduly narrow and inadequate for the vast array of decisions that are involved in content moderation.

Principle 2—Abandon the Aspiration to Universally Consistent Rules. The utopia of a borderless internet was always a fantasy, as Goldsmith and Wu were early to point out.[80] Nevertheless, platforms have generally insisted on having a single set of global rules.[81] Governments are also increasingly enamored with insisting that platforms should apply their laws globally and not merely geoblock content within the borders of their state.[82] But such universalizing ambitions are misguided. As international human rights law recognizes, local context is all-important in understanding the meaning of speech as a factual matter, and there is an important role for regional law to play as a normative matter.[83]

But to accept that there can be no universal rules is only to ask a different, and much more difficult, question: How can platforms enforce speech norms in a way that is sensitive to local context but does not engage in cultural relativism? When is it appropriate for platforms to push back on government demands out of respect for users' speech rights, and when is that recalcitrant disobedience of legitimate governmental power? This is one of the most difficult and important issues for the future of online speech. I cannot answer it here; indeed, it's not clear that a good answer exists. But for present purposes it is enough to note that

the realist is less embarrassed by inconsistencies. A formalist seeks consistency and uniformity, while a realist is happy to take each case on its own terms (or at least is more accepting of the fact that this is what happens).[84]

The idea that any two cases can, in the hands of any sufficiently competent lawyer, be made to be analogous or distinguishable is not new. But content moderation happens on a global scale while also being incredibly granular. Every online utterance is potentially subject to moderation, while the social contexts that content moderation must span could not be more diverse. Step one in resolving this tension is admitting we have a problem.

Completely consistent content moderation is not only practically impossible, but normatively undesirable. Different contexts, places, and times require different rules. Content moderation should not be embarrassed by this but embrace it. Indeed, it is a strength of private content moderation systems that they can be far more fluid and adaptive than state-based governance systems that can be ossified and clunky.

Principle 3—Expect Constant Change and Adaptation. Content moderation is capable of more change and adaptation than offline governance systems, which are known for their stasis and lack of agility, because it is not beholden to the formalism of democratic procedure.[85] A platform's rules can be changed by tapping out a blog post (although putting that change into practice can take longer). Amendments do not have to go through both houses of Congress or a notice-and-comment process. They cannot be subject to challenge before a court (although Facebook's Oversight Board may now weigh in on its policies).

This fluidity and responsiveness of content moderation is a good thing, given the phenomenon that it regulates. The meaning of online speech is in constant flux. The hashtag #ProudBoys can go from being an organizing tool for an extremist group, to something reclaimed by gay men after the president referenced the group in a presidential debate, and back again within days.[86] As platforms devise rules, such as rules for policing manipulative or misleading behavior on their services, nefarious and merely opportunistic actors alike try to exploit loopholes, resulting in a constant game of cat and mouse.[87] Political circumstances can also suddenly but drastically change the meaning of online speech. Harmless satire can morph into a dangerous meme rapidly, and the tipping point is not always easy to identify.[88] Content moderation must respond to these evolutions with a speed that governments could not.

This does not mean that government regulation has no part to play in content moderation. It does mean, however, that regulators must be careful to avoid crafting rules that undermine the responsiveness of the system and disincentivize innovative responses to the ever-developing understanding of how the online information ecosystem works.

Take the example of transparency reports. Even the industry-leading transparency reports (Facebook's is currently the most detailed) focus on a very limited slice of what constitutes content moderation: generally, the number of takedowns of individual pieces of content with some level of detail about how that content was detected and if it was subject to appeal. Legislative enactments and proposals are increasingly looking to mandate transparency reporting by platforms, and the form this takes is often to demand exactly these kinds of reports, which the major platforms already produce.[89] But these reports often lag far behind new product rollouts like the creation of groups, pages, or audio spaces. They also don't acknowledge, let alone provide transparency into, or assess the effectiveness of, other measures platforms take that are core content moderation decisions, like labeling posts that contain misinformation, deamplification of problematic content, or mass sweeps and deletions of certain trends, groups, or users on the basis of "behavioral signals."

Given these constant changes in how platforms moderate, static transparency mandates will both fail to respond to platform product changes *and* fail to incentivize them. Why should platforms keep innovating in content moderation—which they generally see as essentially a cost center and unfortunate by-product of their main offering (the content that creates the pesky problem of moderation)—when static regulation has given them a bar they can more easily meet? Certainty and stability come at the cost of responsiveness and effectiveness.

The content moderation realist is also not embarrassed by the uncertainty caused by the need for a fluid system this creates. The formalist is: It creates inconsistent results over time and fails to set clear governing principles for both platforms and their users when the governance requirements and rules keep changing. These all seem at odds with fundamental principles of the rule of law like stable rules with consistent application. But the realist acknowledges that legal certainty is an instrumental value and an optimization problem. A more rigid decision-making structure based on rules may result in fewer errors or inconsistencies overall and more certainty, but a more fluid one based on standards might result in fewer *serious* errors and more optimal decisions that are context-sensitive.[90] The formalist may be right that a substantively unjust decision in an individual case is likely to have a small social cost, but "from a dynamic perspective, the opportunity costs of formal rules in these cases may be relatively high because these may be the cases in which the potential for learning and systemic gain from contextual examination may be highest."[91] The realist acknowledges that rules will be bent in hard cases regardless, accepts this inevitability—indeed, embraces it—and designs governance around it.

Principle 4—Talk About Errors Differently. The formalistic model has no space for candid acceptance of errors. It's important to be clear what I mean by "error" here: I do not mean substantively unjust outcomes, with which—as has just

been noted—formalism *is* familiar.[92] I mean a decision that is an error on its own terms: The rule and its application are clear, but the outcome does not accord with that application. Formalism's emphasis on clear rules and certainty is in large part to avoid such errors. As such, substantively unjust outcomes might be accepted in the name of the greater good of making it easier to get it right according to rules as written. The idea that the ultimate goal of a system is to minimize errors at all costs does not sit easily with a speech governance system that is and will always be plagued by them. As discussed above, "perfect enforcement" in content moderation is impossible. Content moderation is all about tradeoffs. It's possible to imagine a system of content moderation with much fewer errors: one that commits to taking barely anything down except for the most egregious content, for example, or one that takes down any post with a certain word in it regardless of context. These approaches would be easier to operationalize, and therefore many more decisions would be congruent with the stated rules and so "correct" in the sense I mean here. But such correctness comes at a high cost and elevates certainty, an instrumental value, over substance.

Therefore, a key need in content moderation debates is to be *less* embarrassed by mistakes, so we can have more candid conversations about them. Relative error rates are intrinsically bound up in the rule-formulation process and constrain policy possibilities. A realist has a better language for understanding this than the formalist who would find such a proposition distinctly uncomfortable. As usual, though, the formalist's denial that error rates exist does not change this fact, but merely pushes it beneath the surface. Embarrassment about errors prevents them from being properly evaluated.

Principle 5—Plan for and Channel Disagreement. Content moderation rules will never be settled. Ideal speech rules do not exist in any system, and the closest approximation of them will change over time. Contestation around what people should and shouldn't be allowed to say is not something to be eradicated but made more visible and continuous. Argumentation about these rules is part of how a community constitutes and defines itself.

Once again, returning to the example of the deplatforming of Donald Trump is illustrative. A few months after platforms made the decision to deplatform the president, a survey found that Americans were almost exactly evenly divided on the question of whether now-citizen Trump should be allowed back on social media.[93] Legal experts, politicians, and civil rights figures the world over had different opinions on what the platforms should have done. As long as there are social media platforms there will be politicians on social media platforms that use their online pulpit in harmful ways. The process of arguing about how Donald Trump should have been treated by platforms was not only about a single account, but about how other politicians should be treated both now and in the future. It is through the very process of arguing about what to do that norms can

be created and defended. In a world where there are no settled legal norms about what is allowed on social media and so legal legitimacy is elusive, sociological legitimacy is important. In a deeply divided society such legitimacy cannot be taken for granted but must be fought for. It cannot come from unilateral, opaque decisions, but must be created through open, inclusive discussion and public reason.

The realist accepts that there will be an irreducible zone of discretion for platforms to decide their content moderation policies and enforcement. There is no legal or practical alternative other than accepting that many difficult choices will always be left in their hands. Creating avenues for open and directed disagreement with those choices, information-forcing mechanisms so that the vast information asymmetries between platforms and both their regulators and users are reduced, can make space for the process of generating norms and expectations around content moderation—still in their infancy—to take place.

Putting Principles into Practice

It's beyond my scope here to give a comprehensive account of how these realistic principles are made into reality. But one thing should be clear: It will require multiple stakeholders. Platforms are both a necessary part of the process, given their presence at the front line of content moderation and technical capacity, and also actors with motivations that are unlikely to align with the public interest, given their status as profit-making businesses. Governmental regulation will be important in coercing changes platforms will not adopt on their own initiative and bringing the legitimacy of democratic accountability. But governments are also hampered by constitutional limitations on their ability to regulate speech and the shortcomings of the comparatively slow and rigid nature of regulatory processes. Advertisers have the power to influence platforms to bring about change given that they are platforms' true customers—and indeed have brought about changes such as promises of independent audits of platform practice that other stakeholders had failed to make happen—but their idea of "safe" content is unlikely to align with others' given their priority of protecting their brands from controversy. The voices of civil society and affected users will be essential to understanding the impacts of any content moderation choice, but their power to compel change is weak.

Some of these changes can be brought about by government regulators, but most of the principles here are more negative rather than positive recommendations to regulators. They highlight things to be *avoided* rather than enacted. Rigid transparency and process mandates, punishment for individual errors, restrictions on how often platforms can change their rules—these are all measures that are reflected in regulatory proposals but would not be realistically

beneficial. In some sense, the most important avenue for change is cultural. Broader discourse about what content moderation governance should strive for needs to be informed by the reality of what is possible. It should not compare content moderation to a hypothetical formalist ideal model that isn't possible and wouldn't improve overall effectiveness or legitimacy.

Conclusion

From accidental beginnings, content moderation governance systems are becoming some of the most elaborate and extensive bureaucracies in history. They are deeply imperfect and need reform. But they will never mirror the picture that content moderation formalism—the ascendant philosophy of content moderation—paints for it. The sprawling chaos of online speech is too vast, ever-changing, and varied to be brought into consistent compliance with rigid rules. This is not an embarrassment or admission of defeat. Formalism has always had its limits offline, and it should be no surprise that it will not exceed those limits in the context of online governance.

The myth of content moderation formalism is unrealizable, but myths can have value. Platforms *should* strive for more regularity in their governance systems, and the baseline from which they started was very low indeed. But as these governance systems mature, it is time to be content moderation realists about the task ahead.

10

Free Speech on Public Platforms

JAMAL GREENE*

Social media companies such as Facebook, Twitter, and YouTube moderate the content that appears on their platforms. It is often assumed that the relative absence of legal limitations on their moderation practices rests primarily on their status as private companies, and that if a government were to operate an analogous platform, the First Amendment and international human rights law would condemn most of the content moderation the government did there.

I want to problematize this assumption, but not in the usual way of questioning the application of the "state action" doctrine to market-dominant technology companies. My challenge, rather, is to the premise that content moderation on a hypothetical *state*-operated social media platform would, ipso facto, be legally problematic. I offer the challenge not because I necessarily believe it to be successful but because I believe it to be a challenge, and therefore worthy of careful investigation and assessment.

Consider the following thought experiment. Let's say the US government were to establish and operate a social media platform in order to create a forum for citizen expression. The forum would reside within a broader marketplace in which there are also competing forums. What rules would the First Amendment and international human rights law permit the government to establish on the platform? Could the platform ban nudity? Pornography? Could it prohibit hate speech, for instance, by banning specific racial slurs? Could it require speech to be truthful? Could it apply a warning label or append additional links to posts it deemed to be misinformation? Could it downrank or limit the virality of certain posts? Could it prohibit anonymous posting? Could it ban commercial speech?

* Jamal Greene is Dwight Professor of Law, Columbia Law School. The author is a co-chair of the Oversight Board, which reviews content moderation decisions made by Facebook. The views expressed in this chapter are solely those of the author in his personal capacity. This essay has benefited from generous comments from evelyn douek, Jameel Jaffer, and David Pozen.

Could it suspend users who repeatedly violate the rules of the platform? Indeed, could it deploy rules at all? Content moderation at scale is not possible without rules, but rules are by their nature overbroad, posing a challenge both for strict scrutiny and for proportionality analysis.

The answers to these questions will appear obvious to some, but they are not obvious. It is true, of course, that all the above forms of expression are protected under both the First Amendment and international human rights law. A state-operated space created for the purpose of free expression is what First Amendment law calls a designated public forum.[1] While certain "time, place, and manner" restrictions might apply to such a forum, the government may not discriminate based on the content of speech or, a fortiori, the viewpoint the speech expresses.[2] And yet to take that injunction seriously would mean that content moderation is not possible. A platform whose boundaries were set by the limits of free speech rights would quickly be overwhelmed with spam, smut, and other unwanted speech and therefore be unusable for most users.[3] This analysis suggests that the government is, in effect, simply barred from entering the social media marketplace on anything like the terms of its choosing. We should think carefully before conceding that the Constitution requires this conclusion.

International human rights law also lacks an obvious answer to the hypotheticals above. State limitations on free expression may be justified only when the state can show that it is acting to protect the rights of others or to protect narrowly understood interests in national security, public order, and public health and morals, and when it is doing so proportionately and with adequate clarity and notice.[4] This test is a demanding one in the context of speech restrictions, ordinarily requiring a close and particularized nexus between the restrictions on speech and the rights being protected.[5] For much of the content moderation that occurs on social media platforms, this nexus is attenuated or lacking entirely. That said, the international human rights law regime generally contemplates that state speech restrictions carry penalties of detention, fines, or other coercive remedies. One's assessment of the limitations permitted might differ where the sole consequence is a post being removed or blocked from a platform for violating the platform's preset rules of decorum.

Another way to put the problem is to ask whether and to what extent the content-related regulatory practices of social media companies are *exceptions* to First Amendment and human rights law because the companies are private entities or whether, quite apart from their private status, the law's toleration of these practices simply applies existing legal norms to forums of a certain character.

The answer to this question is significant for at least four reasons. First, some commentators have called for platform content regulation to mirror First Amendment standards,[6] and at least one platform—Parler—has committed to

regulating content within the spirit of the First Amendment notwithstanding its private character.[7] These calls and commitments are not intelligible without clarity about what the First Amendment actually requires of the state.

Second, recent bona fide litigation around the social media practices of government officials using private platforms such as Twitter has assumed that the most relevant question is whether the official who is blocking commenters is acting in a public or private capacity and not, for example, on what basis and subject to what process the official was blocking people.[8] The latter inquiry can help clarify the rights and duties of public users of social media and those who interact with them.

Third, the major social media companies have committed via the United Nations Guiding Principles on Business and Human Rights (UNGP) and the Global Network Initiative to be guided by international human rights norms.[9] The platforms' compliance with those norms will at times depend on whether their moderation practices would be lawful if engaged in by a state.[10]

Fourth, the focus of commentators and regulators on the public-private distinction may tend to relieve platforms of normative obligations that are as important as, if not more important than, respect for their users' free expression, the importance of which the companies' own status as rights-bearers dilutes. In a scaled-up regulatory environment, in which relatively crude rules may be necessary, the norms of greatest interest arguably revolve around transparency, clarity, and reason-giving—what one might call "rule of law" or "due process" norms—rather than around free expression as such. The degree to which free expression is burdened on a platform depends intimately on substantive internal commitments that are almost entirely (and appropriately) within the operator's discretion, but both public and private platforms should be expected to keep their promises.

Turning to the merits, the legality of content moderation on a hypothetical government-run platform should not depend simply on whether regulation involves "content discrimination" or even "viewpoint discrimination." It should depend, rather, on the nature and degree of the burden imposed on speakers, the government's purposes in imposing those burdens, the risk of arbitrary or corrupt enforcement, and—relatedly—the clarity and consistency of the rules the government has put in place. It primarily should depend not on jejune questions of whether the government is making speech distinctions per se, but on whether it is doing so coherently and transparently. Norms around due process, transparency, legal clarity, and procedural regularity more readily translate between public and private forms of regulation than do norms around free expression.

The essay concludes with some brief and necessarily tentative observations on how we might think about some of these questions in the context of public platform regulation of nudity and sexual content, hate speech, and false speech.

First Amendment Doctrine

Users of social media platforms in the United States do not have a legal right to keep their content free of moderation.[11] It is commonly assumed that the reason for this immunity is that platforms are operated by private companies and therefore not subject to the First Amendment.[12] As private entities rather than states, they also are not directly subject to international human rights law. Indeed, social media companies are not only not subject to these legal regimes, but they or their principals might be rights-bearers under them. They have a right to exercise substantial editorial control over the content that appears on their services.[13] With respect to international human rights law, the UN Special Rapporteur on free expression has said that private platforms should regulate content according to human rights norms but has allowed that they may be free of certain obligations that relevant human rights instruments such as Article 19 of the International Covenant on Civil and Political Rights (ICCPR) impose on states.[14]

The assumption that the nonstate character of these platforms enables them freely to moderate content has been called into question. Most commentators who have looked at the problem have focused on the "private" end of the public-private divide, suggesting, for example, that platforms may lawfully be subjected to "must carry" rules or other forms of regulation because of their degree of market power,[15] because they are what Jack Balkin has called "information fiduciaries,"[16] or because they are akin to common carriers or public accommodations.[17]

The other end of the public-private divide has gone less explored but is at least as interesting. Whether or not private companies should sometimes be treated as if they were public, are there circumstances under which public speech regulation should be treated in much the same way it would be if it were private? Thus, if the government operated a social media platform, what limits would the First Amendment or international human rights law impose on its content moderation practices?

Public Forum Doctrine

This hypothetical implicates several different strands of US free speech law. Perhaps most obvious is so-called public forum doctrine. The doctrine arises out of the recognition that the government should be able to place certain kinds of limitations on speech that takes place on government property. Public forum doctrine is notoriously Byzantine. Over the years, at least four categories of public forum have been identified—traditional public forums, designated public

forums, limited public forums, and nonpublic forums. It is worth lingering on which of these categories might apply to a state-operated social media platform, and what the implications of that placement would be.

As it turns out, only the second and third of these categories—designated and limited public forums—could plausibly apply.[18] A designated public forum is "property that the State has opened for expressive activity by all or part of the public."[19] A limited public forum is one that has been reserved "for certain groups or for the discussion of certain topics."[20] Within a designated public forum, the government may not discriminate based on either the content of speech or the viewpoint it expresses.[21] The government may discriminate against certain forms of content in a limited public forum so long as the distinctions it makes are "reasonable in light of the purpose served by the forum," but it may not discriminate on the basis of viewpoint.[22] Think here of a school building that is open for public meetings of particular groups but not for the public's general use, and only at particular times. The state can distinguish between Alcoholics Anonymous (AA) meetings and poker night, but it cannot permit only AA meetings run by Democrats.

A state-run social media platform would be a designated public forum almost by stipulation—it would have been opened up specifically for the purpose of free expression by members of the public. Because content and viewpoint discrimination are forbidden on a designated public forum unless strict scrutiny is satisfied, terms of service that banned pornography or commercial speech or hate speech would need to be narrowly tailored to serve compelling state interests.[23]

Perhaps this test can be satisfied, but the Supreme Court's precedents are not encouraging. The Court in 1978 allowed the federal government to prohibit certain instances of indecent language on broadcast television and radio, but that holding was driven by concerns particular to the media and technological environment of a bygone era.[24] Key to the Court's decision was the difficulty users would have avoiding offensive programming or shielding children from it.[25] Social media users of today are not captive audiences. They have vastly more media options than television viewers and radio listeners did in the 1970s.[26]

More significantly, in cases like *Pacifica*, which allowed the FCC to ban indecent broadcast programming, and *Red Lion*, which upheld the "fairness doctrine," the Court did not find that the government's actions satisfied strict scrutiny. In each case, the Court took the scarcity of frequencies along the broadcast spectrum to be a reason to apply less stringent review to regulation of broadcast content.[27] The privilege of holding a broadcast license meant that the government could impose certain obligations on broadcasters that it could not impose on the general public.[28] It is unlikely that a modern court would view the millions of essentially nonrivalrous users of a public social media platform

as akin to broadcast licensees upon whom the government may impose duties of fairness or decency.

More apt than *Pacifica* and *Red Lion* might be cases holding that a city could not prohibit nudity at drive-in movie theaters or treat wearing "Fuck the Draft" on a jacket in a courthouse as a criminal offense or prevent a publicly leased theater from airing a production of *Hair*.[29] As Justice Lewis Powell wrote in the drive-in case, "when the government, acting as censor, undertakes selectively to shield the public from some kinds of speech on the ground that they are more offensive than others," its restrictions are permitted "only when the speaker intrudes on the privacy of the home, or the degree of captivity makes it impractical for the unwilling viewer or auditor to avoid exposure."[30] These exceptions do not generally apply in the social media context. On a government platform, it is likely the state would need to make available technological options to enable children or other sensitive users to be shielded before being permitted to censor content entirely.[31]

Although my hypothetical state-run social media platform most naturally lends itself to the designated public forum category, it might perhaps be argued that it instead involves a limited public forum. My hypothetical government would not in fact have established a free-for-all. Rather, by hypothesis, it would be trying through its terms of service to create something like a "usable, respectful, family-friendly" space. Spammers, trolls, and smut merchants need not apply.

A court might reach out for this solution if it wished to uphold certain kinds of content restrictions, but it would be a stretch. The limited public forum category has long attracted charges of circularity, as it enables the government to define the forum in terms of the kinds of limits it wishes to impose.[32] But it would be even worse in the case of nudity or hate speech because the limits would not be based on particular subject matter but rather on whether speech is especially colorful or indelicate. Decorum limits are sometimes permitted in a limited public forum, such as a town hall meeting, but only when decorum is germane to the subject matter limitation.[33] To say a user may speak about literally anything, but only with a degree of modesty, is not the kind of limit to which the limited public forum category has traditionally applied.[34]

Moreover, even if the government could argue successfully that a social media platform it operates is a limited rather than a designated public forum, it still would not be permitted to discriminate on the basis of viewpoint. That mandate seems easy enough to justify and apply in the context of partisan distinctions, but viewpoint discrimination is not so limited. Twitter's Rules and Policies forbid "targeting individuals with repeated slurs, tropes or other content that intends to dehumanize[,] degrade or reinforce negative or harmful stereotypes about a protected category," including race, sex, and gender identity.[35] Facebook's

Community Standards prohibit "dehumanizing speech, harmful stereotypes, expressions of contempt, disgust or dismissal, cursing, and calls for exclusion or segregation" on the basis of protected characteristics.[36] Both of these policies constitute viewpoint discrimination within the meaning of the Supreme Court's case law.[37] A user may, for example, call for the inclusion of a racial group but not for its exclusion. The hate speech rules of these platforms do what (it seems) the state may not: "license one side of a debate to fight freestyle, while requiring the other to follow Marquis of Queensberry rules."[38]

Government Speech

There is, however, a second strand of First Amendment doctrine that, if it applies, would be more promising to a government wishing to impose rules of engagement on a public social media platform. In *Walker v. Sons of Confederate Veterans*, the Supreme Court permitted the state of Texas to veto the messages drivers were permitted to display on state-issued license plates.[39] The Texas motor vehicle division rejected an application by the Sons of Confederate Veterans to offer its organizational logo, which included an image of a Confederate battle flag, for inclusion on one of the available license plate options. There is no question that Texas rejected the application because it disapproved of the viewpoint the flag represented, or perhaps worried that others would take offense, but a five-to-four Court majority permitted this judgment as a form of government speech.[40]

Government could not function if public officials could not express and promote its official views.[41] The Supreme Court has taken this truism a long way, holding over the course of several cases that the First Amendment imposes no limitation on what the government may say.[42] The austerity of US First Amendment law in other domains makes "government speech" an attractive safe harbor, even when its fit is as awkward as it was in *Walker*. As Justice Alito noted in dissent, the state had permitted license plates to include messages such as "Rather Be Golfing" and those promoting the Oklahoma Sooners or celebrating the NASCAR driver Jeff Gordon. To say that these were instances of the state of Texas adopting this speech as its own seems counterintuitive at best. Still, the Court's holding might reflect some discomfort at putting the government to a choice between either allowing its license plates to bear emblems of racial terrorism or denying individuals and organizations the freedom to propose license plate designs.

Walker might mark the outer limit of government speech doctrine. License plates are state-issued and state-branded forms of vehicle identification that the state requires motorists to display.[43] Speech on a hypothetical government-run social media platform, if analogous to speech on currently existing platforms,

would not carry the state's imprimatur in nearly the same way. That said, the fact that the *Walker* Court reached for the government speech lifeline in a case involving private choices about speech hints at the shortcomings of public forum doctrine that I discuss further in the section "In Defense of Public Content Moderation."

Unconstitutional Conditions

Before getting there, another line of doctrine that might be relevant to the hypothetical case of content regulation on a public social media platform is worth mentioning. The government doles out lots of money, often attaching conditions to the receipt of those funds. Sometimes those conditions relate to speech. Thus, in *National Endowment for the Arts v. Finley*, the Supreme Court upheld the statutory standards by which the NEA chooses funding recipients even though those standards required the NEA to consider "decency and respect" for the diverse views of the American people in evaluating candidates for funding.[44] And in *Rust v. Sullivan*, the Court upheld a Department of Health and Human Services regulation that forbid physicians receiving federal family planning funds from counseling, referring, or providing information regarding abortion.[45] These cases suggest that there might be some leeway for the government to restrict speech when it offers certain benefits that would otherwise be unavailable.

It is not likely, though, that speakers on a public social media platform would fall into this category. Notwithstanding cases like *Finley* and *Rust*, the presumption is that the government may not condition benefits on the relinquishment of constitutional rights. In cases involving, for example, government benefits such as tax exemptions, welfare, and unemployment insurance,[46] as well as those involving public employment,[47] the Supreme Court has repeatedly said that "even though a person has no 'right' to a valuable governmental benefit and even though the government may deny him the benefit for any number of reasons, . . . [i]t may not deny a benefit to a person on a basis that infringes his constitutionally protected interests—especially, his interest in freedom of speech."[48]

The unconstitutional conditions line of cases can be difficult to navigate, but the decisions make it clear that the bare fact that someone is receiving a benefit does not license the government to curtail their speech rights. The *Finley* Court noted that the esthetic judgments the NEA makes as part of its mandate required it to engage in content discrimination relevant to those judgments, and it construed the relevant statute as not compelling viewpoint discrimination.[49] And although the Court upheld the condition in *Rust*, it did not explicitly endorse viewpoint discrimination as a federal funding condition. Rather, Chief Justice Rehnquist wrote for the majority that the regulatory ban on counseling

and referrals was designed to ensure that the statutory ban on federal funding of abortion was honored.[50]

Rust and *Finley* suggest not that the conferral of benefits licenses the abridgment of rights *tout court* but that insofar as the government itself is speaking through the subsidies it offers, it may restrict speech instrumentally to ensure that its own views are not undermined.[51] This doctrinal proposition is consistent with public employment cases such as *Garcetti v. Ceballos*, in which the Court held that the ordinary rule that the government may not curtail speech as a condition of public employment did not apply to employees who speak while acting within their official duties.[52] The relationship between social media platforms and their users is not nearly as intimate as the relationship between the NEA and its funding recipients or public employers and their employees acting in their official capacities. As the public forum cases make clear, the fact that the government makes available a platform is not an invitation to the government to curtail individual constitutional rights in support of its own objectives.

International Human Rights Law

The First Amendment is not the only legal regime that would be relevant to public content moderation. International human rights law also constrains state behavior in the free expression context. It is important to engage with this question both because the United States has international human rights obligations and because private social media companies themselves have committed through the UNGP to accepting the guidance of human rights norms.

Freedom of expression enjoys strong protection under international human rights law.[53] The Universal Declaration of Human Rights speaks of the "freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers."[54] Article 19 of the ICCPR reiterates and builds on this language. It adds a specific reference to freedom of expression and clarifies that the ideas it protects are "of all kinds, regardless of frontiers, either orally, in writing or print, in the form of art, or through any other media" an individual chooses.[55] The ICCPR acknowledges that the right to free expression is not absolute, but limits on expression must be (1) "established by law" and (2) "necessary" either "[f]or respect of the rights or reputations of others" or "[f]or the protection of national security or of public order, or of public health or morals."[56] The UN Human Rights Committee has interpreted this language as requiring that state limits on expression meet a three-part test of legality, necessity and proportionality, and legitimacy.[57] The United States is a signatory to the ICCPR and did not attach a reservation to

Article 19 on the prevailing assumption that First Amendment law would be more speech-protective than international human rights law.[58]

Because international human rights law lacks a consistent and coherent approach to public forums or government speech, however, it is not clear that this assumption is reliable. In particular, the "necessity" and "legitimacy" prongs of Article 19's three-part test conspire to require that a restriction on expression be the least restrictive means of protecting either the rights or reputations of others or national security, public health, or morals. The rights and reputations of others are understood in terms of the rights protected under international human rights law and do not include, for example, a right not to be offended.[59] "Public morals" is a mischievous phrase on its face, but according to the Committee, the government may not invoke it or any other legitimate ground for limitation of free expression unless it shows "in specific and individualized fashion the precise nature of the threat and the necessity and proportionality of the specific action taken."[60]

Without more, it seems difficult to justify a case like *Walker* under Article 19 principles. Display of the Confederate flag is protected expression. Neither a general concern that it might offend nor the state's own disapproval of the message it conveys qualifies as permissible grounds for limitation under Article 19(3). The Committee's General Comment 34, which offers the most comprehensive guidance on the application of Article 19, does not categorically distinguish speech restrictions on government property from first-order limits on speech, though it does note that "public order" might at times justify regulating "speech-making in a particular public place."[61] In articulating this basis for limiting speech, the Committee referred to an Australian case in which it determined that a permitting requirement for speech at a mall was allowed, though disproportionate in the particular case, which involved political speech.[62] That case, *Coleman v. Australia*, implicitly blessed what US lawyers would call "time, place, and manner" restrictions, not those (as in *Walker*) that were based on content or viewpoint.

On rare occasions, the Committee has specifically endorsed some public forum–like principles, as when in *Zündel v. Canada* it concluded that a Holocaust denier could be denied access to a public press conference site because Article 19 "does not amount to an unfettered right of any individual or group to hold press conferences within the Parliamentary precincts, or to have such press conferences broadcast by others."[63] According to the Committee, even though the man had properly booked the space before being denied access by Parliament, it was sufficient that he "remained at liberty to hold a press conference elsewhere," such as on the street outside the building.[64]

This holding has come in for significant criticism, including from Michael O'Flaherty, the rapporteur for General Comment 34.[65] The Committee's

omission from the General Comment of the principle *Zündel* seems to memorialize was deliberate.[66] *Zündel* suggests a thick border between rights and privileges, recalling Oliver Wendell Holmes's famous quip that "[a] policeman may have a constitutional right to talk politics, but he has no constitutional right to be a policeman."[67] This glib categorical segregation of rights from privileges runs contrary to the animating spirit of proportionality analysis, including as it is more typically used in the Article 19 context.[68] Still, the fact that one is speaking in a state-provided space (and, indeed, that one is a police officer)[69] should be relevant to the scope of permissible free expression limits. Exactly how relevant it should be has gone underexplored in international human rights law, focused as that legal regime has been on state penalties or coercion rather than on access to public spaces.[70]

Where relevant human rights opinions have trained on online speech, they have understandably focused less on putative public platforms than on public regulation of private ones or on the companies' own obligations under international human rights law.[71] In 2018, the Special Rapporteur on freedom of opinion and expression issued a report on public regulation of user-generated online content.[72] That report reiterates Article 19's restrictions on state behavior without offering any systematic analysis of how a state's regulatory interest might vary between online moderation and other exercises of public power.[73] This may be in part because the mine run of government exercises of regulatory power over social media content does not involve direct regulation of a public platform but rather the very different issue of applying civil or criminal sanctions to companies (or their users) that do not play by the government's rules.[74] The Special Rapporteur's report focuses on state laws or enforcement actions that police the entire country rather than just a single platform.

As to the companies themselves, the scope the Special Rapporteur's report gives for them to regulate content on their platforms derives primarily from their private status and the scale of the content they host.[75] The report expresses concern about using "[p]rivate norms [that] vary according to each company's business model" as guidance to how they intend to regulate.[76] The human rights of companies or their directors to moderate content according to those private norms receive no attention at all.[77]

In Defense of Public Content Moderation

Above, I have sketched a domestic and international rights regime that prohibits the government from discriminating among social media content based on its form or the viewpoint it expresses except under compelling and specific circumstances. This result will seem obvious and quite right to some readers,

particularly those trained within the US free speech tradition. Public content moderation on social media smacks of government censorship, indeed of prior restraints, traditionally considered the most brazenly illegal form of content restriction.[78]

And yet the sketch produces an anomaly that requires unpacking. Recall that government speech generally does not implicate First Amendment concerns. And government institutions that explicitly exercise some editorial judgment or other forms of discretion are also permitted to incorporate standards of decency into those judgments. Few would question that a state-run television station may decide not to air programming with nudity, or that a state-run university may refuse to admit students or hire faculty whose writings reveal them to be white supremacists. The government may proselytize without restriction and may selectively open spaces to relatively small numbers of citizens or groups over whom it exercises significant control, but it may not open the same spaces to expression by members of the broader public unless it does so indiscriminately. That is, the less ex ante control the government exerts over speech in a public space, the more tightly the First Amendment restricts its ex post behavior.

An inverse relationship between government domination and constitutional limitation runs contrary to many of our usual instincts, and it is unclear what free speech interests it serves in this context. One possibility is that it prevents the government from discriminating against unpopular speakers or those who would challenge its authority. What space the government makes available to its supporters it must also make available to its opponents. The threat of invidious discrimination of this sort provides good reason to scrutinize the rules the government applies to public platforms, and a fortiori how it applies those rules in practice. But the categories of "content discrimination" and "viewpoint discrimination" make for crude constitutional rules. The government's interest in its platform's discursive norms is itself a viewpoint. That interest would not suddenly vanish just because the government decided to permit the public to access a platform in large numbers.

The discontinuity between government speech doctrine and public forum doctrine encourages the government to protect its interests by exerting influence over forums, lest it be forced to forego those interests entirely. Consider again the Court's decision in the *Sons of Confederate Veterans* license plate case. Had the Court held, as Justice Alito urged in dissent, that Texas had made its license plates into even a nonpublic forum (and, a fortiori, into a limited or designated public forum), existing precedent would have compelled the Court to say that the price of the state's specialized license scheme is the need to permit a Confederate flag logo on state-issued license plates. By the same reasoning, the state would also have needed to permit a swastika. And yet the same First Amendment doctrine enabled the state to avoid these obligations simply by

persuading the Court that the license plates represented government speech, on which the First Amendment imposes almost no restrictions at all.[79]

Cases such as *Walker* speak to the hydraulics of US free speech doctrine. The idea that the government may not open a public space to Democrats but not Republicans, or to artistic but not political speech, seems right to many of us, but the law responds to these intuitions by constructing overly broad, overly rigid categories such as "viewpoint" and "content" discrimination. Likewise, the idea that the government itself may not speak through its public institutions seems wrong to many of us, so the law responds to this intuition by creating an unassailable category of "government speech." As in many other areas of law,[80] we come to view complex disputes over free expression as *about* the categories the law constructs instead of being about the particularized interests on either side of the dispute. When the categories don't fit the facts, judges improvise.

Government speech and public forum doctrine will continue to produce perverse results—and cannot coherently be applied to the social media space—unless and until the intuitions they rely upon are surfaced and integrated into a single regime. Consider prior restraints. American law disfavors prior restraints in the free expression realm for separation-of-powers reasons. The idea is that decisions about whether speech should see the light of day should be made by judges and juries rather than by politicians. But the fact that we also believe that the government should have some say, even on the margins, over what content is displayed on public platforms (such as Texas's license plates) points in the opposite direction. The existence of the limited public forum category speaks to the legal system's ambivalence about prior restraints in contexts in which the government is voluntarily offering a space for public expression.

Prior restraints indeed have the advantage of being susceptible to consistent application, which is the most vexing problem for content regulation at scale. One can evaluate the rules a prior restraint comprises without reference to individual cases, thereby reducing the potential for biased application. Particularized decision-making is not possible at internet scale, leaving a platform's regulators to choose between rules that will necessarily be overbroad or underenforcement that can make the platform unusable or even dangerous. It is unrealistic to suppose that forbidding content-based rules would make a government choose an unregulated platform rather than choosing not to offer the platform at all, and so enforcing conventional designated public forum rules would discourage even the creation of forums that would host the vast majority of speech without restriction.

Perhaps of greater significance, anxiety over abusive or coercive government restrictions on speech are most compelling when those restrictions are in fact abusive or coercive. We should be less exercised by content restrictions

in a context in which the "penalty" is simply that the restricted content does not appear on the platform, with no sanction against the user, and when nearly the exact same content can be reposted without the nipple, or the slur, or the terrorist hashtag. This burden is not in the same universe as being arrested, imprisoned, tortured, fined, or intimidated by government agents, the usual interest of the international human rights regime. To apply the same doctrine, or anything like it, to penalties on the speaker as one does to the application of rules to the speech is difficult to justify.[81] This is not to say that content moderation on a public platform should be treated like government speech, with no constitutional guardrails at all, but it is to say that the reasons we give license to government speech should inform those guardrails even when (again, as in the license plate case) they are intermingled with the speech rights of individuals.

An unregulated public platform might even violate various human rights norms by dint of its *lack* of regulation. Article 20 of the ICCPR imposes upon states a duty to prohibit "any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence."[82] A public platform therefore would need to disallow at least a subset of hate speech. More broadly, as the Supreme Court wrote in *Red Lion*, an unregulated broadcast medium "would be of little use because of the cacophony of competing voices, none of which could be clearly and predictably heard."[83] The social media context, with its problems of scale intersecting with the omnipresence of malevolent actors, can produce darker user experiences than cacophony.[84] In the international sphere, the Human Rights Committee has suggested that Article 19 might impose duties on the state to make platforms accessible to users.[85] But an unruly, indecent, or abusive online environment stands as a practical obstacle to the full range of users having access to social media platforms.

Requiring that platforms tolerate such an environment implicates their rights as well. The Special Rapporteur's downplaying of the "private norms" of social media companies in his 2018 report on online content moderation seems insufficiently attentive to the rights of these companies and their directors to control the character of the communicative services they offer.[86] The notion that the free speech practices of companies should be homogenized in accordance with international human rights norms seems to suppose that the expressive interests of Facebook and Parler, of YouTube and Reddit, must align. Private companies should be permitted to espouse very different values and free expression commitments without having to justify those commitments in terms of what human rights law specially protects. So too with governments, to a degree, but strict application of Article 19's limitations language seems not to permit this kind of variation, arising as it does out of a "government speech" principle that has no well-articulated analogue in international human rights law.[87]

Content Moderation as a Process Problem

A state's own expressive and policy interests complicate analysis of the free speech rights users of a public platform would enjoy. By hypothesis, the state may withdraw the platform entirely at its discretion or may apply restrictive ex ante rules that create a limited rather than a designated public forum, either of which could reduce the amount and diversity of speech available to users and their audiences. This license does not mean free speech norms should have no role to play in curtailing the government's regulatory reach over a public platform—the greater power does not always include the lesser.[88] But it does mean that the kinds of interests these norms protect would differ between on- and off-platform regulation. The government's competing ends would dilute any unqualified interest in "more speech."

It does not, however, dilute a different concern that sometimes shows up in First Amendment law, namely the interest in avoiding arbitrary decision-making. Free speech is not just about the volume of speech individuals are permitted to produce but about who decides, through what process, and according to what standards.

This concern surfaces most directly in vagueness doctrine. Vagueness is a problem in many areas of law,[89] but the Supreme Court has found it especially concerning in connection with laws restricting expressive activities, which can be "chilled" if citizens do not have a firm sense of what is permitted and what is not.

The Court's special interest in consistent application of legal regulations of speech pops up in other realms as well. As we have seen, limited public forum doctrine requires any content-based restrictions the government puts in place in such forums to track "the lawful boundaries it itself has set."[90] Even though the State is not required to create a limited public forum in the first instance, once it does so, it may not adjust the forum's borders on the fly. Likewise, prior restraints on speech have been thought most problematic when the determinations of legality are placed before a licensing board with standardless discretion.[91]

Even if certain forms of content restriction are permissible under the First Amendment, then, a regulatory scheme susceptible to arbitrary and capricious enforcement could nonetheless violate the Constitution. International human rights law recognizes similar principles. General Comment 34 specifies that, quite apart from the substantive requirements of Article 19, a law may not "confer unfettered discretion for the restriction of freedom of expression on those charged with its execution" and must "provide sufficient guidance" for legal decision makers to know when expression may or may not properly be limited.[92]

These commitments, which may at times run orthogonal to substantive free expression norms, carry important implications for which kinds of speech restrictions would and should be permitted on a public social media platform and why. It might mean, for example, that in the platform context, rules are often preferred to standards. Rules typically sacrifice precision in application of the values the law protects in order to achieve consistency and predictability. Standards have inverse costs and benefits. They offer the possibility of getting the decision right by the lights of the motivating legal norm but at the cost of piecemeal and potentially ad hoc adjudication. If what the legal system values above all is making sure that protected speech is always permitted, then rigid rules are intolerable. If, on the other hand, what the legal system values above all is consistency (or even the appearance of consistency), then content moderation through relatively rigid rules has much to recommend it. Optimizing free expression regulation for consistency reduces the benefit of the particularized decision-making often associated with proportionality analysis.

Relatedly, to the degree consistency is a leading First Amendment value in a particular setting, the usual presumption US lawyers apply against prior restraints requires a revisit. Prior restraints are rightly disfavored under a presumption that speech "getting out" is paramount, or if the restraint is enforced through arbitrary standards.[93] But if the most important First Amendment value relates to the discretion of decision makers, there may be good reasons to *prefer* prior restraints.[94] With a prior restraint, the standards associated with the exercise of discretion may be evaluated in advance of any enforcement decisions, and risks of bias can be significantly mitigated.[95]

Recognition of the significance of consistent rules and procedural regularity lends some insight to what I presented in the section "In Defense of Public Content Moderation" as a paradox, namely that ex ante government control and domination of a platform tends to relax rather than strengthen the constitutional constraints that attach to government restrictions on private speech on the platform. If we examine the problem along the dimension of consistency rather than control, the paradox weakens. Greater government control of a platform enables the government to police its rules with greater regularity. The more a platform looks like a free-for-all, the more arbitrary, and therefore suspicious, government restrictions seem to be. The state cannot appear to be making up the rules as it goes along. The key point, though, is that, *in the name of consistency*, the rules that are permitted may not always be speech-maximizing. In the social media context in particular, in light of the scale on which platforms operate, rules designed to maximize consistency in application are unlikely to maximize speech.[96]

The suggestion here is not that a government that operated a social media platform would have unfettered discretion to moderate content so long as it did so according to consistent rules. It is, rather, that the government's own legitimate interest in structuring the platform to ensure its effectiveness should carry some weight—neither absolute weight nor trivial weight—in any legal analysis. As with many other exercises of state power, platform moderation rules motivated by partisanship, caprice, personal self-dealing, bias against particular social groups, vindictiveness, or other markers of distrust, as well as those that silence speech in ways that are not tied to the government's goals for the platform, should not be entitled to deference.

Applications

I have argued above for a normative shift in attention from the volume of speech a platform permits to the coherence and consistency of the rules that govern which speech it permits and which it does not. This shift is consistent with and, indeed, compelled by free expression values as applied to platforms, whether those platforms are operated by the government or by a private company. It is not my aim to apply this analysis to particular companies or conflicts, but I offer some brief observations below about how some of the relevant considerations vary with the type of speech at issue, focusing on nudity and sexual content, hate speech, and false speech.

Nudity and Sexual Content

Most Americans likely agree that the government may limit a platform it owns or operates to nonpornographic content. But it turns out to be somewhat difficult to specify why either the First Amendment or international human rights law would permit this form of regulation, at least as it pertains to a platform that is generally open to public discourse. Sexually explicit images short of obscenity, whose definition is demanding,[97] constitute protected expression whose limitation is ostensibly forbidden on a designated public forum.[98] Content discrimination such as a restriction on nudity or sexual content is permitted within a limited public forum, but as the section "First Amendment Doctrine" explains, a conventional social media platform would not easily qualify.

Nudity and sexual content involve what Article 19 calls "public morals," a legitimate reason for limitation of speech.[99] Although the Human Rights Committee has not as of this writing decided cases that help to flesh out how this limit applies, the European Court of Human Rights has held that certain

forms of sexually explicit artistic expression could be outright penalized without offending Article 10 of the European Convention on Human Rights (ECHR), the analogue to ICCPR Article 19.[100] An important difference between the ECHR and the ICCPR is that the former has been read to allow for a margin of appreciation, which the Human Rights Committee has rejected.[101] Given that there is no margin of appreciation doctrine under Article 19, it is far from clear that "public morals" is sufficient ground to restrict nude expression as a matter of first-order punishment.

Platform restrictions on nudity and sexual content are well suited, however, to the framework this essay articulates. Bans on "adult" content can be controversial, of course, but the fact that they are common on platforms around the world suggests that, in most instances, they are motivated by a loose consensus around public morality, especially in relation to children. The fairest way to be sure of this is to permit such bans to operate according to relatively rigid—and necessarily overbroad—rules. We can concede that rules against the display of explicit sex acts or images of sex organs cannot satisfy a particularized form of strict scrutiny or proportionality analysis. Warning screens, age gates, or parental controls might be less restrictive means of limiting the visibility of such content, and not all instances of nudity or sexual content necessarily implicate the moral concerns that motivate a rule against such content. It is conceivable that an exemption for matters of high educational or political importance would need to be in place. Still, commonsense limitations on nudity or sexual content likely should be permitted on a public platform whose operators wish not to offend the moral sensibilities of its users, even if such limitations could not be imposed *erga omnes* or applied to primary conduct.

Hate Speech

Hate speech can be hard to define—it is not a recognized term either under the First Amendment or under international human rights law—and it can sometimes be difficult to identify in the wild, especially at scale. Platform bans on hate speech also might come closer than bans on nudity or sex to limiting core political expression. For these reasons, entrusting the state with the power to decide which forms of hate speech may appear on a public platform and which may not would carry significant risks.

Still, the fact that defining hate speech occasions difficult judgment calls is not a reason to give up on its regulation on platforms, nor is it sufficient reason to ignore the legitimate interests a government platform operator would have in reducing racist, sexist, or homophobic invective on the platform. As noted, governments in fact have an obligation under international human rights law to

make media and other communication platforms accessible. Fronting a public platform loaded with ethnic slurs, rape jokes, and swastikas would seem to sit in serious tension with that responsibility.

This is where it becomes important to focus less on unhelpful slogans such as "viewpoint discrimination" and more on the safeguards that a platform has in place to ensure that viewpoint discrimination is serving legitimate ends. Such safeguards might include well-publicized and specific rules of engagement based on consultation with stakeholders across partisan divides, the creation of such rules by an entity structured to resist political control (e.g., through a bipartisan or merit-based selection and removal process), and an appeal procedure that is sufficiently transparent to enable public accountability.

The premise, of course, is that viewpoint discrimination is not, per se, illegitimate for a government to pursue. It is reasonable to resist that premise in the context of civil or criminal liability. It would be unreasonable to do so when it comes to the rules of engagement on a public platform, where the only "penalty" is removal of an offending post. The government is permitted to have an interest in civil discourse and to pursue that interest by promulgating rules, such as banning certain slurs. Telescoping out to broader definitions of hate speech—such as by generally banning racist or dehumanizing language—brings with it greater uncertainty in application.

False Speech

Devising standards through which a public platform should be permitted to police false speech presents even greater challenges. For one thing, lying is constitutionally protected speech in the absence of specific harm.[102] And it will often be the case, and not just on the margins, that determining truth or falsity is challenging. Moreover, this determination will in many cases involve speech about the government itself, the regulation of which we have good reason to keep out of the government's hands.

The difficulty here is both epistemic and conceptual. It is impossible for any platform operator to amass the resources needed to reliably fact-check content at scale. Even if one could somehow solve this problem, the reality of human communication is that truth and falsity function more as a spectrum than a binary, with exaggerated or misleading claims of every stripe and in every degree sitting between the poles. Consistent application would be elusive, in just the area where consistent application is most needed to shore up trust in the decision maker.

For these reasons, the interventions we tolerate in this space might be more limited than in the areas of nudity or hate speech. A government could

perhaps explore targeting disinformation, which is intentional and often coordinated, but not necessarily misinformation, which is endemic and unavoidable. An intent and coordination requirement would reduce the "chilling effect" on true speech, unintentional falsehoods, and benign puffery. Another option for lowering the cost of policing false speech would be to impose less drastic remedies than removal, such as downranking or the application of labels or links to ostensibly accurate information, which several private platforms already do in some instances. The Supreme Court imposed the very high "actual malice" standard for a finding of defamation in the famous *New York Times v. Sullivan* case in part because liability involved treble damages and, during the civil rights movement, was imposed by biased southern juries trying to harass civil rights activists.[103] The much lighter consequence of having a label affixed to one's false content on a public platform should get us far less exercised, even if it doesn't leave us unbothered entirely.

Conclusion

Determining how to apply freedom of expression principles developed in the context of first-order lawmaking, law enforcement, and personal liability at the hands of states to private systems of governance on social media platforms involves multiple acts of translation. The analysis cannot simply involve putting the private platform into the shoes of a state regulator of primary conduct. Much commentary has recognized that private platform moderation is distinct insofar as it is private. Commentators have focused less on the ways it is distinct because it involves platforms. Ignoring the ways in which the law distinctly applies to platforms even with respect to public actors risks developing a body of rules that impose *greater* constraints on private companies than they would impose on states acting as market participants.[104]

This essay has analyzed a hypothetical public platform in order to develop some intuitions along these lines. The law of platforms under the US Constitution is mature, but it suffers from some of the paradoxes of the Supreme Court's categorical approach to free speech law and is not well suited to the scale of social media. International human rights law generally is less categorical but has yet to develop a mature jurisprudence around public forums.

The suggestion of the essay has been that it will be more constructive to build a doctrinal architecture around procedural constraints on moderation decisions that ensure consistent application of forum rules than to think of platform governance simply in terms of the quantity of speech available to users. The government's competing interest in the quality of the speech that occurs on the platforms it operates, the risks associated with the proliferation of certain kinds

of expression at internet speed and scale, and the accessibility of platforms to users who find themselves on the business end of harmful content can neither be ignored nor considered an evergreen "trump" of free expression. The right balance will not be the same across all platforms, but we can nonetheless strive for consistency within them.

11

The Limits of Antidiscrimination Law in the Digital Public Sphere

GENEVIEVE LAKIER

Cries of private censorship abound in the digital age. For years now, those on both the left and the right have complained that their voices are being silenced by the private companies that control the social media platforms that provide some of the most important forums for public speech in contemporary society.

What those who decry social media "censorship" mean when they use the term varies. Some critics use the term to accuse the social media companies of applying the rules that govern speech on their platforms inconsistently, to the disadvantage of certain viewpoints or speakers.[1] Others use the term to accuse the companies of applying speech rules that have a disparate impact on the ability of certain groups to communicate.[2] Others use the term to refer generally to the power that private companies possess to decide what counts as acceptable or unacceptable speech on social media.[3]

What unites all these complaints of platform censorship is the fear that underpins them: namely, that the private and almost exclusively for-profit companies that control the platforms may use that control to distort public debate and to deny equal access to the social, political, and economic goods that the platforms provide.[4] This is a fear that cannot be easily disregarded.

While the lack of transparency about the content moderation practices of the social media companies makes it hard to reach general conclusions about whether and to what extent political bias, or other kinds of bias, influences their operation, the broad discretion that these companies currently enjoy to regulate the speech that flows through their platforms makes it entirely possible that they might, or already do, discriminate against certain viewpoints or speakers, either because of their ideological convictions or because it suits their economic or political interests to do so. And there is no question that individual speakers are denied access to the platforms all the time for reasons that are hard

to fathom—although whether that is a consequence of algorithmic or human error (the sheer difficulties of making speech decisions at scale), or some more invidious reason, is not always clear.[5]

Anxiety about what might be motivating the platforms to take down speech, to deprioritize it, or add warnings to it has led lawmakers in the United States to propose—and in some jurisdictions, enact—laws that make it unlawful for the social media companies to discriminate against their users in various ways. In Florida, for example, the legislature recently enacted a law that makes it unlawful for social media companies to suspend the account of any candidate for political office in the period prior to election, and more affirmatively requires all social media platforms that operate in the state to apply their speech rules "in a consistent manner among its users on the platform."[6] In Texas, the legislature recently enacted a law that makes it unlawful to limit the dissemination of their users' speech because of their viewpoint, their message, or their geographic location.[7] Antidiscrimination platform bills have been introduced in Congress as well. For example, Rep. David Cicilline (D-R.I.) has introduced legislation that would prohibit social media platforms from treating similarly situated "business users" of their platforms differently, except when they can show that doing so will not "harm the competitive process" by "restricting or impeding legitimate activity" or is necessary to prevent the violation of a state or federal law or to protect user privacy.[8] Other proposed federal laws condition the platform's immunity from liability on their compliance with nondiscrimination obligations.[9] Meanwhile, scholars and policymakers have argued that the same stringent nondiscrimination duties that federal and state laws impose on common carriers or places of public accommodation should apply to the social media companies.[10]

That legislators and scholars have turned to antidiscrimination law as a solution to the problem of private censorship on the social media platforms is not surprising. Antidiscrimination law has historically provided one of the primary regulatory tools that lawmakers in the United States and elsewhere have used to ensure that the private companies that control important forums of mass public expression do not use that control to exert undue influence over either the political or the commercial realm.[11] In part because it does not require a significant reorganization of the digital economy—as the other tools that regulators have historically used to protect the democratic inclusiveness of the mass public sphere against private power do—antidiscrimination law remains an attractive tool for achieving those same ends today.[12]

Applying antidiscrimination rules to the social media platforms turns out to be quite a tricky business, however, for a number of reasons. First, laws that impose antidiscrimination obligations on the social media companies raise difficult constitutional questions because, unlike the other regulatory tools available (transparency mandates, breakups, and so on), they directly interfere with the

choices private companies make when they regulate speech on their platforms. They consequently implicate the stringent protection the First Amendment provides against compelled speech and compelled access.

Even if one leaves the First Amendment questions to one side, poorly drafted or overly expansive antidiscrimination laws may do more harm than good to the digital speech environment by making it *too* difficult for social media companies to discriminate. This is because discrimination, when it comes to speech, is not always a bad thing. To the contrary, it is the quotidian content discrimination that individuals engage in every day—choosing to read one newspaper over another, to express one viewpoint rather than another, or to associate oneself with one party rather than another—that powers the marketplace of ideas.[13] And while social media companies do not necessarily participate in the speech market in the same way that ordinary speakers and listeners do—instead they operate in many contexts more as regulators of the public sphere than as participants in it—for these "new governors" of the digital public sphere, being able to treat different users' posts differently because of their content serves many purposes: It allows them to direct users to speech that is most interesting to them; to protect users against threatening or harassing speech; and more generally, to exclude from the platform spam, misinformation, and other kinds of what is considered to be cluttering or harmful speech.[14]

The fact that it is by making content-discriminatory decisions that the social media companies are able to provide users with much of the information and connections they come to the platforms to find means that lawmakers who seek to use the tool of antidiscrimination law to make the digital public sphere better, not worse, face a daunting task: They must craft rules that not only can withstand First Amendment scrutiny but that do not undermine the value of the platforms as social spaces. Is it possible to craft such a law?

This essay argues that the answer to this question is yes: Yes, not only would it be *possible* to craft a constitutional nondiscrimination law for the platforms, but it would be normatively desirable for lawmakers to do so. But it also argues that such a law, if it is to be both constitutional and beneficial, rather than detrimental, to the vitality of the digital public sphere, will not be able to address most of what gets described as platform censorship these days. In particular, antidiscrimination law neither should, nor constitutionally will, be able to deny the platforms the ability to set speech rules that differ from those set by the First Amendment. As a result, it will not be able to prevent the social media companies from creating rules that have a disparate impact on the ability of some groups—conservatives, feminist activists, practitioners of New Age medicine—to disseminate their speech.

This is not to say that antidiscrimination law is useless as a mechanism for limiting platform power. What it can do is require the social media companies

to enforce their rules in a nondiscriminatory manner. It could also prevent platforms from discriminating against users because of their race, their gender, and their sexual orientation—and even, potentially, their political affiliation. These are not insignificant accomplishments. But they are not what most of those who advocate for antidiscrimination laws want those laws to accomplish: namely, prevent private social media companies from writing rules that shape the boundaries of public conversation in the United States.

The upshot is that, to the extent that platform power over public discourse is a legislative concern—as it surely should be—lawmakers will have to find tools from other reaches of law to cabin it. Antidiscrimination law can play an important role in the regulation of the digital public sphere. But it is not the be-all and end-all to the problem of social media censorship that some have suggested it might be.

The Janus-Faced Character of Content Discrimination

Legally mandated nondiscrimination obligations have been a part of the regulation of the public sphere in the United States for as long as the First Amendment has been around. Initially, they applied only to government carriers of information—specifically, the post office, which, from the late eighteenth century onward, was required to grant newspapers, and later magazines, equal access to the mails.[15] Beginning in the 1840s, however, states and the federal government began to impose similar nondiscrimination obligations on private companies—namely, the private telegraph companies that were displacing the government-run post office as the primary carrier of news and information around the body politic.[16] By the end of the nineteenth century, a number of states had passed laws that also required the news-gathering organizations that gathered and disseminated news on the telegraph wires to provide their services "to all newspapers, persons, companies or corporations . . . at uniform rates and without discrimination."[17] And similar, although narrower, nondiscrimination obligations were eventually extended to radio and television broadcasters, to cable companies, and newspapers. For example, under the Communications Act of 1934, broadcast radio and television providers were required to grant all "legally qualified candidates for . . . public office . . . equal opportunities to all other candidates . . . in the use of [their] station."[18] The FCC subsequently used its rulemaking authority to extend the same nondiscrimination requirement to cable television providers.[19] Newspapers, meanwhile, are prohibited by federal law from charging more to political advertisers than to commercial advertisers.[20] In some states, legislators also have prohibited newspapers from

charging some political candidates more than they charged others to advertise in their pages.[21]

As these examples demonstrate, antidiscrimination law has long worked to protect the inclusiveness of the mass public sphere against both governmental and private biases. Indeed, antidiscrimination laws represent one of the primary regulatory instruments that legislators have used to ensure that the powerful governmental and private entities that control important channels of mass communication in the United States do not use that control to exercise undue influence over the democratic political process or impede the efficient operation of the market.

What the history also shows is that, when designing nondiscrimination laws for the mass public sphere, lawmakers have recognized the Janus-faced nature of content discrimination and worked to ensure that nondiscrimination duties do not prevent either the government or private entities that disseminate speech to the public from engaging in the kinds of speech-related discrimination that, in the view of those writing the laws, at least, add value to the public sphere. This desire explains why the early postal laws not only permitted but mandated that postal agents discriminate in favor of newspapers, and against letter writers, when it came to the setting of postal rates.[22] Members of Congress believed that this discriminatory system of genre-based cross-subsidization was the only means of giving rural newspapers the postal subsidies they needed to survive—in other words, that price discrimination aided, rather than threatened, the vitality of the news marketplace, given the preeminent role that newspapers played in the circulation of news in the late eighteenth century.[23]

A desire to shield valuable kinds of discrimination from the reach of the nondiscrimination laws also explains why, when it first developed the regulatory framework for radio broadcasting in the United States, Congress rejected a proposal to designate radio broadcasters as communications common carriers.[24] Members of Congress believed, correctly enough, that requiring broadcasters to abide by the broad nondiscrimination duties imposed on telegraph and telephone companies—including the duty to transmit speech in the order it was received—would prevent those broadcasters from being able to provide the "regular entertainment" that "induced the public to listen in."[25] That is to say, it would prevent them from providing the programming content that, by the 1920s, it was widely assumed was the primary contribution of the radio to public life and culture.

When designing antidiscrimination laws for the social media platforms, lawmakers should show similar sensitivity to the two-sided character of discrimination as they showed when designing rules for the postal service and radio stations. This is because speech-related discrimination plays as important a role in the operation of the platforms as it plays in other reaches of the mass public, even

if the kinds of discrimination that the social media companies ordinarily employ differ. Unlike radio and television providers, social media companies do not usually attract users by offering them "regular programming." They therefore do not need to make the inevitably discriminatory choices that content providers have to make when deciding what stories or subjects or kinds of art to create or disseminate, or not to create or disseminate. And unlike carriers of information, social media companies typically do not incur the costs of creating new pathways of communication or transporting information to a destination selected by the user. They consequently do not need to, and indeed, do not, charge some users more than others to move their speech from point A to point B.

The social media companies do, however, discriminate between users and between individual speech acts in at least two other ways—and they do so, at least in part, in order to provide the service that users have come to expect of them. First, the platforms engage in a kind of concierge speech management that remains perhaps the most distinctive feature of the digital public sphere. Unlike carriers of information or, for that matter, the government or private entities that regulate the brick-and-mortar public sphere, the social media companies have from the very beginning of their existence not only provided a forum for public speech but played an active role in connecting participants in that forum to one another.[26] They help users make sense of the vast ocean of speech flowing through the platforms by recommending speech to them or ranking it, and by distinguishing the speech of "friends" from those of enemies; that is to say, by engaging in pervasive speaker- and content-based discrimination.

The social media companies also expend considerable labor and capital policing the boundaries of acceptable speech on their platforms.[27] They are constantly distinguishing, in other words, between speech deemed to comply with the platform-specific service rules and speech that does not. And, to the surprise of many of those involved in the creation and operation of the social media cites initially, these distinctions tend *not* to mirror the distinctions the First Amendment cases make between protected and unprotected speech.[28] Instead, they invariably prohibit more, or much more, speech than the First Amendment allows the government to prohibit when it regulates the brick-and-mortar public sphere.

As any contemporary reader of this essay will know, how much value these discriminatory practices add is today a highly contested question. But it is difficult to argue that the tendency of most social media companies to rank and sort speech and to sometimes aggressively gatekeep the boundaries of their speech forums adds *nothing* of value to the digital public sphere. Indeed, it may add a great deal. By constantly directing the flow of speech throughout the platform, social media companies are able to create new kinds of speech forums in which people feel willing to participate in public expression to a degree, and in ways,

that were previously neither possible nor expected. By excluding speech and harassment from the platforms, the social media companies also work to protect that "public order without which liberty itself would be lost in the excesses of unrestrained abuses."[29]

What this means is that, unless lawmakers are committed to a radical reinvention of the digital public sphere, they should not attempt, as some have suggested, to deny the social media companies all power to discriminate between different kinds of constitutionally protected speech. They should not impose on the social media companies nondiscrimination obligations, like those imposed on communications common carriers, that would prevent them from crafting speech rules that are more speech-restrictive than those the First Amendment applies. Doing so would prevent the companies from engaging in much of the speech management that users today expect of them.[30] It would also reduce the diversity of the digital public sphere by preventing social media companies from tailoring speech rules to the nature of the speech forum: from developing, for example, different speech rules for a speech forum designed to attract knitters or dog lovers than for one designed to host no-holds-barred political speech.

Even if lawmakers *wanted* to radically reinvent the platforms by prohibiting the social media companies from discriminating, they likely could not do so consistent with the First Amendment. This is because the First Amendment ordinarily allows the government to prohibit private speech intermediaries from discriminating against other persons' speech only when those intermediaries hold themselves out as "neutral, indiscriminate conduits" of third-party speech, or, alternatively, when the government employs a content-neutral law to limit the private discrimination and is able to satisfy the intermediate scrutiny that applies to content-neutral laws.[31] Neither of these conditions is likely to be satisfied by a law that extends common carrier duties to social media platforms.

First, it is quite obvious that the social media companies do *not* hold themselves out as neutral, indiscriminate conduits of speech. To the contrary, all of the major platforms publish on their websites detailed descriptions of the various content-based restrictions on what speech they allow on their platforms.[32] Nor is it just a matter of self-description. As noted above, the social media companies expend considerable time and money patrolling the boundaries of acceptable platform speech. They not only do not purport to be neutral, indiscriminate conduits of information. They really are *not* neutral, indiscriminate conduits of information.[33]

Second, a law that imposed common carrier duties on social media companies in order to prevent them from limiting the dissemination of disfavored speech would *not* be considered content-neutral under contemporary precedents, even if it made no content distinctions on its face. This is because such a law would be "designed to favor . . . speech of [a] particular content"—specifically, all of

the constitutionally protected speech that is prohibited, or limited, by the platform rules—and this is not a content-neutral purpose, as the Court made clear in *Turner Broadcasting v. FCC*.[34]

It is true, as Eugene Volokh and others have argued, that common carrier regulations would not necessarily be motivated by a desire to promote *particular* viewpoints, that they might reflect instead, or more fundamentally, a legislative desire to prevent the social media companies from "leveraging their economic power into control over public debate" by disfavoring *any* viewpoints or topics of speech.[35] But, as the campaign finance cases make clear, this also doesn't count as a content-neutral purpose because what underlies it is the assumption that there is something harmful about the content choices the social media companies are making.[36] To put it another way: If the government may not, when writing campaign finance laws, prohibit wealthy corporations from using their economic resources (specifically, their money) to promote particular political messages in an effort to prevent them from "obtaining an unfair advantage in the political marketplace"—as the Court in *Citizens United v. FCC* held it could not, consistent with the First Amendment—then surely the government also may not, when writing internet regulations, prohibit wealthy social media corporations from using their economic resources (specifically, their control over their platforms) to promote particular political messages in order to prevent them from obtaining an unfair advantage in the political marketplace.[37]

It is very likely, therefore, that, were Congress to enact a common carrier law for the social media platforms, its constitutionality would be evaluated under the demanding standard of strict scrutiny. This is true notwithstanding the fact that were Congress to apply the exact same law to a telephone company or an internet service provider, it would likely be reviewed under no more than rational basis scrutiny.

The same is true of more tailored nondiscrimination laws like the Florida and Texas laws. This is obviously the case when it comes to the Florida law, which makes numerous facial content-based distinctions.[38] But it is equally true of the Texas law, which although facially content-neutral, was clearly intended to prevent the social media corporations from wielding their economic power to favor, or disfavor, particular viewpoints or messages.[39]

The analysis gets considerably more complicated, however, when it comes to a law like the one that Rep. Cicilline has proposed. This is because the purpose of that law—as evidenced both by the statements of its sponsors and by its statutory text—is not to prevent social media censorship per se, but instead to prevent the social media companies from using their control of important commercial forums to restrict economic competition.[40] This is a content-neutral purpose, as the Court made quite clear in *Turner Broadcasting* when it concluded that an antidiscrimination law that applied to cable companies was content-neutral

because it worked to prevent "unfair competition by the cable systems," rather than "to favor or disadvantage speech of any particular content."[41] Hence, so long as a law justified by this purpose is *also* facially content-neutral, its constitutionality should be assessed under intermediate scrutiny, not the considerably more demanding standards of strict scrutiny.[42]

Intermediate scrutiny does not, of course, mean that a law like Rep. Cicilline's would necessarily survive constitutional scrutiny. The government would still have to show that the law did not "burden substantially more speech than [was] necessary" to achieve its aims.[43] This would not always be an easy thing to do. In the case of Rep. Cicilline's draft bill, for example, it might be difficult for the government to show that preventing the social media companies from "discriminat[ing] among similarly situated business users" except when they could affirmatively show that doing so was necessary to protect user privacy or prevent illegal activity, or that it would not impede "legitimate activity" (whatever that might mean), was not broader than necessary to further the procompetition objectives of the law. This would be particularly true if that provision were interpreted broadly to prevent the social media companies from in most instances being able to enforce their content-based terms of service against those who use their platforms to make money by, for example, posting a monetized video or acting as a corporate "influencer."

Intermediate scrutiny does not provide regulators, in other words, a blank check to regulate as broadly as they may desire. But it does make it much easier for the government to prevail against a First Amendment challenge, because it does not require the government to identify a compelling state interest or show that the means it has adopted are the least speech-restrictive available.[44]

This suggests, more generally, the kinds of antidiscrimination laws that lawmakers could plausibly enact in the coming months and years, if they wanted to do something about social media censorship without violating the First Amendment as it is currently understood. They could enact laws that, like the Cicilline bill, do not attempt to prevent the social media companies from deciding for themselves what counts as acceptable or unacceptable speech on their platforms, but advance some other, and content-neutral, purpose. In the next section, I suggest what some of these content-neutral antidiscrimination laws might look like and do.

What Can the Legislature Do?

As the previous section made clear, when assessing the normative merits, as well as the constitutionality, of antidiscrimination laws for the social media platforms, it really matters what purposes those laws further, as well as the means

they employ to further them. This fact partly reflects the government purposes–oriented nature of contemporary First Amendment law.[45] More deeply, however, it reflects the democratic commitments that inform (or are supposed to inform) public policy in the United States, as well as free speech law.

Since the early twentieth century, the Supreme Court has insisted that a democratic government like our own has no right—or more precisely, no power—to dictate to its people "what shall be orthodox in politics, nationalism, religion, or other matters of opinion."[46] This is because, the Court explained, it is the proper responsibility of the people in a democratic polity, not their elected representatives, to resolve contested matters of opinion for themselves. This is what it means to say that, in a democratic state, "[a]uthority . . . is to be controlled by public opinion, not public opinion by authority."[47]

This view of the relationship between a democratic government and its people explains why laws that attempt to dictate to the platforms, or to any other private speech intermediaries for that matter, an orthodox view of freedom of speech—laws that require the platforms to agree with the courts, for example, about the value of lies and misinformation, or hate speech—are deeply problematic, both constitutionally and normatively. It also helps explain why laws that promote other goals—individual privacy, for example, or consumer welfare, or market competition—are much less troubling normatively and constitutionally, even when they target very similar kinds of discriminatory acts. This is because a democratic government is not only *not* barred from protecting these interests, it arguably has an affirmative duty to do so.

The result is that there are a good number of laws that legislatures could plausibly enact to limit the expansive power that the social media companies currently possess to discriminate against those who use their platforms that would not run into the constitutional dead ends that would face a platform common carrier law or other content-based regulation. The discussion of the Cicilline bill points to one way in which lawmakers could act: They could prohibit the social media companies from discriminating against the speech of those who use their platforms to compete against them or to compete against their friends. But anticompetitive manipulation of the platforms is not the only evil that lawmakers could combat using the tool of antidiscrimination law.

Another problem that lawmakers could use antidiscrimination law to combat is a problem that pervades the digital public sphere at present: namely, the nontransparent and potentially unprincipled application of the platform's rules to individual speech acts. The fact that social media companies have no existing obligations, legal or otherwise, to provide an explanation to users when they take down, deprioritize, or otherwise limit the dissemination of their speech makes it entirely possible—in fact, quite likely—that all sorts of factors that have nothing to do with the content of the user's speech factor

into their decision-making: that similarly situated users are not treated alike. If so, this is a kind of discrimination that lawmakers can do—and are in fact doing—something about. In Europe, for example, lawmakers are considering a law, the Digital Services Act (DSA), that would, among other things, require social media companies to provide users a right of appeal when their speech is discriminated against or their account suspended and to provide prompt redress if and when they conclude that the speech was in fact "not incompatible with [the platform's] terms and conditions."[48] Similar provisions are included in the Platform Accountability and Consumer Transparency Act (PACT Act) that Senators John Thune and Brian Schatz introduced in Congress in 2021 and in other proposed laws.[49]

Unlike those discussed in the previous section, these laws would not prevent those companies from making their own decisions about what kinds of speech to disseminate on their platforms. Instead, they would merely require the companies to enforce those rules in a more transparent and more principled manner than they might otherwise choose to do. As such, they do not pose the same normative difficulties a common carrier law would, or the nondiscrimination provisions in the Texas and Florida laws do. Nor do they pose the same constitutional difficulties. This is because a desire to ensure that commercial service providers accurately represent their terms of service and apply those terms in an evenhanded manner is a content-neutral rather than a content-based motivation for government action: It does not reflect even a "subtle . . . content preference."[50]

How effectively these laws will be able to achieve that purpose will depend on their details. One might in this respect profitably compare the PACT Act—which grants platform users a right of appeal but provides virtually no constraints on how that appeal should occur—with the DSA, which grants platform users many of the same procedural due process rights (a neutral arbitrator, notice, and an opportunity to be heard) that courts in the United States have recognized must be provided before individuals can be deprived of important social and economic goods by government bodies.[51] To the extent that lawmakers believe that the denial of access to the social media platforms threatens goods of roughly commensurable value to those implicated by the procedural due process cases—and it is difficult to understand why they would grant users procedural rights of this kind if they did not think so—the approach taken by the DSA makes much more sense than that employed by the PACT Act.

But even the more bare-bones procedural rights provided by the PACT Act would presumably do *something* to prevent arbitrary content moderation decision-making on the platforms' part, by making clear that the right of users to a nonarbitrary content moderation decision is not something the social media companies have the freedom to grant or deny at will but is a legally enforceable right. And they would do so without raising insuperable constitutional

difficulties or demanding that the platforms abide by an "orthodox" view of freedom of speech.

The same is true of another kind of content-neutral antidiscrimination law that might constrain the discretionary power of the social media platforms: namely, public accommodations law. Federal law grants all persons the right to "full and equal enjoyment of the goods, services, facilities, privileges, advantages, and accommodations of any place of public accommodation, as defined in this section, without discrimination or segregation on the ground of race, color, religion, or national origin."[52] State public accommodations statutes, meanwhile, grant individuals a right against discrimination on the basis of many more identity categories than those protected by federal law.[53]

Scholars and advocates have argued that laws of this kind should be applied to the social media platforms to prevent actions by either the companies or their users that deny racial minorities, women, the disabled, or other groups full access to the goods the platforms offer.[54] These arguments are persuasive. As a policy matter, there is no question that race, gender and antigay discrimination is a problem in the digital public sphere—perhaps a greater one than it is in other reaches of public life.[55] Therefore, the same normative commitments that justify the application of public accommodations law in other sectors of public life apply equally here.[56]

Of course, for all the reasons discussed earlier, the reach of public accommodations rights will have to be limited by the First Amendment right of the social media companies to decide for themselves what speech to host, or not to host, on their platforms. To put it in a more doctrinal form: The Court has held that, although public accommodations laws are generally considered to be incidental regulations of speech that trigger only very deferential First Amendment scrutiny, they become content-based when they are interpreted to require private speakers to communicate a message they do not wish to communicate.[57]

In the context of the social media companies, what this means is that some applications of public accommodations law are likely to be constitutionally barred. For example, it would be extremely difficult for the government to justify a public accommodations law that prohibited the social media companies from crafting or enforcing speech rules that resulted in the disparate takedown of conservative, radical, or feminist speech, so long as those rules were motivated by something other than a desire to exclude certain groups from the platform.[58] This is because a law of this kind would clearly interfere with the right of the platforms "to choose the content of [their] own message"—in this case, their message about the relative value or valuelessness of different kinds of speech.[59]

Many other applications of a public accommodations law would not, however, trigger the same degree of constitutional scrutiny. Consider, for example, a

public accommodations law that was interpreted to prohibit the platforms from discriminating on the basis of race, gender, or sexual orientation—or political affiliation, even—when it came to the *application* of their rules. There is plenty of evidence that the social media companies apply their rules differently to more and less powerful actors—that there are, as the Brennan Center recently put it, "double standards" in social media content moderation.[60] A law that made it unlawful for the platforms to employ these kinds of double standards when the intent or the result was to disparately disadvantage racial minorities, women, or any other protected group would not interfere with the platforms' ability to communicate a message. It would merely require them to live up to the promise that most already make: namely, that they apply the same rules to all users, regardless of their national or sexual or political identity.[61] Therefore, such a law should be entirely permissible under the First Amendment, so long as the government was able to show that the law was not enacted or applied "for the purpose of hampering the [social media companies'] ability to express [their] views" and there was no less-restrictive means of ensuring equal access to the public goods that the platforms provide.[62]

Enforcing such a law would require courts to distinguish cases in which a social media company's speech policy was applied in an evenhanded manner but nonetheless had a disparate effect and cases in which the companies' rules enforcers in fact applied the same rule differently to different users. This might not be an easy task. Because the rules that social media companies rely upon to justify the takedown or deprioritization of users' speech are often very broadly worded, the companies might be able to rather easily cloak what is in fact identity-based discrimination as simply the nuanced application of the rule.[63] Meaningful enforcement of the public accommodations law might therefore require more than just a prohibition on discrimination; it might also require the disclosure of information from the platforms about the outcome of similarly situated cases, about internal platform policies, and about the procedures of content moderation.

This is not necessarily a drawback, however. The fact that enforcing a public accommodations law might require the platforms to disclose more information to the public—or at least to the courts—about their internal practices than they might otherwise be willing to do might instead be a secondary benefit of the law by forcing the social media companies to be more transparent, and therefore more accountable, about their speech-related decision-making. This suggests in turn the complementary role antidiscrimination laws might play within a broader regulatory framework as one, but only one, of the instruments by which regulators work to limit the threat that platform power poses to democratic values, including perhaps the core value that democratic government enshrines: the equality of the individual.

Conclusion

The nondiscrimination laws canvassed in this essay do not exhaust the available options. There might be—likely are—many other kinds of pernicious kinds of discrimination, either enacted by the platforms or enabled by them, that lawmakers could and probably should do something about by imposing on the social media companies reasonable, content-neutral antidiscrimination obligations. As has been true with earlier technologies of mass communication, antidiscrimination law can help ensure that access to that technology is equitably distributed and that those who control it do not exercise undue influence over the content of public debate.

But, as this essay has suggested, antidiscrimination law cannot and should not provide a complete solution, or anywhere close to that, to the problem of platform power. It will not radically transform how speech decisions on the platforms get made or deny the social media companies the power to favor or disfavor, in all kinds of ways, particular voices and viewpoints. To render platform power more amenable to democratic values, lawmakers will have to rely upon—and devise—solutions other than the familiar device of the antidiscrimination mandate.

This is because the central problem with the operation of the digital public sphere today is not the problem of discrimination (although that is a problem, too, as this essay has suggested). It is the tremendous power that a few large and nontransparent companies possess over some of the most vital arenas of public discourse in contemporary society, by virtue of their ownership and control. To solve this problem will require more than simply the imposition of nondiscrimination norms on the social media companies, since nondiscrimination law is too blunt and too weak an instrument to ensure, as a general matter, equitable access for all to the digital public sphere. To ensure *that* goal will require much more ambitious, and structural, reform of the digital public sphere as it currently exists. Nondiscrimination laws represent, in other words, merely a first step to solving the problems that those who decry the problem of social media censorship are pointing to. Those who advocate for these kinds of laws, or who enact them, should not forget this crucial fact.

12

Platform Power, Online Speech, and the Search for New Constitutional Categories

NATHANIEL PERSILY

What position should large social media platforms occupy in American constitutional law?

Are they like political parties[1] or one-company towns?[2] In that case, they would be state actors that must respect the First Amendment rights of their users. Or are they like railroads,[3] telephone companies,[4] or Federal Express[5]—common carriers that ordinarily cannot discriminate among their customers in delivering goods or information from place to place? Or perhaps, as Justice Clarence Thomas suggested in a recent opinion,[6] they are like "places of public accommodation,"[7] akin to hotels and restaurants that are regulated by antidiscrimination laws? Or are they like a shopping mall[8] that must allow protesters to hand out leaflets in its parking lot, or a law school[9] that must allow the military to recruit students on campus? Or are they public utilities, like an electric company,[10] which enjoy First Amendment rights but also can be subject to considerable regulation? Or are they like a broadcast television station,[11] a cable provider,[12] or a newspaper?[13] If they are like one of those, the government may or may not have the power to require them to carry diverse and contrary points of view. Or are they simply like other corporations with First Amendment rights against government speech regulation, even including the corporation's right to spend unlimited amounts of money seeking to influence election campaigns?[14]

All such analogies have something to offer, but none fits perfectly. The internet and the rise of social media require the creation of new legal categories even as we struggle to fit these new institutions and relationships into old conceptual boxes. That each analogy to an earlier form of expression may fall short should not surprise us: Each generation tends to view technologies of its day as

unprecedented and in need of new legal definitions. The question that these new platforms pose, however, is: How can we create a legal environment that protects individuals' ability to reach an audience on the internet while also recognizing the twin dangers of unfettered expression that leads to real-world harm and unconstrained platform power that allows for consolidated corporate domination of the speech marketplace?

While we struggle to characterize the nature of social media platforms under the Constitution, courts and even the platforms themselves have described the environment these platforms have created as "the new public square."[15] As the Court majority put it in *Packingham v. North Carolina*, "[t]hey allow a person with an Internet connection to 'become a town crier with a voice that resonates farther than it could from any soapbox.'"[16] The Court was considering a law that restricted access to social media, not a practice by a social media company regulating content on its platform. However, even the leaders of Facebook and Twitter have joined in the chorus to characterize their platform environments in the same terms.[17] The "public square" metaphor has had predictable consequences, leading some to argue for increased regulation of social media companies given the outsized importance of the speech domain they control and others to argue in favor of treating their own self-regulation as akin to government regulation of individual speech.[18] Facebook has as much power as any government (and more than most), the argument goes, so we should consider its regulations of hate speech, incitement, and so on, as we would those of the government. Similarly, if the Facebook Newsfeed is now the public square, any regime of speech regulation (e.g., political advertising laws) which leaves out social media is ignoring the principal arena in which speech (dangerous or otherwise) is taking place.

Finding the right legal category for new social media platforms will not dictate how they should be regulated or how they should behave if unregulated. However, it represents a first step toward defining a permissible range of regulatory outcomes. As we explore different models, it is important to account for the unique features of social media platforms, in general, or even a given platform, in particular. Assuming we can identify the universe of "large social media platforms" to which these new rules should apply (itself a challenging task), these new rules ought to account for the (1) revolutionary speed of communication on the internet, (2) the need for automated filtering that operate as prior restraints, and (3) the global scope of these platforms. They also must recognize the distinctive characteristic of these social media platforms—one that makes them different from the traditional public square—namely, (4) the algorithms that organize and prioritize communication by making judgments that are inevitably content- and viewpoint-based.

This essay first analyzes and rejects the arguments that internet platforms are state actors, common carriers, or places of public accommodation, in the traditional sense. The next section explains why neither First Amendment law nor international human rights law—developed to deal with government speech suppression—has much to offer when it comes to guiding content moderation decisions. To be sure, case decisions in whatever forum dealing with general issues of free speech can be helpful (just as a good law review article would not be irrelevant), but there are critical differences between the law developed in these contexts and the analysis needed to deal with platform content regulation. The final section explains why and what to do about it. Certainly, we need to be worried about the power these private monopolies have over the speech marketplace. But different rules should apply to different products, and platforms should have a wide degree of latitude to impose the types of content- and viewpoint-based regulations that would be off-limits to government. The Conclusion discusses these issues in light of recent disclosures made by the Facebook whistleblower Frances Haugen and recent actions taken in Europe to regulate large online platforms.

This essay oscillates between positive arguments about the nature of platforms (constitutionally or jurisprudentially speaking) and normative arguments about how governments should regulate them and how platforms should regulate speech. To be clear from the outset, my view is that very few rules that have bound government speech regulation should bind platforms. This means that (1) the First Amendment (and international human rights law applicable to governments) does not and should not constrain platform content moderation; (2) the government should *not* pass a statute that requires those same legal principles to be applied by platforms on their own (as the state of Florida has and some in Congress have urged); and (3) the platforms, even absent government compulsion, should not, on their own, constrain themselves according to those principles, even if they have a "right" to do so. This may seem, at first, like the most extreme version of an argument for platform autonomy. However, at the same time as I would urge rejection of the platforms-as-governments analogy, I would also give very little credence to the platforms-as-speakers argument—that is, the notion that the government cannot and should not regulate large platforms for fear of violating the platforms' First Amendment rights. Indeed, neither the Constitution nor some other abstract theory related to constraints of government power over corporate speakers should stand in the way of government regulation of "large online platforms" in the area of content moderation or in the fields of antitrust, privacy, transparency, advertising, and taxation. Indeed, regulation in these putatively nonspeech arenas may go a long way toward addressing

some of the concerns about corporate power that undergird the worries about the damage platforms cause democracy through their content moderation policies. Government regulation of social media might implicate the constitutional rights of users, but the platforms themselves (contra *Citizens United*) should not be viewed as having rights that stand in the way of regulation of content moderation or other practices.

One final prefatory note as to the universe of actors considered in this analysis: One of the problems with commentary related to the dangers of "big tech" and social media is that several different types of communication platforms with very different "affordances" are grouped together as if their content moderation or free speech impacts are the same. However, as scholars such as evelyn douek have observed, content moderation questions affect everything from video games to restaurant reviews to home exercise equipment like Peloton.[19] Anytime a platform allows a user to "speak" to other users, the platform must come up with rules about what is permissible speech.

Moreover, some "platforms" comprise many different products. Not only does Facebook own Instagram, WhatsApp, and Oculus, but even the Facebook platform itself can include Messenger (which is also available as an independent application) and other features. Similarly, Google Search presents very different speech questions than do Gmail, Google Chat, YouTube, or the Android App Store. We should not expect the same rules of content moderation (as prescribed by governments or platforms) to apply to every mode of communication that a platform offers. WhatsApp is basically a glorified text messaging service, but the fact that Facebook owns WhatsApp probably does not mean that that messaging service should be subject to rules different than those that govern Apple's iMessage or even text services enabled by a cell phone provider.

Nor should we expect the same rules to apply to every layer of the internet stack. In other words, social media companies may have different obligations than hosts (e.g., Amazon Web Services, DreamHost), transit providers (e.g., Level(3), NTT), reverse proxies/CDNs (e.g., Akamai, Cloudflare), authoritative DNS providers (e.g., Dyn, Cloudflare), registrars (e.g., GoDaddy, Tucows), registries (e.g., Verisign, Afilias), internet service providers (e.g., Comcast, AT&T), recursive DNS providers (e.g., OpenDNS, Google), and browsers.[20] Each of these types of companies must decide what types of "speakers" can use its services. Each also has the potential to "silence" or amplify certain voices.

Finally, even if we can define the class of social media applications to which certain speech rules should apply, we inevitably must distinguish, on a somewhat arbitrary basis, which are large enough to warrant the special rules necessary to protect democracy and the speech marketplace. Facebook and Twitter should be subject to different rules than an upstart like Gettr or Parler, let alone the comment sections of a website like the *New York Times*, Fox News, Breitbart, or DailyKos. But what about TikTok, Reddit, Clubhouse, Next Door, or Discord?

Should the rules (again, either as to the propriety or constitutionality of government regulation or the rules the platform itself should enact) depend on the size of the user base, the professed openness of the platform, or some other metric associated with the danger the platform poses? At a certain point, the speech discussion begins to bleed over into a similar discussion occurring among antitrust experts concerning platform power—that is, once a platform achieves certain scale due to network effects it should be subject to greater regulation by government and its content moderation rules take on greater significance. For purposes of this essay, though, assume that the universe of relevant actors includes (at least) the Facebook newsfeed, YouTube, Twitter, and TikTok, based simply and arbitrarily on the size of the US user base, which for all of these is in excess of fifty million monthly active users.

Platforms as State Actors, Common Carriers, Public Accommodations, and Essential Facilities

State Action

If large internet companies are state actors, then they have no rights to object to government regulation, and they must respect the First Amendment and other rights of their users. Companies can become state actors if they perform a function that has been "traditionally, exclusively reserved to the state"[21] or if they are so intertwined with the state (through funding or joint performance of state-directed policy) that they are, in effect, agents of the state. Given the often adversarial relationship of the platforms to the government, it would be difficult to consider such corporations state actors in all their functions and incarnations. That a firm is powerful—even a powerful monopoly—does not mean it suddenly converts into the state. More is necessary. We can fear the power of Facebook and Google and even conclude they should be regulated into oblivion, but they are still private companies unless certain other factors are present. State action depends not merely on the size and importance of the company but also on the type of power it exercises and its proximity to formal organs of the state or state officials.

In the class action complaints he filed against Facebook, Google, and Twitter, following his deplatforming in the wake of the attempted insurrection on January 6, 2021, President Trump argued that the companies are state actors violating his First Amendment rights.[22] The suit is frivolous, but it raises important questions about what, if anything, might convert these companies into institutions that must obey the First Amendment. Although, like so many politicians and pundits do when they attack the platforms for viewpoint discrimination, the complaint emphasizes the platform's scale and the "public square" argument from *Packingham* and elsewhere,[23] it also grasps for other more traditional state action

arguments. It tries to latch onto the series of cases related to government entanglement with private actors, to suggest that the federal government encouraged, coerced, or partnered with these platforms to silence his and others' speech.[24]

Specifically, the complaint argues: "In censoring the specific speech at issue in this lawsuit and deplatforming plaintiff, defendants were acting in concert with federal officials, including officials at the CDC and the Biden transition team."[25] Whether through a series of implied threats from Democrats (what Genevieve Lakier discusses more generally as "jawboning"),[26] partnership with the government through free advertising, or reliance on agencies as official sources of information against which private posts would be judged as disinformation, the platforms, under this view, have become handmaidens in a government censorship program. They are all the more state-like, the complaint maintains, because Section 230 of the Communications Decency Act (which, they also argue, happens to be unconstitutional) grants platforms the benefits of immunity from liability for user-generated speech that they host and for takedowns of speech deemed "obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable."[27] This "benefit" of Section 230 immunity, like generous government funding or using government facilities, the argument goes, converts their content moderation decisions into state action and censorship in violation of the First Amendment.

The lawsuit is absurd for a number of reasons, not the least of which is that the plaintiff in this case happened to be the elected leader of the government alleged to have censored him at the time. Government officials can use their bully pulpits to excoriate internet platforms without turning them into state actors, and they can run advertisements or even give money to corporations (whether internet platforms, TV networks, hospitals, private schools, or restaurants) without those recipients turning into "the government." Indeed, if the government is coercing the platforms, as suggested, then a lawsuit against the United States or perhaps individual officeholders might be appropriate. Moreover, if Section 230 is a sufficient "hook" for state action, then any website that moderates user comments is a state actor, as they all benefit from the immunities that the statute provides. Section 230 is not limited to "big tech" or monopolistic websites, as impoverished popular commentary suggests. It applies to all websites that allow for user-generated content. If mere protection by Section 230 creates state actors, then much of the web is "the government."[28]

Common Carriers or Public Accommodations

A more serious analysis of the constitutional position of large internet platforms suggests that they be treated akin to common carriers (telephones or railroads) or places of public accommodation (hotels and restaurants).[29] There are strong

and weak forms of these arguments. The strong form is not dissimilar to the state action argument—suggesting that, as common carriers, the platforms have a constitutional obligation to welcome all comers. Just as a telephone service could not allow only "Democratic" phone calls, so, too, Facebook could not only allow Democrats to have accounts in its service.

The weak form of the common carrier argument maintains that the platforms can be forced to allow all comers onto their service. In other words, the state may force them, in essence, to comply with First Amendment precedent that would constrain a government—for example, to prohibit them from engaging in content- or viewpoint-based discrimination. Just as a common carrier (or place of public accommodation) could be forced to provide equal access to people of different races or political parties, the same obligation could be imposed on the platforms. Of course, were the state to impose such an obligation on the platforms, much of their efforts at policing disinformation (banning groups like QAnon or antivaxxers) or addressing hate speech (banning white supremacists and terrorist sympathizers), let alone garden-variety moderation of bullying or depictions of violence and nudity, might thereby be preempted.

One key question here is whether some right of the platform—either the platform's own First Amendment rights or perhaps its property rights—should prevent government treatment of them as common carriers or places of public accommodation. Do large platforms have a constitutional right to engage in moderation efforts to create an experience for their users that might exclude certain types of objectionable (but constitutionally protected) content?

Of course, *some* platforms must have such a right. The First Amendment would certainly protect me in my efforts to start a chat group with my friends or those of the Republican Party (or a conservative publication like Breitbart) creating a website limited to Republican users or content. But these are not examples of *common* carriers. Indeed, Facebook itself could not plausibly have been considered a common carrier when Mark Zuckerberg developed it in his Harvard dorm room and limited it to his classmates. The question is whether, once achieving some level of scale or importance for the communication ecosystem, Facebook became a common carrier. Facebook's implicit promises of allowing broad participation on its service irrespective of party, race, and so on might also be important (although that promise is itself cabined by the Community Standards and Terms of Service). The implication here, of course, is that a social media service, which might originally have been conceived as a closed and highly regulated environment, could become a common carrier once it achieves a certain scale and importance.

This is a powerful argument with serious real-world implications, as a recently passed[30] and rapidly enjoined[31] law from Florida highlights. That law imposes significant fines on platforms that deplatform or shadow ban candidates. It also

punishes platforms for taking "any action to censor, deplatform, or shadow ban a journalistic enterprise based on the content of its publication or broadcast."[32] It exempts any platform that operates a theme park (Florida's convenient "Disney exception"), but otherwise applies to any platform with one hundred million global monthly users and brings in one hundred million dollars annually.[33] Florida argued that the platforms were common carriers, making arguments echoing *Pruneyard* and *Rumsfeld*.[34] The court sided with the platforms, though, preliminarily enjoining the law. It held that the law was a content-based restriction on the platform's own First Amendment rights, failing either strict or intermediate scrutiny. Because the Florida law was so poorly written, riddled with content-based exceptions, and publicly described as an attempt to control liberal tech companies,[35] the court did not feel the need to engage deeply with the common carrier argument.

Justice Clarence Thomas has given a more thoughtful, even if still cursory, judicial treatment of the platform-as-common-carrier or public accommodation argument in his concurrence from a vacatur of the decision that held that President Trump's Twitter account was a public forum.[36] Describing the platforms as "networks within a network," Thomas argues that platforms with "substantial market power" that "hold themselves out as open to the public" might be considered common carriers. Gesturing toward the arguments in the Trump lawsuits arising from the Facebook, Twitter, and YouTube deplatforming, Thomas also finds the benefits bestowed through Section 230 and the specter of government coercion of the platforms as suggestive of their common carrier status. He throws Google Search and Amazon into the mix, as well, suggesting the argument extends well beyond social media.

Essential Facilities

What is missing from these accounts, as well as from a series of *Wall Street Journal* op-eds from Professor Phillip Hamburger,[37] is an appreciation and understanding of the different functions that social media platforms perform. Specifically, most of what they do is not "carrying" in the same way as railroads and telephone companies. They are not merely hosting speech, but organizing it. The most important feature of the platforms is the algorithms they employ to structure a unique "feed" for every individual user.

In what is the most sophisticated treatment of the subject, Professor Eugene Volokh develops a more modest common carrier argument restricted to the particular functions of the platform.[38] He considers the platforms as "common carriers" merely in their "carrying" functions—namely, hosting an account, posting content, and allowing users to follow other accounts. On this view, the state could prevent Facebook from deplatforming President Trump (and anyone else violating its community standards) or from preventing

people from following him. It could also prevent takedowns of content from the platform. If it were truly a *common* carrier, then the platform might be forced to keep on all hate speakers, disinformation purveyors, terrorist or supremacist sympathizers, bullies, self-harm videographers, and any number of other speakers that, for example, could not be discriminated against by phone companies or railroads.

However, when it comes to the presentation of information to users in feeds, Volokh (as well as the court that struck down the law in Florida) emphasizes the First Amendment rights of the platforms. The algorithms that organize what users see are, under this view, like the newspaper editors in *Miami Herald v. Tornillo*[39]: exercising legitimate content-based judgments as to what appears where on the page (or in this case, screen). Facebook and Twitter's Newsfeeds, YouTube's recommendation system, and Google Search would be beyond the reach of these common carrier regulations. When a platform organizes, prioritizes, and packages communication, it "speaks" rather than "carries."

This distinction seems plausible, even if, as I will argue later, it is undesirable. It would also allow for common carrier regulation of messaging apps and email. In other words, the state can require a platform to allow all comers to exist on the platform and communicate directly with a willing audience, but it cannot prevent a platform from discriminating against speakers by limiting the reach or visibility of their posts. The Florida law, which prohibited "shadow banning"—the use of algorithms to deprioritize a speaker's content in users' feeds—would go well beyond this minimalist view of common carriers. Even so, the implication of this weak form of the common carrier argument is that the platforms might not be able to deplatform or delete the posts of purveyors of hate, disinformation, violent imagery, and even advocacy of some offline harm.

If the common carrier argument only extends to the hosting function, then the platforms become indistinguishable from the internet itself. The ability to appear on a site and to have users see your content is indistinguishable from the right to have one's own separate website. The platform's added value comes from its algorithms. Facebook reports that when it demotes content it reduces reach by 80 percent.[40] Similarly, 70 percent of YouTube views come from the recommendation or autoplay features.[41] Merely hosting an account with content on it for users who seek it out does not afford the speakers much more than what they would get from having their own URL. In other words, deeming them common carriers for the accounts they host might protect "freedom of speech," but what these users are really fighting for is "freedom of reach,"[42] namely, the likelihood that their content would land in the feeds of a large audience on the platform.

The Place of Free Speech Law in Platform Regulation and Policy

Should the First Amendment Govern Platform Content Moderation?

The more limited common carrier argument is both too restrictive of platform autonomy to regulate speech and too protective of the First Amendment rights of platforms with respect to their algorithms. Platforms need to take action against harmful but constitutionally protected speech, and governments need to be able to regulate more than the mere ability to host accounts. For the largest platforms with the most influential algorithms, the state should enjoy wide latitude to regulate the ingredients of such algorithms to ensure both fairness and transparency.

Just as the state action argument goes too far, so, too, does the strong common carrier argument that the platforms should use the First Amendment as the guidepost for their content moderation decisions (or be compelled to do so). There are certain absurd implications that should be recognized from the outset. *Citizens United* and the constitutional restrictions on political advertising regulation are cases in point. If Twitter were "the state" and the newsfeed were the "public square," its ban on political ads would be unconstitutional,[43] as would Facebook's recent ban on new political ads being launched in the week before an election.[44] Indeed, many of the restrictions Facebook places on ads would be unconstitutional,[45] including its ban on ads with sensational content[46] or those that lead to "low-quality or disruptive" landing pages.[47] Some of these rules are not dissimilar to what a television or radio station might require, but the argument is common among different types of media entities: The legitimate interests and methods of a company to determine the boundaries of speech-for-purchase are different from those that a government can lay down for all of society.

This argument applies to the prioritization of content, as well as the rules for purchasing audience reach through ads. If one were to apply the First Amendment to platform algorithms, then content- and viewpoint-based prioritization of content would be prohibited (or at least subject to strict scrutiny).[48] Yet, in a very real sense, that is what a social media algorithm is. It makes personalized decisions about what you should see based on a host of characteristics about you (most importantly, your viewing history and friend networks), as well as decisions by the company about what kind of product and experience it wants to deliver to the users. If it could only include content-neutral factors in the algorithm, then only a chronological feed (or something similarly ringing in the spirit of "time, place, and manner" regulation) would be allowed. (This is one of the recommendations made by Frances Haugen, the Facebook whistleblower, in her Senate testimony.)

One consequence of content neutrality for social media algorithms is that they would be easily gamed and overrun by spam. The platforms' most aggressive content moderation programs, by far, concern the takedowns of accounts and content deemed to be spam. For example, a chronological feed basically favors speakers who post as often as possible. If you have one hundred Facebook friends, ninety-nine of whom post one post per day, but one of whom (perhaps due to automation, perhaps due to overcaffeination) posts one thousand posts per day, a chronological feed will favor the information superspreader. Of course, there may be content-neutral ways of dealing with this particular example, but the point remains that, for the algorithm to function, it must make "decisions" about speakers and content quality. It is no wonder, then, that Facebook takes down close to four billion accounts per year, mostly due to its antispam enforcement.[49]

Most spam is generated by automated or anonymous accounts. Anonymity, however, is generally protected under the First Amendment.[50] Were that to be the decision rule applied by platforms, real-names policies (like Facebook's) would not be allowed. To be sure, there are instances when disclosure can be mandated by government—as in, for example, campaign spending or advertisements. But if Facebook were a state, it could not require all organic content to be identified with a speaker. As it stands, some platforms, like Twitter, allow for anonymity and others, like Facebook, try to avoid it. But if the First Amendment were the lodestar for content moderation, anonymity would need to be generally protected. Doing so would seriously threaten the platform's ability to police online harm, foreign election interference, and a host of other problems.

It also remains far from clear what it might mean for an algorithm to "obey the First Amendment." Anyone who interacts with these firms quickly appreciates that there are few, if any, people who fully understand all the components of the Facebook newsfeed algorithm or Google search, for example. These "speech regulations" have now evolved over decades and are basically millions of lines of code, some of which are the product of machine learning that is not amenable to the kind of First Amendment analysis one finds in a judicial proceeding.

As compared to the algorithm, however, the platforms' community standards would be amenable to such an analysis, and they would fail miserably. Trump's complaint against the platforms did have this right: The platforms' rules against hate speech, glorification of violence, dangerous individuals, and most other standards, such as nudity, bullying, suicide, and self-injury, would be deemed unconstitutionally vague and overbroad. Almost any restriction on hate speech,[51] disinformation,[52] and glorification of violence[53] would violate the First Amendment. Even in areas where the government has greater latitude, as with obscenity and indecency, the platforms' rules on nudity go well beyond the boundaries the courts have allowed for government.[54] At times, the platforms' rules are both overly specific and overly broad. Facebook, for example, has a

series of very specific rules when it comes to comparisons between certain racial or religious groups and certain types of animals, which means that (given historical and cultural context) some groups, and only some groups, can be compared to some animals, but not others.[55] At the same time, its hate speech rules prevent users from calling for "social exclusion" of protected classes of people, "which means things like denying access to spaces and social services."[56]

One of the key reasons the platforms' rules are sometimes overly specific and overly broad (and often "unconstitutional") is that many such standards must be applied through automated filtering. A computer will have great difficulty, for example, in distinguishing positive (e.g., group-reaffirming) from negative uses of racial epithets or between satire, news coverage, and harmful speech (e.g., reporting on or mocking a hateful speaker while also restating the hate speech). The platforms simply are not in the position to make judicial-style judgments for each violation of the community standards.

In short, the platforms and their automated filters are *in the business* of prior restraints, which is about as electric a third rail as exists in First Amendment law.[57] Speech is, necessarily, prevented on the platform even before it reaches its audience. Both the scale and the speed of online communication make such filtering—which is inevitably overinclusive and underinclusive of the targeted content category—inevitable. Facebook must "adjudicate" millions of pieces of content every day.[58] No number of human moderators would be able to eyeball every post for community standards violations. Moreover, only automated filtering can respond with the speed necessary to enforce the community standards before the violative content goes viral. As any number of examples—such as the Christchurch massacre livestream[59] or the spread of Election Day disinformation—demonstrates, waiting for judicial or even executive-style enforcement action is a luxury the platforms do not enjoy. Either their systems pick up the offending content and limit its distribution, or it quickly may reach an audience of millions or even hundreds of millions of users.

One final note regarding the relevance of scale and whether the First Amendment can guide platform content moderation: Although in certain circumstances platforms can limit their enforcement to a particular speaker and audience in a specific geographic territory (so-called geofencing), their content moderation rules necessarily apply to a global audience. It would be near impossible to have a different hate speech or obscenity policy for every country in which Twitter, Facebook, and YouTube operate. The First Amendment of the US Constitution, awesome though it may be, does not represent a universal norm as to how to draw lines between permissible and impermissible speech. Most of Facebook's audience—indeed, well over 90 percent—live outside of the United States. If Facebook's Community Standards are to represent the Facebook community, imposing one country's

view of speech seems inconsistent with the nature of the platform, its reach, and its diverse user base. Leaving aside the fact that some countries actually ban some speech that the First Amendment clearly protects (potentially pitting a platform's permissive content guidelines against a country's regulation of speech), the platform's rules need to be sensitive to the charge of American First Amendment imperialism over the world's speech marketplace. Our First Amendment tradition is exceptional in a number of respects (defamation, hate speech, obscenity, and incitement are just a few examples). Universal speech rules for platforms based on the unique American tradition would be in persistent tension with the global nature of the platform.

Should International Human Rights Law Govern Platform Content Moderation?

Given the global reach of the platforms, should international human rights law, perhaps, guide platform content moderation decisions? Indeed, many NGOs and scholars have argued as much.[60] The Facebook Oversight Board, pursuant to the Charter developed by Facebook itself,[61] also has applied international law in its decisions on content moderation. If the Oversight Board is going to act like a global speech court, the argument goes, then it stands to reason that it should adopt international speech norms as guideposts for its decision.

However, for the same reasons that the First Amendment—designed as a limit on government control of speech—serves as a poor guide for content moderation, so, too, international human rights law—designed for the same purpose—fails as well. Even if international law is not generally as speech protective as the US First Amendment, it still provides an answer to a different question than the one platforms are answering in developing their content moderation policies.

Two cases from the Facebook Oversight Board demonstrate the mismatch. The first involved a Brazilian breast cancer examination video on Instagram originally taken down by an automated filter under the platform's nudity policy.[62] Facebook, which owns Instagram, later realized this video probably should be allowed under an applicable exception to its rules, as the board eventually held. However, the case exemplifies the difference between human rights law and rules for content moderation.

Were a government to ban all expression involving nudity, health education or otherwise, it would certainly be violating international human rights standards (let alone the First Amendment). But is the harm really the same when a platform applies an automated filter for nudity? Because it will often be difficult to determine "context" and "intent" on a grand scale in a split second, automated filters that will necessarily overenforce will need to be employed.

Given the importance of trying to prevent indecent material, child sexual imagery, and nonconsensual pornography, is the cost of preventing some legitimate and important nude imagery really one a platform should be forced to bear? It is not as if there is a shortage of places on the internet where one might find nudity. Whereas a government should not be able to eradicate nudity from the speech marketplace, a platform operating globally that takes down or filters out millions of nude images per month ought to be able to decide that a clear rule, easily enforced by automation, might be the best way to preserve the platform environment.

The same point could be made with respect to a case involving borderline COVID disinformation that the Oversight Board ordered reinstated onto the platform.[63] The post in question included a video that indirectly questioned the French government's policy of banning the use of hydroxychloroquine in treating COVID, saying it was a harmless drug with some evidence of success. It was removed under Facebook's incitement rules, which also prevent certain types of disinformation posts that could cause offline harm. The Oversight Board, basing its decision in part on international human rights law, ordered the post to be reinstated because Facebook had not demonstrated the post caused a risk of "*imminent* harm."[64] And indeed, were a government to have a blanket ban on borderline posts such as this, it might run afoul of applicable international law and certainly the First Amendment.

But Facebook ought to be able to strike the balance in a different way. It is being blamed for being a cauldron of COVID misinformation, leading people to doubt vaccine effectiveness, to experiment with dangerous "cures," and to believe a host of conspiracy theories. If Facebook decided to filter out all talk of hydroxychloroquine, given the risk that some people might die from prophylactic use of an unapproved drug, it would certainly prevent some valuable medical information from reaching its audience (such as warnings not to take the drug). But again, Facebook is not the only place on the internet where one might find information or misinformation on COVID. Its cost-benefit analysis on the harms of overcensoring through automated filtering versus potential lives saved through avoidance of COVID disinformation should be different than that of a government.

Evelyn douek is one of the few scholars who have recognized this point.[65] She argues, at a length unavailable here, that the international law rule of "proportionality" in speech regulation needs to be adapted to "probability" determinations that platforms can use in their automated and human moderation systems, especially including the algorithms that determine the reach of content. As a result, platforms' overbroad rules more easily adapted to automated global enforcement require, at least, a rethinking of what might be a new international human rights norm for platform content moderation.

Platform Rights and Platform Obligations

How, then, should we think about the rules that platforms should impose on users and that governments should impose on platforms? As should be clear from the discussion presented here, both of these decisions need to be informed by the unique features that distinguish large platforms from governments: namely, the global environment in which they operate, the speed with which they need to make decisions, the fact that prior restraints (usually through automated filtering at the source) represent the only effective way of preventing many online harms, and the special character of algorithms as a source for organizing and prioritizing speech.

As a threshold matter, we should cast to the side the notion of strong First Amendment rights for large platforms. They may not be common carriers, as argued above, but so too they are not like a newspaper or private club. These corporations' First Amendment rights (in the *Citizens United* sense) should be limited by important state interests in preserving democracy, preventing online and offline harms, and ensuring a competitive and healthy online marketplace of ideas. The First Amendment should be read as preventing many broad regulations of speech on the internet, but when it comes to regulating the large platforms themselves, the government should have much greater latitude. Many regulations of platforms might violate users' First Amendment rights (e.g., forcing Facebook to take down all insults, would, in fact, make them the state's agent in doing so),[66] but the platforms' own rights should not prevent these regulations.

Even when constitutional, such regulations might not be wise, of course. For reasons explained above, an "all comers" policy based on common carrier theories would cause more problems than it would solve. However, more modest antidiscrimination approaches could succeed. For example, whereas small platforms and websites have the right to discriminate on the basis of race or gender, it seems reasonable and constitutional to prevent the large platforms from doing so. The same could be said with respect to declared candidates and parties: The law could prevent the large platforms from using their power to exclude one or another candidate. Of course, that does not mean they can't disproportionately take down the accounts or content of Republicans or Democrats if, for example, one party tends to engage in disinformation and hate speech. Nor does it mean that each user must receive an equal amount of Democratic and Republican content; the platform could still bias the delivery of content based on user behavior and search history. It just means that platforms of a given size cannot use their monopoly position to intentionally favor or disfavor a candidate or party without some other community standard enforcement reason for doing so.

Moreover, a whole suite of regulations of advertising on large online platforms should be seen as presenting no constitutional problems. Requiring total transparency of purchasers and advertising contracts, bans or limits on microtargeting, and unique regulations for political advertising are examples of reasonable regulations. Indeed, a host of transparency rules, which might be constitutionally problematic if applied by government to the entire internet, should be seen as constitutionally unproblematic when applied to the large platforms. This would include compelled access to aggregated platform data by outside researchers, as well as periodic audits of algorithms to avoid (or at least be transparent about) bias and to ensure safety. I have authored one such piece of legislation, the Platform Transparency and Accountability Act,[67] which would empower the Federal Trade Commission to compel platforms to develop secure, privacy-protected pathways for independent research on platform-controlled data. The goal of such legislation is not to provide a luxury good to academics granted access but, rather, to alter platforms' behavior by removing their ability to operate in secret and to provide reliable information to policymakers for further legislative initiatives.

When it comes to moderation of content such as disinformation and hate speech, there is not much the government can do without running afoul of the First Amendment rights of users. A White House Office of Information Integrity is something that should send shivers down bipartisan spines. If the government is going to regulate harmful but legal content, it will need to do so indirectly. For example, it might require large platforms to comply with their own stated policies regarding content moderation. This suggestion may be a bit more controversial, given that many of their policies would violate the First Amendment if applied by government. Indeed, this principle was in part the basis for President Trump's controversial executive order following Twitter's actions against his posts.[68] As loony as the executive order was in other respects, its suggestion that repeated misapplication of explicit content moderation rules might constitute a kind of fraud presents a creative potential middle path for enforcement.[69] It might lead the platforms to promise less in terms of content moderation or perhaps all their rules would become mere suggestions. But given how difficult (and as argued above, wrongheaded) legislation on content moderation might be, this attempt at "co-regulation" (to borrow from the European approach) may hold some potential for striking the right balance.

When it comes to the platforms themselves, we should not expect or desire each platform to resolve controversial issues in the same way (in terms of both process and result). Indeed, as evelyn douek warns, there are good reasons to fear "content cartels."[70] Moreover, the "affordances" of the platforms, let alone the resources they have to dedicate to content moderation, differ considerably. Video platforms like YouTube and TikTok face different challenges than a social

media company like Facebook, which, given its size and wealth, has very different capabilities than does Twitter.

The decisions of the Facebook Oversight Board are important because they are the first official body with some distance from the platform itself (but with real-world impact) to grapple openly with the challenges of content moderation. We need more institutions like it, whether as a product of industry associations, corporate-government partnerships, or single-firm initiatives. The more transparency we have around resolution of the hard cases of content moderation, the more we might be able to lurch toward something resembling a common law in this space.

As described above, I think the common law here looks different from what we might see from constitutional courts. It will still balance individual rights of expression against public interests in preventing harm, but it will also include legitimate corporate interests tailored to the functioning of the product. These are not the raw financial interests of the company (although one can legitimately debate whether those are in scope as well) but, rather, the kind of speech environment that the platform seeks to cultivate. Are children using the platform? Are there closed groups, or are all posts viewable by all users? How much power is given to individual users to determine the content they see? (Indeed, this last question may hint toward a future of distributed content moderation, in which a market for different rules develops and users can opt into different community standards or different algorithmic prioritization rules. This, along with larger moves toward interoperability between platforms, might change the way we think of the impact of platform rules on speech.)

This common law, however, must be tailored to the unique challenges of rapid, automated content moderation on a global scale. Those features of platform speech regulation, as compared to that done by governments, necessarily translate into error rates at levels that constitutional courts would not and should not tolerate. This warranted tolerance for higher error rates comes, in part, from the fact that when a government (unlike a platform) bans speech or speakers, it often (though not always) removes their ability to utter those words anywhere in public or at least from the *truly* public square.[71] Decisions by firms and outside oversight bodies, let alone chatter from the pundit class, need to account for the place of a platform's speech rules in the larger context of online speech. The reason why large platforms' own speech rights should be given short shrift (as I argued above) is the enormous (arguably oligopolistic) power a few firms wield over the information ecosystem. Nevertheless, when certain content disappears from Twitter, for example, it does not disappear from the internet. Even when speech is taken down, it can almost always be placed elsewhere for users to view.

Of course, loss of an audience on one of the major platforms can be devastating for speakers. Creators on YouTube can spend years cultivating a following,

only to see it vanish overnight if they get deplatformed or otherwise demoted. Similarly, President Trump's efforts to re-create the audience he enjoyed on the major platforms fell flat. Costs measured by reach, instead of speech, are still significant costs, and may mean the loss of livelihood for creators and publications that have become dependent on a platform. But for most enforcements of community standards, as in the Oversight Board cases described above, the platform's decision does not prevent the speakers from getting their message out; it merely prevents them from taking advantage of the platform algorithm that amplifies it.

In the end, though, rising platform power may present the greatest challenge to the speech marketplace. That problem needs to be addressed head-on through antitrust and competition regulation. Indeed, the debate over content moderation cannot be divorced from the antitrust debate: If there were twenty small Facebooks, we might be less concerned over the community standards enforcement of any given one. For now, though, we should aim for a more transparent and experimentalist enforcement regime from the large platform from which a common law of content moderation might develop.

Conclusion

The summer of 2021 may be remembered as a turning point when it comes to regulation of social media. The Facebook whistleblower Frances Haugen captivated the public and Congress, first with a series of disclosures of Facebook's internal documents published in the *Wall Street Journal* and then with testimony before a Senate committee. She alleged that Facebook put profit before safety, and came with examples drawn from internal studies and documents relating to everything from Instagram's effect on teen girls' health to special content moderation appeals processes for influencers to negligence in policing offline harm in the global South. Haugen's testimony has lit a spark that may finally lead to comprehensive legislation to deal with the assorted problems afflicting the large platforms. At a time of historic partisan polarization, it appears that mutual hatred of "big tech" may be one of the few things that unites Democratic and Republican lawmakers.

Then again, the parties remain far apart when it comes to their characterization of "the problem" and therefore the desirable solutions. Democrats tend to want the platforms to take down more content, especially disinformation, hate speech, and incitement, and Republicans worry that aggressive content moderation policies lead to censorship of conservatives. Finding a bipartisan, constitutionally viable regulation of harmful content will not be easy, except maybe in a few narrow contexts such as child endangerment. Rather, we should expect

legislation on privacy, competition, taxation, advertising, and transparency to have a greater likelihood of passage.

In contrast to the typically sclerotic US policymaking process, Europe may be a first mover when it comes to tech policy. As scholars have shown in other contexts of "the Brussels Effect,"[72] European regulators often set rules with global extraterritorial application. We have seen this in the tech context with the General Data Protection Regulation (GDPR), Europe's privacy law, which companies like Facebook have said they will apply worldwide. The same might happen with content moderation as Europe begins to draft the specifics for enforcement of the Digital Services Act, which contemplates broad regulation of American tech companies. Without speech protections as robust as those in the US First Amendment tradition, Europe may be able to impose more significant content moderation standards. In turn, the US tech companies may (as with GDPR) decide that it is easier to apply such rules worldwide than have different standards by region.

For now, we need a better understanding of what is happening on these platforms and how platform policies are shaping the information ecosystem.[73] We should not need to wait for whistleblowers to blow their whistles in order to get glimpses of the insights platform researchers have culled from company data. The flip side of freedom of expression is a right to information. The public has a right to be informed about what is happening on these large internet platforms. However one might characterize these new, powerful entities for purposes of constitutional law, these platforms have lost their right to secrecy. Congress can and should require that they open themselves up to outside scrutiny.

13

Strategy and Structure

Understanding Online Disinformation and How Commitments to "Free Speech" Complicate Mitigation Approaches

KATE STARBIRD

Introduction

The pervasive spread of disinformation online is a critical societal issue, limiting our ability to respond to collective challenges such as pandemics and climate change. Though it can be difficult to quantify the reach and impact of disinformation,[1] we don't need to look far to see evidence of the problem. In 2020, as the COVID-19 pandemic pushed people away from physical social gatherings and into digital ones, antivaccine conversations expanded from the margins of the internet through mom blogs and Facebook groups and eventually out through cable news pundits and into the broader population, likely playing a role in large numbers of Americans refusing potentially life-saving vaccines. Currently, nearly a third of Americans believe—contrary to the evidence and in the midst of an ongoing disinformation campaign targeting them—that the outcome of the 2020 US election was fraudulent.[2] Conspiracy theories that once seemed confined to the margins of public discourse are now mainstream.

Researchers continue to investigate and debate the causes and potential remedies, but most agree that there is something about our current moment—some intersection of our human vulnerabilities, our collective challenges, and the dynamics of our now mostly digitally mediated information spaces—that seems to be enabling this sort of mass manipulation (and the resulting mass delusion) at scale. Though the examples in this essay will mostly focus on the US context, the phenomenon is international, with researchers documenting similar issues all around the globe.[3] Scholars have argued that, considering the complexity and incredibly high stakes of human interaction at scale (including

the challenge of addressing disinformation), the study of collective behavior in a digitally connected world needs to be approached as a "crisis discipline."[4]

Pervasive disinformation is of particular threat to democratic societies. Farrell and Schneier warn that disinformation can function as an attack on the shared understandings that hold political systems such as democracies together.[5] Without this kind of common ground—for example, around the rules guiding political participation, including how elections work and who the winners are—citizens of democratic countries may be unable to come together to govern themselves. The events of January 6, 2021, at the US Capitol underscore this danger, revealing how a disinformation campaign can motivate a direct attack on democratic processes by a country's own citizens.

In this essay, I aim to define disinformation, contrasting it with its more familiar cousin, misinformation, and situating it within a historical context that sheds light on how—and why—it works. I'll describe disinformation as *intentional*, as *misleading* but often built around a true or plausible core, and as functioning as a *campaign* rather than a single piece of content. I'll explain how online disinformation exploits social media and other information systems (both online and off) in complex ways, and how disinformation can become embedded in the structures of online ecosystems—for example, the social networks and algorithms that shape how content flows through these systems. I'll also explain how online disinformation is *participatory*—incorporating "unwitting" crowds of sincere believers who routinely amplify and even produce false and misleading narratives.

Once I've established this understanding of the nature of online disinformation, I'll explore how our commitments to "freedom of speech"—a core tenet of democracy—complicate the challenge of addressing disinformation and other online toxicities. The strategies of disinformation have long exploited, both legally and rhetorically, "freedom of speech" to manipulate democratic discourse and justify their actions. Examining the history of these strategies reveals the persistent challenge of addressing disinformation while preserving our democratic commitments to free speech. Finally, I'll offer a few pointers for how platforms can navigate these challenges, focusing on the structural dimensions and attending to the participatory dynamics of disinformation.

Defining Disinformation

To understand the unique threat of disinformation, it's important to differentiate it from its less insidious cousin, misinformation. Misinformation is information that is false, but not necessarily intentionally false. Misinformation includes, among other information types, organic rumors that turn out to be untrue as

well as information that was once true (or likely true) but continues to spread after it is no longer valid. Disinformation, on the other hand, is *false or misleading* information that is *intentionally seeded or spread* for an *objective*.[6]

Let us unpack this definition—and provide some context. The current wording was achieved by bringing recent definitional work on "problematic information"[7] and "information disorder"[8] into conversation with historical understandings of specifically Soviet-style disinformation.[9] In particular, this definition relies on the work of Lawrence Martin-Bittman (originally Ladislav Bittman), a former Czech intelligence officer and practitioner of disinformation who defected to the United States in 1968 and later became an author and university instructor specializing in the study of propaganda and disinformation. Bittman's writings provide insight into Soviet-style "active measures" that can help us understand modern disinformation. Active measures are intelligence operations that actively manipulate, rather than passively monitor, information environments for geopolitical gain. And disinformation is a primary tool in the active measures toolbox. Though the KGB were early experts in disinformation (and their descendants within the Russian government continue to practice it), the tools and techniques of disinformation are now widely available and, in the connected era, easier than ever to employ by foreign intelligence officers, domestic political campaigns, and online activists alike.

False or Misleading Information

The most effective disinformation is not necessarily false but, rather, *misleading*, built around a kernel of truth—a true or plausible core—then layered with distortions and exaggerations intended to shape how others perceive reality. As Bittman explains, "Most campaigns are a carefully designed mixture of facts, half-truths, exaggerations, and deliberate lies."[10] Aligned with this historical view, Aral describes how a much more recent (2014) Russian disinformation campaign targeting Ukraine—part of its successful hybrid-war effort to annex Crimea—was detectable in online data from a spike in news that mixed both true and false information.[11]

As Bittman's quotation suggests, to do the work of blending fact and fiction, disinformation most often functions not as a single piece of misleading content but as a *campaign*. Disinformation is therefore difficult to reduce to a single article or tweet or even a single narrative, but takes shape through numerous and diverse actions. For example, the Russian disinformation campaign targeting the US election in 2016 included, among other tactics, hacking and leaking emails through friendly organizations, strategically amplifying emergent conspiracy theories about those emails through state-controlled media, and operating hundreds of fake social media profiles that pushed out a range of messages targeting

"both sides" of the US political discourse.[12] Due to these characteristics—its blending of fact and fiction across many different information actions, often by different actors (many unwitting, as I will describe below)—disinformation can be resistant to simple fact-checking.

Seeded and/or Spread

Disinformation can be created from whole cloth by those with strategic intentions. Or it can consist of organic information (e.g., a conspiracy theory) opportunistically amplified by strategic actors. A classic example of this latter dynamic is Operation InfeKtion (actually Operation Denver), a Soviet active measures campaign which asserted that the US government had created HIV/AIDS as a biological weapon.[13] In his 2020 book *Active Measures*, Thomas Rid traces that disinformation campaign back to its roots, revealing that the first claims of US government involvement in the creation and spread of HIV/AIDS originated in American gay rights activist communities (in 1983) as they struggled against catastrophic inaction by the US government during the early days of the epidemic.[14] The KGB built their (eventually very successful) campaign around that initially organic conspiracy theory, shaping the narrative and catalyzing its spread through the introduction of forgeries and by planting stories in various media outlets around the world. Eventually, the claims would begin to spread organically, through media outlets beyond those initially targeted. This interplay between organic (unintentional) activities and the orchestrated (intentional) activities of disinformation campaigns continues—and is in many ways supercharged—in the online realm, where global communities can connect to rapidly generate the "fittest" conspiracy theories in response to any and every breaking news event and where strategic actors can effectively amplify and shape these theories to advance their goals.

For an Objective

Central to its definition is the fact that disinformation is employed to support an objective (or objectives). Historically, that objective was primarily a political one. Bittman described Soviet "active measures"—of which disinformation is a common tool—as an instrument of foreign policy, used to "disrupt relations between other nations, discredit Soviet opponents, and influence the policies of foreign governments in favor of Soviet plans and policies."[15] Though they were leaders and innovators, the Soviets were not alone in using these techniques during the twentieth century.[16] And with expanded access to information production and dissemination, in the twenty-first century the tools and techniques of disinformation are now usable for a range of goals, by diverse actors.

Modern disinformation campaigns are often organized around political objectives as well,[17] though the perpetrator may not be a rival country. Private individuals and organizations can also create and spread disinformation to attempt to shape desired political outcomes, in their home countries or abroad—either in support of their own objectives or as proxies for other political actors including governments.[18]

Financial gain is another common objective of disinformation. In recent years, we have seen disinformation about corporations used to move financial markets[19] and medical disinformation leveraged to sell alternative treatments and "cures."[20] Other motivations for creating and spreading disinformation online include reputation gains (i.e., accumulating online attention in the form of engagements or followers) and entertainment (i.e., "for the lulz"). Reputational gains can be converted into financial and political gain. For example, there are individuals who strategically create and amplify rumors and conspiracy theories during breaking news events to gain followers—and then later leverage their accounts to sell products or spread political messages.[21]

Many intentional spreaders of disinformation have multiple complementary motivations. In our research team's first studies of online disinformation,[22] we discovered networks of "alternative news" websites that repeatedly spread conspiracy theories and disinformation. Many appear to be operated for purely political purposes—for example, RT (formerly Russia Today) and Sputnik News are integrated into these networks in various ways. But a large number of websites, including ones that repeatedly reshare articles originating in state-controlled media outlets, seem to be financially motivated. Some generate their own content, but many simply republish articles produced elsewhere, using the free and sensationalist content of political disinformation campaigns to pull readers to their pages. Advertisements for nutritional supplements, drones, and "prepping" supplies (for a post-disaster or post-apocalyptic world) often run alongside the browser margins of online articles spreading disinformation.

A single disinformation campaign can work toward multiple parallel objectives, and these can play out on similar or vastly different timescales. For example, the multiyear (2014–17) RU-IRA campaign targeting the 2016 US election likely had short-term goals (such as denigrating a specific candidate and shifting the outcome of an election toward another), midterm goals (like eroding trust in democratic governance), and the long-term goal of weakening an adversarial country. The infrastructure—including cultivated audiences—that a campaign puts down for one objective can be leveraged later for others. For example, Russia's online disinformation efforts (2014–21) often build upon old narratives—for example, around racial tensions, antiwar sentiments, and distrust in the US government—that echo back to campaigns run before the fall of the Soviet Union. Similarly, some of the domestic networks of astroturfed

accounts that were established between 2011 and 2013 to support the Tea Party later functioned to spread disinformation related to the 2016 and 2020 elections.

Intentionality

The primary difference between disinformation and misinformation is the *intention* to deceive. However, this essential element of the disinformation definition gets unwieldy when we start to unpack how disinformation campaigns work, because many people who spread disinformation are not aware that they are part of a campaign or that the content they are spreading is false or misleading. Bittman explained how Soviet-era disinformation campaigns consisted of operators (those perpetrating the campaigns), targets (adversaries who will be affected by the campaigns), and unwitting agents (who help carry out the campaigns without understanding their role).[23] In our research on several distinct disinformation campaigns, we estimate that most spreaders are what we call "sincere believers." Though the definition of disinformation maintains that it involves intentional deception, not everyone who spreads disinformation does so intentionally.

Bittman noted that journalists, in particular, frequently served as unintentional conduits of disinformation. In some cases, journalists are impersonated, allowing intelligence agencies to plant articles within otherwise independent media. In other cases, real journalists are cultivated, lured in by the promise of scoops or other financial or reputational rewards. While Bittman documented these practices in the 1980s (drawing from his experiences in the 1960s), these strategies of planting articles and cultivating journalists are still in use today.[24] Bittman stressed how the KGB explicitly exploited the legal and ethical frameworks of "freedom of speech" and "freedom of the press" to carry out their disinformation campaigns, revealing this as an inherent vulnerability of the West—and particularly of the United States. I'll return to this later in this essay.

Drawing from historical accounts and interviews with Bittman, Thomas Rid[25] explored how Western political activist communities were also leveraged by Soviet-era disinformation campaigns. Relating several historical case studies of intelligence agents infiltrating existing movements and attempting to shape their activities toward the objectives of the infiltrators, Rid explains how active measures and organic activism often became entangled in ways that are difficult to fully unwind. In research on online activism around the 2016 US election and the civil war in Syria, my colleagues and I documented similar entanglements between political activism and modern disinformation campaigns.[26] Online activists may be especially susceptible to being mobilized as unwitting participants in these campaigns due to the relative anonymity afforded by the digital environment and the "networked" nature of digitally connected movements.

Participatory Disinformation

For many of us living in the United States, we can think about our recent awakening to the problem of disinformation as a play with two acts: the 2016 and 2020 elections. The story of online disinformation and the 2016 US election was one of foreign agents at Russia's Internet Research Agency (RU-IRA) using "inauthentic" accounts to coordinate an attack on the US electorate. This was not the *whole* story, of course. The reality was much more complicated and included US citizens and international actors performing a variety of roles, both witting and unwitting. But the "foreign interference" framing made for a straightforward story with clear delineations of right versus wrong, victims versus villains. Eventually it became the story that media,[27] government reports,[28] and the social media platforms themselves[29] focused on. In the aftermath and reckoning with foreign interference in the 2016 election, platforms organized their responses to disinformation around this type of threat. For example, Facebook focused its mitigation efforts on "coordinated inauthentic behavior"[30] and over the next four years, took down numerous networks of such accounts operated by governments and private organizations all over the world.

The story of online disinformation and the 2020 US election was starkly different. Instead of being foreign, inauthentic, and explicitly coordinated within a top-down organization (as in 2016), this disinformation campaign was largely domestic, perpetrated by real people—often blue-check, verified accounts—and only loosely coordinated. The most impactful disinformation campaign centered around narratives meant to sow distrust in the voting process and, eventually, the election results. This campaign began in earnest months before the election and took shape through many and diverse narratives, eventually encompassing false and exaggerated claims about the legality and security of mail-in voting, about the vote-counting process, about "statistical anomalies" in the vote counts themselves, about dead people voting, about voting software changing votes, and many more.[31] These narratives were repeatedly shared by right-wing media and online influencers, members of the Trump campaign, and even the president himself.[32] And they were extremely effective. A survey conducted in the months after Joe Biden's inauguration[33] reported that nearly two-thirds of Republicans believed that the 2020 election had been stolen from former president Trump.

This disinformation campaign—first labeled as one by Benkler and colleagues in a report published in October 2020[34]—was very much a top-down effort, seeded and shaped by right-wing political and media elites who repeatedly pushed false, exaggerated, or otherwise misleading claims of voter fraud. However, it was also a bottom-up phenomenon, with everyday people using social media to spread and in some cases produce false narratives about voter fraud.[35] This

campaign took shape as a collaboration between politically motivated elites (in media and government) and their audiences. Disinformation, in the case of the 2020 US election, was *participatory*.[36]

With the rise of social media, everyday people are not merely consumers of information but are producers—and spreaders—of information as well. Early internet scholars heralded the positive potential for this kind of mass participation and the "democratization of information production."[37] Seizing on this potential, social media companies boasted about their capacity for making connections, that is, "connecting the world."[38] And indeed, in the early years of social media (2007–12), the world witnessed people coming together, in some cases from across the globe, to (among other things) assist in disaster response efforts and demonstrate solidarity for political protests during the Arab Spring. Reflecting some of the optimism of that time, my own lab, created in 2012, is named the "emergent capacities for mass participation" (emCOMP) lab. But mass participation has its downsides, including harassment, hate speech, and the proliferation of mis- and disinformation.

While the production of information has been democratized, so has the production of disinformation and other forms of propaganda. In a chapter introducing their conceptualization of participatory propaganda, Wanless and Berk wrote:

> Modern propaganda can no longer be viewed as a traditional top-down process alone. Target audiences are no longer mere passive consumers of such targeted persuasive content, but are also active in its creation, modification, spread and amplification, often inadvertently furthering the agenda of propagandists whose messaging resonates with their worldview.[39]

This conceptualization applies to the modern version of propaganda's cousin, disinformation, as well. My colleagues and I have characterized diverse online disinformation campaigns, from RU-IRA interference in the 2016 election to the state-sponsored campaign against the "White Helmets" (officially named the Syria Civil Defence), as collaborations between witting agents and unwitting crowds of online political activists.[40] Even in cases where state-controlled media and intelligence organizations play a clear role, online disinformation campaigns are *participatory*.

As I noted in the introduction to this section, the disinformation campaign that sowed doubt in the 2020 US presidential election was intensely participatory as well, with online activists helping to amplify and even in some cases produce false narratives of voter fraud. A particularly insightful example is the #SharpieGate narrative. This narrative emerged on Election Day, as a small number of voters in several parts of the country noted that the Sharpie pens

provided by poll workers were bleeding through their ballots. Though election officials assured voters that this would not affect their vote,[41] reports of Sharpie bleed-through began to spread on social media. Initially, these posts were accompanied by messages of concern and directives to "bring your own pen." But later they would take a more suspicious tone, and eventually—as the narrative began to "go viral" on the day after the election—the Sharpie bleed-through was framed as an intentional effort to disenfranchise specifically Trump voters.

On Twitter, early claims about concerns with Sharpie pens originated in accounts with fairly low follower counts. Eventually, though, they were amplified by right-wing media and political influencers, including President Trump's two adult sons. And though the first tweets about Sharpies bleeding through were in reference to issues in Chicago, shortly after Arizona was called by Fox News for candidate Biden, the narrative converged around concerns in Maricopa County, Arizona. That county would later become the focus of a politically motivated "audit" that attempted to prove some of the myriad false claims of fraud.[42]

One of the most interesting elements of the SharpieGate narrative surfaced on November 4, after it had begun to go viral and focus on Arizona. Republican voters were being encouraged, within a number of social media posts, to check the status of their ballots and report (to various entities, including lawyers and the Republican National Committee) anything suspicious. Several social media users shared screenshots—accompanied by expressions of anger—of an online tool that reported their ballot had been cancelled. A few of these posts circulated widely. The only problem was that the online tool reported the status of their *mail-in ballot*, which had been cancelled when they voted in person, *not the status of their in-person ballot*, where they had used a Sharpie. In other words, after repeatedly hearing the refrain that they would be and then had been cheated, these voters misinterpreted information they found online, and even their own experiences of voting, in ways that aligned with the false voter fraud narrative. Through their online participation, they were able to feed those (false) experiences back into the disinformation campaign, where they were passed up the chain to influencers who could spread them widely.

SharpieGate is just one of dozens of similar examples of the participatory dynamics within the 2020 voter fraud disinformation campaign, a campaign that culminated in the January 6, 2021, attack on the US Capitol by a group of people who believed that Donald Trump was the legitimate winner of that election. In our research of social media activity around this campaign, we find that many, perhaps even most, of those spreading—and in many cases even those creating—disinformation are sincere believers of the content that they are sharing. They may be sharing "lies" but they aren't lying. They are misinformed. Their social media posts reveal some to be misinterpreting their own experiences, victims of intentional efforts to reshape their realities. Participatory disinformation is a

powerful, two-way dynamic that incorporates its target audiences into its production and spread.

Online Disinformation as a Threat to Democracy

Scholars, journalists, and political leaders alike have noted that disinformation presents a threat to democracy.[43] Acute disinformation, like what we saw with the disinformation campaign attacking the integrity of the 2020 election, can attack democracy directly, sowing distrust in democratic processes and election outcomes. But pervasive disinformation—continuous falsehoods that slowly erode our trust in information—undermines democracy from another direction. Pomerantsev and Weiss describe how disinformation is not necessarily meant to convince but to confuse, to create muddled thinking across a targeted population, so that they are unable to discern what is true.[44] In the case of democratic societies, this may render citizens unable to make the decisions they need to make to govern themselves.

From a related perspective, disinformation can be viewed as a kind of attack on "common knowledge"—the consensus beliefs that hold political systems such as democracies together.[45] In the US 2016 election, the Russian government and collaborating organizations perpetrated a multidimensional attack on the "common knowledge" of democracy in the United States. It included efforts to undermine trust in election infrastructure (through hacking of physical systems), traditional media (calling them "fake news"), social media (through widespread impersonation of US political activists using "troll" accounts), and the legitimacy of outgroup political actors (by sowing division and impersonation). In 2020, we witnessed some of these same kinds of attacks being carried out by domestic groups—citizens, motivated by domestic political goals, using disinformation to attack their own institutions from within. Though they may have hoped to achieve short-term political gains, the widespread use of disinformation as a tool of domestic politics has the potential to critically weaken the system over which it aims to win advantage. Disinformation chips away at the foundations of democracy—our trust in information (including traditional and social media), government institutions (including election systems), and each other. Over time, it may destabilize the common ground citizens of a democracy need to stand up to govern themselves.

The Strategic Social (Re)Construction of Reality

One of the more concerning aspects of pervasive disinformation is the intentional distortion of reality—at scale. At the end of Rid's (2020) book about

Soviet active measures (and other Cold War information tactics used by Western countries),[46] he begins to flirt with a compelling idea about the "social construction" of reality and its relationship to disinformation strategy. In the social constructionist view, reality is subjective and constructed through human experiences and social interactions.[47] Extending from a social constructionist view, postmodern scholars often work by "deconstructing" knowledge to unpack how these realities have been produced. Rid suggests that the perpetrators of active measures (including disinformation) have taken postmodern deconstruction and turned it on its head, shifting social constructionism from a descriptive theory to a prescriptive one. He writes:

> Most academic critical theorists were, however, only studying and deconstructing the "practices" of knowledge production to shape intellectual discourse, to interpret the world. Meanwhile, in the shadows, intelligence agencies were actually producing knowledge, constructing new artifacts, shaping discourse in order to serve their tactical or strategic purposes—changing the world as they went.[48]

Disinformation "works" by strategically shaping perceptions of reality within a targeted population. Once embedded in those realities—and consequently in the histories, social norms, and collective consciousness of a society—it may be difficult to undo, address, or even identify the manipulation.

Structure and Sociotechnical Systems

In the connected era, this reshaping of social realities is taking place to a considerable extent within the online media ecosystems that mediate so much of our social and political interactions. There has been some debate about the role of social media, specifically, in the spread of disinformation and other forms of propaganda within these ecosystems. Benkler and colleagues have repeatedly argued that while social media get a lot of the attention, other media outlets—especially hyperpartisan news outlets—are perhaps even more responsible for the current spread of propaganda and disinformation.[49] Benkler's point is a valid one; we should not overlook the role of partisan news, including conservative cable news leader Fox News. But perhaps a more productive way to approach understanding modern disinformation is to think about social and other media as inextricably connected. Through the actions of their users to move content from one site to another, social media platforms are deeply integrated with other media, including traditional media (e.g., newspapers and broadcast media), internet-first alternative media, and even cable news. From this view, and in consideration of recent examples of disinformation campaigns exploiting these

platforms[50] to achieve their goals, criticism—and interrogation—of the role of social media in facilitating the spread of disinformation is warranted.

One way to understand the dynamics within these complex sociotechnical systems is through the lens of structuration theory, which helps us uncover and examine the bidirectional relationships between social structure and human agency.[51] Social structure, in this view, encompasses all of the "rules and resources"[52] that guide human action, for example, social norms, social relationships and hierarchies, and institutions. Structuration theory holds that social structure both shapes and is shaped by human agency. In other words, the decisions we make are constrained and guided by the social structures embedded within our society. And at the same time, those structures are reified, stressed, and reconfigured by the choices we make and the actions we take.

In the social media realm, "structure" includes the networks, algorithms, and platform features that shape our interactions there. Platform features determine which actions are available to us. For example, on most platforms we can post messages, follow other accounts, and interact with other content through "like" buttons and comment or reply features. Many platforms allow us to reshare (e.g., retweet) content. Some allow us to embed links to articles posted elsewhere on the internet. On some platforms we are able to see certain signals about content or users, like how many engagements a certain post has or how old an account is, which can guide our decisions about how to interact with them. For example, we may be more likely to form a positive opinion of content when we see it has a large number of "likes,"[53] while a "disputed" label may make us less likely (or in some cases more likely) to engage with a piece of content.[54] All of these features profoundly shape how we encounter, interpret, and share content as well as how we interact with others on platforms.

To a large extent, our networks—that is, the accounts and pages we choose to follow or friend—determine what content is made most visible to us. Most social media platforms have some kind of "feed" where we can see content shared by the accounts we follow. However, successful platforms learned early on that the content produced exceeds our capacity, as individuals, to attend to it. And so they employ algorithms, pieces of computer code that make decisions based on some set of heuristics or on "machine learning" processes, to filter that content, deciding which messages from which friends will be most visible to us.

Algorithms also play a role in shaping our networks, giving us recommendations for whom to follow. The algorithms are, in turn, shaped by our actions, which feed them the raw materials (our decisions) they use to do their "learning." Most social media algorithms are black boxes, at least to those of us outside their companies, so it's hard to know exactly how they work. But it's likely that algorithms take into account aspects of social networks when they make decisions about which accounts and what content to make most visible.

That's why we are more likely to get a recommendation to follow someone who is a friend of a friend, and more likely to see a piece of content that has been liked hundreds of times than one that has few engagements.

So we can see these complex and mutually shaping relationships between network structure, algorithms, platform features, and human behavior in online systems. For those of us who spend time on these platforms, these sociotechnical structures determine what content is most visible to us, when. They also overlay that content with various social and informational signals, from veracity to political valence to popularity. And this content shapes what we think about the world, including our understandings of political debates and social norms, both extremely relevant to concerns about disinformation. For those aiming to reshape reality for their political, financial, and other objectives, these sociotechnical structures and affordances—and the interactions between them—represent an opportunity.

Online disinformation campaigns piggyback on the infrastructure of social media, and they can become embedded in the structures there. They exploit the affordances of social media to spread their messages. As they operate, they also reshape the networks and the algorithms that determine how information flows. Examining the operations of the RU-IRA leading up to the 2016 election,[55] my colleagues and I noted that a large portion of the activity was not focused on spreading "fake news" or a specific piece of propaganda but instead on gaining followers. First, the "trolls" attempted to embed themselves into the networks. Then, they used their positions there to shift the discourse toward their objectives. Eventually, a few of their accounts gained enough visibility to become highly retweeted, followed, and likely recommended as well (researchers are not able to retroactively study the recommendations that platforms may have given at the time).

Similarly, automated and astroturfed activities do not just shift the content of conversations, they reshape the underlying structures of those conversations. Research suggests that automated (bot) activity on social media platforms like Twitter may be just as much about manipulating networks and algorithms as about directly reaching other users with specific messages.[56] These strategies shape what posts are most visible, for example, by making a certain topic make it into "trending topic" lists, as well as determining who gets recommended and followed.

Manual tactics allow users to quickly grow vast, dense networks. On Twitter, this can be seen through "follow-back" mechanisms whereby users reciprocally follow others and "train" techniques where users mass-mention other accounts that all subsequently follow each other. Researchers have documented these practices in pro-Trump, QAnon, and left-leaning "resistance" communities.[57] The resulting networks—which have an interesting property whereby users

typically have both a high follower count and something close to a one-to-one ratio of friends to followers—routinely facilitate the flow of disinformation and political propaganda. On Facebook, using a similar tactic, a few savvy organizers employed large "invite lists" to bring participants into private groups during the #StopTheSteal efforts, allowing them to rapidly grow their groups and then utilize those groups to organize physical events protesting the election.[58]

In our research on disinformation around the 2020 election,[59] we describe the networks as being "wired" for the spread of mis- or disinformation. The same groups of accounts repeatedly activated to share dozens of false and misleading claims of voter fraud. In many cases, these claims flowed across network connections (and perhaps algorithmic structures as well) that had formed and become reinforced through the spread of misinformation about COVID-19. The activities of past disinformation campaigns are now embedded in the networks and algorithms that shape how information flows online.

Addressing Disinformation/ Protecting Freedom of Speech

To summarize: The proliferation of disinformation is a threat to democracy. Sociotechnical systems have become a catalyst for the spread of disinformation, enabling actors (old and new) to employ old methods in new ways, reaching vast audiences with limited investment. Disinformation isn't merely perpetrated *against* the citizens of democratic countries, it works *through* us, exploiting our biases and vulnerabilities and leveraging us in its production and spread. As it spreads, disinformation becomes embedded in our sociotechnical systems and distorts perceptions of reality in ways that may be impossible to unwind.

One potential solution to this societally critical challenge is more moderation by the platforms—for example, creating and enforcing policies that mitigate mis- and disinformation by removing problematic content and suspending or banning the accounts that spread it. However, moves from platforms in this direction have often been countered by concerns about infringement on freedom of speech. Those who intentionally spread disinformation (the "witting" agents) have repeatedly invoked these concerns about freedom of speech strategically, attempting to diffuse potential moderation by claiming censorship. For example, Russia's state-controlled media outlet, RT, recently published one opinion piece arguing that support of Facebook's suspension of President Trump was "un-American"[60] and another calling out Facebook for taking action against hate speech and arguing that the platform needed to choose "freedom of speech over censorship."[61] But there are more sincere concerns, for

example, that through their moderation practices platforms already have too much power in deciding what speech and which voices are heard,[62] and that solutions addressing misinformation in one part of the world may have detrimental effects elsewhere.[63]

The challenge of addressing disinformation has long been complicated by concerns about freedom of expression. Disinformation attacks democratic societies at specific weak points. For the United States, our freedoms of speech and press have historically been two of those vulnerabilities. According to Bittman, Soviet Colonel (later General) Ivan Ivanovitch Agayants, referring to active measures against the United States, once proclaimed that "if they did not have press freedom, we would have to invent it for them."[64] This was in the context of the Soviet Union's successfully exploiting journalists as unwitting agents in the spread of disinformation. Bittman remarked that American journalists—and US society more broadly—seemed unaware of the risks.

> But it is disturbing that relatively few American journalists recognize the significant potential for abuse offered by current interpretations of freedom of the press to the Soviet bloc. Communist disinformation campaigns not only injure the United States, they represent violations of First Amendment rights and sometimes place American journalists in the invidious position as unwitting messengers and even victims of hostile propaganda.[65]

In the midst of the Cold War, Bittman appeared here to be arguing for a weakening of interpretations of the extent of free speech in order to defend against these kinds of attacks. But giving up democratic freedoms for this kind of defense would seem like just as significant a loss in these "games" (as Bittman referred to them).

However, disinformation is on the rise again, perhaps more of a threat now than it was before, and these core tenets of democracy remain vulnerabilities. Disinformation campaigns still exploit the freedom of the press, for example, using fake media outlets and planted stories to spread disinformation.[66] In the age of the citizen journalist,[67] Bittman's admonitions about being more aware of our vulnerabilities should perhaps apply to all of us (those who use these platforms to share information), and to the platforms—and potential government regulators—who have to contend with these tensions between free speech and strategic manipulation at scale.

These tensions and trade-offs crystallize as we consider the participatory nature of disinformation. In the online realm, many (and in some cases most) spreaders of disinformation are sincere believers of the falsehoods they share. Though they may be driven by political motivation to participate in the

production and spread of these falsehoods, they are often not aware that what they are sharing is false. It may be difficult to align values, norms, and legal frameworks of free speech with policies that would censor these sincere believers from sharing those beliefs.

Considering the top-down and bottom-up dynamics of participatory disinformation, perhaps there is potential to differentiate between everyday people caught up in these campaigns as unwitting agents and the high-profile public individuals—including elected leaders and media pundits—who play an outsized role in directing the attention of their audiences to disinformation. Within (and occasionally across) certain contexts, researchers have been able to identify a small number of "superspreader" accounts that repeatedly help to catalyze the spread of disinformation online, for example, that related to the 2020 election[68] and COVID vaccines.[69] Typically, these are high-profile accounts with large followings that can instantly reach large audiences with their content. Instead of giving high-visibility accounts such as those of politicians a pass for being of "public interest," platforms may benefit from creating and enforcing policies that hold accounts with larger audiences to a higher standard. The suspension of President Trump from platforms such as Facebook and Twitter after the events of January 6, 2021, suggests that platforms are taking steps in this direction. But recent revelations about an internal "Whitelist" at Facebook suggest that there is still a long way to go.[70]

Platforms also need to consider the structural impacts of algorithmic and network manipulation that have taken shape through years of mostly unmoderated disinformation campaigns. Many of the high-profile, superspreading influencers who played a role in helping to spread harmful disinformation in 2020 around COVID and/or the US election gained their prominent position in the larger information ecosystem by exploiting the platforms. As the platforms work on addressing disinformation (and other toxicities), they may want to consider how to both prevent future exploitation and mitigate the impact of past exploitation. Updating features, recommendation systems, and moderation policies to improve future interactions is only the first step. Platforms also need to take note of the structural effects of previous disinformation campaigns. If the impacts are embedded in the structures, then the structures themselves need to change. Suspensions are one lever. But short of mass suspensions, network connections can be trimmed and the algorithms themselves can be inspected, reconsidered, and adjusted to take into account (and counter) the structural impacts of disinformation.

Surely, these kinds of mitigating actions will be met, especially by those caught up in the enforcement efforts, with complaints about censorship and violations of free speech. Many politically active social media users already complain of being "shadow-banned" by the platforms—that is, having their content

demoted by the algorithms and therefore receiving less visibility on the platform. These structural approaches to addressing online disinformation bring to the fore questions about whether or not freedom of speech extends beyond the ability to post content to the right to benefit from the algorithms and networks that allocate visibility to certain content. As disinformation researcher Renée DiResta has explained, freedom of speech doesn't necessarily mean freedom of reach.[71] Are certain users, based on their position in the networks, entitled to have their messages spread to millions of people? Are they entitled to maintain their position in the networks regardless of how they arrived there?

Designers of online systems have long employed a variety of different levers to shape how communities form on their platforms, from seeding early content and norms, to incentivizing participation, to moderating content.[72] For prominent social media platforms like Facebook, Twitter, and YouTube, addressing how the structure of their networks and the design (and training) of their algorithms are (and have been) manipulated to give outsized visibility to certain users and certain kinds of content is well within their purview as designers, as well as their rights and responsibilities as "good" corporate citizens in democratic societies that are increasingly threatened by the pervasive spread of disinformation.

PART FOUR

OTHER POSSIBLE REFORMS

14

To Reform Social Media, Reform Informational Capitalism

JACK M. BALKIN*

Introduction

Social media today contains lots of speech that is protected by the First Amendment but that people argue is harmful to individuals, to groups, and to democracy itself. Social media companies already remove or demote much of this speech but not all of it. Should we change First Amendment doctrines to give social media companies incentives to remove it?

Minor changes in the First Amendment doctrine—even those I think unnecessary—will not destroy our system of free expression. First Amendment doctrine is a moving target; it has been changing continuously from the beginning of the twentieth century to the present. And some proposals for increased content regulation, such as laws aimed at revenge pornography, don't even violate existing First Amendment doctrine, or at most require very minor adjustments.[1]

But for the most part, these debates focus on the wrong issue. The problem is not First Amendment doctrines that protect harmful speech. The problem is the health of the digital public sphere: in particular, whether the digital public sphere, as currently constituted, adequately protects the values of political democracy, cultural democracy, and the growth and spread of knowledge. Instead of tinkering with First Amendment doctrines at the margins, we should focus on the industrial organization of digital media and the current business models of social media companies.

Only a handful of social media companies dominate online discourse. In addition, the business models of social media companies give them incentives

* My thanks Lee Bollinger, Robert Post, and Geof Stone for their comments on previous drafts.

to act irresponsibly and amplify false and harmful content. The goals of social media regulation should therefore be twofold. The first goal should be to ensure a more diverse ecology of social media so that no single company's construction or governance of the digital public sphere dominates. The second goal should be to give social media companies—or at least the largest and most powerful ones—incentives to become trusted and trustworthy organizations for facilitating, organizing, and curating public discourse. Social media companies serve an important social function in the digital public sphere, but their business models prevent them from serving this function adequately and responsibly.

Given these goals, the focus should not be on First Amendment doctrines of content regulation, but on digital business models. To the extent that First Amendment doctrine requires any changes, one should aim at relatively recent decisions concerning commercial speech, data privacy, and telecommunications law that might make it harder for Congress to regulate business.[2]

Models of Regulation

Changing First Amendment doctrines of content regulation for social media would almost certainly require changing them for other media as well. If you want to ban conspiracy theories on Facebook, you will be required to ban them on cable television, email, and podcasts as well. There are three reasons for this.

First, a focus on social media is seriously underinclusive. The major sources of disinformation and propaganda in the United States are pre-digital media: talk radio and cable channels like Fox News.[3] The most important propagandist in the United States in the past five years has been Donald Trump, who, as president of the United States, was relentlessly covered by every form of media, not just social media. This is hardly surprising. In general, one should expect that the most powerful forms of propaganda and falsehood in society will often be spread by the most prominent and powerful actors in society. They will draw the most attention in many different forms of media.

Second, the internet is protean. It contains many different forms of communication and allows a wide range of different affordances. Social media constitute simply one phase in the continuous evolution of digital communications, and the applications we call "social media" today are rapidly evolving in their features and affordances.

One might argue that social media have special communicative power because they so greatly lower the costs of transmitting content to vast audiences. But the same could be said for many other internet media—email, podcasts, text messaging, blogs, and websites. Some of these, like social media, are push

media—they push content to end users (rather than requiring end users to search for content or travel to specific sites). Push media make it easier for content to spread widely and quickly. But email is also push media, and one can easily add push features to other digital media or create feeds with similar effects. The problem with treating social media as special objects of content regulation is that many different types of internet content can spread easily and rapidly to large numbers of people, and many more such applications will no doubt be designed in the future. That is the whole point of the digital revolution.

Third, content regulations that apply only to a specific type of media require a special regulatory hook. For example, government licenses the right to broadcast because (given broadcast and reception technologies) spectrum is scarce. But there is no special regulatory hook that allows government to regulate content on social media more closely than other digital media—such as free-standing webpages, email, text messaging, or podcasts—or pre-digital media, for that matter.

Government justifies broadcast licenses because spectrum is scarce, so that only a limited number of people can broadcast without interference. Hence broadcasters are treated as trustees for a larger public. But the public trustee model justifies exposing audiences to a broader range of differing views, not constricting the number of viewpoints. So even if we adopted this model for social media, it would not justify most types of content regulation.

In any case, the model doesn't make much sense for social media. Social media feature easy and abundant access for speakers, not scarcity. This abundance produces a different kind of scarcity—the scarcity of attention in people's minds, which drives our current attention economy. But the scarcity of attention in human minds does not justify stricter government regulation of the content of speech. Where public discourse is concerned, government cannot restrict content because it wants people to think about some things more than other things.

One might argue that, quite apart from scarcity, social media, like broadcast television, are uniquely pervasivemedia.[4] The Supreme Court has used this argument to justify banning indecency on broadcast media in order to protect children who might be watching or listening at home.[5] But (1) what worries people about social media is not primarily indecency; (2) people are worried about the viewing habits of adults, not children; and (3) people access social media on phones they use everywhere, not just at home. In any case, whether the "pervasiveness" theory made any sense concerning television in the 1970s, it makes little sense today. Pervasive media are media that comprehensively shape one's culture and that one cannot easily avoid. That argument, taken seriously, would also apply to cable television, a far more pervasive medium that has always been treated differently than broadcast television, and to many other digital media, including email.

Still another approach would treat social media as public utilities. But the public utility model does not justify greater content regulation—although, as I describe later on, it might be useful for other purposes.

Governments treat companies as public utilities because they have something close to monopoly power *and* they provide an essential service that virtually everyone needs. Classic examples are companies that provide water, telephone services, and electrical power. A small number of social media companies—Facebook and its affiliated applications, and Google's YouTube—are very large, but most are not. Twitter, Telegram, TikTok, Snapchat, Reddit, and a host of small far-right social media companies that have come and gone into existence (like 8chan) do not have anything close to monopoly power. Nor are Facebook and YouTube essential in the same way that water and electricity are essential. I can get along pretty well without a Facebook account (I do not have one), but I could not do without access to water and electricity.

There *is* a plausible candidate for a digital public utility, but it is not social media. It is broadband access. Internet communication is structured in layers.[6] At the bottom are basic internet services: TCP/IP and other basic internet protocols; digital broadband and cellular companies (and their investments in hardware and infrastructure); the domain name system; and caching and defense services like Cloudflare and Akamai. Above these basic internet services sit the many different applications that people use to create and distribute digital content, including websites, blogs, e-mail applications, and social media. Without these services lower in the digital stack, social media speakers—like all other digital content providers—would be unable to reach their audiences.

There is a good argument that broadband internet access—whether mobile or landline—has become essential for modern commerce, education, and political and social life, and therefore broadband companies, like telephone companies before them, should be treated as public utilities. Indeed, they already fall under the jurisdiction of the Federal Communications Commission, and if the FCC treated broadband as a telecommunications service (as it once did), it would have many public utility obligations.[7]

The standard reasons to treat an enterprise as a public utility flow from its quasi-monopoly power and the universal need for what it produces. The goals of public utility regulation are to maintain reasonable prices, to secure universal access, and to ensure the quality of continuous service for consumers. That is why water, power, and telephone services are usually treated as public utilities, and why adding broadband to the list makes sense.

Given these criteria, social media are a bad fit. The price of social media is free. Access is universal. And companies happily provide continuous service because they want as much attention from end users as possible. More to the point,

the public utility model does not justify content regulation. It justifies universal access, reasonable prices, and quality service.

One can try to shoehorn content regulation into the public utility model by arguing that an uninterrupted quality of service really means an uninterrupted quality of content moderation. That is because, as I'll argue later, content moderation is a crucial element of social media services. But this fact actually leads us away from a public utility model. In the communication and transportation industries, public utilities are usually treated as common carriers that may not discriminate in terms of what they transport or transmit. In the context of broadband, this would mean something like network neutrality rules—which is why advocates of network neutrality have pushed for classifying broadband as a telecommunications service under the Federal Communications Act. The argument for network neutrality is that broadband companies shouldn't be able to block or slow down particular content or applications in favor of other content or applications—unless companies are just applying neutral network management rules designed to ensure quality of service for everyone.

Apply this feature of public utility regulation to social media and you don't get a justification for increased content regulation. You get no content regulation at all. This explains why Justice Clarence Thomas, who, like many conservatives, is concerned about social media censoring conservatives' speech, has recently been attracted to a common carriage model for social media.[8] Thomas reasons that if social media are common carriers, they can't exclude users and block or demote content.

Of course, laws preventing social media from moderating any content would also make them useless for most people, as social media would quickly fill with pornography and spam. To be sure, we could avoid this result by separating out the various functions of social media into separate companies—assuming, of course, that there were viable business models for each. The existing companies would become common carriers; they would simply accept, collate and deliver posts when requested. Other companies would arrange for individualized end-user feeds and engage in content moderation, and still others would buy and sell advertisements and insert them into moderated social media feeds. But this arrangement would not get rid of content moderation; it would simply shift it to another set of companies.

The fact that content moderation is an important function of social media, however, does not mean that government should require it. This misunderstands how a system of free expression handles harmful, false, offensive, or uncivil expression that the state may not reach. The role of institutions is central to that story, and as I shall now explain, the problem we face today is that institutions in the digital public sphere have failed to do their jobs.

The Public Sphere and Institutions

A system of freedom of expression operates in a public sphere—the sociological phenomenon in which people exchange ideas, knowledge, and opinions with others. Whether the public sphere functions well or poorly depends on whether it furthers the central values that justify and support the freedom of expression. In this essay, I will emphasize three of these values: the protection of political democracy, broad opportunities for cultural participation (what I call cultural democracy), and the growth and spread of knowledge.

A well-functioning public sphere involves much more than individual speakers. It also features a wide range of different institutions for circulating ideas and opinions and disseminating knowledge, information, and art. These include book publishers, broadcast media, movies, and every sort of communications medium.

A well-functioning public sphere also needs lots of institutions that produce and curate knowledge according to professional standards, because otherwise the public will be uninformed. These organizations include universities, research institutions, museums, archives, and journalistic institutions.

The institutions that circulate ideas and opinions overlap with the institutions that produce knowledge. Media organizations promulgate ideas and opinions, and within these organizations journalists research and investigate news according to professional standards. Universities produce knowledge according to professional standards and also disseminate it to their students and to the general public. But the two kinds of institutions are not identical in all respects, and both are necessary.

Without trustworthy and trusted professional institutions for producing knowledge, public discourse will just be opinions chasing other opinions. But knowledge production by itself is not enough. A well-functioning system of free expression also needs lots of different institutions for disseminating ideas and opinions, and the more the better. If only a small group of people have access to knowledge, the health and vibrancy of the public sphere will also suffer. And if there are only a few institutions for disseminating ideas and opinions, the system of free expression may be skewed and there may be bottlenecks to both information and knowledge. To borrow a phrase from Justice Hugo Black, a well-functioning system of free expression needs "the widest possible dissemination" from lots of "diverse and antagonistic sources" of information.[9]

Although most of the institutions that enrich the public sphere are privately owned, government may also contribute through investments in and subsidies for information infrastructure, educational and scientific institutions, and scientific research. Government may also fortify and enrich the public sphere through

collecting, collating, and curating knowledge in public libraries, museums, and archives, and through producing its own knowledge—for example, about public health, weather, and agriculture.

A system of free expression, in short, is more than a set of laws that guarantee free expression, and more than a set of speakers. It also needs institutions for producing, curating, and disseminating knowledge. And it needs a multiplicity and diversity of these institutions.

These institutions have important roles to play in a well-functioning system of free expression. Collectively, these institutions should help further the values of political democracy, cultural democracy, and the growth and spread of knowledge—and thereby promote the health of the public sphere. Therefore, in this essay, when I speak of the public function or the social function of a certain kind of institution or a certain kind of media, I am making a normative claim: I am asserting that this is the role that a particular institution or medium *should* play in promoting the health and vibrancy of the public sphere, whether or not it actually does so.

The digital revolution shook up the existing system of free expression. It disrupted and transformed older pre-digital institutions and created a new kind of digital public sphere with a new set of digital institutions.

Social media form one of these new institutions. They have three basic functions in the digital public sphere:

First, social media *facilitate public participation* in art, politics, culture, and the spread of knowledge and information.

Second, social media *organize public conversation* so people can easily find and communicate with each other.

Third, social media *curate public opinion*, not only through individualized results and feeds, but also through enforcing community standards and terms of service. Social media curate not only by taking down or rearranging content, but also by regulating the speed of propagation and the reach of content.

Unlike broadband companies and telephone companies, whose function is to be neutral and efficient carriers of content, social media are curators who organize and moderate content. Content moderation is not an optional feature of social media. It is central to its social function in a well-working digital public sphere. And it is what makes social media valuable to end users.[10] In the earliest years of the internet, the importance of moderation quickly became obvious in bulletin boards and newsgroups. In any digital medium open for public discussion in which strangers could interact with each other, moderation was necessary or the site would be flooded with spam and pornography or the discussion would quickly degenerate into trolling and abuse. The problems usually came from only a relatively small percentage of users, but digital technology gave them the power to make everyone else miserable.

Social media companies learned the same lessons as internet pioneers. They originally imagined themselves solely as technology companies rather than as curators of content. But they were soon inundated with complaints from end users about bad behavior by other end users. Faced with increasing costs to their reputation, and in order to keep people from leaving their sites, social media companies like Facebook and Twitter slowly began to create bureaucracies and algorithms for content moderation.[11] People may well criticize companies for performing these functions poorly. And I will criticize them in this essay for the ways in which their curation is skewed by their business models and their incessant quest for the monopolization of audience attention. But the fact that social companies organize and curate end-user contributions badly does not mean that this is not part of their appropriate function in the digital public sphere.

The Role of Private Content Regulation in a System of Free Expression

Well before the digital age, institutions that curated content and made judgments about quality and civility were important elements of the public sphere. During the twentieth century, for example, the people who owned and ran movie production studios, broadcast media, book publishers, and newspapers regularly curated public discourse. They picked which topics to emphasize and downplayed others. They restricted and sanitized the content they published and broadcast for mass audiences. Shaped both by professional standards and market incentives, twentieth-century media companies usually limited speech viewpoints far more than the law required. Movie studios did not produce, and book publishers did not publish, every manuscript they received. Newspapers did not run every letter to the editor or publish every story pitched to them by their reporters. Nor did they treat every argument made in the public sphere as equally worthy of publication or amplification.

I do not mean to suggest that the twentieth century was a golden age. In many respects its system of private speech regulation was seriously defective. Many valuable viewpoints and ideas were downplayed or excluded, and many valuable speakers and artists were ignored or censored. My point, rather, is that a system of free expression has lots of content curation and content regulation. But in a free society, most of it is not done by the "negative" state, by which I mean the aspects of government that regulate expression through criminal and civil penalties. Most regulation and curation of content is performed by members of civil society; by institutions like mass media that create and disseminate content; and by institutions that produce and disseminate knowledge like schools, universities, libraries, archives, museums, and research institutions. In addition, the "positive" state promotes and privileges certain ideas and certain kinds of

information through public education, through building or subsidizing communication infrastructures, through subsidies to media, science, and the arts, and through producing its own information.

People who defend the freedom of speech often invoke an abstract notion of "counterspeech" by fellow citizens that will somehow secure the promotion of knowledge or the protection of democracy. But that is not how the public sphere actually works. The counterspeech may not occur; it may not occur in sufficient amounts or with sufficient eloquence; and it may not occur quickly enough (for example, to deal with lies and conspiracy theories). There may not be enough counterspeech, it may not be efficacious, and in some cases, it may be irrelevant because the damage to privacy or self-worth may already have been done.

Whether your favorite justification for free speech is that it leads to truth, secures self-government, or realizes cultural democracy, you will need far more than a bevy of individual speakers offering contrasting views. A well-functioning public sphere relies on multiple institutions and devices, some private and some public, to set agendas, judge assertions, produce and preserve knowledge, curate content, promote civil discourse, and protect democracy. Without these institutions, the system of free expression will degenerate and fail to produce the goods that justify it.

Three Ingredients for Success

In sum, a healthy public sphere needs three things: First, there must be public-regarding institutions governed by professional norms for curating information and producing knowledge. Second, there must be a wide variety of different institutions and actors in society engaged in content production, curation, and regulation. There must be many sources of knowledge and cultural production. When only a few large institutions effectively control what people see, hear, or read, this undermines the health and vibrancy of the public sphere. Enforcement of cultural norms by only a few powerful groups or organizations, even with the best of intentions, can be oppressive.

Put another way, there need to be diverse and antagonistic sources of knowledge production and dissemination, which means there must be diverse and antagonistic curators and content regulators. As noted above, this requirement means more than simply having lots of voices that disagree with each other. There must be also different institutions for knowledge production that are public-regarding and that have professional norms that guide how they produce, organize, and distribute knowledge and opinion. Moreover, because professional standards are no guarantee of sound judgment, and because institutions, no matter how well-intentioned, may fail to recognize their own limitations, biases, and flaws, there must be many different players who can check and criticize each other.

Third, taken as a whole, institutions and professions for producing and promulgating knowledge must be generally trustworthy and trusted. When they prove untrustworthy or are not trusted, the public sphere will decay, even if legal rules protecting freedom of speech are still in place. Weaken the institutions or destroy trust, and the public sphere becomes a rhetorical war of all against all, where no one is believed except the members of one's own tribe, and people cleave to whatever beliefs are most comforting to them.

Bad Incentives

In a relatively short period of time, social media have become some of the most important institutions in the digital public sphere. They are different from twentieth-century mass media because they don't produce their own content. Instead, they facilitate, organize, and curate content provided by end users. They are key institutions for democratizing freedom of expression.

The early promise of social media, like the early promise of the internet generally, was that they would promote a diversity of views and offer alternatives to dominant cultural gatekeepers. They would also support the growth and spread of knowledge by lowering the costs of knowledge production, dissemination, and acquisition. To some extent, this promise has been realized. Widespread access to digital communications has also helped people scrutinize professions and institutions and disclose their flaws and failings. But social media, like the internet more generally, have also disrupted norms of civility, undermined professionalism, and helped people spread distrust in knowledge-producing institutions and in democracy itself.

Even though the public function of social media companies is different from that of newspapers and broadcasters, the previous analysis applies to them as well. For the digital public sphere to work properly, social media organizations must be guided by public-regarding and professional norms. And they must also be both generally trusted by the public and trustworthy.

In fact, the largest social media companies understand that the public expects them to act in the interests of society (and democracy). They engage in public relations campaigns asserting that they are trustworthy actors who are acting in the public interest. One can also see the beginnings of the development of professional norms in the creation of distinctive content moderation bureaucracies in the largest social media companies. Facebook has also created an independent Oversight Board to advise it about content moderation.

Of course, despite all this, large social media companies regularly fail to act in the public interest. My point, however, is that these companies understand the public's expectations about their appropriate social role as trustworthy facilitators, organizers, and curators of end-user content.

The problem is that the business models of these companies systematically undermine these goals. Social media companies are among the most prominent and powerful examples of what Shoshana Zuboff has called "surveillance capitalism"—providing free services in order to collect data about end users and others who interact with them.[12] Companies then monetize these data through sales of behaviorally targeted advertising, and they use data about end users to prod and manipulate end users to serve the company's ends.

Of course, media and advertising attempt to engage and persuade people all the time. By "manipulation," I mean more than simply persuading someone to do something they would otherwise not have done. Manipulation means using a person's emotional vulnerabilities or cognitive limitations against them to benefit the manipulator (or the manipulator's contractual partners) and harm the person manipulated. One also manipulates if one uses people in this way without concern for the harm that results.

There is an inherent conflict between business models based on surveillance capitalism and the social functions of media in the public sphere that I described earlier in this essay—the promotion of political democracy, cultural democracy, and the growth and spread of knowledge. The pursuit of profit has always put media organizations potentially at odds with their public functions in the public sphere. But the rise of surveillance capitalism and the fact that only a few large companies structure online discourse have made these conflicts particularly severe.

First, social media companies have economic incentives to maximize end-user engagement and attention because this helps them sell advertisements. To be sure, social media companies do not have incentives to allow and promote every kind of content that might attract attention, because some content repels end users and scares off advertisers. Nevertheless, companies have incentives to promote highly engaging material even if some of that material turns out to be false or misleading, undermines trust in knowledge-producing institutions, incites violence, or destabilizes democracies. The most engaging material is content that produces strong emotions, including negative emotions like fear, anger, and prejudice. This includes conspiracy theories and political propaganda that undermine trust in democratic institutions, and false information about public health.

Second, social media companies have incentives to collect as much information as possible about end users. This allows them to predict what will engage their users and maintain their attention. And the ability to predict the tastes and preferences of consumers and keep them engaged allows social media companies to sell more advertising and make more profits. Despite companies' repeated assertions that they respect privacy, they have few incentives to protect it all that much, because privacy conflicts with their business model. For the

same reason, they also have little incentive to educate end users, who may have little idea about what data are collected and how they are collated and used.

Social media companies collect data not only about their own end users, but also other people who are not even members—for example, by placing trackers in third-party sites. In addition, collecting data about some individuals allows companies to make inferences about others who interact with these individuals or are similar to them. Thus, a person's privacy is affected not only by what they do but by what other parties do or fail to do, creating a collective action problem for individuals who want to protect themselves. Thus, social media have incentives to take advantage of individuals' lack of understanding and the limitations of their individual decision-making.

Third, social media companies have incentives to underinvest in content moderation. Social media companies impose significant externalities on the societies in which they operate, including political unrest, democratic decay, ethnic violence, the spread of false information about public health, addiction, increased rates of depression and harms to mental health, and manipulation by political demagogues and unscrupulous advertisers.

Content moderation that avoids or mitigates these harms is difficult and expensive to perform quickly, accurately, and at scale. Because social media companies do not fully internalize the social costs of their activities, they will tend to skimp on content moderation that does not increase their profits. And when companies have monopolistic or oligopolistic power, like Facebook and Google, they can devote even less effort to content moderation and still retain their base of end users.[13]

To save money, Facebook uses algorithms for content moderation that are often imperfect. It contracts out content moderation to third-party contractors that employ low-wage laborers who often work under very difficult and stressful conditions. Or it relies on subsequent complaints by end users, governments, and civil society organizations to spot problems.

Because social media companies do not have to internalize the social costs of their business practices, they will amplify too much content and take down too little content that generates these externalities. Moreover, as Facebook's history demonstrates, social media companies have incentives not to discover the problems they cause, much less to be transparent about the effects of their policies or to allow others to investigate them.

Despite all of these bad incentives, social media companies do attempt to limit some of the damage they cause. They also work with governments around the world to identify and remove abusive and illegal content. But content moderation systems, even if staffed by well-meaning individuals, cannot overcome the basic incentive structure created by the underlying business model of surveillance capitalism.

Facebook offers a good example. Significant parts of the company have adopted quasi-professional norms and are more or less public-regarding. Company personnel work at content moderation and cooperate with civil society organizations, governments, and other social media companies to prevent harms caused by social media use, both to end users and to societies. Facebook has also created an independent Content Oversight Board to review some of its content moderation decisions. This part of Facebook puts out fires. The problem is that the other half of Facebook—the operations that make the company money—helps start the fires.

Devoted as it is to growth at all costs, Facebook has repeatedly treated the damage its business causes primarily as a public relations problem. It apologizes profusely and promises to do better, but the engine of surveillance capitalism proves stronger than professions of good intentions. Economic incentives inevitably lead to more unexpected damage and a new round of apologies by the company, and the cycle repeats itself.

Earlier I said that a healthy public sphere requires trusted and trustworthy private institutions that produce knowledge and curate content according to public-regarding and professional standards. The digital age has severely weakened the institutions that serve these functions. And no new comparable institutions have arisen to take their place. The most powerful new players in the digital public sphere—social media—have not become the trusted and trustworthy organizers and curators of public discussion they need to be.

The result is that we have created a new kind of digital public sphere that lacks the connective tissue of institutions and practices necessary to sustain the underlying values of freedom of speech. We have the formal liberty of speech without the accompanying institutions that help protect and promote the values that justify that liberty: cultural and political democracy and the growth and spread of knowledge.

One might blame this institutional failure simply on the democratization of free expression. I think this is too easy an answer. Democratization is inevitably disruptive to existing institutions and norms. But the more important accelerant is the development of powerful models of informational capitalism, which occurred only in the last fifteen years. These developments, in turn, were not simply the result of entrepreneurial genius and vigor. They were made possible by a host of deregulatory changes in different areas of law that made it possible to construct digital behemoths like Facebook, Google, and Instagram.

The Goals of Social Media Reform

Social media reform requires *both* a less concentrated media ecology *and* the promotion of professional and public-regarding norms. Each of these goals requires

changes to current industry structure and current business models. A handful of players will continue to dominate online discourse as long as law allows them to do so. And social media will not become trustworthy institutions that can play their appropriate social role in the digital public sphere as long as they retain their current business models. To reform social media, we must reform informational capitalism.

Changes in many different areas of law made today's version of surveillance capitalism possible.[14] Therefore no single tweak in the law will be sufficient. Instead, multiple reforms are required. In the remainder of this essay I will focus on two central areas of reform. The first is competition law. The second is privacy and consumer law (in the digital age, the two are increasingly inseparable). At the close of the essay, I will briefly mention how we might employ a third major lever of reform—intermediary liability and intermediary immunity rules—in the service of competition and privacy reforms.

Competition Law

By competition law, I mean more than litigation under current US antitrust laws. I also include new laws and administrative regulations that seek to regulate industry structure. The purpose of such laws need not be limited to increasing consumer welfare, to promoting innovation, or even to the broader purpose of making markets more efficient. In American telecommunications law, competition regulations (for example, media concentration rules) have traditionally had the additional purpose of promoting democracy and the health of the public sphere.[15] That is their central purpose in the discussion that follows.

Competition law reforms aim at a less concentrated social media industry with many more different kinds of companies. Increasing competition among social media companies promotes diversity and pluralism rather than making social media more professional and public-regarding. (In the next section I will discuss reforms directed to that second purpose.)

Suppose that instead of one enormous company, Facebook, with its subsidiaries Instagram and WhatsApp, there were fifty little Facebooks. Imagine also that competition law prevented large companies from easily buying up smaller startups, co-opting their features, and forestalling their rise as potential competitors.

Increased competition among social media companies would encourage them to develop new innovations and affordances to gain customers. They might develop different moderation practices, which would help produce diverse content.

Some companies might specialize and cater to certain kinds of content or certain kinds of audiences. They might develop special kinds of expertise in moderation.

Smaller companies might specialize in quality content moderation to attract end-users. Some companies might be able to devote more attention to specialized audiences, particular languages, or specific geographical regions. Large companies like Facebook may struggle to understand and effectively moderate posts in multiple languages from a wide variety of cultures around the world. The problems of global content moderation may become even severe as Facebook attempts to incorporate ever more different forms of media (for example virtual and augmented reality). For these reasons, it is possible that some (but not all) aspects of content moderation do not scale very well.[16] And to the extent that is so, an industry with many smaller companies might be better than one large company that insists on promulgating and enforcing a single set of content moderation rules around the globe.

One might worry that too many social media companies would fragment the digital public sphere and lead to increased polarization. In fact, the largest companies already fragment and polarize the public sphere through their algorithms and individualized feeds.

Social media companies offer social media services in exchange for end-user attention (and end-user data). More competition for scarce end-user attention means that social media companies will have greater incentives to give end users what they want from social media.[17] This might include content moderation policies that end users prefer, greater transparency, less confusing and manipulative interfaces, more control over the content end users receive, and more procedural protections.

Greater competition might also make the digital public sphere less vulnerable to the decisions of only a few companies. Currently fake news and propaganda operations only have to learn how to manipulate Facebook's and YouTube's algorithms to gain access to vast audiences.[18] They might find it more difficult to manipulate fifty different social media sites, and competition among these sites might improve the quality of social media algorithms in identifying and removing such content.

But fifty little Facebooks might be less efficient at moderating at scale. They might have incentives to create what evelyn douek calls "content cartels,"[19] sharing blacklists and applications for content moderation. Some kinds of cooperation—involving spam and child pornography—might be beneficial. But others might reduce valuable forms of diversity.

Finally, if you think that cooperation between government and social media is helpful in identifying and curating certain kinds of harmful content—for example, terrorist recruitment, child pornography, and false public health

information—then it is harder for governments to deal with fifty companies than with just a handful. On the other hand, if you fear that government will co-opt social media companies for improper purposes, having many different companies is an advantage, because it is harder to co-opt them all.

Network effects may undercut the goals of a competitive social media industry. Facebook is large because it benefits from network effects: People want to be on the site because other people are on the site. The greater the number of people who join an app, the more valuable it is, not only to the end users but also to advertisers. If network effects are strong, a world of fifty little Facebooks may soon evolve into a world with only one very large company, surrounded by many little companies struggling to survive competitively.

But the existence of network effects, and, more importantly, who benefits from them, is a result of legal rules. Law can change the distributive consequences of network effects to make smaller companies viable.[20] For example, governments could require interoperability between social media networks. Thus, when you sign up with Facebook, you can also connect with all of the social media networks that interoperate with Facebook. This approach treats Facebook a bit like an early twentieth-century telephone company that was required to connect with other telephone companies. A slightly different approach is to change intellectual property law and the Consumer Fraud and Abuse Act to allow programmers access to social media companies' application programming interfaces (APIs). This would allow entrepreneurs to build super-applications that can create feeds of multiple social media applications, even those with different features and affordances. These super-applications, in turn, would make it easy for people to join many different applications and, in some cases, communicate between them.[21] The point of these reforms is to redistribute the benefits of network effects from a few large companies to smaller companies and to the public as a whole.

Vertical integration of digital businesses is just as important to the health of the digital public sphere as horizontal integration. Facebook and Google do far more than social media and searching. They are also the world's two largest digital advertising brokers. This creates a serious conflict of interest. These companies compete with other media organizations for audience attention and advertising revenue. But they also run the advertising auctions. Not only do Facebook and Google set the rules of digital advertising, they also have access to more data than any other companies. This amplifies their competitive advantage.

The largest social media companies should not also be the largest digital advertising brokers. Separation of functions might help newspapers and other media companies to compete more effectively with social media and negotiate better bargains with the largest digital companies.

Here one might return to the idea of treating social media companies as public utilities, but for a different purpose. Governments may require that public utilities operate only one kind of business and not vertically integrate. But since governments can independently pass laws requiring separation of functions without labeling a company a public utility, the term does little additional work. Moreover, it has a signal disadvantage. Treating social media companies as public utilities accepts that the largest companies will remain digital behemoths and exercise monopoly or quasi-monopoly status for the foreseeable future. It essentially throws in the towel on regulating the size of these companies.

Privacy and Consumer Protection Law

Competition law is not a complete solution to social media regulation. It does not address many of the issues that arise from data collection and end-user manipulation. In some cases, competition law remedies can make these problems worse. Even if we broke up Facebook, each of the fifty little Facebooks would still be practicing surveillance capitalism. Because they would be competing for a smaller share of audience attention—and end-user data—they might adopt increasingly manipulative and abusive practices. And open APIs might allow lots of different companies to gain access to information about end users, including end users' social graphs and contact history, enabling even more manipulation and abuse.

If we want social media companies to become responsible organizers and curators of public discourse, we must impose obligations on how they collect, analyze, and use data. This is the job of privacy and consumer protection law.

Unlike Europe, the United States still lacks a comprehensive digital privacy law. The European Union's General Data Protection Regulation (GDPR) is a complex regulatory system. But its application to surveillance capitalism is not yet fully worked out. Depending on how the European courts interpret it, the GDPR could force companies to abandon many troubling features of surveillance capitalism or leave most of those features untouched.

My own contribution to these issues is the concept of information fiduciaries.[22] I've argued that the digital age has created great asymmetries of power and knowledge between the digital businesses that collect data from end users and the end users themselves.[23] These asymmetries of power and knowledge create special vulnerabilities for end users that are the traditional concern of fiduciary law.[24] Therefore, I've argued that businesses that collect data from end users must assume fiduciary duties of confidentiality, care, and loyalty to the people whose data they collect and use. They also must ensure that these duties

"run with the data," so that if companies share the data with other companies the latter must also assume the same fiduciary obligations.

These fiduciary duties apply not only to social media companies, but to any companies that collect and monetize end-user data. This is important because the internet of things allows many different objects and appliances to collect personal data. Fiduciary duties must also apply to smart homes, self-driving cars, and personal digital assistants. Moreover, social media companies collect data everywhere they can, and not just from the people who use their social media applications. Even if you have no Facebook account, Facebook still collects data about you from trackers it places on many of the websites you visit. Even if you never visit YouTube, Google can still collect data from anyone who uses an Android phone.

Facebook and Google are not simply social media companies. They offer multiple services that allow them to collect and analyze consumer data. In fact, it is better to think of Facebook and Google as surveillance companies that offer free social media services—among many others—to collect and monetize end-user data.

The information fiduciary model would disrupt these practices and cause digital companies to alter their ways of making money. Information fiduciaries may not use the data they collect about people to manipulate them or harm them. The information fiduciary model argues that data privacy is a matter of trust between companies that collect data and those who use their services. The central obligation of an information fiduciary is not to abuse or betray that trust.[25]

Suppose that the information fiduciary concept were implemented in federal law—for example, through a framework statute that set out basic principles of data collection and use and gave a federal agency power to implement them through regulations and enforcement actions. Then significant features of current digital business models would have to change. That is because digital companies currently have few qualms about externalizing the social costs of their activities onto the people whose data they collect. Under an information fiduciary model, social media companies would have to take reasonable steps to organize their digital advertising system in ways that would not foreseeably cause harm to or create conflicts of interest with their end users. This advertising system includes both the collection of end-user data and the algorithms that companies use to engage audiences and maintain their attention in order to sell advertisements.

The information fiduciary model will only have indirect effects on social media companies' editorial decisions. Its goal is not to specify particular content moderation policies but, rather, to promote the internalization of professional norms and public-regarding behavior. Under current business models, social media companies care most about advertisers and advertising revenue. End users are

valuable primarily as a resource for collecting data that feeds algorithms and as a product to be sold to advertisers. But if social media companies are information fiduciaries, they must put the interests of end users first, ahead of advertisers—and they must restrict their profit-seeking strategies so that they are consistent with their duties of care, confidentiality, and loyalty.

Generally speaking, fiduciaries may not use information they receive about their clients or beneficiaries in ways that are likely or calculated to harm or undermine their clients' or beneficiaries' interests. Because information fiduciary rules are designed to ensure privacy and consumer protection, they will aim primarily at data collection and use.

First, fiduciary duties will limit how and when companies collect data; the conditions under which they distribute or sell data to third parties; and the ways they use data in amplifying content, organizing feeds, and making recommendations to their end users.

Second, fiduciary duties will alter how companies employ targeted advertising. Not all targeted advertising is alike. For example, *contextual* advertising—advertising targeted based on the particular site one is visiting—does not require the collection of very much information about end users. In contrast, *behavioral* advertising, which depends on prediction models, targets consumers based on a rich dossier of information collected about them. Business practices that encourage the collection of as much data as possible about end users also encourage manipulation of end users. One goal of the new federal agency mentioned above would be to specify fair practices for targeted advertising consistent with fiduciary obligations.

Third, fiduciary duties will change how companies design interfaces that coax end users into disclosing data about themselves or making choices that benefit the company at the expense of the end user.

Fourth, algorithmic advertising and recommendation systems that are self-consciously designed to addict or manipulate end users would be a breach of the duty of loyalty.

Companies will probably challenge regulations that require companies to change their business practices as unconstitutional under the First Amendment. But privacy and consumer protection regulations are not aimed at the particular editorial decisions in the company's content moderation system. They are aimed at how companies collect and use data gathered from end users. Restrictions on data collection and use have a long history in privacy regulation, and courts generally do not regard them as raising First Amendment problems.[26] Fiduciary requirements of confidentiality have an equally long history.[27]

Earlier I noted that one half of Facebook sets out norms of appropriate content and behavior on their site and then enforces those norms through a combination of algorithmic filtering and human moderation practices. The company's

content moderation system acts like a fire department that puts out fires. But I also pointed out that the other half of the company, which collects data and deploys algorithms to amplify content and maximize engagement, has incentives to start fires.

If social media companies become information fiduciaries, they will have legal incentives to act more like fire departments and less like arsonists. That is the predictable consequence of regulations designed to change how social media companies collect and use data to make money. In fact, if social media companies cause fewer fires to spread, that may actually ease burdens on their content moderation systems and make them more effective.

Information fiduciary obligations would also require social media companies to be consistent between the value commitments they make to the public and the values they actually enforce through their algorithms. For example, if social media companies publicly represent that antivaccination propaganda is harmful to their end users—either in their statements to the public or in their community standards—they may not turn around and employ end-user data to amplify or recommend antivaxxing sites. Nor may they proclaim that they respect their end users' privacy and then design their software interfaces to encourage end users to disclose data that make it easier to manipulate them. Once a social media company decides to collect and monetize end-user data under conditions of severe asymmetry in power and knowledge, it becomes an information fiduciary, and it takes on relational obligations of trustworthiness toward its end users that it might not otherwise have.

Intermediary Liability and Intermediary Immunity

Intermediary immunity is a familiar target of proposals for reforming social media. Section 230 of the 1996 Telecommunications Act[28] immunizes interactive computer service providers from liability for certain kinds of content that appear on their platforms; it also immunizes them for their decisions to remove content.[29]

Critics argue that social media companies are abusing their immunity from liability, and that Congress should withdraw the immunity to make them behave properly. But if the goal is to make social media companies more trustworthy curators and organizers of public discourse, it is a bad idea to abolish intermediary immunity entirely. That will simply give companies incentives to remove too much content.[30] Rather, government should condition intermediary immunity on social media companies' adopting business practices that ensure their trustworthy and public-regarding behavior. In other words, we

should leverage intermediary immunity to further the kinds of reforms I've just described.

The full scope of the Section 230 immunity is not required by the First Amendment.[31] The difference between what the First Amendment requires and what Section 230 provides is, in effect, a regulatory subsidy. It not only eliminates potential damage awards against companies but also lowers the costs of repeated litigation over content moderation policies.[32]

Instead of abolishing intermediary immunity entirely, government should employ the techniques of the positive state. It should condition the immunity on social media companies' agreeing to accept a new set of public interest obligations for the digital age. Instead of imposing the public interest obligations of twentieth-century broadcasters—such as the long-abandoned fairness doctrine—government should articulate a new set of public interest obligations appropriate for digital platforms.[33] Unlike the twentieth-century model, these obligations should extend beyond social media companies to other companies that practice surveillance capitalism, including personal digital assistants, search engines, commercial platforms, internet-of-things businesses, and robotics companies.[34] Most digital companies—including social media companies—will have incentives to accept the deal, but those who decline remain fully protected by the First Amendment.

What should these digital public interest obligations include? Here are three examples, drawn from the discussion above:

First, digital businesses who want the Section 230 immunity must agree to be regulated as information fiduciaries. (The condition could be limited to companies of a certain size or with a certain number of end users.)

Second, social media companies (and, where relevant, other digital businesses) must allow interoperability for other applications, as long as those applications also agree to act as information fiduciaries.

Third, digital businesses must allow government regulators to inspect their algorithms at regular intervals for purposes of enforcing competition law, privacy, and consumer protection obligations.

Conclusion

The issues discussed in this essay—competition law, privacy law, and consumer protection—form the real battleground for First Amendment doctrine in the years to come. That is because the federal courts have decided a series of recent cases about commercial speech and data privacy that, read broadly, might restrict Congress's ability to regulate surveillance capitalism.[35] Key First

Amendment fights of the next several decades will concern whether Congress can undo horizontal and vertical integration in social media, whether commercial speech will receive ever greater protection from consumer protection regulation, and whether comprehensive data privacy laws are constitutionally permissible. These questions will prove crucial because the problems of social media today are not primarily problems of inadequate content regulation by states. They are problems of informational capitalism.

15

Follow the Money, Back to Front

YOCHAI BENKLER

> In the words of Chesterton, a journalist became one who wrote on the back of advertisements.
>
> Harold Innis

Social Media Are Not the Problem

American democracy does not have a social media problem. American democracy has an institutional and political-cultural problem manifested in and reinforced by media market imperatives since the 1980s. Beginning with televangelism and Rush Limbaugh in the 1980s and supercharged by Fox News since the 1990s, a large and discrete media audience comprised of White identity and Christian voters presented a lucrative business opportunity for outlets in a changing technological, regulatory, and business media environment. The new multichannel environment offered rich rewards for those willing to sell vitriol, outrage, and alienation from Black Americans, immigrants, and women claiming their independence, as well as from the professional and managerial class that held the high ground of what Limbaugh called "the Four Corners of Deceit": government, academia, science, and the media. Since the late 1980s, selling right-wing outrage has been big business, and its commercial success enabled it to take over the conservative media ecosystem.[1]

The present epistemic crisis is the product of the interaction between the political strategy of Richard Nixon and Ronald Reagan and the economics of media markets no longer dominated by one-newspaper towns and three broadcast networks. The former forged a distinctive, relatively homogeneous media audience alienated by the victories of the civil rights movement and the women's movement and reacting to the transposition of the New Left into the Me Generation with its focus on self-actualization and career success as a measure of worth. Nixon's Southern Strategy, combined with Reagan's racialized attack

on "the Welfare Queen" and his embrace of the Moral Majority, harnessed these alienated voters to the chariot funded by the 1970s emergence of Organized Business.[2] For forty years, until 2016, the business wing of the Republican Party successfully leveraged the outrage and fear of White and Christian identity voters to keep the base turning out for elections but kept control over actual government, repeatedly electing presidents who knew how to blow the dog whistle that kept the mass of voters turning out but who were members and loyal servants of the business wing of the party. Their strategy was only "new" on the background of the unusually controlled media environment typified by one-newspaper towns and three television networks that marked the height of "high modernism" in American media. The audience that Pat Robertson and Rush Limbaugh had rediscovered, and the paranoid style that Limbaugh bequeathed to Hannity, Carlson, and others, has a long and deep history in American political culture.[3] And the politics of hatred they stoke today is weak tea by comparison to its antecedents, from the conspiracies of treason undergirding the Alien and Sedition Act, through the exhortations to civil war and later lynching in the South, to the justification of industrialists shooting workers demanding an eight-hour workday and later yet Southern resistance to desegregation in the 1950s. All that has happened in the past thirty years is that these deeply antidemocratic streams in American political culture were harnessed by one part of the professional and managerial class against another, and that part of the elite lost control over its populist base as the business dynamic fed by this old-new audience spun out of control.

The business opportunity that alienated White identity and Christian identity voters emerged when a combination of new technologies and antiregulatory ideologies made targeting discrete, large specialty audiences the dominant business strategy relative to the strategy of targeting broad, centrist content that had typified media markets when readers and viewers had nowhere else to go. Changes started with UHF stations and the All Receivers Act in the 1960s, combined with deregulation of public interest obligations that allowed Evangelical Christian broadcasters to outbid mainline Protestant broadcasters. These formed the foundation of televangelism. These initial changes were complemented by dramatic technical improvements in cable systems' channel capacity and ground-to-satellite retransmission that underwrote the first superstations and the Christian Broadcasting Network (CBN), coupled with significant cable deregulation, and crystallized with the emergence of FM radio to its full potential in the 1970s, leaving AM radio looking for a low-audio-quality, low-cost format, for which Rush Limbaugh and talk radio were the answer. In a one-newspaper, three-TV-network market, the dominant business strategy had been to broadcast inoffensive materials to the center and aim for a share of the whole.[4] It allowed elites to more or less limit what the population

at large considered acceptable beliefs for well-socialized people to what elites themselves, for all their internal divisions, considered acceptable for a well-socialized member of the professional and managerial class to believe. In a multichannel environment, identifying a particular large-enough audience that desired unique content became the dominant strategy. And Pat Robertson and Rush Limbaugh road-tested and proved the strategy that could leverage the audience that Nixon and Reagan had assembled, the business strategy that Rupert Murdoch and Roger Ailes would then perfect in Fox News.

The outrage industry, once so forged, became a catalyst of increasingly asymmetric polarization in American politics. It is important not to misunderstand my focus on larger and longer-term political and cultural dynamics as implying that media did not, and are not, playing a critical part in the conflagration engulfing American democracy, even if media were not the initial accelerant. By supplying and competing on serving outrage, the media ecosystem created an ever-more-detached-from-reality audience demanding and rewarding increasingly strident, hate-filled media personalities and politicians. The result was a propaganda feedback loop: Media drew audiences with stronger identity-confirming assertions and brought them out to the polls to reward politicians who were present on these media,[5] and competed for audience share by policing each other and the politicians for identity consistency, not for truth. It became consistently harder for conservative media that sought to continue to be anchored in some semblance of reality and a shared polity to survive, as the demise of the *Weekly Standard* exhibited, and harder yet for conservative politicians to survive without toeing the increasingly unhinged line. Senator Richard Luger, who joined the Senate as a conservative Republican and left it as a moderate without changing his own votes, marked the long-term transition. Senator Pat Toomey, who replaced the avid but not unhinged conservative Arlen Specter as a Tea Party darling in 2010, marks the continuation of the same process, as he became the measured voice on his way out the door as the Tea Party was succeeded as the right-wing marker by the Q Shaman of 2021.

Social media came into being long after the propaganda feedback loop had already taken hold of the right-wing media ecosystem. In the first few years of the blogosphere, it was the left wing of the blogosphere that was more mobilized, but the asymmetry manifested in organizational forms and technology deployment, not in political extremism.[6] When social media arose to integrate decentralized authoring with a platform for delivering advertising and making money from decentralized authorship, it opened opportunities for new entrepreneurs seeking to take advantage of the same strategy that talk radio and Fox had already exploited effectively for twenty years. It was then that asymmetric polarization online aligned with the already asymmetric polarization on mass media.[7] The highest-quality data studies published in the past few years converge on the

finding that most social media sharing of false and hyperpartisan news is concentrated in a small minority of mostly over-sixty-five-year-olds with mostly conservative ideology[8]—that is, in the typical Fox News demographic—and that they do so because they seek reinforcement of their already existing views, not because they innocently surf social media and are exposed to false or extremist information.[9] Consistent with these scientific studies, a May 2021 survey by the Public Religion Research Institute found that consumption of far-right television sources Newsmax and One America News Network, followed by Fox News, was more highly associated with holding core beliefs of the QAnon conspiracy theorists than was true of people who did not rely on television news.[10]

The drivers of the worst forms of epistemic crisis in America, then, are not politically neutral, technologically mediated processes hitting unsuspecting citizens. Nor are they Russian operatives, though they gladly take advantage of the unbalanced American media ecosystem to project the appearance of more power than Russian efforts have ever been actually shown to exercise. Rather, the drivers are a combination of profit seeking and political opportunism, taking advantage of the particular historical confluence of political culture and market structure that has characterized the United States in the past forty years, to harness a large, socially alienated population to an advertising-supported business model that relies on intense engagement through identity confirmation and hatred of racial and gender minorities, on the one hand, and expert elites, on the other hand.

There is no empirical evidence supporting the proposition that regulatory solutions that require platforms to monitor and suppress discrete illegal posts or statements on social media have had a measurable impact on dissemination of disinformation and hatred.[11] There is evidence that sustained efforts by platforms at "deplatforming" some of the worst offenders do reduce the visibility of those specific speakers on social media,[12] but no study to date has shown that these kinds of speech and hatred as a whole are depressed as a result of these individualized decisions with respect to individual items of speech, or even broader decisions to exclude a particularly harmful individual speaker from social media platforms more generally.

The lack of evidence is hardly surprising if one understands the problem of hate speech and disinformation not as a problem of discrete bad actors, or as a problem of technologically mediated confusion for users who are good-faith truth-seekers, but as a problem driven by media market dynamics with deeper and longer-term sources and drivers. Just as the fast-food and packaged-foods industries optimize the fat, salt, and sugar contents of their products to keep customers buying more of their products, obesity epidemic be damned, and the tobacco industry packed addictive nicotine to make consumers dependent, killing millions, so, too, it is the profitability of stoking anger and providing

easy, identity-confirming outrage that underwrites much of the disinformation and polarization that pervade American media. On the background of that understanding, solutions must focus not on the whack-a-mole process of batting down falsehoods as they arise, or periodically deplatforming a bad actor, but on changing the payoffs to selling outrage and hatred as a business model.

Below I outline two approaches aimed to cause the outrage industry to internalize the externalities of its business model. The first, a reconsideration of defamation law, is limited and likely to be somewhat effective but is very much a double-edged sword that may do more harm than good, particularly when applied by a politically appointed judiciary such as the United States has in the early twenty-first century. The second is a hybrid regulatory–civil society approach, one that uses the power of the state to impose meaningful transparency on market actors but relies on decentralized peer production or on civil society organizations to harness that transparency to impose meaningful costs on the outrage industry. Like the first approach, this hybrid approach is far from a silver bullet, and it, too, has potential to go wrong. But it does have the benefit of harnessing more democratic processes to contain the worst abuses, and offers an example of an approach designed to avoid the inevitable fallibility of not only the state, but also the market.

Did *New York Times v. Sullivan* Give Us Alex Jones, Sean Hannity, and QAnon?

Alex Jones of Infowars is a snake oil salesman who draws buyers to his online store by feeding them outrageous rhetoric and hate-filled narratives. In the past few years, he has claimed that Hillary Clinton ran a pedophilia ring out of the basement of Comet Pizza in Washington, DC;[13] that a DNC staffer named Seth Rich, who had been the victim of a murder-robbery in Washington, DC, had in fact been murdered because he, not Russian operatives, had leaked the DNC emails to WikiLeaks in 2016;[14] that the Sandy Hook murders were a "false flag" hoax, and the grieving parents were lying about their loss;[15] and more. According to Similarweb, the site received about eight million visits a month as of the middle of 2021, and about one tenth of that number enters the Infowars store through which Jones sells various twenty-first-century "cures" with names like "X-3 Bio-True Selenium Combo" and "Survival Shield X-3."

Sean Hannity makes millions of dollars as both the host of the show with the highest or second-highest rating on cable television (he shares that position with Tucker Carlson, who offers a similar show on Fox News) and one of the leading stars of talk radio. His draw is an aggressive presentation style, often in long monologues, sometimes through the selection of interviewees, that draws

audiences by reinforcing their deeply held beliefs and fears about "others": the media, academia, the elites, immigrants, "the left," or the deep state. These broader narratives are sometimes punctuated by false stories that make diverse inflammatory statements about specific people—from claims associating John Podesta, then Hillary Clinton's campaign manager, with a "spirit cooking" ritual, through aggressive promotion of the Seth Rich conspiracy even after the Fox News network itself had retracted its hyped-up version of the story, to assertions that top prosecutors involved in the Mueller investigation had conspired to sell 20 percent of US uranium to Russia,[16] or that Dominion Voting Systems, Inc., altered the election results in the 2020 US presidential elections. Indeed, Dominion is now pursuing defamation suits against several public figures who made such allegations, including attorney Sydney Powell, MyPillow CEO Mike Lindell, and Fox News, whose various personalities named in the complaint—Tucker Carlson, Sean Hannity, Lou Dobbs, Maria Bartiromo, and Jeanine Pirro—spread the falsehoods about Dominion up and down Fox News' broadcasting schedule.[17]

Much of the time, these programs offer pure opinion or vague references to "the mainstream media" as a whole or to "the left" or "critical race theory" or similarly abstract objects of hate and derision. But I picked the handful of examples in the prior paragraph because they did, or could, result in defamation or other tort lawsuits as the more abstract exhortations were interwoven with specific, concrete narratives making derogatory and inflammatory statements about concrete individuals. Often these are general-purpose public figures. Indeed, Donald Trump's rise to political prominence was largely built on the strength of being the most prominent celebrity to embrace the false claim that President Obama was constitutionally ineligible to be president because he was not born in the United States. Sometimes—as were the cases of the owner of Comet Pizza, Seth Rich, and Chobani yogurt[18]—the victims are simply unlucky enough to be the wrong person (or business) in the wrong place so that they become a target of convenience in the quest for more money.

Even QAnon, often thought of as the clearest example of online conspiracy theory spinning out of control and infecting millions of people, is itself fan fiction of a narrative spun by Fox News and propagated by such diverse sources and personalities as former National Security Adviser Michael Flynn and Blackwater founder Erik Prince in support of Donald Trump's 2016 election.[19] If QAnon is fan fiction,[20] it is fan fiction based on a mass-media, elite-driven narrative originating in reporting by Fox News reporter Malia Zimmerman, using flight logs from billionaire pedophile Jeffrey Epstein's private plane, and asserting that Bill Clinton had flown to Epstein's "pedophilia island." That story, picked up and amplified for several days in the spring of 2016, was Fox News's ticket back to the hearts of Trump devotees after the network lost online prominence during

the primaries to more strident pro-Trump platforms like Breitbart. It became the Fox story most widely shared on Facebook in the entire 2016 election cycle.[21]

Ever since *New York Times v. Sullivan*, publishers of such hate-filled drivel are quite safe from suits in torts, particularly defamation.[22] The standard that plaintiffs have to fulfill, showing that the publisher acted with actual malice, including reckless disregard of the truth, has been applied strictly. One particular line of the progeny of *New York Times v. Sullivan* that is applicable to the outrage industry has been courts' willingness to rely on the very outrageous nature of the speaker or speech to find that no reasonable person could believe that what was being said was an assertion of fact, rather than frothing-at-the-mouth invective. To be defamatory, a statement has to have been a provably false assertion of fact,[23] and even a specific accusation that a person has committed a crime is not defamatory if the context suggests that the statement is "rhetorical hyperbole."[24] Indeed, as Judge Mary Kay Vyskocil of the Southern District of New York wrote when she dismissed Karen McDougal's lawsuit against Fox News, "the 'rhetorical hyperbole' normally associated with politics and public discourse in the United States" means that immunity from defamation for "rhetorical hyperbole" is "especially true in the context of commentary talk shows like the one at issue here, which often use 'increasingly barbed' language to address issues in the news." Judge Vyskocil was referring to Tucker Carlson's statements that McDougal was guilty of "extortion" when she "approached Donald Trump and threatened to ruin his career and humiliate his family if he doesn't give [her] money."[25] In an ironic inversion of the common law maxim attributed to Lord Mansfield that "the greater the truth, the greater the libel," contemporary American First Amendment defamation law seems to have gravitated toward a rule where "the greater the liar, the lesser the libel."

This particular logic seems to get the incentives exactly backward. The more hyperbolic and untrustworthy a source, the less liable it is in defamation, and the more careful and trustworthy a publication is normally, the more liable it will be on the rare occasions that it does allow a defamatory falsehood through its editorial filters. One can understand the sources of the logic. At least since *Cohen v. California*,[26] American First Amendment law has respected Justice Harlan's assertion that "one man's vulgarity is another man's lyric." One may well be wary of policing tone, particularly in political news commentary. And yet, in repeated surveys, viewers and listeners of outrage media report that they believe as true various assertions of fact that are false and defamatory. Forty-six percent of Trump voters, for instance, reported in a December 2016 Economist/YouGov poll that they believed that emails from the Clinton campaign talked about pedophilia and human trafficking.[27] Forty percent of respondents who watch Newsmax or One America News Network surveyed in May 2021 reported that they believed that "the government, media, and financial world in the U.S. is controlled by

a group of Satan-worshipping pedophiles who run a global sex-trafficking operation."[28] The distance between Cohen's jacket slogan "Fuck the Draft" and detailed monologues by Hannity alleging that the leading Justice Department officials Robert Mueller, Rod Rosenstein, and Andrew McCabe worked with Hillary Clinton to get the Obama White House to sell 20 percent of American uranium to Russia in exchange for donations to the Clinton Foundation is vast. The fact that a reasonable person who is *not* a consumer of outrage media would obviously understand that these assertions are political rhetorical hyperbole is irrelevant when the primary target audience for these words is also the audience that consistently reports that it ranks Sean Hannity and Rush Limbaugh (when he was still alive) as its most trusted sources of news behind Fox News more generally.[29] What seems like political hyperbole to the reasonable person who does not watch outrage media is received as trustworthy factual news by the audiences that consume those media outlets and commentators. And as the lone gunman who walked into Comet Pizza to investigate "Pizzagate" makes clear, the consequences for the victims of the business model of outrage media can be dire.

Several instances of outrage, particularly when directed at nonpublic figures, did result in meaningful settlements. Fox News settled with the bereaved family of Seth Rich for its defamatory statements about the murdered young man.[30] Alex Jones was forced to retract or settle statements he made about Comet Pizza and the shooting in Sandy Hook. But these are relatively rare events in the normal course of the business of selling outrage.

It is difficult to evaluate how much of an impact such a doctrinal change would have. At the broadest level, the United Kingdom has a lower standard of care, more friendly to plaintiffs, and its tabloids yet thrive. More generally, even if the standard I propose does make direct personal attacks on prominent politicians a touch harder, it may simply shift the content of the outrage-stoking media to more diffuse, abstract targets. It would be more a case of "these liberal elites want your son to marry a man" or "the FDA is lying to you, and hiding the successful use of Hydroxychloroquine or Ivermectin to treat Covid-19" than "Hillary Clinton procured thirty-three Haitian children for Bill's pleasure." And yet many of the most prominent and politically targeted attacks, particularly in a political system where individual candidates and actors, rather than parties, are central, a rule that makes it easier to sue habitual repeat offenders for defamation may offer some remedy to the dynamic. A wholesale overturning of *New York Times v. Sullivan*, abandoning the reckless disregard standard altogether along the lines Justice Thomas proposed in his concurrence with denial of certiorari in *McKee v. Cosby*,[31] may, however, chill too much speech. Peter Thiel's funding of Hulk Hogan's lawsuit against Gawker offers one recent example of how such a broad reassessment could be used by the growing number

of American billionaires, many with strong political orientations, to suppress media they dislike.[32] But a more narrowly tailored revision of the rules within that line of cases, one that takes into consideration the reality of outrage media and the actual false beliefs and hatred it instills in the minds of millions of people, could be administrable: for example, introducing survey evidence about audience beliefs, rather than making bright line judicial rules of what a well-educated legal professional would believe. Whether one is willing to entrust that kind of power to the McConnell Court and a judiciary that is the product of an ever more politicized appointment process in an increasingly dysfunctional political system is another matter.

Creating Transparency Regarding the Profitability of Selling Alienation and Hate-Mongering

A major challenge for any regulatory efforts designed to contain disinformation and propaganda that are fundamentally political speech and reflect a political perspective is the fallibility of both state and market institutions. Before we get to the question of how one might design a regulatory framework based on content (falsehood) that could resist First Amendment scrutiny, there is the question of how anyone who values democracy could take the risk of passing such regulation in the aftermath of the 2016 and 2020 elections. Any currently proposed design must assume at least in the 2020s, and possibly into the 2030s, episodic control of the federal government by a party that has been taken over by an antidemocratic, illiberal faction willing to deny facts and peddle lies, suppress opposition voting, and change the rules of democratic elections and governance to maintain minority rule whenever it reaches power in a state or in the federal government. It is precisely for times such as these that a robust First Amendment is most critical.

The "solution" of relying on commercial platforms to censor disinformation and propaganda is neither available nor desirable. It is unavailable because the worst abuses, with the largest influence, are the core product of the most influential platforms—Fox News's, Newsmax's, OANN's, and iHeart Radio's talk-radio coast-to-coast bile. Even on social media, keeping the eyeballs and engagement drawn by the hate-filled rhetoric is directly in the interest of the social media companies, such that the efforts of these companies will always be governed by a tension between the desire to contain the dissatisfaction of consumers upset by the outrage and the profit-driven need to keep this attractive-to-many content on the platform. The interests of the social media firms are less clearly aligned with selling outrage than are the incentives of the mass media platforms, but the bad incentives are there nonetheless, given the actual structure of the audience

for American political media. Moreover, the asymmetric architecture of outrage media in the United States leaves social media companies with the unappetizing prospect of regulating media in ways that will be perceived as favoring the Democratic Party, and we have seen examples of social media dealing with that political risk by intentionally tweaking their algorithm to downgrade left-leaning outlets that do not propagate outrage or falsehoods, simply so they can point to their own neutrality when challenged by Republican politicians.[33] Moreover, in an era of high concentration of wealth, after decades of increasing concentration in markets, legitimating the intensive engagement of commercial platforms in regulating political speech would be to leap out of the frying pan and into the fire, as far as preserving a democratic speech environment is concerned. Beginning at least with the National Association of Manufacturers in the 1930s,[34] American companies have invested in sowing doubt and propagating falsehoods that serve their bottom line: from lies with global effects like denying climate change, to whole industries distorting public health research like the carcinogenic effects of tobacco, to confounding discrete profit-threatening effects like the role of sugared soft drinks in the obesity epidemic.[35] The idea that Big Tech will be systematically better behaved than Big Oil, Big Tobacco, or Big Sugar requires a significant leap of faith.

One proposed alternative is a revival of the regulatory power of the Federal Communications Commission (FCC) to contain the worst of the abuses—in particular the fairness doctrine. At one level, the fact that the primary sources of disinformation and media polarization are mass media—television and radio—makes that option more available than is commonly recognized in debates that focus solely on social media platforms. Such a revival of direct federal regulation would be neither feasible nor advisable. It would be infeasible because First Amendment doctrine has become more robust in its defense of business. As a purely predictive matter, efforts to reinstate the fairness doctrine would more likely result in overturning the exceptional treatment of over-the-air broadcasting under the First Amendment, and its assimilation to cable broadcasting. *Red Lion*[36] was a product of its time, the tail end of high modernism and its belief in enlightened regulation in the name of the public interest, a time when it seemed possible that the First Amendment permitted, perhaps even required, newspapers, too, to come under right-of-reply regulations.[37] Its original logic was as much based on a non-technology-specific conception of a First Amendment designed to protect speech "from diverse and antagonistic sources" as it was narrowly specific to spectrum scarcity.[38] When *Miami Herald v. Tornillo*[39] didn't even bother to cite *Red Lion*, it left it to later decisions, beginning with *League of Women Voters*, to reify the "spectrum scarcity" rationale for regulation of over-the-air television in a fundamentally more invasive manner than the courts permitted for cable, telephone, or later the internet. That distinction was barely

a fig leaf in the 1980s, as cable TV penetration began to grow, and is much less in an era of the internet and over-the-top apps on digital TVs. Even without the *Lochner*ization of the First Amendment of the past thirty years, any current Supreme Court is pressed to justify broadcast exceptionalism when over-the-air TV is just one of many channels available to all, and when the technical "reality" of spectrum scarcity has long been bypassed as a technological matter.

Worse, there is no reason to trust the FCC to be a well-functioning, independent regulator. Different agencies have different institutional structures and cultures. The Bureau of Labor Statistics produces some of the most politically charged information of any government agency, and yet has succeeded in maintaining its political independence throughout its existence. The Federal Drug Administration affects billions of dollars' worth of companies' value, and yet has succeeded in preserving its independence from industry. The FCC, by contrast, has a long and sustained history of both politicization and a remarkably well-oiled revolving door to industry. In the 1920s and 1930s, under a still-nascent and limited modern First Amendment doctrine, the FCC clamped down on socialists (WEVD) and labor (WCFL), and on supporters of Hitler objecting to US support of Britain (Father Coughlin). As the 1970s and 1980s saw more robust constitutional constraints placed on the commission, its enforcement of those areas where it still had power to regulate—non-obscene sexual content—mostly followed the election returns, with enforcement of the loose *Pacifica* standard ramping up during Republican administrations as cheap signals to the Christian fundamentalist base, and ramping down during Democratic administrations. The record on revolving doors is no better. Michael Powell, son of then-serving Secretary of State Colin Powell, gutted the open access provisions of the Telecommunications Act of 1996 by reclassifying broadband as information, not telecommunications service. He later became the chief lobbyist for the cable industry. Meredith Atwell Baker, James Baker III's daughter-in-law, became NBC Universal's chief lobbyist four months after voting as FCC commissioner to approve the Comcast-NBC merger, and later became the chief lobbyist for the wireless industry. Of the six chairs of the FCC appointed to a full term since 1992 other than Powell, three (two Democrats and one Republican) moved on to private equity firms, one (Tom Wheeler) had already been chief lobbyist for both the cable and the wireless industries and could afford to be aggressive against industry interests, and two, Reed Hundt (a Democrat) and Kevin Martin (a Republican), who had taken aggressive positions in regulating industry, did not move on to positions in business. Hundt, Martin, and Wheeler all saw their temporary victories reversed. Hundt's efforts regarding open access to telecommunications infrastructure were reversed by William Kennard (a Democrat formerly with the Carlyle Group and now chairman of the board of AT&T) and Powell, Martin saw his efforts to

impose net neutrality obligations on the cable companies nullified by the courts based on Powell's reclassification, and Wheeler saw his effort to fix the constraint Powell had created by reclassifying broadband as telecommunications approved by the courts but reversed by his Trump-appointed successor. The FCC's entire institutional structure and culture would likely need to be revised to turn it into a genuinely independent agency (independent both from political parties and from the industries it regulates) before one could reasonably rely on it to regulate misinformation and propaganda. Such a restructuring, however, is only likely to follow statutory changes that could themselves only be the result of an already transformed US political system.

An alternative would be to leverage the newly possible decentralized social action, both peer production and civil society organizations, in partnership with a state whose role is limited to enabling social production rather than taking on the regulatory role itself.[40] Rather than seeking to regulate hate speech and disinformation on social media, regulatory efforts should be aimed at leveraging the new affordances that undergird surveillance capitalism toward providing broad and deep transparency about the money flows associated with selling hatred and outrage. Rather than working against the powerful technological forces of surveillance and ubiquitous data collection and analysis, the regulatory framework should focus on turning those capabilities into a publicly curated database that makes transparent which companies advertise on, and which companies profit from, advertising associated with hateful content. While this approach will not contain hatred and disinformation on its own, it will provide a platform for consumers and shareholders to hold the brands they identify with and the companies they own to account for the outrage their advertising dollars support. And this approach could apply not only online, but actually to the outlets that are the most effective disseminators of outrage and disinformation—on television.

The objective would be to construct a system of accountability for advertisers who support false and radicalizing content. It should be designed to be neutral among competing judgments as to what is false and what is true, as well as to what is radicalizing, as opposed to informing or mobilizing. The government's role would not include deciding whether Black Lives Matter online protests are "mobilization," while White supremacist bile is "radicalizing." We simply cannot trust American politics in the coming decade or two to deliver governments that can be trusted to make such choices in ways that support, rather than undermine, democracy. Rather, the government's role should be limited to requiring disclosure, collecting and managing the database, and making it freely accessible to the public. It would be up to networks of peer volunteers, civil society organizations, activist shareholders, academics, or journalists to develop the insights from the data and to mobilize to put economic pressure on outlets and advertisers they deem to be supporting falsehood and radicalization. This system would be far

from foolproof. It would be as available to right-wing consumers or shareholders who want to prevent advertisers from advertising in the *New York Times* (much less on the Rachel Maddow show) as it would be for those who wish to boycott advertisers who support the most popular providers of hate and outrage, Tucker Carlson and Sean Hannity. It should cover social media platforms, but also online sites that receive most of their traffic from direct access, rather than through social media referrals, like Alex Jones's Infowars. Because it is neutral, and because there is a large and willing audience that seeks out and is gratified by the outrage industry, it is entirely possible that such a transparency regime, even if perfectly implemented, will have no impact on the profitability of selling outrage. Instead, there will be sorting, with advertisers (like MyPillow) whose core audience aligns with the audience segments that pay attention to outrage industry sources supporting those programs, while advertisers aiming for broader market segments do not.

The natural home for designing a comprehensive database of advertising and its relationship to content is the Federal Trade Commission (FTC), which has tended to be less of a captured agency than the FCC. While the FTC's Division of Advertising Practices focuses its enforcement on misleading advertising, it is a reasonable application of that jurisdiction that the FTC may collect information about who is advertising what on which channels and programs. As technology has evolved to make finer-grained distinctions, such pervasive transparency requirements can be implemented in fine-grained increments, such as tying the advertising to specific stories in online media alongside which the advertisements appeared, or specific time stamps in television programs to enable data analysis to pin down what specific stories on television the advertising supported. It is entirely possible, for example, that an advertisement that by itself does not claim that it can cure a particular condition may carry that implication if run during and after a news segment on that condition. Or, an advertisement that does not directly address children may nonetheless merit review under rules applicable to advertisements to children if it runs in a program, or on a website, that carries children's programming.

In substance, transparency rules could be addressed to advertising delivery systems (websites, social media platforms, search engine advertising tools, and online ad exchanges) but also mass media (television channels, radio stations, and newspapers). Rules could, in the alternative or additionally, be imposed on advertisers themselves, who would be required to report monthly to the FTC where their advertising appears—compelling them to contract with advertisement delivery systems to provide them with that information. Most advertising online is not necessarily matched to content but to specific users or, rather, user profiles. To preserve individual user privacy, no information about the specific characteristics of the user may be included in the disclosure, but purely

the final piece of content, usually in the form of a URL when it is online, that was associated with a given advertisement. For mass media audiences in mass media formats that do not personalize advertising (on paper or on televisions that do not provide addressable advertising capabilities), the reporting would simply be the advertisement, advertiser, name of the program, and time stamp, to the minute, of when during the program the advertisement aired. While such requirements would have imposed insurmountable administrative and technical burdens in the past, today these data are collected in online media regularly, as a normal part of the operations of the data-informed preference manipulation industry (a.k.a. "behavioral" or "persuasive" advertising), and by television and radio outlets, who necessarily must record data on which advertisements they ran at what times for billing purposes. Including additional bits of data, such as the specific minute of airing in what is already a digital video stream on television, should be relatively trivial as a technical matter given that all television is, at this point, digital streams.

It is important to emphasize that such a requirement would not permit disclosure of any individual's reading and viewing patterns, and would insist that any possible personally identifying information be stripped from the data included in the advertising dataset. Websites have long inserted personally identifiable information into the specific URLs they return to users in order to circumvent users' efforts to preserve their privacy. To avoid returning personally identifiable information, content providers would be required to return "canonicalized" URLs, those stripped to the bare minimum necessary to access the content to which the advertising was attached, but without any of the added codes websites use to track their readers. Moreover, such a requirement neither prohibits nor regulates any expression. It is applied neutrally to all content and all advertising and prevents no content, whether editorial, entertainment, or advertising, from reaching its intended willing audience.

Creating such a database would help consumers hold the brands they buy, often associating their identity with that of the brand, accountable for the social and ecological impact of their products. It would help advertisers hold the advertising delivery systems they use accountable to place their advertisements in contexts of which the advertiser approves. And it would help shareholders hold the companies they own to account for the kinds of content the company supports with its advertising dollars. It would not help government or private parties identify who is reading or watching which programs. It would not help the government censor any content. The database would not be associated with a government-run program designed to assert that this or that program published truth or falsehood. It would not help social media platforms gain any information they do not already have to the extent they want it. It would not, in other words, exacerbate risks of either public or private censorship, whether online

or in mass media, nor would it require collection and curation of any data that might compromise user privacy beyond what companies already collect, nor make any such identifiable personal data available to parties that do not already have legal access to such data.

Instead, the database would be available to the public through published application programing interfaces to be developed with the expected community of user organizations—particularly academics and civil society organizations. This would allow both individual users, using publicly developed tools, to seek out advertising information, and organizations who seek to act as trusted sources to publish the information for users who trust them. A journalistic truth-oriented organization, say, the Poynter Institute's PolitiFact, might take its own individual story fact-checking approach of giving sources and individual media and political personalities "Truth-O-Meter" scores, and attach to those stories or media personalities a list of the advertisers who were aired or shown alongside the identifiable falsehood. Advertisers could subscribe to such a Truth-O-Meter service and insist that their advertising not be aired next to false information that could tarnish their brand. Businesses such as NewsGuard that sell a service through a browser plug-in to rate the reliability of websites' subscribers' visits could extend their offerings to include rating advertisers based on their relationship to the amount of advertising they do on sites that the company rates as unreliable. Consumer or other civil society organizations could develop rankings and ratings for companies similar to those that various sites provide for adherence to fair trade practices, ecological sustainability, or some index of a range of ethical commitments that consumers may support.

A system that would be sufficiently neutral to resist abuse by an antidemocratic administration, however, would necessarily be neutral regarding the organizations that publish ratings of advertisers. It could, therefore, be misused to mobilize political boycotts for purely political purposes, further entrenching the polarization dynamic. A politically oriented organization could look for politically oriented content and publish a list of advertisers who advertise on programs known to support the opposing party—the Republican Party could publish a list of firms that advertise on MSNBC generally, or on the Rachel Maddow show in particular, while the Democratic Party could do so for One America Network, Newsmax, or Fox News. Viewers or readers with well-defined and active political identities could subscribe to such lists and choose to support only brands that do not advertise on the opponents' networks or websites. While this may make advertising on any such networks undesirable, and therefore lead to a reduction in the commercial incentives to publish outrage-stoking content, it would be impossible to limit the judgment of what counts as "political outrage-mongering." Right-wing media do not have a monopoly on hyperpartisan content, though they do operate under a different economic

dynamic, with a pronounced propaganda feedback loop. PolitiFact rates Rachel Maddow stories as being as likely to make mostly false or false statements as those of Sean Hannity. But there is no reason to think that partisans will boycott only opposing hyperpartisan programming. Survey evidence from Pew suggests that viewers who trust Sean Hannity distrust CNN more than they distrust MSNBC, and distrust the *New York Times* not much less than MSNBC.[41] A fully decentralized approach that was designed to facilitate use of collective economic tools against the opponents of truth could well be repurposed to facilitate collective economic tools against any opponents.

How likely is such a database to address and contain the incentives to publish disinformation and stoke anger? While there are broad claims within the marketing literature about the increasing importance to consumers, particularly younger consumers, of incorporating ethical choices into their consumption patterns, sustained studies of the actual impact are lacking, and research recognizes a gap between survey responses that state intent to consume ethically and actual consumption habits.[42] Organic food sales make up no more than 6 percent of all food purchases, and possibly as much as 15 percent of fruit and vegetables.[43] This overall share of the market is similar to the market share of fair trade coffee,[44] while nationalistic ethical consumption, like the drive to "Buy American," appears to suffer a similar "intention gap" between popularity of the policy in survey responses and actual change in consumption habits.[45] There is, therefore, a significant risk that developing a system to render transparent the flow of funds that support outrage will not, in fact, lead to any consumer action that would impact the profitability of stoking hatred. On the other hand, unlike other ethical consumption choices, such as fair trade or organic products, there is no systematic reason to believe that brands that advertise on sites that disseminate falsehoods and hatred are systematically cheaper than brands that refuse to publish on such sites. Focusing one's consumption patterns on brands that do not financially benefit the outrage industry need not, systematically, involve an ethics/price trade-off in the way that many other ethical consumption efforts do. Moreover, while as a media market the consumers of outrage media are a relatively large market segment, they are still a minority of consumers in the economy as a whole. Furthermore, conversion rates on television (the conversion of exposure to marketing and actual purchase) are not high, and online they are practically unmeasurable.[46] The negative impact of consumer boycotts need not be large to outweigh the small and uncertain gain advertisers can expect from continuing to advertise on outrage media.

Another potential vector of influence is socialization and elite cultural pressures. The broad target audience of outrage media is older White users, with a relatively high proportion who have high school education or less.[47] Executives at firms that buy advertising are more likely to be influenced by elite attitudes of

others in the professional and managerial class and, to the extent that the transparency can put a spotlight on the complicity of certain firms in financing outrage media, these executives may be subject to some within-class social pressure. While such social pressure is unlikely to affect executives at the companies that themselves sell outrage, for others in the same class, who have no particular economic interest in selling outrage, shaming campaigns within the professional and managerial class may outweigh the uncertain gains from marketing on outrage media. In a way that is more formal than merely social cultural pressure, one might imagine shareholders raising objections to harmful advertisement investments that may prove costlier to deal with than the advertising choices are worth.

* * *

Media outlets that sell outrage are not the driver of epistemic crisis or the prevalence of hate-filled public discourse in America in the twenty-first century. But commercial media are caught up in a competitive dynamic created by the political-cultural dynamic that undergirded the Reagan revolution and have, over the past forty years, substantially reinforced and amplified the alienation and hatred that resulted in the takeover of the Republican Party by its Trump wing. What started as a business opportunity for Pat Robertson and Rush Limbaugh became a market imperative strictly enforced by the competitive pressures of the lucrative market in selling outrage to a large minority of Americans. In the long term, fundamental solutions will have to address the legitimate underlying pain of middle- and working-class, often less well-educated audiences who have borne the brunt of the alliance between the neoliberal business wing of the Republican Party, which rendered them economically powerless, and the neoliberal Clinton wing of the Democratic Party, who traded victories against identity-based domination in exchange for embracing neoliberal economic extraction. But in the shorter term, efforts to contain the propaganda feedback loop and contain outrage, falsehood, and hatred in media should focus on the market dynamic that makes selling outrage so profitable and kills any efforts to reach the audience of these media with truth-seeking, rather than identity-confirming, programming.

16

The First Amendment Does Not Protect Replicants

LAWRENCE LESSIG

Imagine a platform called CLOGGER which offered computer-driven, or more precisely, artificial intelligence–driven, content to political campaigns. The technology would automatically craft content, both text and video, in the form of tweets, Facebook posts, Instagram posts, and blog entries. The substance of that content would be determined algorithmically, within broad constraints set by the campaign initially, and then developed over the course of the campaign based on the responses that the content produced, as well as the support the candidate is receiving in regular, computer-driven polling. Think of it as a campaign in a box—an automatic, human-free, and AI-driven political content-spewing box.

Nothing about the details of this platform is hard to imagine. Brad Pascale, the first digital director for President Trump's campaign, reported sending fifty thousand to sixty thousand distinct messages to Facebook users every day during the 2016 campaign.[1] CLOGGER takes Pascale's technology to a new level. Millions of items of content crafted algorithmically, based on constant feedback from voters within the relevant jurisdiction, would be spread by CLOGGER across the course of the campaign. Beyond a few initial specifications about the candidate—Republican, a lawyer, female, forty-five years old—the campaign need do nothing else. All content would be developed based on actual evidence of what moves voters, either to participate in the election or not, depending on the strategy most likely to produce victory.

Should such a technology be regulable?

It's trivial to build a First Amendment argument against its regulability. CLOGGER utters political speech—the speech most protected under the First Amendment. Though no human crafts that speech, there is plenty of authority for protecting speech regardless of its source. As Justice Scalia remarked in *Citizens United*:

> The Amendment is written in terms of "speech," not speakers. Its text offers no foothold for excluding any category of speaker, from single individuals to partnerships of individuals, to unincorporated associations of individuals, to incorporated associations of individuals. . . . We are therefore simply left with the question whether the speech at issue in this case is "speech" covered by the First Amendment.[2]

Scalia was not the first to make that move. "The proper question," Justice Powell had declared thirty-two years before, "is not whether corporations 'have' First Amendment rights, [but whether the law] abridges expression that the First Amendment was meant to protect."[3] From this perspective, the question of regulability hangs not upon the speaker—here, an AI-driven algorithm—but upon the speech—here, the most protected speech within the constitutional canon. From this perspective, CLOGGER is not regulable.

Yet this conclusion should give us pause. We've seen enough to see why an ecosystem of political speech like CLOGGER's could well be harmful to democracy. As messaging is personalized and rankings are effectively anonymous, political discourse changes. Its character changes. As the *Wall Street Journal* has reported, political parties in Europe have complained to Facebook that their algorithms were driving the parties to change the content of their digital campaigns from substance to sizzle. As one internal document reports about Poland, "One party's social media management team estimates that they have shifted the proportion of their posts from 50/50 positive/negative to 80% negative, explicitly as a function of the change to the algorithm."[4] That dynamic would only increase with the spread of CLOGGER-like platforms. And though there are many who are skeptical about the absolute effect such technologies would have on voting or political attitudes,[5] we can bracket the empirical question of the magnitude of any such effect. If it is not negligible, it is a fair question to ask: Why would our Constitution prohibit us from protecting our democracy in this way?

* * *

The history of constitutional law is the story of struggles around latent ambiguities. At least it is in part. When the Framers of our First Amendment wrote "Congress shall make no law . . . abridging the freedom of speech," not one of them contemplated a world in which political speech could be crafted by a machine. Even if they did contemplate what those words literally describe—"political speech . . . crafted by a machine"—they had no way to understand what the machine that I am describing would be. Even we don't understand the dynamics of AI-driven content: We don't understand how it affects people, or political communities; we don't understand what strategies are most effective when its

content is deployed; and we have no understanding of whether the strategies that are most effective for it are strategies that are simultaneously democratically edifying or destructive. It is perfectly plausible that most effective CLOGGER-like platforms for winning elections are also effective technologies for destroying democracy. We—and our Founders—just don't know.

Given these things that they plainly didn't understand, why should we interpret their Amendment as if it covered "the freedom of speech by machines" as well as "the freedom of speech by humans"? I'm not saying it would be crazy to protect the poetry of *Blade Runner*–like replicants. I'm not even sure I'd be against it. I am instead asking whether in fact we have. Or more precisely, why does it make sense to believe that we have? Why should we read words written and enacted by a people who could not begin to contemplate CLOGGER as protecting the speech of replicants?

This concept—the "speech of replicants"—is central to the argument I am making here. It evokes an idea suggested in many science fiction contexts, none more powerfully than Ridley Scott's film *Blade Runner* (1982). In that film, replicants were synthetic humans that had evolved sophisticated intelligence. While the entities had been bioengineered by humans, no one could be said to have programmed the intelligence that they eventually manifested. Instead, that intelligence had evolved through experience, in a process similar to how machines within machine learning contexts are said to learn today. "Replicants" in the sense that I will use the term in this essay are thus *processes that have developed a capacity to make semantic and intentional choices, the particulars of which are not plausibly ascribed to any human or team of humans in advance of those choices*. Humans may build the systems, but humans do not make or directly control the decisions that those systems manifest—including the decisions about "their" "speech."

"Replicant speech" is thus a subset of the "machine speech" that Tim Wu has described.[6] If the president prerecords an address explaining his reasons for a new direction in foreign policy, and then that message is broadcast on all major networks, we could understand his speech as "machine speech," since, at least from the audience's perspective, it is being produced, literally, by machines. But that speech is not "replicant speech," as a human crafted the president's words. The speech that I am describing is different: It is crafted or originated algorithmically, with the substance of that algorithm not in any meaningful sense programmed by any individual in advance. Like parents with kids, a programmer may have started the system off. But at some point, the speech is no longer the programmer's.

Coders may resist this description. The replicants, they might believe, are theirs, and therefore, they may insist, their words are theirs. But the distinction I'm drawing is familiar. One may enable something, but that doesn't mean one is

cognizant of the things the entity so enabled might produce. Gun manufacturers make weapons; they don't choose who uses the weapons or how they are used. Parents make children, but the college essay of an eighteen-year-old is not the parents' speech. For the speech of a machine to be the coder's or programmer's speech, the coder must in a significant sense craft or control that speech. While that's an easy test for most machine speech, for replicant speech, there is no such link. At some point along the continuum between your first program, "Hello world!," and Roy Batty (*Blade Runner*), the speech of machines crosses over from speech properly attributable to the coders to speech no longer attributable to the coders. Or at least, everything in the argument I am making presumes there is such a line, and that line can be roughly discerned.

So, are there replicants in our world today? Obviously not in the Ridley Scott sense. But there is certainly "replicant speech" in the sense that I have defined here. Facebook, for example, deploys replicant intelligence in the spread of content across its platform. The platform makes choices about the priority, kind, and nature of the content that it will expose its users to, for the purpose of inducing those users to behave in a particular way—whether that is to click on a link, or reveal more information about themselves, or share the content with others, or even to vote in a particular way. How the platform does each of these things, often through the use of "speech," is beyond the ken of any particular individual or team of individuals in advance of those acts and speech. The machine learns how to do its task best, and it produces content as it learns.

That no one knows quite how it will evolve was made clear by a particularly embarrassing example revealed in the fall of 2017.[7] At that time, Facebook started offering a new category of users to interested advertisers—"Jew haters." That category apparently collected users who had expressed anti-Semitic ideas or followed anti-Semitic content. Yet that category had not been created by any human. There was no person or department of people who had decided it was a good idea for Facebook to profit from people who "hated" Jews. Instead, that category was crafted by a replicant: an AI-driven algorithm for developing potentially profitable advertising categories had made the "judgment" that "Jew haters" as a category would increase Facebook's profit. Obviously, once the executives at Facebook had discovered what their machines had done, they quickly removed the category.

I don't doubt Facebook's responsibility for the harm caused by their machine's "creativity." But I also believe that we should not ignore that the machine had creativity. It is that creativity that I mean to identify as replicant speech. And it is the content of that category of speech that I don't believe we can ascribe to any particular human or team of humans.

No simple line will distinguish replicant speech from Wu's "machine speech." Simple algorithms don't transform machine speech into replicant speech;

complex algorithms do, but only if they are of a certain self-learning character. Yet the intuition should be clear enough to support the argument in the balance of this essay and to guide the process of line drawing in application. The question in each case is not simple responsibility—rules assigning responsibility are focused elsewhere. Nor is it a question of specific intent—that my machine speaks without my meaning it to speak does not make my machine a replicant. Instead, the line must track a sufficiently sophisticated system of technology, the actions of which no individual can understand or predict in advance. If we can draw that distinction usefully,[8] then the question I mean this essay to address is whether we must protect replicant speech under the First Amendment. Why, again, must the Amendment apply, given that its authors could have had no idea of the technology that would produce this potentially protected speech?

* * *

The traditional answer to this obvious question is that our Constitution is a set of words, and those words have meaning.[9] The issue is not what the Framers "imagined" or "contemplated" or, God forbid, "intended." The question for constitutional interpretation is what their words mean or, for an originalist, meant.[10] How do their words, as interpreted through the doctrines of constitutional law, apply to the world as we find it today?[11] The Framers never imagined an economy as large or as integrated as the national economy; the Court has held repeatedly that that doesn't mean that the Commerce Clause should be read to secure to Congress a power less than its words direct. As the Supreme Court declared in *United States v. Classic* (1941):

> We may assume that the framers of the Constitution in adopting that section, did not have specifically in mind the selection and elimination of candidates for Congress by the direct primary any more than they contemplated the application of the commerce clause to interstate telephone, telegraph and wireless communication which are concededly within it. But in determining whether a provision of the Constitution applies to a new subject matter, it is of little significance that it is one with which the framers were not familiar. For in setting up an enduring framework of government they undertook to carry out for the indefinite future and in all the vicissitudes of the changing affairs of men, those fundamental purposes which the instrument itself discloses. Hence we read its words, not as we read legislative codes which are subject to continuous revision with the changing course of events, but as the revelation of the great purposes which were intended to be achieved by the Constitution as a continuing instrument of government.[12]

So much is settled doctrine. As applied to replicant speech, the question is thus how "the great purposes which were intended" by the First Amendment should apply to speech uttered by replicants.

I'm a believer in the value of settled doctrine. I agree that at the margin, the question is always and only how existing doctrine applies to a new set of facts. Fidelity to role requires that courts minimize interpretative depth.[13] Rules must be clear and applied in obvious ways if the very act of applying rules is not to be read as inherently political.

But there are exceptions to this conservative principle, at least when technology has changed fundamentally, and when the consequences of that change for important institutions or traditions are severe. In this case, technology has changed fundamentally.[14] And in the context of replicant speech, the consequences of that change are severe. An unthinking application of ordinary First Amendment doctrine to efforts to regulate platforms like CLOGGER could have profound consequences for our democracy. And rather than a jurisprudence on autopilot, we need to rethink exactly how broadly the principles of the First Amendment must apply here. Or put differently, we must consider seriously whether the First Amendment must be read to disable democracy from protecting itself from the threat of replicant speech.

* * *

We've long mythologized the depth of the Framers' conception of a theory of free speech. Yet as writers for generations have noted, their "freedom of speech" was an untheorized set of practices, enforced through republican institutions, not courts.[15]

But however articulated their theory was, it was tied to a very practical understanding of how a democracy functions. That practical understanding constituted, in *Classic*'s language, "the great purposes which were intended" by the First Amendment. According to those purposes, those vying for power within a democracy should, on this account, offer reasons. Those reasons were public. Those reasons were then meant to persuade or not within what a century later would be called a "marketplace of ideas."[16] Those playing the persuasion game within this market were themselves subject to persuasion. The very idea of a debate is that those debating might become convinced that the other side has a point.

This ideal operates despite a gaggle of psychological barriers to rational thinking. Humans are bad reasoning machines (at least compared to machines). Confirmation bias, identity bias, and a host of other flaws together mean that the effect of any process of persuasion is marginal at best.[17]

Yet history is marked by our minds being changed. President Nixon was loved by Republicans as much as President Trump was—until he was not, because at least half of Republicans had been brought to see that Nixon was indeed a crook.[18] White Americans believed in segregation and laws that prohibited the mixing of the races—until they did not, because a wide range of culture remade those White American ideas.[19] These changes were not the product of debating clubs. But they were consistent with the ideals being expressed within debating clubs of every kind—including the halls of Congress. The very idea of a system of free speech embeds the rare possibility that those participating might be brought to see the world differently. "The freedom of speech" is a shorthand for that ecology of understanding.

It follows that to protect that freedom, the government must treat all of us as equals. It does that by standing to the side of our political debates. The domain of democratic deliberation, as Robert Post has so powerfully described it,[20] must be free of the government's coercion or control. If we are indeed equal citizens, then we must preserve an equal right for all of us to engage the rest of us with our own ideals. We each, that is, must have the freedom to engage others if the dignity of our status as citizens is to be preserved.

None of this is to imagine that we all possess equal power within this domain of deliberation. Tucker Carlson has wildly more power than Bill Moyers or Elizabeth Warren. But he doesn't have more power because the government has given him more power, at least directly. Inequality may be inevitable, but inequality produced by the coercion of government is not allowed.

Yet there is no democratic reason or reason of dignity to extend this privilege beyond our species. We must protect Tucker Carlson's freedom as well as Bill Moyers's because both are citizens entitled to persuade the rest of us to their very different views. But CLOGGER can claim no equivalent entitlement. CLOGGER is a replicant.

This conclusion follows a fortiori from Justice Kavanaugh's opinion in *Bluman v. F.E.C* (2011).[21] In that case, foreign nationals residing temporarily within the United States sought the freedom to make political contributions to candidates and political action committees. Their claim had been buttressed by the Supreme Court's decision in *Citizens United*.[22] If corporations—which are not citizens—get to speak politically, why can't people—who happen also to not be citizens—speak politically? If, as Scalia insisted in *Citizens United*, the question was not the status of the speaker but the character of the speech, then it would seem clear that political speech by immigrants should be protected just as the political speech of unions and corporations was in *Citizens United*.

Judge Kavanaugh disagreed. On behalf of a three-judge court, Kavanaugh rejected the immigrants' First Amendment claim. That conclusion was later

summarily affirmed by the Supreme Court. "The Supreme Court has long held," Kavanaugh observed, "that the government . . . may exclude foreign citizens from activities that are part of democratic self-government in the United States."[23] More specifically, they can be excluded from "activities intimately related to the process of democratic self-government."[24] This does not mean that the government could prohibit "foreign nationals from engaging in [any] speech." But speech intimately tied to the democratic process was speech that foreign citizens had no constitutional right to engage.

If *Bluman* is right, then CLOGGER can be regulated. CLOGGER would be uttering replicant speech. It would be deploying that speech to affect elections. Those two facts together—a noncitizen engaging in political activity—draw it within the scope of *Bluman*'s reasoning.

This move—let's call it the identity move in First Amendment jurisprudence—is familiar. Talking cats have no First Amendment rights, as the Eleventh Circuit has informed us.[25] Likewise, following the brilliant insight of Kate Darling, if we see replicants as a kind of animal, then their speech, too, should be entitled to no strong First Amendment protection.[26] Children and young adults have fewer First Amendment rights, as the Supreme Court has instructed.[27] These beginnings suggest the obvious step to excluding replicant speech as, well, based solely on the identity of the speaker.[28]

Wu and others don't take that step, however, at least this directly, because the same jurisprudence also reserves significant First Amendment protections for corporations. Corporations are not citizens (even if they are "persons"); why should they get rights when machines do not?[29]

Yet while corporations are certainly not people, they are associations of people. And while I certainly agree with many that the corporation speaks in a voice distinct from the voice of the people it comprises,[30] it still is an entity comprised of people. It wasn't a necessary step to conclude that association speech should be treated in the same way as people speech. But it's not a crazy step, either. Corporations are the institutions through which some people express their ideas. Whatever ideas they express, they are ideas crafted by humans.

Not so with CLOGGER. CLOGGER may share the objective of a traditional campaign manager. But the way CLOGGER does its work is radically different. There is no human making any judgment about what should be said or not said. There is no process by which there is any reflection by any individual or group upon what should or should not be said. Whatever this replicant is, it is not a person, let alone a citizen of the United States, with any entitlement to participate in our political process.

My point is not that such technology should be banned. Or that such technology is necessarily corrupting. My point is simply that such technology does not deserve the full protections of the First Amendment. If it is the dignity of

democratic citizenship that the First Amendment is protecting,[31] there is nothing in the replicant's algorithms that deserves any such dignity. No American ever voted to secure to replicants any First Amendment protection. There is no reason in logic to extend regular First Amendment protection to them.

None of this is to say that such speech is entitled to no protection at all. This is the insight in Justice Scalia's opinion in *RAV v. City of St. Paul* (1992).[32] We could well conclude that replicant speech is entitled to no protection but also conclude that the government is not free to discriminate among replicant speech. From this perspective, the replicant targeting the ads in Facebook's algorithm would have no presumptive constitutional protection. But the government couldn't decide to ban Republican targeting, but not targeting for Democrats. As in *RAV*, that is not because the underlying speech is protected. It is because a second value within the contours of the First Amendment is the value of government neutrality.

* * *

The argument that ordinarily resists the position I am advocating for here is put most famously by Eugene Volokh. In an essay commissioned by Google, Volokh argues that the results of search engines are protected by the First Amendment, even if those results are produced by algorithms.[33] "Google, Microsoft's Bing, Yahoo! Search, and other search engines are speakers," Volokh declares, because "each search engine's editorial judgment is much like many other familiar editorial judgments."[34]

Throughout his argument, Volokh considers the "speech" of search engines to be the same as the speech of newspaper editors, for example, when they arrange stories on the front page, or select which op-eds they will publish, without ever addressing the question whether the fact that humans are behind one should distinguish that speech from the speech that humans are not behind. He—and the courts that have followed this position[35]—can make that move because they believe it follows from the principle that the First Amendment protects listeners as well as speakers,[36] as well as the fact that humans wrote the code that selects some speech over other speech.[37]

The latter point is certainly true, at least for some code and some automation. My claim is simply that we have no good reason to extend that conclusion to all code, regardless of its automation—specifically, replicant code. The former point simply ignores—literally, the case is not even cited—Justice Kavanaugh's opinion in *Bluman*. Whatever else *Bluman* means, it must mean that the fact that there are listeners does not elevate all political speech into protected political speech. I very much would like to see the political ads paid for by legal immigrants to the United States. Despite that wish, their speech, at least through

donations, in the context of political campaigns, is not protected. Likewise with CLOGGER: Whether or not Google's ranking of Volokh's blog over Balkin's is protected by the First Amendment, a Google-deployed CLOGGER is not.

* * *

Replicant speech, I've argued, as a subset of "machine speech," is entitled to no presumptive constitutional protection. So too might "functional speech," as Wu defines it, also be exempt from First Amendment "coverage."[38] But both claims are distinct from the claim that "social media" are entitled to no constitutional protection. Social media include content crafted by humans. That portion, at least, is entitled to ordinary or full First Amendment protection. But not all content from social media is content crafted by humans. The part that isn't and that is crafted by replicants—including targeting and other techniques of spiking—is not entitled, on this account, to those same protections.

Yet beyond constitutional entitlement, we should be clear about precisely why it would be a bad idea to extend presumptive First Amendment protection to the products of these machines. What's the harm from this kind of content? What's the reason a government might prefer to regulate it?

This question forces us to be clear about what we mean by "speech" in this context. In the sense in which the First Amendment is currently being invoked, speech is broader than words. It includes the editorial judgment in the selection and ordering of content, as well as emphasis and placement. It includes, in other words, not just words, but acts which are meant to convey or affect meaning and understanding.

All of these are not new, even if their sophistication and pervasiveness is. Shoshana Zuboff describes the outrage expressed by Congress in the 1970s at the emergence of techniques of "behavioral modification."[39] Inspired by the work of B. F. Skinner, "behavioral modification" used a wide range of techniques to "manipulate" individuals into behavior preferred by the modifier. Such manipulation could be opposed in general. But for obvious reasons of social psychology, it should raise special concern in the context of political speech. We know enough about "us" to know exactly why such manipulation can be so harmful.[40] Among teens, the use of social media has been linked to substantial and predictable harms. After two decades of decline, high depressive symptoms for girls thirteen to eighteen years old rose by 65 percent between 2010 and 2017.[41] Social media use is a predictor of that depression.[42] As Tristan Harris, former Google engineer and founder of the Center for Humane Technology, puts it,

> The problem is [the] attention-harvesting business model. The narrower and more personalized our feeds, the fatter [the technology

> company's] bank accounts, and the more degraded the capacity of the American brain. The more money they make, the less capacity America has to define itself as America, reversing the United States' inspiring and unifying motto of E Pluribus Unum or "out of many, one" into its opposite, "out of one, many."[43]

The point might be made best by analogy to food. We know enough about the human body to know that it is vulnerable to certain kinds of food. Evolution has left us craving a mix of salt, fat, and sugar that, at least unmoderated, is not in fact healthy for the human body. The companies that supply us with this unhealthy mix do so not because they want to make us unhealthy. They do so because our cravings will make them rich, or more charitably, because the market demands they maximize profit, and the addictive version of their food is an effective way to maximize that profit. Michael Moss tells a powerful story about processed food companies struggling to make their products healthier—with each quickly discovering the market's punishment for their virtue.[44] That punishment has the predictable effect of driving those companies back into their old ways. The obvious response to this competitive dynamic is for government to intervene in the most effective way possible. Nutrition labels are an obvious step. Quantity regulations for certain fast foods might also be appropriate.

The same happens at the level of the mind. We know from psychology that we all yearn to have our view of the world confirmed, and that we are more likely to react to titillation and outrage than reason or facts. These facts about us are inputs into the replicants' algorithms. They direct how the replicant will respond. Like the replicant in the film *Ex Machina*, it knows our weaknesses and our vulnerabilities. It exploits them. It rewards discourse that is extreme and false over that which is balanced and true. It leverages insecurity and emotion rather than understanding and empathy. It does this not because it is evil or programmed against us. It does this because this is who we are, and it aims to maximize something given the constraints of who we are. Were we all versions of Spock, we would not need to fear replicants. As we are not, we should.

These facts about us should allow us to protect us—or at least our democracy—from the consequences of these manipulations. My claim in this essay is that we can do that constitutionally—at least so long as we are resisting the manipulations of replicants rather than machine speech more generally. The Constitution complicates any defenses against machine speech that is not replicant speech. Some machine speech, driven by authors ranging from Donald Trump to Tucker Carlson, is powerful and effective at exploiting the same psychological weaknesses that replicants do. But Trump and Carlson have the First Amendment on their side. Replicants should not. And thus, even though our protection will be incomplete, that is no reason

not to do what we can to protect ourselves and our democracy. Nothing in our Constitution should be read to require us to remain vulnerable to the dangers from replicants. To the contrary, everything in our Constitution should direct us to protect against their harmful effect.

Acknowledgments

I am grateful to Amanda Shanor for extremely helpful comments on an earlier draft, and to Mark Poston for the exceptional research that helped frame this argument.

17

Social Media, Distrust, and Regulation

A Conversation

NEWTON N. MINOW, NELL MINOW, MARTHA MINOW, AND MARY MINOW

Mark Twain said, "A lie will fly around the whole world while the truth is getting its boots on." He also said, "It ain't so much the things that people don't know that makes trouble in this world, as it is the things that people know that ain't so."

Actually, he didn't. They sound like him and both statements are widely attributed to him, but there is no evidence that he said either one.[1] It is a lot easier to keep attributing those statements to Twain than to verify them. This epitomizes the problem of the age of digital media, where truth is the needle in a haystack of misinformation.

The good news is that the gatekeepers are disappearing, giving voice to people who could never get an op-ed published in the *New York Times* or a seat at the table on *The View*. The bad news is that the gatekeepers are disappearing: Even though the media gatekeepers kept out people who did not look and sound like a small segment of society, they also ensured basic standards of fact-checking, disclosing sources, and conflicts of interest. Most troubling, there is a new Gresham's law of information: Instead of bad money driving out good, it is avalanches of misinformation, some of it coming from ignorance and some from untraceable, malevolent sources, drowning out accurate, fact-based information in oceans of misdirection, misinformation, and lies.

Who can and who should do something about it? Is it possible to protect privacy while making individuals and organizations responsible for what they publish? Our family has been debating these issues for decades. We four are lawyers bringing different areas of expertise, involved with media but with different perspectives. This essay is adapted from a recent series of conversations sorting through the layers of law, politics, business, and government.

Martha Minow: Dad, when you were chairman of the FCC, an equal time complaint came to your attention through an unusual source.

Newton N. Minow: In the fall of 1962, former first lady Eleanor Roosevelt called me and said, "Why aren't you responding to Reverend Smith?" I said, "Mrs. Roosevelt, I'm sorry, I don't know anything about this. Who is Reverend Smith? What is this about?" She told me Reverend Smith was a Black minister in Jackson, Mississippi, where he was running in the Democratic primary for Congress against the incumbent, Congressman John Bell Williams. He went to the local television station, WLBT, with a check, trying to buy time on television to make his case to the voters. For weeks, he was told to come back, and when he came back, he was told, "Come back next week." Now the election was next week. Mrs. Roosevelt said he had complained to the FCC, but got no response.

I checked and found the complaint sitting on the desk of one of our staffers. Staff had checked with the other candidate, the incumbent John Bell Williams, and he was not buying any time. The equal time law, Section 315 of the Communications Act, provides that a broadcaster must treat candidates equally. If time is sold to one, time must be sold to the other. If time is given to one, it must be given to the other. But since Reverend Smith's opponent was not buying any time and since we were not giving any time to any candidate, the station and our staff believed that there was no violation of law.

I saw the situation differently. While that may be a technical interpretation of Section 315, it overlooks the overriding principle of the Communications Act: A broadcaster must serve the public interest. How can it be in the public interest to have no discussion on television of this political race in this community? At that time, we still sent telegrams. I sent a telegram to WLBT saying, "Explain today why you think it's in the public interest to have no discussion of this congressional race." Within hours, a lawyer for WLBT came to my office and said WLBT would put Reverend Smith on the air.

And I asked, "When? The election is next week."
"We're putting him on soon."
"What does that mean?"
"Tomorrow."
And he went on television.

Myrlie Evers, the widow of Medgar Evers, said that seeing a Black candidate for public office on television was "like the lights going on." Reverend Smith lost the election. But my point is that public interest is the guiding principle of American regulation of communications. What does the public interest mean in

our new technological advances into the digital world? I'm turning to our three daughters—who make me and my wife so proud—and asking the next generation to tell us whether the public interest is still relevant.

There's an extraordinary postscript to the story. Many years later, in 1996, I was a delegate to the Democratic Convention, which was in Chicago. I was invited to meet with the Mississippi delegation, which was at that time an all-Black delegation. One of the delegates said, "Are you the guy that was chairman of the FCC?" I said, "Yes, sir, I am." He said, "I'm Aaron Henry. I was the campaign manager for Reverend Smith. I'm the one who called Mrs. Roosevelt. After you left the FCC, the station continued its anti-Black policy and the United Church of Christ came to us and organized a challenge to the license. Our group applied. And the station lost its license because it didn't serve the public interest. And now I am the chairman of WLBT."

I recognize it was simpler then. The number of broadcast stations was limited and had to be licensed to have access to the public airwaves, and our government enforced the broad requirement that they serve the public interest. Station WLBT lost its license when the FCC and the courts decided it was not serving the public interest. However, these days, new technologies are not licensed. And that is why I'm turning to the next generation, who are much more knowledgeable than I am, to figure out what the public interest means today.

Nell Minow: That is exactly my question. The premise for the public interest requirement was based on scarcity. The government had to allocate licenses to broadcasters so their signals would not interfere with each other. You fought very hard at the FCC for expanding viewer and listener choice through cable and satellite. Then DARPA, the Defense Advanced Research Projects Agency, opened up the internet. Everyone who wants to use it has close to the same access to an audience that the established media does. When you were at the FCC, television came over the airwaves and phones were plugged into the wall, and the number of newspapers was much greater than the number of television stations. Now the opposite is true. More than twenty years after you left the FCC, you argued that the KTTL radio station in Kansas should lose its license because of its virulent racist and anti-Semitic content. Today, that content would be considered mild in comparison to what is consumed by Americans via the internet and talk radio.

Newt: That's correct. You can't make the scarcity argument today. Yet the issues today are urgent for democracy, and indeed, public safety. What happened on January 6 [2021] with the riots and attempt to overthrow our government by force was in part a result of misinformation and incitement. I believe that we have to find a way to reinstate a commitment to the public interest in our world of nonscarcity.

Martha: Look at Canada, which regulates digital media with a public interest focus, not predicated on scarcity. Of course, it is a different constitutional system. Canada's analogue to our First Amendment appears in its Charter of Rights and Freedoms.[2] It broadly protects expression, but the Canadian courts tie those protections to democratic governance and individual dignity[3] and permit limits on hate speech in ways that the US courts do not. There is no country in the world that has the extreme speech protections that our US courts have developed through their interpretation of the First Amendment. Actually, a public interest requirement could be compatible with the First Amendment if extreme interpretations of the amendment are moderated, as many commentators urge.[4]

The "public interest" *is* hard to define, however. I bet we can make more progress by discussing a specific example. Let's contrast the immunity granted to digital media from the well-established liabilities of newspapers, publishers, broadcasters, cable television, and individuals for content that is defamatory or that incites violence or illegal activity. Section 230 of the Communications Decency Act granted an immunity to digital platforms from these familiar liabilities in order to nurture new internet businesses. Internet businesses are now among the largest and most profitable businesses in the world. I think the public interest calls for reexamining a law exempting platform companies like Facebook from liability for what they spread. People face harms. Internet platforms are getting a subsidy and unfair advantage compared with other media, which already face severe financial challenges. Actually, the public interest has long existed compatibly with the First Amendment. Courts must balance competing interests. New challenges call for new efforts such as dealing with anonymous tweets. As of 2017, a quarter of Twitter users opted for full or partial anonymity.[5] Also, Twitter is enabling access to the entire world; it is not the individual. Why should Twitter be given a subsidy compared to a newspaper or radio station, which are not shielded by Section 230?

Mary R. Minow: That's not true. If a newspaper has an open comment section, they have the same protections. It's a matter of who's doing the speaking, not who the sponsor is. Section 230 protects user-generated content. When the newspaper publishes something on its own, it is subject to publisher liability. When it permits user-generated content on its website or social media page, it has the same protection as any website.

Nell: Reporting abusive tweets and responding to public officials directly on Twitter gives me a sense of agency. On the other hand, it makes me very uncomfortable to think that Twitter is relying on its users to police the tweets. Twitter, Facebook, Reddit, even Parler all have lawyers; those lawyers have

drafted user agreements.[6] These include indemnification and promises not to violate copyright or post pornography or spam. They can delete your posts or kick you off if you do. These are corporations, not the government, so the First Amendment does not apply. They have advertisers and customers and shareholders. If they make a market-based decision that what is best for their business is to create, if not a safe space, at least a safer one than one that has no limits at all, we all agree that is fine, right? If, say, a new competitor to Twitter decided that it would only publish grammatically correct tweet equivalents in iambic pentameter, they could do so. An individual or a corporation can impose almost any restrictions they want.

But what can the government do? Banning certain topics or individuals only sends them to places that are (1) harder to monitor and (2) assumed to be even more credible because they are more raw and unmonitored. See the excellent documentary *Feels Good Man*: An innocent cartoon image of a frog was co-opted as a symbol of the most virulent racist and insurrectionist groups. Matt Furie, who created the image, found there was no way to stop his Pepe the Frog from being appropriated.[7]

Martha: The government can create incentives and it can tax platform companies using targeted advertising—a practice associated with misinformation campaigns and discriminatory displays of employment opportunities.[8] A tax might even be placed on companies that do not develop a fairness policy or that do not report their criteria for content moderation. The government can require and enforce consumer protections; the government can say to platform companies, "If you don't come up with a way to enforce your own user requirements and service agreements, which you currently do not do, you will not have this immunity." The government can break up large platform companies with antitrust tools or condition a decision not to break them up on greater assurance of protections for users. There are many steps that government can take within the Constitution.

Mary: Is there a way that the government could, in fact, say "this is a public forum," just as California said that private shopping centers are public forums and, therefore, subject to the First Amendment?[9]

Nell: What do you do then about Parler, which is deplatformed for violating the user agreements of the companies carrying it, and then opens up a week later operating out of Russian servers? Like environmental regulation and tax regulation, rules issued according to physical geographical borders are less and less effective.

Martha: The borderless global nature of social media and digital communications is the bigger challenge. The European Union, Australia, and China are

governing social media already. Understandably, companies are searching for some common framework; it's too hard to comply with competing and even clashing laws from different countries. We can continue to talk about the First Amendment, but then the actual governance of social media is going to be decided elsewhere.

Newt: Another way to look at the concept of "public interest" is to look at the origins of the internet. It would not exist without the US government's investment. The government put up the money through the Defense Department for the technology that led to the internet. Isn't there a way to condition access to the internet on some set of requirements? For example, let's just take one thing. Let's take hate speech. Only hate speech, which is outlawed in most countries other than the United States. Couldn't there be a requirement that nobody can be on the internet if they use hate speech?

Nell: Unfortunately, in 2016, the United States turned over control of the internet to an international authority made up of stakeholders from government organizations, members of private companies, and internet users.[10] Also, nobody agrees on what hate speech is. And I guarantee you within thirty seconds of any kind of a prohibition on hate speech, extremist Christians are going to say it's hate speech for people to make pro-LGBTIA comments because it makes them feel attacked. Mary has seen that kind of complaint in libraries.

Martha: Beyond hate speech, there's genocidal speech. Fomenting genocide or mass violence violates international human rights. Some speech genuinely and directly leads to the deaths of human beings. Facebook posts are linked with genocidal killings in Myanmar.[11] And under any constitution, messages inciting imminent violence should not be shielded from removal by a private company. Deciding what is and is not incitement can be challenging; local political contexts and code words matter. Here the insulation offered to social media platforms by Section 230 should be a benefit; the private companies have latitude in guarding against violence—and in forbidding harassment, hate, and disinformation. Ultimately, the Supreme Court of the United States is not the ultimate authority for decisions by private companies—they can make their own rules—and countries other than the United States are regulating. I do wonder about what kinds of norms young people ought to learn about internet use?

Mary: That presumes that you're willing to make everyone identify themselves before they go on, which really precludes people who are abused, undocumented workers, etc.

Martha: Well, you're an expert in this, Mary; isn't there some way to create secure privacy and anonymity, while still making sure that there is some accountability if needed?

Mary: You can use tokens to get through that first hoop to say, "Yes, I am legitimate." For example, libraries are able to authenticate a user's eligibility to use licensed databases, and then let the user through. Bitcoin and other cryptocurrencies can authenticate without tracing users. Maybe I passed that driver's license test, and then continue to be anonymous. But you want to trace back.

Martha: Why can't a company say it will only use the information under these very limited circumstances, but otherwise you cannot use our service? Why can't there be some requirements to be allowed to use the service, like sharing your personal identifying characteristics, which the company will not use except in urgent situations? Some companies already do this.

Mary: Many people come into public libraries, sometimes specifically, so they can use an internet terminal that isn't tied to their name and their identity. Maybe they want to check out drug abuse, or some other sensitive question. I want that door open for people who need it. Also, people who are searching from their own devices could be prompted to click on their local library to get online to chat or email with their library to ask a question. Or if they are using Facebook, wouldn't it be great if there was a button to click in order to ask their local librarian if a post is true or partially true? Google has already geocoded where users are coming in from and can easily direct users to their closest library website, so this is already feasible. The big platforms could also tabulate the referrals and send monetary support for library staffing!

Martha: Well, serious privacy protections are needed. It would be great if the platforms respected library privacy policies and deleted the identity of a questioner and the questions once answered unless the patrons indicate that they are fine with keeping it up. But people who organized the January 6 events leading to the Capitol assault and other assaults on the safety and lives of public officials use the privacy protections. We overcome them when we deal with foreign terrorism. Why not with domestic terrorism?

Mary: Public libraries actually have a click-through agreement that you're not going to do something illegal, you're not going to incite, you're not going to violate copyright.

Nell: What do we do about people who create many accounts to make it look like they have more support than they do,[12] or that the people they attack have more critics than they really do?[13]

Mary: It's like many other agreements. Violations must come to the other party's attention. Without tracking, it must be noticed in real time. How can libraries help law enforcement when there is a legitimate request, going through the right process, and yet still protect everyone? It's a challenge, if you don't trust law enforcement—and there are many who do not, often with good reason. Libraries track some user data, but as a matter of professional ethics, try to

keep as little as possible. Generally, the name of a user checking out a book is kept only until the book is returned.

Nell: Let's take the shareholder perspective. Most shareholders—whether through activist hedge funds or large institutional investors—care about a sustainable business model and want digital media companies to be safe and credible. It's like real estate. An investment in a safe neighborhood is worth more than one in a treacherous, badly maintained area. The State of New York's pension fund submitted a shareholder proposal to Facebook asking for reporting on its terms of service enforcement and risks from posted content.

Mary: You would think that there'd be a privacy bonus—that the consumer would go to the companies that have protections in place—but that has not been the case. It's because the general public is not educated as to how their information is being used. There are search engines that try to distinguish themselves as having privacy protections, unlike Google, but they are terrible and not much used.

Martha: The market isn't working well; concentrated economic power suppresses innovation on the supply side. On the demand side, as Mary notes, people may not know or care about options.

Nell: Twitter already has the blue check for verified people and other sources. I hope they develop an even more effective system to help users evaluate the trustworthiness of what they see.

Newt: Let's go back to basic concepts. You can buy a car, but if you want to drive it, you have to get a license for the car and a license for yourself because we recognize, as a society, that cars are dangerous. If you want to get on the internet, why shouldn't there be some sort of driver's license test?

Nell: Yes, and we require extensive disclosures for purchases of consumer goods, credit cards, and investment instruments, though there is a lot of evidence that no one reads them. Guns are very dangerous and there are very few restrictions on buying or using them. Laws are not always consistent. But I think it's fair to say that if a website can make you prove that you can tell what is a bus and what is a taxi to show that you're a real person, they could probably give you a test to say that you understand the difference between fake news and real news.

Mary: How does that help the person who is undocumented, the abused woman? If they need to pass some sort of test, they're not going to be able to go online for help.

Martha: Good point. What if the platforms have a duty and an incentive to create instructions so people can learn how to pass the test? Some campaign organizing sites already have a version of this—volunteers have to pass a quiz after reading an informational document or taking a training course before the volunteers can join the effort and send messages. People can retake the

test if they fail; sometimes there are observation and audit efforts by experienced leaders to follow up. Let's take a page from environmental policy and use regulation to tap into innovation—and here, push information technology toward learning?

Mary: Libraries already offer online instruction and have since the dawn of the public internet. It has ramped up in recent years to focus on helping users identify factual misinformation.

Nell: Who reads the user agreements?

Martha: I agree. But I do think that this underscores the divide in the world between people who are tech savvy and people who are not. But, are there more wholesale solutions than relying on the end users? Elements in the hardware and software can incorporate checks; the design can create speed bumps to slow down the transmission of unverified information. For example, a site can limit the number of people with whom content can be shared if it is not actually read, if its author is anonymous, or even if the site has not subjected its content to a review. People are more likely to share fake or false information when they are not thinking about it. The platforms prompt people to think about whether what they are sharing is true; that can slow the spread of misinformation. The architecture of sites has effects; designers should have responsibilities. Putting the whole burden on the end user seems unfair and ineffective.

Newt: I believe there is a need to think about the public interest when we're dealing with this technology. Or should we just let the market decide?

Martha: The public interest matters and can be defined at least in some instances, such as protection of children. The market matters too and has a role—but if the big players are insulated from market pressures, it is an inadequate avenue. Addressing the misinformation will require the involvement of every part of our culture. A 2020 study showed that exposure to inaccurate or misleading information about COVID-19 contributed to people's disinclination to seek out more information.[14]

If we analogize the social media misinformation problem in an area like the pandemic or exposure to toxic waste in the physical environment, there are public safety grounds for government mandates on private entities to remove misinformation or hateful material and report on the methods used and quantity of problematic content removed. To those who object that "this is censorship," let's be clear, moderation is happening every day by private platforms, and just like editing of traditional media content, it is lawful and even constitutionally protected. The big platforms remove massive amounts of spam—including fake accounts and bot-produced messages—and no one objects. Otherwise we'd be "flooded"—overwhelmed with more content. That would interfere profoundly

with speech, with listening, and with the trust needed for both. In the spirit of environmental protection, why can't government require platform companies to report on their methods and rates of such removal efforts—and their research and development even in trying to detect and control disinformation?

Nell: One thing we've learned is that the market is not a good mechanism for evaluating information. The market for slanted or bad information is very robust, as we see with the financials of FOX and its subsidiaries. Barriers to entry are lowering all the time and "channels" like Newsmax and OANN (funded at least in part by Russia) are out-FOXing FOX. What has been most effective at pushing back has been the defamation lawsuits filed by Dominion for the unfounded claims of the failure of their voting machines. That is a combination of market forces and government, with the laws and court system. The Dominion lawsuit got Newsmax and FOX to back down. On the other hand, I am also aware of strategic lawsuits against public participation (SLAPP) claims, which are filed to suppress or chill legitimate speech because even if they are frivolous it is too expensive to fight them.

Newt: Justice Holmes said in *Schenck v. the United States* that the First Amendment does not protect someone for "falsely shouting fire in a crowded theatre and causing a panic."[15] The two key elements there are that the speech has to be false—obviously if there really is a fire you want someone to let everyone know—and that the false information can cause a panic, where people can get hurt. The head of the FBI testified in the Senate about what we know about why the January 6 attack on the Capitol happened. Social media was a crucial part of not just organizing and getting the mob in place, but of spreading the lie that the election was "stolen," which led to disaster. I was a law clerk at the United States Supreme Court in 1951, when the *Dennis* case was decided.[16] My boss, Chief Justice Vinson, decided that speech advocating the violent overthrow of the government was not protected speech.

Nell: The difficulty is the pervasiveness of media content. NewsGuard on my browser does a good job of identifying the reliability and objectivity of news sources that I view, including links on Twitter. People like Alex Jones and Sidney Taylor admit in court filings that no one should believe what they say, but their listeners just think that makes them more credible. They keep saying it, and the press keeps amplifying it.

Martha: Lies are hard to identify and regulate. In *United States v. Alvarez*, the Court protected lying about having a military medal and struck down a portion of the Stolen Valor Act because it failed to pursue less restrictive measures.[17] But I think Dad's point is: Can prevention of falsehoods that carry risks of violence and havoc be compatible with the First Amendment? I think the answer is yes.[18] Changing technologies and understandings of human

behavior may call for adjusting what counts as the required proof of "imminent harm." Cass Sunstein analogizes regulation of climate change to the incitement of violence as less-than-imminent, but potentially devastating harms could be preventable in both cases.[19]

Nell: But keep in mind the Arab Spring and remember how excited and happy everybody was that social media made it possible for people to rebel against a totalitarian regime. Believe me, that is the way that the January 6 insurrectionists and whoever the next group is are going to characterize themselves. The consequence is that we treat them all the same. One thing we've learned is that the less "authorized" a source is, the more certain kinds of people find it believable, especially those who feel overlooked or condescended to by traditional sources.

Mary: When there is regulation of the platforms, the platforms just go underground or abroad, making it even harder for authorities to monitor.

Martha: Counterterrorism experts should be consulted about how to weigh that risk, perhaps by using a cost-benefit calculus. Standards should play a role.[20] Meantime, to improve trust in information, can the librarians do something? Mary?

Mary: I've always thought that there should be a social media alternative to Facebook that's been created by librarians and others in the public sphere, a public option. However, libraries and local governments don't have the resources. They're not in the business of making scads of money by monetizing everybody's private information. How about taxing the platforms to support the creation of the public internet option and also seek public contributions (as the Corporation for Public Broadcasting receives) and private contributions from individuals and foundations?

Newt: People have asked me from time to time, why should there be any public funds for television? And I say, "We have private parks. Why should there be public parks? We have private schools. Why should there be public schools? The same with hospitals and libraries." PBS was a pioneer in the kind of programs other broadcasters did not think would make money, like educational programs for children and the Ken Burns series on American history and culture. The closest thing we have now to the Walter Cronkite news programs of the '60s and '70s is the *NewsHour* on PBS. PBS and NPR consistently rank as the most trusted news sources in the United States, or among the most trusted.[21] There are some things the market doesn't provide. I believe what's missing today is that no one I know of who's in the new technology world is talking about something that is important, namely, two words: public interest. Nobody is thinking or talking about that. Those two words are absent.

Martha: What does public interest mean to you?

Newt: Public interest means to me that we—the public, the government—create certain things that the market does not provide. I believe in the marketplace because the market has produced more good things for more people than any other system of economic activity. But at the same time, the market doesn't provide everything. And when there are things that the market does not provide, it is there where I believe government comes in.

Martha: Ethan Zuckerman, for example, calls for public investment to build a public service web.[22] We don't seem to have much trouble recognizing that there's a public interest in maintaining the essential and vital services of electricity, water, and gas. And these days, isn't communications media one of those vital goods or services that everyone needs? The FCC could mandate "net neutrality" rules to prevent internet service providers from blocking applications and websites or from granting priority access and hence treating different consumers and different providers differently. The FCC ultimately did use its statutory authority to designate the internet as a "public utility."[23] Then, with a change in leadership, the FCC deregulated broadband. Yet the language of public interest remains in place; the basic idea is that if the internet is open to all—because it operates as a "common carrier"—then it can be subject to regulation accordingly.[24]

Newt: That goes back to the days of Louis D. Brandeis, and reformers more than one hundred years ago, who realized that certain things not provided by the market require that the government intervene. It seems to me that when the internet arrived, that thinking stopped.

Mary: So who would be the agency? Would it be the FCC that would apply the public interest to the internet?

Newt: I've seen all kinds of proposals. And I know Congress has before it, at the moment, many proposed legislative actions. Tom Wheeler, former chairman of the FCC, has written that a new agency dealing only with digital communication should be created. Given the lack of oversight that went into the transfer of authority for the internet out of the United States, I would like to feel more confident that this multifactorial problem is being addressed in a thoughtful way.

Nell: I still think the internet is a net benefit and very much in the public interest—its democratizing effect, where anyone can start a blog or write a tweet and get a job or a book deal or an answer about a consumer problem without being kept on hold for hours with customer support. It's remarkable to be able to give voices to people who didn't own a newspaper and didn't have access to television. And so I think we want to talk about how beneficial that's been, and that we don't want to impede that in any way.

Newt: You're right. And I think that's a very important part of the balance.

Martha: That's right. There are inevitably trade-offs. It's often easier to talk about what we think is the big worry, rather than getting our arms around everything. So what is the biggest worry you have, the worst thing that's going on with social media that you would hope government would regulate?

Newt: I would say it's the advocacy of violence. I turn to what the Europeans are doing through the European Union and also national laws, for example, in Germany, and in England. The EU's Digital Services Act develops comprehensive regulation of social media and other digital services, with particular interest in improving the media ecosystem and guarding against disinformation.[25] Germany, which emerged from the Nazi era with explicit protections against hate speech, has stepped up amid rising right-wing extremism and the digital context with a twenty-four-hour deadline before fining platforms that allow distribution of hate speech.[26] England is debating a law that would punish platforms for keeping up "lawful but awful" content, though critics warn of government censorship.[27] I see the need for some interventions involving the things that are dangerous to the existence of democracy in this country.

Mary: From what I understand, the neo-Nazi movement is alive and strong in Germany. I don't know that having the capacity to regulate hate speech is actually working.

Martha: We also don't know if it would be worse without it. The question has to be, compared to what?

Mary: The ACLU makes the argument that when you suppress speech it only makes the people who have been suppressed stronger. They've done the research to show that people find themselves as martyrs, and they organize better, especially today when we've got all these platforms offshore. I'd rather have the speech be more transparent. What am I most worried about? I'm worried about the next January 6. But I would narrow it to incitement, not hate speech. I would want the big private entities that operate like a new public square—so much that you can't participate in your local political party chapter without joining the Facebook group—to use the same standards that apply to the public square.

Nell: I worry about two things. The first is directly related to what we're talking about. I worry less about what individuals have to say on Twitter or Reddit than about "news" outlets, like Newsmax and OANN, which people assume meet the traditional standards of journalistic integrity and give weight to them that they have not earned. They are fly-by-night organizations that set up a channel on YouTube and make stuff up that has nothing to do with reality. And they ask questions at White House press conferences, and congressional representatives and senators go on camera and legitimize them. So I worry

about the devaluing of empiricism, the blurring of lines between legitimate, fact-based media and the twenty-first-century version of the *Weekly World News* tabloid with front-page stories about aliens and Elvis being spotted at a Burger King.

Second is the problem of dark money in politics. This is not a case of journalism failing to report on a topic of vital importance; it is a case of that information not being available. Dad, you and your long-time friend, the late Henry Geller, who was general counsel at the FCC, tried to get the FCC to use its authority to force sponsors of political television commercials to disclose who paid for them. Your efforts were thwarted when Congress prohibited the FCC from pursuing any disclosure requirements. I believe broadcast stations and other resources would gain a lot of credibility if they made it clear who was behind the ads they run, and if "networks" like OANN revealed that they are funded by Russians.[28]

Martha: I agree. Government censorship is bad, and so is quickly spreading hateful expression. Frankly, I think the Enlightenment itself is in trouble. Ideas traceable to Europe in the seventeenth and eighteenth centuries—that there is a truth to seek, that science is more reliable than antiscience, that reasoned debate can resolve or manage disagreements and prevent violence, that the dignity of every person is sacrosanct—have been called into question. These days, some of the same people paid by corporate money to create pseudoscientific articles questioning the link between tobacco and cancer later moved over to write pseudoscience on climate change. Journalists, scientists, and academics built up systems for testing claims and building paths to knowledge.[29] Unfortunately, the internet allows so many falsehoods, doubts, and conspiracy theories to circulate widely while knocking down the vehicles of verification.

Newt: Justice Jackson said, "The Constitution of the United States is not a suicide pact."[30] When I think about what happened on January 6, I think it was madness. Why are the country and Congress not standing up and saying, "This has got to stop"? Where is Edmund Burke when we need him? He said, "The only thing necessary for the triumph of evil is for good men [and women] to do nothing."[31] Our courts have drawn a clear line: Speech advocating the violent overthrow of our government is not protected when it creates a likely and imminent danger of serious harm. That line should apply to our new technologies, and our government should enforce it.

Mary: I concur about the dark money. I think the disclosure rules that Dad's pointed to and written about should be adhered to.

Martha: I agree, but a Fourth Circuit decision struck down mandated disclosure of campaign contributions as violating the First Amendment.[32]

Nell: But *Citizens United* itself specifically says that corporate political contributions are legitimized by shareholder oversight: "The First Amendment

protects political speech; and disclosure permits citizens and shareholders to react to the speech of corporate entities in a proper way."[33] When the contributions are not disclosed, much less approved by shareholders, the contributions should not be considered free speech. Reporter Judd Legum has done some great work on exposing the hypocrisy of companies' political contributions. Of course they routinely give to everyone—state, local, national, Democratic, Republican. But following the #BLM protests, Legum's reporting led to companies withdrawing support from elected officials who had failing grades from the NAACP, and his reporting after January 6 led to withdrawal of support from representatives who refused to acknowledge the legitimacy of the election results. Indeed, in a few cases, including JP Morgan Chase, they halted all contributions while they reconsidered their policies.

Martha: The SEC has failed to exercise its authority to regulate in this area. This may be an area of action sought by the new commissioners.

Nell: This is another issue that comes up often in shareholder proposals, though even a 100 percent vote is nonbinding; shareholder proposals are advisory only. In 2021, thirty-three shareholder proposals on political spending were submitted at companies including Netflix, Delta, Kimberly-Clark, and Nike, Inc. There are other kinds of market-based pressure: boycotts, which have not been effective, and asking companies to stop advertising on certain "news" programs, which has been more effective.

Mary: Putting money into public libraries, into local journalism, and into education is the answer. You fight bad speech with good speech.

Martha: We do not all agree about every aspect of the problem or about potential responses, but we converge on some ideas.

Concluding Comments: Here are ideas that we hope gain traction; some involve purely private action, and others require government action, and hence could trigger constitutional concerns. We think that these at least deserve discussion:

1. Major internet and other media companies should support the articulation of a core set of journalistic principles they will adhere to in anything they label "news"—principles such as requiring at least two sources for anything reported as a fact; committing to reflect all sides and draw on diverse sources; and committing to check and report on credentials and conflicts of interests of sources.
2. Significantly enlarge public support for public libraries, for media literacy education, and for public media. Incentivize the platform companies not only to direct users to the public libraries, but also to contribute the resources libraries need to answer user questions.
3. Establish a nonprofit public internet initiative to receive public funds while insulated from government control.

4. Large internet platform companies, functioning like the public square, should be required to make readily available current statements of their policies and practices with regard to their "community standards" governing removal or demotion of objectionable content and should be required to respond promptly to reports of false or harmful content.
5. Internet platform companies should be allowed—as they are currently by US law—to adopt and implement standards for removing or demoting content based on their own community standards and for suspending or terminating accounts that violate those announced standards. We urge resisting any regulations or judicial interpretations that would prevent such private content moderation.
6. Internet platform companies should be required to adopt policies to halt or at least slow the distribution of content identified by expert reports as inciting imminent violence.
7. The FCC, FTC, and SEC should be authorized to and should exercise the authority to require disclosure of the names of people and actual companies (not special-purpose opaquely named entities) paying for political ads.
8. A bipartisan national commission should review and analyze emerging social media regulations from other countries and self-regulation by companies, and assess what ideas, consistent with the US Constitution and existing international law agreements, should be adopted here. We hope others will join this conversation!

18

Profit Over People

How to Make Big Tech Work for Americans

AMY KLOBUCHAR

"Facebook has realized that if they change the algorithm to be safer, people will spend less time on the site, they'll click on less ads, they'll make less money."[1] *If they allow people to be safer, they'll make less money.*

Those were the words of Facebook (now known as Meta) whistleblower Frances Haugen from her groundbreaking October 2021 Senate hearing testimony, in which she—armed with mountains of internal company documents—thoroughly described the harms Facebook is inflicting on our children, families, and society.

In both Haugen's testimony as well as the subsequent release of the "Facebook Papers,"[2] we learned how Facebook's research showed how the platform consistently feeds harmful content to its users, putting profit over people time and time again.

For example, we learned that more than 13 percent of teenage girls reported that Instagram made their thoughts of suicide worse.[3] We also found out that, despite the internal research that would have led most responsible companies to reconsider the impact of their products, the company still proposed "Instagram Kids" (which it recently suspended following significant public outcry).[4]

We also learned how Facebook's algorithms were fostering a cesspool of misinformation and hate and how the company's automated systems continued to allow 99 percent of violent content to remain on the platform.[5]

At the time, I called Frances Haugen's testimony a catalyst for change.

Since her testimony, I have met with parents from around the country who are concerned about the impact of social media on their children. Parents are worried; they see their kids getting addicted to these platforms, being fed misinformation and harmful content with no recourse for help. I heard from a mom in Minnesota who described social media like an overflowing sink that she couldn't

turn off: When she goes and mops up the water, the faucet keeps flowing. The harmful impacts of these platforms don't stop.

Like Frances Haugen, these parents are catalysts for change.

As we have seen throughout our history—from worker strikes in Chicago against monopoly robber barons, to the Granger Movement of America's farmers—Americans across the political spectrum have come together to drive the reforms needed to meet the great challenges facing our nation. And now our country's parents are saying enough is enough.

Today, with the rise of the digital economy and all of the opportunities and problems it has brought, we are at another critical moment in our history and an inflection point for reform. And there is no time like the present—our privacy laws and competition policy haven't been meaningfully updated since before the internet was created. Now, faced with sprawling digital monopolies, including but not limited to Facebook, that harvest and use our data with impunity, target vulnerable users with harmful content and toxic disinformation, and control access to key online markets, it has become clear that America needs laws that are fit for today's digital economy—laws that are just as sophisticated as the tech giants that increasingly dominate American life. That means a federal privacy law, provisions to better protect children online, more transparency in social media algorithms, and yes, updates to our competition laws.

Just as they have in the past, Americans of all political stripes are coming together to confront the challenges of this era. And each American stepping up and speaking out is a catalyst for change.

The Rise of Digital Technology

Over the last several decades, American society has been transformed by the rise of digital technology. Today, we have routine access to powerful tools that enable us to perform tasks that would have seemed extraordinary just twenty-five years ago.

We have gone from the days of *Wall Street*'s Gordon Gekko with his cell phone—affectionately known as "the Brick," which weighed two pounds and was thirteen inches long—to smart phones the size of a wrist watch packed with apps that can do a multitude of things. From a device that monitors our vital signs and instantly sends alerts if something is off to an app that connects to a hearing aid and can instantly translate speech into more than thirty languages, technological advances have improved our lives in many ways.

I can connect with my daughter instantly, via video, even if she's across the country or on the other side of the world.

People in underserved medical areas can consult with top medical experts via the internet, with X-rays getting transmitted in a nanosecond.

Due to the COVID-19 pandemic, we are raising a new generation of kids who learned to read during Zoom video conferences, sometimes with their parents working remotely in the next room.

And upcoming advances in computing power, wireless networks, connected devices, and artificial intelligence promise to make these technologies an even greater part of American life in the years to come.

The Bureau of Economic Analysis estimates that the digital economy accounted for nearly 10 percent of the United States' gross domestic product in 2019, with a growth rate that was more than double that of the economy as a whole.[6] That figure is certainly higher today, but even an updated percentage would understate the significance of digital technology. As digital advancements spill over into other sectors, they have the potential to drive innovation, growth, and employment across our economy.

The growing importance of digital technology in America, however, has raised serious concerns from the public and a host of challenging issues that policymakers must confront. At the root of these concerns is the fact that these transformative technologies are controlled by a handful of dominant technology companies. These companies, which started out as scrappy innovators, have grown to become some of the most valuable companies the world has ever seen, operating powerful online platforms and acting as digital gatekeepers to millions of businesses and billions of consumers around the world. Although these platforms provide valuable services that many of us use every day, these benefits have come at a cost. And every day, those costs are becoming more and more obvious.

These gatekeeping platforms have amassed unprecedented power over our everyday lives. That includes power over Americans' personal data, over how businesses reach customers, and over the advertisements and news that people see, as well as the unfettered ability to amplify information—or disinformation—to an audience of millions in the blink of an eye. It also includes the power to dominate key digital markets by suppressing or buying out any potential competitors.[7] Since the dawn of the digital economy, most of this power that has been exercised by these platforms has gone unchecked.

The result has been highly concentrated digital markets where consumers' personal data is free to be sold or monetized as we all become profit centers for big tech companies, profit-maximizing algorithms force feed harmful misinformation and disinformation to millions of consumers, and small businesses trying to compete online are boxed out of true competition by self-preferencing behavior.

The current state of affairs not only harms consumers, businesses, and technological innovation, it threatens our free press, public health, and our democracy. Local newspapers around the country continue to close and newsrooms continue to shrink at an alarming rate.[8] Even after more than 987,000 Americans have lost their lives to the coronavirus,[9] the major digital platforms are still rife with vaccine disinformation.[10] And the January 6, 2021, insurrection at the US Capitol demonstrated how digital platforms can be easily co-opted by extremists to stage an assault on the temple of our democracy.[11]

The companies running the dominant digital platforms—Amazon, Facebook, Google—are American companies. In fact, they are some of this generation's business success stories. Over the years, they have innovated, created new markets, employed thousands of workers, and helped drive our economy forward. While many Americans are understandably proud of the success of American technology companies on the global stage, we must recognize that no company—no matter how successful or innovative—can be permitted to run roughshod over the interests of the public.

As these new digital markets evolve, our laws must evolve along with them to promote the public good and ensure opportunities for the next generation of American success stories. We must shape the technology landscape to ensure that these digital platforms are agents of progress and public good, not tools of destruction. Where the law does not address the novel problems caused by these digital giants, Congress must act. The problem? In the past fifty years as our economy changed dramatically with the internet, not one major federal law to protect consumers has passed.

America needs a comprehensive approach to reining in the excesses of Big Tech. That includes setting rules of the road to prevent the spread of harmful content,[12] protecting privacy and promoting transparency, strengthening our antitrust laws, preventing anticompetitive discrimination, and empowering our enforcers to hold even the most powerful digital platforms accountable when they engage in harmful conduct.

Cracking Down on Harmful Online Content

The internet is rife with toxic, hateful, and inaccurate content that causes harm in the real world: harm to our health, children, and our democratic institutions. And the examples continue to multiply.

The *first* problem that came into sharp focus in the recent past is the issue of vaccine misinformation. The numbers are revealing: In May 2021, 30 percent of adults said that they were unwilling to be vaccinated or uncertain about doing so.[13] In many cases that was because of misinformation they read online.

Two-thirds of unvaccinated Americans believed common myths about the vaccine[14]—myths that went viral on platforms like Facebook and Instagram.

I remember talking to a café worker early on when, only seniors were eligible for the vaccine, and he said, "You know, my mother-in-law was going to get it, but she decided not to because she read online that they would implant a microchip in her arm." I asked where his mother-in-law had read that lie. His answer: "Facebook."

This is far from surprising. The *New York Times* reported that an internal Facebook study found that the most viewed post on Facebook for the first three months of 2021 led to an article suggesting that the COVID-19 vaccine may have been involved in the death of a "healthy" doctor in Florida. The post was viewed by nearly fifty-four million Facebook users, but far fewer people saw the update reporting that the medical examiner did not have sufficient evidence to conclude whether the vaccine had actually contributed to the doctor's death.[15]

As we look at all the needless COVID-19 deaths in the past months, primarily among people who are unvaccinated,[16] it is clear that misinformation about the safety of vaccination has had devastating real-world consequences. That is why I introduced the Health Misinformation Act to help hold digital platforms accountable for the spread of health-related misinformation online during public health emergencies.[17]

Second, we have to address the fact that digital platforms' algorithms are facilitating the spread of harmful content to users. That means we need transparency in how these algorithms work *in order to* stop the unchecked spread of harmful content.

As I noted, Frances Haugen explained that Facebook has known for years that its algorithm promotes content to entice users, particularly young users, to spend more time on the platform.[18] But the algorithm optimizes engagement without sufficient regard to whether content could be addictive or harmful to the user, with predictably devastating results. For example, Ms. Haugen highlighted the spread of eating disorder–related content, which can have particularly dangerous effects on young women, testifying that "Facebook knows that they are leading young users to anorexia content."[19] In fact, when my colleague Senator Richard Blumenthal's office created an Instagram account posing as a thirteen-year-old girl who noted interest in dieting and weight loss, Instagram's algorithm quickly began promoting content that glorifies eating disorders, referring the user to accounts with titles like "eternally starved," "I have to be thin," and "I want to be perfect."[20]

I have introduced legislation to combat eating disorders over the years because I know they can be fatal.[21] Anorexia has an estimated mortality rate of around 10 percent, making it among the deadliest of all mental health disorders.[22] But that was not enough to convince Facebook to take action to ensure its algorithms

were not leading susceptible users to this harmful content or to build an effective system for removing it from the platform.

And this is not just a problem with Facebook. Children also risk exposure to toxic content on TikTok, YouTube, Snapchat, and other platforms. Some platform features can effectively encourage reckless behavior by teens. For example, Snapchat's now-discontinued "speed filter"—which allowed users to capture how fast they were moving and share it with friends—has been linked to a number of deadly or near-fatal car accidents involving teens.[23]

Why would some of the richest, most sophisticated companies in the world fail to do everything in their power to ensure that their products were not harming children? The answer is simple: User engagement is more important for these businesses and their bottom line than the safety and well-being of their users, even their youngest users. Their singular focus is on fostering a large base of engaged users to whom they can target advertising, no matter the human costs.

This is unacceptable.

We cannot let the most powerful tech companies get away with doing less than the bare minimum. Systems that catch less than 20 percent of harmful content are inadequate. The dangers are simply too great. That's why I am fighting to increase transparency about how their existing systems work so we can figure out what improvements are needed.

Third, it is essential that we address the use of digital platforms by extremists, both abroad and at home. We saw this on January 6, 2021, when a violent mob staged an attack, not just on our Capitol, but on our democracy.

I know. I was there.

I saw Democrats and many Republicans come together on that tragic day to protect our democracy. But the ease with which the insurrectionists used digital platforms to build support and plan their attack should give every American pause about how these technologies can be used to threaten our system of government.

That is why holding these digital platforms accountable for content that can lead to real-world harm is critical. People spend many hours online every day,[24] and we have to find a way to ensure that the fundamental laws governing our society—like civil rights protections guarding against discrimination—are fully enforced online. While we know we need reform, we must allow for freedom of speech and maintain and foster innovation at the same time.

One proposal that strikes the proper balance is the SAFE TECH Act, which I introduced with Senators Warner, Hirono, and Kaine. It would allow social media companies to be held accountable for enabling cyber-stalking and civil rights violations, targeted harassment, and discrimination on their platforms.[25]

These are simple, straightforward proposals that directly address the most harmful content to protect the most vulnerable users. But we also know that

limiting harmful content isn't a silver bullet. That's why we need comprehensive privacy protections that give all users greater control over their personal information.

Online Privacy and Data Protection

Dominant digital platforms, such as Google, Facebook, and Amazon, collect an enormous amount of information about our daily activities in real time.[26] And the big tech companies are not the only ones keeping tabs on us. Data brokers also buy, process, and sell massive amounts of personal information about consumers. They collect information from the Department of Motor Vehicles, from public records, from our grocery store loyalty cards, and even from other data brokers.[27] They know what we buy, who our friends are, where we live, work, and travel, and more. Though services such as Google Search and Facebook's Instagram are offered to us for free, these companies use the data they harvest from us to make billions of dollars by enabling advertisers to target us directly—we are the product.

But these companies leave consumers in the dark about what personal information is collected, how it is being secured and used, and with whom it is being shared. The simple act of a consumer visiting a utility company's website to pay a monthly gas bill can allow dozens of companies to profit from her, for the most part without her knowledge. Facebook and Google are likely to know about that consumer paying her bill—even though they had nothing to do with the transaction. If the gas company runs advertisements on Facebook—as many do—Facebook would have trackers embedded on the gas company's website.[28] And if she uses the world's most popular web browser, Chrome, Google would know what websites she visited.[29] Both companies collect and analyze this kind of information, building a detailed profile of her and giving advertisers access—for a price, of course.[30]

Consumers should have transparency and control over what is being done with their data, and companies should have clear obligations about how they can collect consumer data, how it can be used, and what they have to disclose to consumers whose data fuels these multi-billion-dollar industries. Unfettered access to personal data is what gives these platforms so much of their power and drives their businesses and their profits. For example, Facebook has estimated that each quarter, it earns $51 on each US user.[31]

We know that when consumers have a choice about their data, they decide to protect it. That is what happened when Apple let its users decide whether they wanted to allow apps to track their cross-app online activity within the Apple ecosystem: Sixty-five percent of users declined to allow the tracking.[32] That says something.

Even though the dominant platforms have some business incentives to protect user privacy, we cannot just trust them to act in consumers' best interests when it comes to the collection and use of consumer data. Even Apple's decision to ask users whether they wanted to allow third-party apps to track their cross-app activity did not limit Apple's own ability to track users' activity across the Apple ecosystem that it controls. The one thing we can trust is that all of these companies will act in their own best interests, which don't always align with those of consumers. That is why we need laws to protect consumers' personal data.

I am working with my colleagues on the Commerce Committee, including Chair Cantwell, on the Consumer Online Privacy Rights Act. That bill would establish digital rules of the road for all companies to follow and ensure that consumers can access and control how their personal data gets used. It also gives the Federal Trade Commission and state attorneys general the tools they need to hold big tech companies accountable.[33] With tech terms of service many people don't know what they're agreeing to or what data they are letting the platform use or sell. The bill requires greater transparency to make sure that when consumers give consent for their data to be used, it is informed consent. As of the time of this writing, bipartisan negotiations continue on this legislation along with other proposals to improve online privacy, including Senator Markey's legislation to update the Children's Online Privacy Protection Act, which passed into law more than twenty years ago. His bill, the Children and Teens' Online Privacy Protection Act, would prohibit internet companies from collecting personal information from anyone thirteen to fifteen years old without the user's consent,[34] which is something the whistleblower endorsed in her testimony. Current law only applies to children under thirteen.

I have also introduced bipartisan legislation to protect consumers' private health data, which is increasingly tracked by connected devices. Health data tracking apps have given tech companies access to unprecedented amounts of consumer health data,[35] but current law does not adequately address the emerging privacy concerns presented by these new technologies. The Protecting Personal Health Data Act, which I introduced with Republican Senator Lisa Murkowski of Alaska, addresses these health privacy concerns by requiring the Secretary of Health and Human Services to promulgate regulations for new health technologies such as health apps, wearable devices, and direct-to-consumer genetic testing kits that are not regulated by existing laws.[36]

There is strong bipartisan consensus on the need to protect online privacy. We have been working on these issues for many years, and the time has come to get something done.

Reviving Antitrust Enforcement

At their core, many of the problems that plague the digital economy are caused or exacerbated by the unchecked power of the dominant digital platforms. The monopolization of large swaths of America's digital economy has deprived consumers of real choice and suppressed the competition that would drive platform innovations to address problems like the spread of harmful content or the lack of online privacy. As we confront these issues, it is important to recognize that our problems with Big Tech are also symptoms of America's larger monopoly power problem.

The spectacular growth of big tech companies has come with an enormous societal price tag: consolidation of unprecedented power within a handful of companies that exercise control over our personal data, control over the ads we see and what news we watch, and monopoly power in key digital markets like search, social media, online advertising, commerce, and software app distribution.

As I have discussed in this article, gatekeeper platforms use their power to impose their will on smaller rivals—sometimes excluding them altogether. Such platforms collect as much consumer data as they can, and they buy out their competitors, including potential future competitors.

An email from Facebook's Mark Zuckerberg commenting on emerging rival Instagram—before he bought it—highlights the attitude of these modern-day robber barons to competition: "These businesses are nascent but the networks [are] established, the brands are already meaningful, and if they grow to a large scale *they could be very disruptive to us*."[37]

Those words from Zuckerberg encapsulate the problem at hand. When they were scrappy start-ups, these platform companies were once proud "disrupters" of the status quo. But now that they have reached the top of the mountain, disruption (meaning competition) can no longer be tolerated. Instead, it must be suppressed or bought out. We will never know what bells and whistles Instagram could have developed to improve privacy or combat misinformation. Why? Because Facebook bought it.

We know that it is competition that gives consumers lower prices and forces manufacturers to constantly innovate to improve their products.[38] Competitive markets force companies to pay workers fair wages and improve working conditions.[39] They provide opportunities for entrepreneurs to start and grow new businesses across the country, fueling future economic growth.[40] What the United States of America needs now is a renewed antitrust movement—one that is grounded in a pro-competitive economic agenda that will actually help capitalism and innovation, not just among tech companies, but across the economy.

People of all political stripes who truly believe in the promise of our free enterprise system should come together to encourage greater antitrust enforcement. Rigorous antitrust intervention is a must in our capitalist system because it levels the playing field for free and fair competition.

While there is still a lot of work to do, we are making some progress. The Justice Department is finally taking action to stop Google from engaging in anticompetitive and exclusionary practices in the search and search advertising markets.[41] The Federal Trade Commission has filed a complaint against Facebook for engaging in an illegal buy-or-bury scheme to maintain its dominance.[42] And there are a number of significant actions against these companies from large coalitions of state enforcers.[43]

When I was a young lawyer, I worked for a law firm that represented MCI, which at the time was an innovative young telecom company that was being held back by local monopoly carriers. Along with Republican and Democratic administrations, MCI took on AT&T and ultimately broke up the Ma Bell monopoly, lowering long-distance calling prices for consumers across the country, revolutionizing the telecom industry, and launching a highly successful cellular phone company. Many—including a former AT&T president—have suggested that AT&T's famous breakup made the company stronger because it forced it to be competitive.

Divesting assets is one of many ways to deal with competition issues. It doesn't mean the companies go away, it simply means people are allowed to compete against them.

But this moment calls for more than just potential breakups. To truly rein in the power of Big Tech, we need a comprehensive update to our antitrust laws that addresses the conduct that is harming competition. That's why I've introduced comprehensive legislation to reinvigorate America's antitrust laws and restore competition to our markets.[44]

First, the legislation empowers our antitrust enforcement agencies by providing much-needed resources. For years, enforcement budgets at the Justice Department's Antitrust Division and Federal Trade Commission have failed to keep pace with the growth of the economy, the steady increase in merger filings, and increasing demands on the agencies' resources.[45] To enable the agencies to fulfill their missions and protect competition by bringing enforcement actions against the richest, most sophisticated companies in the world, this bill would authorize increases to the agency's annual budgets. Our enforcers can't take on some of the biggest companies in history with duct tape and Band-Aids.

While we must increase funding, we also must update the laws governing mergers. This legislation restores the original intent of the Clayton Act to forbid mergers that "create an appreciable risk of materially lessening competition" rather than mergers that "substantially lessen competition," where "materially" is defined

as "more than a de minimis amount." By adding a risk-based standard and clarifying the amount of likely harm the government must prove, enforcers can more effectively stop anticompetitive mergers that currently slip through the cracks.

The bill also clarifies that mergers that create a monopsony (the power to unfairly lower the prices a company pays or wages it offers because of lack of competition among buyers or employers) violate the statute.

Certain categories of mergers also pose significant risks to competition but are still difficult and costly for the government to challenge in court. For those types of mergers, the bill shifts the legal burden from the government to the merging companies, which would have to prove that their mergers do not create an appreciable risk of materially lessening competition or tend to create a monopoly or monopsony.

Decades of flawed court decisions have weakened the effectiveness of Section 2 of the Sherman Antitrust Act to prevent anticompetitive conduct by dominant companies.[46] The bill creates a new provision under the Clayton Act to prohibit "exclusionary conduct" (conduct that materially disadvantages competitors or limits their opportunity to compete) that presents an "appreciable risk of harming competition."

Competition and effective antitrust enforcement are critical to protecting workers and consumers, spurring innovation, and promoting economic equity. While the United States once had some of the most effective antitrust laws in the world, our economy today faces a massive competition problem. We can no longer sweep this issue under the rug and hope our existing laws are adequate.

Digital Platform Legislation

We have also seen in recent years that the rising market power of the dominant digital platforms is uniquely and particularly harmful to consumer choice and requires a specific response.

One big area of needed reform is with "self-preferencing" by dominant tech companies such as Amazon, Apple, and Google.[47] While we know these companies have created many useful innovations, we also know that today they are able to exclude their rivals and use their market power to promote their own products and services over those of other sellers, hurting small businesses, discouraging innovation, and reducing benefits to consumers.[48]

Numerous business owners have come forward to report instances of dominant platforms pressuring them to use extra services, like shipping, to get preference on the platform, or using nonpublic data—such as information gathered from users and small businesses—to build knockoff products and then compete against the companies that were paying to sell on their platform.[49]

This isn't your local grocery store chain creating its own potato chips to compete with a brand-name product. This is Amazon using incredibly detailed information that they gather from sellers on their platform to create a copy-cat product and box out competition.

In one case reported by the *Wall Street Journal*, an employee of Amazon's private-label arm accessed a detailed sales report, with twenty-five columns of information, on a car-trunk organizer produced by a small Brooklyn company called Fortem. In October 2019, Amazon started selling three trunk organizers of its own. When shown the collection of data Amazon had gathered about his brand before launching its own competing product, Fortem's co-founder called it "a big surprise."[50]

Recent investigations found that Amazon "ran a systematic campaign of creating knockoffs and manipulating search results to boost its own product lines" in India, one of the company's fastest-growing markets.[51]

It's not just Amazon. At our hearing on app stores, we heard how Tile has been blocked from using the chip technology that Apple utilizes for Apple products that compete with Tile.[52] Google puts its own products first in the rankings, even if competitors have better products: Yelp has noted for years that Google puts its own review content above that of Yelp and other local review sites in search results.[53]

There's nothing stopping Google from displaying its travel booking service ahead of some competitors that are more preferred by consumers—Expedia, for example. This doesn't only hurt the travel booking services—it hinders consumers from finding the best deals.

This is about getting consumers the best deals and ensuring that small businesses have opportunities to succeed. It's as simple as that. And for every business that has come forward, I know there are many more that will not go public for fear of retaliation by one of the companies.

That's why I worked with Republican Senator Chuck Grassley and a bipartisan group of nine other co-sponsors to introduce legislation that restores competition online by establishing commonsense rules of the road for dominant digital platforms to prevent them from abusing their market power to harm competition, online businesses, and consumers.[54]

The bill sets clear, effective rules to protect competition and users doing business on dominant online platforms. It prohibits platforms from abusing their gatekeeper power by favoring their own products or services, disadvantaging rivals, or discriminating among businesses that use their platforms in a manner that would materially harm competition.

It bans specific forms of conduct that are harmful to online businesses and consumers, including preventing another business's product or service from interoperating with the dominant platform or another business; requiring a

business to buy a dominant platform's goods or services for preferred placement on its platform; misusing a business's data to compete against it; and biasing search results in favor of the dominant firm.

The legislation will give enforcers strong, flexible tools to deter violations and hold dominant platforms accountable when they cross the line into illegal behavior, including significant civil penalties, authority to seek broad injunctions, emergency interim relief, and potential forfeiture of executive compensation.

Finally, the legislation is tailored to target bad conduct by the most powerful digital gatekeepers. The bill would prevent self-preferencing and discriminatory conduct by the most economically significant online platforms with large US user bases that function as "critical trading partners" for online businesses. For such platforms, the rules target only harmful conduct, allowing the platforms to innovate, do business, and engage in pro-consumer conduct, including protecting user privacy and safety, preventing unlawful behavior, and maintaining a secure online experience for users.

In other words—Apple won't be able to stifle competition by blocking other companies' services from interoperating with their platforms. Amazon won't be able to misuse small businesses' data in order to copy their products and compete against them. And Google won't be able to bias their platform's search results in favor of their own products and services without merit.

As a result, we will support a market that is fairer for small- and medium-sized businesses, giving them more options and more flexibility, and offering more choice for consumers. It's about helping foster entrepreneurship and allowing more businesses to compete on a level playing field with gatekeeping platforms.

Conclusion

We all want American companies to succeed and prosper, but even the most successful, innovative, and popular companies have to play by the rules. The goods and services provided may change pretty drastically over time, but the threats to competition from monopolies remain the same. And if we continue, as a country, to ignore those threats, it will be consumers, workers, businesses, and our economic growth that will suffer. That is why we must reinvigorate American antitrust enforcement across our economy by funding our enforcement agencies and updating our antitrust laws for the twenty-first-century economy. Maintaining the status quo is no longer an option.

The need to act is particularly urgent when it comes to the digital economy. When our health, children, and democratic institutions are threatened by unregulated and unaccountable algorithms designed to prioritize profit over safety, it is time to take action. When consumers unknowingly sacrifice their most personal

information just so they can share photos with their family and friends, it is time to take action. And when companies have become so powerful that they can buy out their rivals and promote their own products in a way that eliminates all competition in the market, it is time to take action. The power of Big Tech needs to be reined in. That means setting new expectations and putting new rules of the road in place.

I am not naïve about the challenges ahead. The halls of Congress are filled with lobbyists attempting to keep these important reforms at bay. They don't want a free and fair marketplace, they want to block out competition and boost their own profits.

But this is one of those extraordinary moments when the truth breaks through.

Nearly every day, I am finding new colleagues ready to get to work. And with the American people—each one a catalyst for change—behind us, we will ensure that our laws respond to the challenges we face.

REPORT OF THE COMMISSION

KATHERINE ADAMS, JELANI COBB, MARTIN BARON,
RUSS FEINGOLD, LEE C. BOLLINGER, CHRISTINA PAXSON,
HILLARY CLINTON, AND GEOFFREY R . STONE

Introduction

Social media have enabled the dissemination of ideas and information at a rate and scale unprecedented in human history. They have allowed ordinary citizens to spread awareness, offer support, organize, and hold their leaders accountable like never before. But there is a clear and growing consensus that social media also present real dangers both to individual users and to society as a whole.[1] In particular, proprietary algorithms designed to promote popular content and display information based on user preferences have facilitated the viral spread of false or misleading information—with serious, real-world consequences.

In its most insidious form, false or misleading information can be weaponized to manipulate the public, particularly in the political sphere. Agents of the Russian government, for example, launched a coordinated campaign on social media aimed to influence the 2016 US presidential election.[2] Many other state actors have reportedly used social media to fuel conflicts abroad and within their own borders.[3] Individuals or small groups of users have also used social media to amplify wild conspiracy theories, such as Pizzagate and QAnon, to sow distrust in the outcome of the 2020 US presidential election and to undermine public health efforts to combat COVID. The ease with which altered images or "deepfake" videos can be shared makes it increasingly difficult for users to separate fact from fiction.

These are by no means the only examples of harmful speech on social media: Hate speech, speech promoting violence or self-harm, defamation, harassment, and invasion of privacy are all too common. The prevalence of these

kinds of speech on social media has unquestionably harmed users, particularly children and young adults.[4]

Social media companies have failed to rein in harmful speech on their platforms. The sheer volume of false or misleading information on social media—not to mention the difficulty in distinguishing truth from falsehood—is enough to overwhelm even the most technologically sophisticated and well-resourced social media companies. And other forms of harmful speech, such as conspiracy theories, hate speech, and incitement to violence, can be difficult to define in a principled manner. Social media companies' profit motives may also discourage them from taking action to stop the spread of harmful speech, either because of a reluctance to modify the algorithms that draw users and advertisers to their platforms or a fear of being accused of censoring users.

The spread of false or misleading information poses a grave threat to our democracy and public health, which cannot survive without shared confidence in institutions and a common factual framework. Action is urgently needed before it is too late. Stemming the flow of misinformation and disinformation on social media will be no easy feat. It is fraught with complex legal, regulatory, and technological challenges.

This dire problem and the manifold challenges of solving it prompted the creation of the Commission for this book. The Commission's recommendations, outlined below, result from a dialogue among current and former public officials, scholars, journalists, researchers, and members of the media and tech industries, and are influenced by the essays in this volume. The Commission's recommendations also are informed by its recognition of the following overarching considerations:

First, the First Amendment significantly limits the direct regulation of most social media content. Conspiracy theories, hate speech, and false or misleading speech are constitutionally protected, except in limited circumstances such as actionable defamation, misleading advertisements, or incitement to violence. Any attempt by the government to broadly regulate such speech on social media would certainly be challenged as a violation of the platforms' and/or users' free-speech rights. The Commission has considered the existing First Amendment jurisprudence and concluded that there is ample room in First Amendment doctrine for regulation of certain content moderation practices.

In this respect, the evolution of First Amendment jurisprudence to accommodate new technologies such as broadcast media provides an instructive model. In *Red Lion Broadcasting Co. v. FCC*[5] and again in *FCC v. Pacifica Foundation*,[6] the Supreme Court upheld more stringent regulation of broadcasters on the grounds that the broadcast medium's particular attributes—the scarcity of available frequencies and its "uniquely pervasive" presence in American homes—justified a departure from the broader First Amendment protections that had

been applied to print media. The unique attributes of social media demand a similar approach today. In particular, the Commission's view is that platforms' use of algorithms or artificial intelligence to replicate and amplify speech should be considered distinct from the content of speech and therefore subject to lesser First Amendment protection. Placing limitations on such algorithms or artificial intelligence is best considered content-neutral regulation, and the public interest in such limitations is overwhelming.

Second, social media technology and the methods that bad actors use to spread harmful speech are constantly evolving. Any regulatory framework risks quickly becoming obsolete. The Commission considered that challenge—and the related need for up-to-date expertise about social media platforms and technology—in shaping these recommendations. Its view is that the consequences of doing nothing are too dire to let fear of obsolescence delay action, particularly when the most significant legislative framework applicable to social media—Section 230 of the Communications Decency Act of 1996—is decades old. Moreover, as discussed below, social media companies themselves have the requisite expertise and a significant role to play in shaping a constructive solution. That is why the Commission's recommendations attempt to draw on that expertise and to set forth a framework for regulatory action that would allow evolution and modification over time as technology evolves and as the impact of regulation is studied.

Third, although social media companies can and should be part of the solution, self-regulation alone is insufficient. Social media companies are profit-making enterprises with an interest in increasing user engagement to maximize advertising dollars. This means that they are disincentivized from adopting reforms that might slow the spread of viral content and promote transparency and user privacy. Some sort of government regulation is necessary to overcome these incentives. In addition to governmental action, civil society groups can play a role by mobilizing advertiser pressure to curb social media companies' conduct—bringing to bear the power of the purse in addition to governmental powers.

Fourth, promoting user privacy and transparency on social media are important goals for reform (and such reforms would be largely unconstrained by the First Amendment), but will not by themselves address the problem of harmful speech on social media. With that said, greater transparency regarding social media companies' methodology and approach in deciding what content is promoted and how will be crucial to effective regulation. In addition, although data privacy is outside the scope of the Commission's recommendations, that area too is ripe for reform. Social media companies have strong incentives to maximize the amount of information that they gather and that users share because that can increase advertising revenue. But most users do not fully understand the amount of data they are sharing and how their data is being used.

Indeed, social media platforms not only collect their users' data, but when users share information with friends and family, that data can also be used by social media companies for secondary purposes. Policymakers should consider reforms to give users more information about and control over how social media companies use the data they provide.

Fifth, any solution must be international in scope. Because social media technology is unconstrained by national borders, and misinformation and disinformation wreak havoc both inside and outside the United States, US policymakers and regulators should work with allies and international bodies to ensure that any reforms have as much global reach as possible.

Finally, although the term "social media" can be used colloquially to refer to a variety of online media and there are many websites that allow users to post content, the Commission's recommendations focus on first-party services that have a direct relationship with users. These include, for example, Facebook, Twitter, TikTok, YouTube, Instagram, Parler, Reddit, Clubhouse, and Gab.

Condition Section 230's Applicability on a Notice-and-Takedown Process and Industry Organization Membership

Most efforts at legislative reform have focused on Section 230 of the Communications Decency Act, the landmark legislation that gave social media companies immunity from many civil suits arising from (*i*) hosting user-generated content on their platforms and (*ii*) modifying or taking down user-generated content.[7] Doing away entirely with these protections would not meaningfully reduce harmful content on social media. Instead, it would incentivize platforms to over-moderate content to avoid lawsuits and impose potentially massive compliance costs that would deter new entrants to the social media market, thereby increasing the power of the biggest incumbents.

Rather than doing away entirely with Section 230, the Commission proposes making Section 230 immunity contingent on social media companies' (*i*) adherence to a notice-and- takedown regime and (*ii*) participation in a self-regulatory organization (SRO).

Notice and takedown. Section 230's grant of immunity should be made contingent on social media companies' implementation of a statutorily prescribed notice-and-takedown process for unlawful content. Such content would include speech that violates state statutes (such as those criminalizing revenge porn) or tort law (such as content that meets the legal standard for defamation). This approach would mimic the notice-and-takedown regime of US copyright law[8] in

that social media platforms would be immune from suit for unlawful content as long as they remove any content that is subject to a valid takedown notice containing

- Sufficient information to identify the content;
- An affirmation of the submitter's good-faith belief that the content meets the criteria for takedown;
- A notification to the user who uploaded that content of a pending challenge; and
- An opportunity for the user to submit a counter-notice explaining why the content is not unlawful or tortious. If the counter-notice justifies the content, the platform would be required to decide in good faith whether to take down or retain the content.

Platforms would specify processes and decisionmakers for adjudicating disputes over the content's lawfulness. Platforms above a certain number of daily active users should be required to apply a clear and convincing standard of proof before removing content alleged to be unlawful. Those larger platforms are more likely to have the resources necessary to assess and adjudicate whether content is lawful.

By placing the burden of identifying many forms of unlawful content on users, this approach would alleviate some concern about the compliance costs for companies to monitor the enormous volume of content posted on their platforms. This approach has also enjoyed fairly widespread support among legislators: A notice-and-takedown regime has been incorporated in proposed legislation in the United States[9] and in the European Union.[10] The establishment of these processes would also enable platforms to apply this approach, at their discretion, to other types of content that are protected under the First Amendment, such as hate speech and conspiracy theories.

SRO membership. Section 230 immunity should be conditioned on social media companies' membership in an SRO. As described further below, the SRO would impose disclosure obligations on its members in key areas such as algorithms and the use of user data.

Platforms would be free, of course, to decline to maintain a notice-and-takedown regime or to join or comply with the SRO. In that event, however, they would face potential liability for user-generated content on their platforms or for their own content-moderation decisions, subject to First Amendment protections for those decisions. The compliance costs required to monitor and remove potentially unlawful content in order to avoid litigation risk in the absence of Section 230 immunity may be enough to motivate most platforms to agree to these conditions.

Create a Self-Regulatory Organization

While some social media platforms have independently taken steps to address harmful content on their platforms,[11] the Commission believes more coordination is necessary to meaningfully limit harmful speech on social media. An industry-wide SRO can be highly effective and feasible to implement in the near term. Social media platforms themselves have advocated for an industry-led approach.[12] Moreover, the creation of an SRO would not require congressional action.

The SRO would be governed by management selected by a board containing members from industry, government, and academia, and it would be subject to oversight by a federal agency, as described below. The Commission recognizes that because of the federal government's oversight, the SRO's regulatory authority would be limited by the First Amendment. Therefore, the SRO's requirements likely would need to be content-neutral in order to survive judicial scrutiny.

The SRO also should solicit input on regulatory measures from experts from the industry, government, and academia. In relying on their knowledge, the SRO should be able to develop structures capable of adapting to changing technologies while safeguarding users and maintaining high ethical standards. These advisory functions could even be formalized to resemble Federal Drug Administration advisory committees, which bring to bear significant expertise in decision-making.

At a minimum, the SRO should (*i*) establish minimum standards for terms of use; (*ii*) require increased transparency regarding the use of data and algorithms; (*iii*) mandate that platforms conduct annual risk assessments regarding content and content moderation; and (*iv*) impose penalties for noncompliance. Each of these requirements is discussed in further detail below.

Terms of use. The SRO should require social media platforms to adopt terms of use that make clear how the platform intends to address the following topics:

- Content moderation designed to prevent the amplification of knowingly false information that is intended to manipulate the public, including misinformation about voting methods and deepfake videos;
- The influence of foreign money and foreign actors on the spread and consumption of political content, including detection and monitoring systems and communication with US intelligence agencies;
- Removal of content that violates the platform's terms of use. Platforms' terms of use could include prohibitions on misinformation, disinformation, hate speech, or other forms of protected speech;
- The types of user data that platforms can collect and use and how platforms can use that data; and

- User anonymity, specifically whether users may remain anonymous to other users or to the platform itself. Although anonymity is important in many contexts, such as to encourage whistleblowing and protect those who come forward, in order to limit the spread of misinformation and disinformation by bots or foreign actors, it might be necessary to make identities of users known at least to the platform, if not to the public more generally.

The SRO should also establish minimum standards for terms of use. This will help ensure that the requirement is meaningful and members do not resort to the lowest common denominator. The SRO should also encourage platforms to share the strategies that have proven most effective, with the goal of being able to evolve to combat increasingly sophisticated bad actors.

Transparency. The SRO should require greater transparency in the following key areas:

- The operation and goals of the platform's algorithms, including the ultimate goal to keep users' attention and market that attention to potential advertisers;
- The platform's largest advertisers;
- The collection and use of user data, including whether and how it is made available to third parties; and
- Users' control over their data, including the ability to opt out of the collection of certain information, request removal of certain information, and decline to authorize the sale or transfer of their data to third parties.

To be sure, not all users will notice or appreciate such disclosures. But, just as requirements to post calorie counts on restaurant menus have helped at least some consumers make more informed choices about their diets, increased disclosures on social media can help some users make more informed decisions about how they interact with social media.

Risk assessments. The SRO should require annual risk assessments by participating platforms that (*i*) identify risks from their user-generated content and their own content-moderation policies and (*ii*) outline their plans for mitigating such risks. Risks from user-generated content might include, for example, the incidence of unlawful content, deepfakes, hate speech, misinformation regarding vaccines or voting methods, and content promoting violence or self-harm. Risks from content-moderation decisions may include, for example, amplification of false information or conspiracy theories without sufficient balancing information.

Penalties. The SRO must have the ability to penalize its members for noncompliance with terms of use, transparency requirements, and mitigation plans. There are several examples of successful SROs in other industries that could

provide useful models. For instance, the Financial Industry Regulatory Authority, under the supervision of the US Securities and Exchange Commission, oversees the broker-dealer industry, creating rules, conducting examinations and market surveillance, and bringing disciplinary proceedings.[13] The National Advertising Division of the Better Business Bureau resolves disputes over advertising.[14] The Digital Advertising Alliance establishes privacy principles for the digital advertising industry, thereby enhancing transparency for consumers and allowing them to control their privacy settings.[15]

The penalties must be significant enough to deter bad behavior, up to and including expulsion from the SRO (which would have the consequence of eliminating Section 230 immunity). At the same time, penalties will need to be sensitive to the size of the platform to avoid forcing out smaller players that otherwise would have chosen to remain in the SRO and increasing the power of entrenched incumbents.

Create a Federal AgencyThe SRO should be overseen by a federal agency with expertise in social media technology. This could be accomplished by creating a new federal agency. Such a new agency would have leadership selected to be as nonpartisan and impartial as possible. One way to achieve this would be to have a two-tiered nomination process. As with the Federal Reserve, this agency could have the members of its board of governors nominated by the president and confirmed by the Senate and have them serve for a set number of terms.[16] The terms of such members would span administrations. Alternatively, if a new agency is not viable, Congress could delegate authority to a division of an existing federal agency, such as the Federal Trade Commission.

Whatever appointment structure or agency is selected, the agency should have a diversity of views.

Such an agency should not be tasked with determining what types of content meet the criteria for false or misleading information that can be subject to regulation. Identifying misinformation and disinformation is a task best left to social media platforms to perform in good faith. Even an agency insulated from direct political control should not be assigned such a task, as it risks becoming co-opted for political ends.

Instead, the agency's primary role would be to oversee and monitor the SRO, and pursue enforcement actions against platforms in the event of persistent abusive practices, such as a failure to follow disclosed practices or follow risk mitigation measures identified in their risk assessments. For example, if a platform's terms of service state that hate speech will be removed but the platform persistently fails to remove such content, the agency could pursue an enforcement action against that platform. The agency could also evaluate whether and to what extent it could, under the First Amendment, regulate platforms' use of algorithms to promote the viral spread of harmful content.

Target Foreign Influences

The Commission is increasingly concerned with the ability of foreign actors to manipulate population behavior, for example, by undermining democracy or public health. Mitigating that threat will require action from social media platforms, US law enforcement agencies, Congress, and our foreign allies.

Our intelligence agencies and social media platforms must communicate to prevent such foreign interference. Platforms should share with law enforcement any threats they identify from foreign actors, and law enforcement and intelligence agencies should provide information to social media platforms on foreign actors' attempts to spread disinformation relating to elections or public health.

Focus on Advertisers

While the Commission believes the creation of an SRO and a new federal agency would be positive steps in limiting the spread of misinformation and disinformation on social media, there is also a role for civil society to play. Public pressure on social media companies' primary source of revenue—advertisers—could incentivize platforms to jettison their most harmful practices, even in the absence of regulation. Advertiser influence can be highly effective. For example, Major League Baseball moved its 2021 All-Star game from Atlanta after Coca Cola and Delta Airlines voiced their strong disapproval of Georgia's restrictive voting laws.[17] Accordingly, the Commission encourages civil society groups to explore ways to leverage advertisers to limit the viral dissemination of harmful content. Presumably some of the disclosures contemplated under the above recommendations will provide additional information that will help put pressure on social media platforms through their major revenue source.

Appendix: Summary List of Recommendations

Encouraging Self-Regulation and Reducing the Harmful Spread of Misinformation

Amend Section 230

I. **Make Section 230 immunity contingent on the notice and takedown of unlawful content.** Section 230(c)(2)'s protections should be modified by Congress so that they are contingent on a social media platforms' implementation of a notice-and-takedown process for unlawful content. This approach would be based on the notice-and-takedown regime of US copyright law.[18]

II. **Provide Section 230's immunity only with participation in an SRO.** In establishing this condition on Section 230, social media platforms would obtain the statute's protection only if they join the SRO and meet their obligations to adopt terms of use and disclose information regarding data practices and algorithm usage.

Create a Self-Regulatory Organization

I. **Create governance structure.** The SRO should be governed by management selected by a board, containing members from industry, government, and academia. The SRO should then be subject to oversight by a federal agency, as described below.
II. **Require participating platforms to adopt terms of use.** The terms of use would, at a minimum, require platforms to state how they intend to handle (*i*) content moderation; (*ii*) the influence of foreign money and actors on US politics, including detection and monitoring systems and communication with US intelligence agencies; (*iii*) notice-and-takedown provisions, providing for removal of content that violates the terms of use; (*iv*) the types of user data that platforms can collect and how platforms can use that data; and (*v*) whether users may remain anonymous.
III. **Require disclosure standards for participating platforms.** In addition to terms of use, platforms should be required to provide increased transparency regarding, at minimum, (*i*) the goals of the platform's algorithms; (*ii*) major advertisers; (*iii*) collection and use of users' data; and (*iv*) users' control of their data. There is currently an information asymmetry, but with more disclosure, the Commission believes that individuals can make more informed decisions about how they interact with social media platforms.
IV. **Require annual risk assessments by participating platforms.** These assessments should require platforms to identify potential risks from content and content-moderation decisions and how they will mitigate such risks.
V. **Rely on experts to develop standards.** Technical expertise, including outside technical expertise, will be critical to the success of any SRO. The SRO should receive and consider the recommendations of advisory groups, drawing on academia and industry leaders to stay on top of current issues and developments.
VI. **Create an enforcement mechanism with authority to issue penalties.** Much like the Financial Industry Regulatory Authority does for broker-dealers, the SRO should be able to issue fines or other penalties in the event of violations, such as violations of disclosed policies, failure to conduct risk

assessments, or failure to implement risk-mitigation measures. The penalties would have to be meaningful enough to have a deterrent effect, and might include the loss of Section 230 immunity for significant violations.

Create a Federal Agency

I. **Agency authority.** Either create a new federal agency or assign responsibility to an existing federal agency that would oversee issues relating to regulation of social media.

II. **Foster a nonpartisan leadership structure.** The agency's leaders should be appointed in a manner that preserves their independence and nonpartisanship as much as possible.

III. **Define its main role as oversight of the SRO**. The agency would not be regulating the platforms directly but would rather be overseeing the SRO, reviewing its requirements, and monitoring its overall activities.

IV. **Empower the agency to pursue enforcement actions.** Empower the agency to pursue enforcement actions against platforms for repeatedly violating SRO rules, such as by failing to follow disclosed policies and practices or implement risk-mitigation measures.

Target Foreign Influences

I. **Increase communication with US intelligence agencies.** Given the significant issues stemming from foreign interference, especially in the US democratic process and in the implementation of public health measures, the Commission believes there should be more communication between US intelligence agencies and social media platforms.

A. In particular, platforms should share with intelligence agencies, and vice versa, any threats from foreign governments or nonstate actors that might interfere in elections or exacerbate public health issues.

Focus on Advertisers

I. Increase pressure from advertisers on social media platforms. Civil society groups should explore ways to encourage advertisers to pressure social media platforms to take steps to limit the spread of misinformation and disinformation and other forms of harmful speech.

CONCLUDING STATEMENT

LEE C. BOLLINGER AND GEOFFREY R. STONE

Over the last two decades, we have published four volumes of essays by distinguished constitutional scholars, journalists, and public officials dealing with core First Amendment issues: *Eternally Vigilant: Free Speech in the Modern Era* (2001); *The Free Speech Century* (2018); *National Security, Leaks and Freedom of the Press* (2021); and now *Social Media, Freedom of Speech, and the Future of Our Democracy* (2022).[1] Although we have both devoted most of our scholarly careers to issues posed by the First Amendment, the challenge of putting these volumes together has taught us a great deal about how to deepen our understanding of the interests at stake and of the varied possibilities for resolving them. We have found that a combination of approaches works best, and is in many ways preferable to what a single voice might offer.

In putting together each of these volumes, we began with a general sense that there are serious problems confronting our society that will ultimately challenge our understanding of the First Amendment. Going into these efforts, we are ourselves uncertain about what should be done, and we are confident that we should know more than we do. Such an approach to knowledge and understanding, including the recognition that one never knows it all, is of course at the very heart of open inquiry and freedom of speech. Thus, in each of these volumes we bring together a broad range of experts with competing views and ask them to provide their best judgment about how these issues should be addressed.

In our two most recent volumes, including this one, we convened a "Commission" of experts from different disciplines and backgrounds and asked them, after reviewing all of the essays, to meet with us for a day to see whether it was possible to reach a consensus about the nature and scale of the problems and to identify recommendations for resolving the competing interests at stake. We have been fascinated to find that the diversity of backgrounds and

perspectives in the Commission leads to a discussion that, over many hours, forges a shared sense about what should be done, and then results in a concrete set of recommendations.

How to understand the benefits and risks for democracy posed by the newest technology of communication—that is, the internet and social media platforms, in particular—and what might or should be done to enhance the former and limit the latter is one of the greatest challenges of our era. Although we are only at the beginning of this new transformation of human communication and the public forum, some things are already clear. The views expressed in this volume, both in the essays and in the Commission's report, reveal a broad agreement around three principal propositions.

The first proposition is that, while the benefits of expanding the opportunities for individuals to communicate and of increasing the range of available voices in the marketplace of ideas are tremendous and must be preserved, the structure of the system that has evolved poses extraordinary risks to sound public discussion and decision-making. Near monopoly control by for-profit corporations with insufficient development of responsible intellectual and moral standards (compare the current world of social media to the evolution of standards in the fields of journalism or the academy)—and perverse incentives to avoid such standards—risk yielding a world in which the worst rather than the best of the human mind will triumph.

The second proposition is that this state of affairs will not solve itself and will require interventions from forces outside the technology sector. Whether this takes the form of public pressure arising from organized private citizens or of changes in policies or laws including some forms of government regulation is keenly debated in these essays. But it is undisputed that a business-as-usual approach in this setting will not work and may even undermine democracy itself. In the twentieth century, the new technology of communication of broadcasting was thought to pose similar challenges for society.[2] The way our nation responded to those challenges, with a system of government oversight that forbade government censorship but enabled government regulation to protect the public's interest in knowing about the issues and the conflicts of its time, is instructive for us today.[3] While the particular rationale for that system was different and perhaps peculiar to the circumstances of that technology, the spirit of the congressional response and the Supreme Court's unanimous approval of it provide a potentially useful analogy for the challenges of the present.[4]

The third, and final, major proposition is that this is a time for profound self-reflection about how to address our current challenges in order to achieve the fundamental values and goals of our constitutional system of government. Constitutional law is forged in the context of an era, and we are and should be proud of what the United States has achieved in its protections of the freedoms

of speech and the press. What we have learned in the past, particularly from the free speech conflicts during such periods as World War I, the McCarthy era, and the Civil Rights era, guides us in addressing our current issues. We are now in the midst of another major era that will have enormous consequences for how we define our most fundamental values and how, and whether, we protect and preserve our system of self-governing democracy.

NOTES

Opening Statement

1. Chaplinsky v. State of New Hampshire, 35 U.S. 568 (1941); Virginia v. Black, 538 U.S. 343 (2003); Brandenburg v. Ohio, 395 U.S. 444 (1969); THE FREE SPEECH CENTURY (Lee C. Bollinger & Geoffrey R. Stone eds., New York: Oxford University Press, 2018).
2. *See, e.g.*, Lee Rainie et al., *The Future of Free Speech, Trolls, Anonymity and Fake News Online*, PEW RES. CTR. (Mar. 29, 2017), https://www.pewresearch.org/internet/2017/03/29/the-future-of-free-speech-trolls-anonymity-and-fake-news-online/.
3. *Id.*
4. *Id.*
5. Communications Act of 1934, 48 Stat. 1064 (1934).
6. *Id.*
7. LEE C. BOLLINGER, IMAGES OF A FREE PRESS (Chicago: University of Chicago Press, 1991); LEE C. BOLLINGER, UNINHIBITED, ROBUST, AND WIDE-OPEN: A FREE PRESS FOR A NEW CENTURY (New York: Oxford University Press, 2010).
8. Communications Decency Act of 1996, 47 U.S.C. § 230 (1996).
9. *See* Daisuke Wakabayashi, *Legal Shield for Social Media Is Targeted by Lawmakers*, N.Y. TIMES (May 28, 2020), https://www.nytimes.com/2020/05/28/business/section-230-internet-speech.html.
10. ROBERT S. MUELLER, THE MUELLER REPORT: REPORT ON THE INVESTIGATION INTO RUSSIAN INTERFERENCE IN THE 2016 PRESIDENTIAL ELECTION (Washington, DC: WSBLD, 2019); U.S. SENATE SELECT COMMITTEE ON INTELLIGENCE, RUSSIAN ACTIVE MEASURES CAMPAIGNS AND INTERFERENCE IN THE 2016 U.S. ELECTION, VOLUME 2: RUSSIA'S USE OF SOCIAL MEDIA WITH ADDITIONAL VIEWS (Washington, DC: Government Publishing Office, 2019).
11. *See* Elisa Shearer, *More Than Eight-in-Ten Americans Get News from Digital Devices*, PEW RES. CTR. (Jan. 12, 2021), https://www.pewresearch.org/fact-tank/2021/01/12/more-than-eight-in-ten-americans-get-news-from-digital-devices; Elisa Shearer and Amy Mitchell, *News Use Across Social Media Platforms in 2020*, PEW RES. CTR. (Jan. 12, 2021), https://www.pewresearch.org/journalism/2021/01/12/news-use-across-social-media-platforms-in-2020/.
12. THE FREE SPEECH CENTURY (Lee C. Bollinger & Geoffrey R. Stone eds., New York: Oxford University Press, 2018).
13. U.S. CONST. amend. I.
14. Schenck v. United States, 249 U.S. 47 (1919); Abrams v. United States, 250 U.S. 616 (1919); Debs v. United States, 249 U.S. 211 (1919); Daniel Baracskay, *Bill of Rights*, FIRST AMENDMENT ENCYCLOPEDIA (Feb. 2018), https://www.mtsu.edu/first-amendment/article/1448/bill-of-rights.
15. *See* Geoffrey R. Stone, *Free Speech in the Modern Age: What We Learned in the Twentieth Century*, 36 PEPPERDINE L. REV. 273 (2009).

16. New York Times Co. v. Sullivan, 376 U.S. 254 (1964).
17. *Id.* at 270.
18. Beauharnais v. Illinois, 343 U.S. 250 (1952); Brandenburg v. Ohio, 395 U.S. 444, 447 (1969).
19. *See* Maja Adena et al., *Radio and the Rise of the Nazis in Prewar Germany*, Q. J. OF ECON., 1885–1939 (2015).
20. Communications Act of 1934, 48 Stat. 1064 (1934).
21. Radio Act of 1927, 44 Stat. 1162 (1927); Radio Act of 1934, 48 Stat. 1088 (1934).
22. Red Lion Broadcasting Co. v. F.C.C., 395 U.S. 367 (1969).
23. *Id.*
24. Miami Herald Pub. Co. v. Tornillo, 418 U.S. 241 (1974).
25. THE NEW YORK TIMES GUIDE TO ESSENTIAL KNOWLEDGE: A DESK REFERENCE FOR THE CURIOUS MIND 886 (John W. Wright et al. eds., New York: St. Martin's, 3d ed., 2011); AMERICAN TELEVISION: NEW DIRECTIONS IN HISTORY AND THEORY 13–14 (Nick Browne ed., Chur, Switzerland: Harwood Academic, 1994).
26. Communications Act of 1934, 48 Stat. 1064 (1934).

Regulating Harmful Speech on Social Media: The Current Legal Landscape and Policy Proposals

1. *See, e.g.*, Jane Hu, *The Second Act of Social Media Activism*, NEW YORKER (Aug. 3, 2020), https://www.newyorker.com/culture/cultural-comment/the-second-act-of-social-media-activism; Bijan Stephen, *Social Media Helps Black Lives Matter Fight the Power*, WIRED (Nov. 2015), https://www.wired.com/2015/10/how-black-lives-matter-uses-social-media-to-fight-the-power/; ZEYNEP TUFEKCI, TWITTER AND TEAR GAS: THE POWER AND FRAGILITY OF NETWORKED PROTEST (New Haven: Yale University Press, 2017); Natana J. DeLong-Bas, *The New Social Media and the Arab Spring*, OXFORD ISLAMIC STUDIES ONLINE, http://www.oxfordislamicstudies.com/Public/focus/essay0611_social_media.html.
2. *See, e.g.*, Brooke Auxier, *Social Media Continue to Be Important Political Outlets for Black Americans*, PEW RES. CTR. (Dec. 11, 2020), https://www.pewresearch.org/fact-tank/2020/12/11/social-media-continue-to-be-important-political-outlets-for-black-americans/.
3. *See, e.g.*, Keith Kloor, *Inside the COVID-Denialist Internet Bubble*, POLITICO (Mar. 22, 2020), https://www.politico.com/news/magazine/2020/03/22/inside-fringe-internet-coronavirus-bubble-142960; Rebecca Heilweil, *Coronavirus Scammers Are Flooding Social Media with Fake Cures and Tests*, VOX (Apr. 17, 2020), https://www.vox.com/recode/2020/4/17/21221692/digital-black-market-covid-19-coronavirus-instagram-twitter-ebay.
4. *See generally* Chris Jackson, *More Than 1 in 3 Americans Believe a "Deep State" Is Working to Undermine Trump*, IPSOS (Dec. 30, 2020), https://www.ipsos.com/en-us/news-polls/npr-misinformation-123020; Kerry Howley, *QAnon and the Bright Rise of Belief*, N.Y. MAGAZINE (Jan. 29, 2021), https://nymag.com/intelligencer/article/qanon-capitol-riot-bright-rise-belief.html (remarking on the belief among followers that QAnon brings good news and a plan to "clear the world of wicked-doing").
5. *See, e.g.*, Erin Ailworth et al., *Lost in Life, El Paso Suspect Found a Dark World Online*, WALL ST. J. (Aug. 8, 2019), https://www.wsj.com/articles/lost-in-life-el-paso-suspect-found-a-dark-world-online-11565308783; Rachel Hatzipanagos, *How Online Hate Turns Into Real-Life Violence*, WASH. POST (Nov. 30, 2018), https://www.washingtonpost.com/nation/2018/11/30/how-online-hate-speech-is-fueling-real-life-violence/; Lizzie Dearden, *Gab: Inside the Social Network Where Alleged Pittsburgh Synagogue Shooter Posted Final Message*, INDEPENDENT (Oct. 28, 2018), https://www.independent.co.uk/news/world/americas/pittsburgh-synagogue-shooter-gab-robert-bowers-final-posts-online-comments-a8605721.html.
6. 47 U.S.C. § 230(c), (e).
7. 47 U.S.C. § 230(c), (e). Courts have also found that Section 230 does not preclude claims based on contract or antitrust law.
8. 1995 WL 323710 (N.Y. Sup. Ct., May 24, 1995).
9. *Id.* at *5.
10. *Id.*; *see also* Smith v. California, 361 U.S. 147 (1959).
11. 47 U.S.C. § 230(f)(2).

12. *See* 47 U.S.C. § 230(f)(3).
13. *See* Force v. Facebook, Inc., 934 F.3d 53, 64 (2d Cir. 2019); Klayman v. Zuckerberg, 753 F.3d 1354, 1357 (D.C. Cir. 2014); Pennie v. Twitter, 281 F. Supp. 3d 874, 888 (N.D. Cal. 2017); *see also* Herrick v. Grindr LLC, 765 F. App'x 586, 589–90 (2d Cir.), *cert. denied*, 140 S. Ct. 221 (2019) ("As the district court observed, courts have repeatedly concluded that the definition of an ICS includes 'social networking sites like Facebook.com, and online matching services like Roommates.com and Matchmaker.com,' which, like Grindr, provide subscribers with access to a common server.").
14. For example, with respect to a platform's responsibility for third-party content, the Ninth Circuit denied immunity for a feature that allowed or encouraged users to limit their roommate searches by sex, family status, or sexual orientation on Roommates.com. But the court upheld immunity for the "Additional Comments" section of user profiles, finding that, there, the website was just a passive conduit of third-party information. Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, 521 F. 3d 1157, 1168 (9th Cir. 2008). Similarly, a Georgia state appeals court held that a Snapchat filter modified the user's content enough to make the company an information content provider under Section 230. Maynard v. Snapchat, Inc., 816 S.E.2d 77, 79–81 (Ga. Ct. App. 2018). In a case brought against Facebook alleging that it allowed predators to exploit, extort, and recruit children into the sex trade, the Texas Supreme Court, citing *Roommates.com*, dismissed common-law claims for negligence on Section 230 grounds; however, it allowed the state law human trafficking claim, which alleged "overt acts by Facebook encouraging the use of its platforms for sex trafficking," to proceed to trial. In re Facebook, Inc., 625 S.W.3d 80, 2021, WL 2603687 at *1 (Tex. June 25, 2021).
15. The seminal decision in this area is Zeran v. America Online, Inc., 129 F.3d 327 (4th Cir. 1997). Since then, other appellate courts have agreed with *Zeran*'s broad interpretation. *See, e.g.*, Fed. Trade Comm'n. v. LeadClick Media, LLC, 838 F.3d 158, 173 (2d Cir. 2016); Jones v. Dirty World Entm't Recordings LLC, 755 F.3d 398, 408 (6th Cir. 2014) (noting that "close cases . . . must be resolved in favor of immunity").
16. *See* Jane Doe No. 1 v. Backpage.com, LLC, 817 F.3d 12, 16–21 (1st Cir. 2016) (case involving allegations of trafficking); Lemmon v. Snap, Inc., 440 F. Supp. 3d 1103 (C.D. Cal. 2020) (wrongful death action bought against Snap, Inc., for "speed" filter that allegedly encouraged a user to engage in reckless driving ending in a deadly crash); Force v. Facebook, Inc., 934 F.3d 53 (2d Cir. 2019) (alleging Facebook provided material support for terrorist organization).
17. *See, e.g.*, Barnes v. Yahoo!, Inc., 570 F.3d 1096, 1109 (9th Cir. 2009); Doe v. MySpace, Inc., 528 F.3d 413, 422 (5th Cir. 2008).
18. *See* Sikhs for Justice "SFJ," Inc. v. Facebook, Inc., 697 F. App'x 526 (9th Cir. 2017) (affirming dismissal of claims against Facebook under Title II of the Civil Rights Act of 1964).
19. Section 230 has been interpreted to preempt state criminal laws that are inconsistent with its text. *See, e.g.*, Backpage.com, LLC v. McKenna, 881 F. Supp. 2d 1262 (W.D. Wash. 2012).
20. *See, e.g.*, Zango, Inc. v. Kaspersky Lab, Inc., 568 F.3d 1169, 1174 (9th Cir. 2009); Mezey v. Twitter, Inc., No. 1:18-cv-21069-KMM, 2018 U.S. Dist. LEXIS 121775, at *3 (S.D. Fla. July 19, 2018); Darnaa, LLC v. Google, Inc., No. 15-cv-03221-RMW, 2016 U.S. Dist. LEXIS 152126, at *23 (N.D. Cal. Nov. 2, 2016); Sikhs for Justice "SFJ," Inc., v. Facebook, Inc., 144 F. Supp. 3d 1088, 1095 (N.D. Cal. 2015), *aff'd*, 697 Fed. App'x 526, 526 (9th Cir. 2017).
21. NetChoice LLC et al. v. Moody et al., 4:21-cv-220-RH-MA, 2021 WL 2690876, *6 (N.D. Fl. June 30, 2021).
22. *See supra* notes 15–18.
23. E-ventures Worldwide, LLC v. Google, Inc., 188 F. Supp. 3d 1265, 1273 (M.D. Fla. 2016); other cases cited in Valerie C. Brennon, *Free Speech and the Regulation of Social Media Content*, CONG. RES. SERV., 31 (Mar. 27, 2019), 14 n.134.
24. *See, e.g.*, Enigma Software Group USA v. Malwarebytes, Inc., 946 F.3d 1040 (9th Cir. 2019), *cert. denied*, 141 S. Ct. 13 (2020) (finding that immunity is unavailable when a plaintiff alleges anticompetitive conduct).
25. *See* Brandenburg v. Ohio, 395 U.S. 444, 447 (1969).
26. *See, e.g.*, Roth v. United States, 354 U.S. 476, 485 (1957) ("[O]bscenity is not within the area of constitutionally protected speech or press.").

27. *See, e.g.*, Cent. Hudson Gas & Elec. Corp. v. Pub. Serv. Comm'n, 447 U.S. 557, 562–63 (1980) ("The Constitution . . . accords a lesser protection to commercial speech than to other constitutionally guaranteed expression.").
28. *See, e.g.*, Citizens United v. FEC, 558 U.S. 310, 340 (2010).
29. *See* United States v. Alvarez, 567 U.S. 709, 710 (2012) ("[T]he Court has instructed that falsity alone may not suffice to bring the speech outside the First Amendment."); NAACP v. Button, 371 U.S. 415, 445 (1963) ("For the Constitution protects expression and association without regard to . . . the truth, popularity, or social utility of the ideas and beliefs which are offered.").
30. *See* Matal v. Tam, 137 S. Ct. 1744, 1764 (2017) ("Speech that demeans on the basis of race, ethnicity, gender, religion, age, disability, or any other similar ground is hateful; but the proudest boast of our free speech jurisprudence is that we protect the freedom to express 'the thought that we hate' " (quoting United States v. Schwimmer, 279 U.S. 644, 655 (1929)); United States v. Stevens, 559 U.S. 460, 470 (2010) ("The First Amendment's guarantee of free speech does not extend only to categories of speech that survive an ad hoc balancing of relative social costs and benefits.").
31. *See* Heather Whitney, *Search Engines, Social Media, and the Editorial Analogy*, KNIGHT FIRST AM. INST. AT COLUMBIA UNIV. (Feb. 27, 2018), https://knightcolumbia.org/content/search-engines-social-media-and-editorial-analogy (arguing that there are weaknesses to the editorial function rationale and that the simple act of "convey[ing] a wide range of information" does not necessarily constitute speech).
32. *Cf.* U.S. Telecom Ass'n v. FCC, 825 F.3d 674, 742–43 (D.C. Cir. 2016) (distinguishing between exercising editorial functions, which may be entitled to First Amendment protection, from indiscriminately transmitting "any and all users' speech," which may not be entitled to First Amendment protection).
33. *See* Arkansas Educ. Television Comm'n v. Forbes, 523 U.S. 666, 674 (1998) (holding that selection and presentation of programming is protected speech activity); Hurley v. Irish-American Gay, Lesbian, and Bisexual Group of Boston, 515 U.S. 557, 569–70 (1995) ("[A] private speaker does not forfeit constitutional protection simply by combining multifarious voices, or by failing to edit their themes to isolate an exact message as the exclusive subject matter of the speech."); City of Los Angeles v. Preferred Communications, Inc., 476 U.S. 488, 494 (1986) ("[C]able operators exercise editorial discretion regarding their programming."); New York Times Co. v. Sullivan, 376 U.S. 254 (1964) (invalidating, on First Amendment grounds, a libel award against the newspaper).
34. *See* Sorrell v. IMS Health Inc., 564 U.S. 552, 567 (2011) ("It is true that restrictions on protected expression are distinct from restrictions on economic activity or, more generally, on nonexpressive conduct."); Ashcroft v. Free Speech Coal., 535 U.S. 234, 253 (2002) ("[T]he Court's First Amendment cases draw vital distinctions between words and deeds, between ideas and conduct."). The district court in *Paxton* rejected the view that the involvement of algorithms mattered from a First Amendment perspective, holding that the "core question is still whether a private company exercises editorial discretion over the dissemination of context, not the exact process used." Paxton, 2021 WL 5755120, at *8.
35. 564 U.S. at 570 (rejecting the First Circuit's determination that data miners' sale, transfer, and use of pharmacy records that identified doctors' prescribing practices was conduct and not speech).
36. *See, e.g.*, Zhang v. Baidu, Inc., 10 F. Supp. 3d 433, 438 (S.D.N.Y. 2014) ("[A] 'search engine's editorial judgment is much like many other familiar editorial judgments[.]" (quoting Eugene Volokh & Donald M. Falk, *Google First Amendment Protection for Search Engine Results*, 8 J.L. ECON & POL'Y 883 (2012) at 884)); e-Ventures Worldwide v. Google, Inc., 2:14-cv-646-FtM-PAM-CM, 2017 WL 2210029, at *4 (M.D. Fla. Feb. 8, 2017) ("Google's actions in formulating rankings for its search engine and in determining whether certain websites are contrary to Google's guidelines and thereby subject to removal are the same decisions as by a newspaper editor regarding which content to publish[.]"); Langdon v. Google, Inc., 474 F. Supp. 2d 622, 630 (D. Del. 2007) (holding that requiring Microsoft, Yahoo!, and Google to re-rank the plaintiff's search results and feature its ads prominently would contravene the defendants' First Amendment rights); Search King, Inc. v. Google Tech., Inc., No.

CIV-02-1457-M, 2003 WL 21464568, at *4 (W.D. Okla. May 27, 2003) (holding Google's page rankings were protected under the First Amendment).
37. La'Tiejira v. Facebook, Inc., 272 F. Supp. 3d 981, 991–92 (S.D. Tex. 2017).
38. Fla. Stat. Ann. §§ 106.072, 287.137, 501.2041; Tex. Bus. & Com. Code Ann. § 120 et seq.
39. *See* NetChoice, 2021 WL 2690876, at *11.
40. NetChoice, LLC v. Paxton, No. 1:21-cv-840-RP, 2021 WL 5755120, at *1 (W.D. Tex.) (Dec. 1, 2021).
41. *Id.* at *8.
42. *See, e.g.*, Reed v. Town of Gilbert, 575 U.S. 155 (2015).
43. *See, e.g.*, Burson v. Freeman, 504 U.S. 191 (1992) (applying strict scrutiny to a Tennessee law prohibiting posting of campaign materials near polling places).
44. *See, e.g.*, R.A.V. v. City of St. Paul, 505 U.S. 377, 395 (1992) (in order to uphold a law under strict scrutiny analysis, the government must prove that the restriction is narrowly tailored and is necessary to serve compelling state interests).
45. *See, e.g.*, Members of City Council of City of Los Angeles v. Taxpayers for Vincent, 466 U.S. 789, 792 (1984) (holding that a city ordinance prohibiting posting of signs on public property was a permissible content-neutral restriction, even if it burdened political speech).
46. *See* Ward v. Rock Against Racism, 491 U.S. 781, 791 (1989).
47. NetChoice, LLC, 2021 WL 2690876, at *10.
48. Paxton, 2021 WL 5755120, at *11.
49. Reno v. ACLU, 521 U.S. 844, 868 (1997).
50. *See* FCC v. Pacifica Found., 438 U.S. 726, 748 (1978) ("[O]f all forms of communication, it is broadcasting that has received the most limited First Amendment protection.").
51. *See* 47 U.S.C. §§ 151 et seq.
52. 395 U.S. 367 (1969).
53. *See id.* at 387 ("Although broadcasting is clearly a medium affected by a First Amendment interest, differences in the characteristics of new media justify differences in the First Amendment standards applied to them." (citation omitted)).
54. *See* Red Lion, 395 U.S. at 392 (observing that there is a "First Amendment goal of producing an informed public capable of conducting its own affairs").
55. *See* Pacifica, 438 U.S. at 748 (upholding FCC's decision to discipline a radio station for airing George Carlin's "Filthy Words" monologue because the "uniquely pervasive" nature of broadcast media allows them to seep into "the privacy of the home" without the viewer's consent and in light of the government's interest in protecting children from "offensive expression").
56. *See, e.g.*, Miami Herald Pub. Co. v. Tornillo, 418 U.S. 241 (1974) (invalidating, on First Amendment grounds, a Florida statute requiring newspapers to publish without cost the reply of candidates whose integrity they criticize); New York Times Co. v. Sullivan, 376 U.S. 254 (1964) (invalidating, on First Amendment grounds, a libel award for a public official against critics of his official conduct); Near v. Minnesota, 283 U.S. 697 (1931) (invalidating, on First Amendment grounds, a state statute that allowed enjoining a newspaper from publishing malicious, scandalous, or defamatory material).
57. *Compare* United States v. Playboy Entertainment Group, Inc., 529 U.S. 803 (2000) (invalidating, on First Amendment grounds, a federal statute restricting sexually explicit cable TV shows), *with* FCC v. Pacifica, 438 U.S. at 745–51 (upholding sanction of a broadcast containing constitutionally protected but "patently offensive words" because broadcasting's pervasiveness and its availability to children "justify special treatment").
58. 418 U.S. 241 (1974).
59. 512 U.S. 622, 638 (1994).
60. Reno, 521 U.S. at 868–70.
61. Red Lion, 395 U.S. at 399.
62. *See* Lee C. Bollinger, *The Free Speech Century: A Retrospective Guide*, 2018 Clare Hall Tanner Lectures (Nov. 5, 2018), https://president.columbia.edu/content/tanner-lectures-cambridge.
63. *See id.*
64. *See* Brennon, *supra* note 23 (describing arguments that would need to be made to apply rationale for broadcast regulation to the internet).

65. *See* Ramya Krishnan, *The Pitfalls of Platform Analogies in Reconsidering the Shape of the First Amendment*, KNIGHT FIRST AM. INST. AT COLUM. UNIV. (May 19, 2021), https://knightcolumbia.org/blog/the-pitfalls-of-platform-analogies-in-reconsidering-the-shape-of-the-first-amendment.
66. *See* 47 U.S.C. § 201. These obligations arise under common law. *See also* People v. Brophy, 49 Cal. App. 2d 15, 22 (1942) ("[C]ommon carriers are not the censors of public or private morals, nor are they authorized or required to investigate or regulate the public or private conduct of those who seek service at their hands[.]").
67. *See* Open Internet Order, 30 FCC Rcd. 5601, 5765 (2015). In 2017, the FCC repealed the Open Internet Order and net neutrality rules. *See* Restoring Internet Freedom Order, 33 FCC Rcd. 311 (2017) (restoring broadband internet access to its prior classification as an information service). That order was largely upheld by the D.C. Circuit. *See* Mozilla Corp. v. FCC, 940 F.3d 1 (D.C. Cir. 2019).
68. U.S. Telecom Ass'n v. FCC, 825 F.3d 674, 741–42 (D.C. Cir. 2016).
69. *Paxton*, 2021 WL 5755120, at *6.
70. 141 S. Ct. 1220 (2021).
71. *Id.*
72. For example, what started off as a dozen Facebook employees following a one-page content moderation policy in 2009 turned into an enterprise involving over 7,500 persons in the intervening ten years. *See* Paul M. Barrett, *Who Moderates the Social Media Giants? A Call to End Outsourcing*, NYU STERN CTR. FOR BUS. AND HUM. RTS., 7 (Jun. 2020), https://issuu.com/nyusterncenterforbusinessandhumanri/docs/nyu_content_moderation_report_final_version?fr=sZWZmZjI1NjI1Ng; Alexis C. Madrigal, *Inside Facebook's Fast-Growing Content-Moderation Effort*, ATLANTIC (Feb. 7, 2018), https://www.theatlantic.com/technology/archive/2018/02/what-facebook-told-insiders-about-how-it-moderates-posts/552632/.
73. *See* discussion in the section titled "Statutory and Constitutional Barriers to Addressing Harmful Speech on Social Media," *supra*.
74. *See* Nellie Bowles, *The Complex Debate Over Silicon Valley's Embrace of Content Moderation*, N.Y. TIMES (June 5, 2020), https://www.nytimes.com/2020/06/05/technology/twitter-trump-facebook-moderation.html.
75. *See, e.g.*, Andrew Marantz, *Why Facebook Can't Fix Itself*, NEW YORKER (Oct. 12, 2020), https://www.newyorker.com/magazine/2020/10/19/why-facebook-cant-fix-itself.
76. Jeff Horwitz, *Facebook Has Made Lots of New Rules This Year. It Doesn't Always Enforce Them*, WALL ST. J. (Oct. 15, 2020), https://www.wsj.com/articles/facebook-has-made-lots-of-new-rules-this-year-it-doesnt-always-enforce-them-11602775676.
77. Ariana Tobin et al., *Facebook's Uneven Enforcement of Hate Speech Rules Allows Vile Posts to Stay Up*, PROPUBLICA (Dec. 28, 2017), https://www.propublica.org/article/facebook-enforcement-hate-speech-rules-mistakes.
78. Mark MacCarthy, *A Dispute Resolution Program for Social Media Companies*, BROOKINGS (Oct. 9, 2020), https://www.brookings.edu/research/a-dispute-resolution-program-for-social-media-companies/ (looking at applying "the FINRA model of a non-governmental regulator" to the social media realm).
79. *National Advertising Division,* NAT'L PROGRAMS BBB, https://bbbprograms.org/programs/all-programs/national-advertising-division.
80. *Digital Advertising Accountability Program*, NAT'L PROGRAMS BBB, https://bbbprograms.org/programs/all-programs/daap.
81. MacCarthy, *supra* note 78.
82. *See, e.g.*, International Covenant on Civil and Political Rights, art. 19, Dec. 19, 1966, 999 U.N.T.S. 14668; European Convention on Human Rights, art. 10, Nov. 4, 1950, https://www.echr.coe.int/Documents/Convention_ENG.pdf. *Cf.* Reporters Without Borders World Press Freedom Index (2021), https://rsf.org/en/2021-world-press-freedom-index-journalism-vaccine-against-disinformation-blocked-more-130-countries
83. Paxton, 2021 WL 5755120, at *11.

84. *See, e.g.*, Derek E. Bambauer, *What Does the Day After Section 230 Reform Look Like?*, BROOKINGS (Jan. 22, 2021), https://www.brookings.edu/techstream/what-does-the-day-after-section-230-reform-look-like/.
85. *See* Online Freedom and Viewpoint Diversity Act, S. 4534, 116th Cong., § 2 (2020); Limiting Section 230 Immunity to Good Samaritans Act, S. 3983 (116th Cong., 2d Sess., § 2, 2020).
86. *See* SAFE Tech Act, S. 299, 117th Cong., § 1 (1st Sess., 2021); *Warner, Hirono, Klobuchar Announce the SAFE Tech Act to Reform Section 230*, Mark R. Warner (Feb. 5, 2021), https://www.warner.senate.gov/public/index.cfm/2021/2/warner-hirono-klobuchar-announce-the-safe-tech-act-to-reform-section-230; Health Misinformation Act, S. 2448 (117th Cong., 1st Sess. § 3, 2021).
87. *See* Platform Accountability and Consumer Transparency Act, S. 797 (117th Cong., 1st Sess., 2021), available at https://www.congress.gov/bill/117th-congress/senate-bill/797/text.
88. *See* Phillip Takhar, *A Proposal for a Notice-and-Takedown Process for Revenge Porn*, HARV. J. LAW & TECH (June 5, 2018), https://jolt.law.harvard.edu/digest/a-proposal-for-a-notice-and-takedown-process-for-revenge-porn.
89. Indeed, the notice-and-takedown procedure for copyright has been controversial for many reasons, some of which could be relevant to the Section 230 context: (*i*) the volume of notice-and-takedown claims submitted to larger service providers has been enormous and requires automated processes to handle; and (*ii*) much litigation has ensued over the quality of the takedown notices, the definition of "expeditious," and the implementation and application of the policy against repeat infringers. *See, e.g.*, *Section 512 of Title 17: A Report of the Register of Copyrights*, U.S. COPYRIGHT OFFICE 127–28 (May 2020), https://www.copyright.gov/policy/section512/section-512-full-report.pdf.
90. *See FTC Imposes $5 Billion Penalty and Sweeping New Privacy Restrictions on Facebook*, FED. TRADE COMM'N (July 24, 2019), https://www.ftc.gov/news-events/press-releases/2019/07/ftc-imposes-5-billion-penalty-sweeping-new-privacy-restrictions. *Cf. FTC Requires Zoom to Enhance Its Security Practices as Part of Settlement*, FED. TRADE COMM'N (NOV. 9, 2020), https://www.ftc.gov/news-events/press-releases/2020/11/ftc-requires-zoom-enhance-its-security-practices-part-settlement.
91. *See The FCC and Freedom of Speech*, FED. COMMUNICATIONS COMM'N (Dec. 20, 2019), https://www.fcc.gov/sites/default/files/the_fcc_and_freedom_of_speech.pdf.
92. Federal Trade Commission Act, 15 U.S.C. Sec. 45(a)(1).
93. In *Paxton*, however, the court held that statutory disclosure and transparency requirements were unconstitutional because they both were "inordinately burdensome" and compelled speech. *See* 2021 WL 5755120, at *11.
94. *See, e.g.*, General Data Protection Regulation, Regulation (EU) 2016/679, O.J. 2016 (L 119/1), https://gdpr-info.eu/art-3-gdpr/.
95. *See* paragraph 23, Directive (EU) 2018/1808, amending the Audiovisual Media Services Directive, https://eur-lex.europa.eu/eli/dir/2018/1808/oj.
96. *Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken*, Netzwerkdurchsetzungsgesetz (Sept. 2017), https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/BGBl_NetzDG.pdf?__blob=publicationFile&v=2.
97. "The EU Code of conduct on countering illegal hate speech online," European Commission, 30 June 2016, https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en.
98. *Id.*
99. *See, e.g.*, CEJI—A Jewish Contribution to an Inclusive Europe (Belgium); Freiwillige Selbstkontrolle (Germany); ENAR Ireland (Ireland); Community Security Trust (United Kingdom); Amnesty International Italia (Italy).
100. *See, e.g.*, Finnish Police Academy (Finland); Platforme PHAROS (France); Centre interfédéral pour l'égalité des chances (Belgium); Spanish Ministry of the Interior (Spain).

101. *See, e.g.,* 5th Evaluation of the Code of Conduct on Countering Illegal Hate Speech Online, European Commission (June 2020), https://ec.europa.eu/info/sites/info/files/codeofconduct_2020_factsheet_12.pdf.
102. Regulation on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC, https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM%3A2020%3A825%3AFIN (hereinafter Digital Services Act). On October 1, 2021, the European Parliament's Committee on Legal Affairs adopted the Committee's recommendations on the Digital Services Act. The Council of the European Union and the European Parliament adopted their respective negotiating positions on November 25, 2021, and January 20, 2022. On April 23, 2020 the European Parliament and EU Member States reached a consensus on the final wording of the Digital Services Act. Once formally adopted by the EU co-legislators, the Digital Services Act will be directly applicable across the European Union and will apply from the later of fifteen months after it comes into force or from January 1, 2024. For very large online platforms and search engines, however, the Digital Services Act will apply from a potentially earlier date, being four months after their designation.
103. In the United Kingdom, for example, hate speech regarding disability, race, religion, sexual orientation, or transgender identity is a criminal offense. Furthermore, all EU Member States have introduced offenses to criminalize hate speech regarding racism and xenophobia, pursuant to the EU Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law.
104. Digital Services Act, at Recital 68. Similarly, Sections 453(4) and 187(2) of the UK Draft Online Safety Bill define content as harmful to children if it presents a material risk of significant physical or psychological harm to an appreciable number of children in the United Kingdom. This language is mirrored in Sections 54(3) and 187(2), which define content that is harmful to adults. As with illegal content, a limited number of priority categories of legal yet harmful content for children and adults will be set out in secondary legislation. The definition of harmful content was amended following the publication of the first draft of the legislation to cover a broader range of harms following recommendations in the December 2021 UK Joint Parliamentary Committee on the Online Safety Bill report.
105. Council of Europe, European Convention for the Protection of Human Rights and Fundamental Freedoms 12 (1950), *as amended by* Protocols Nos. 11 and 14, https://www.echr.coe.int/documents/convention_eng.pdf.
106. *See* Pastors v. Germany, App. No. 55225/14 (Oct. 3, 2019), https://hudoc.echr.coe.int/eng?i=001-196148.
107. Digital Services Act, at Article 14.
108. *Id*. at Article 12.
109. *Id.* at Article 13.
110. *Id.* at Article 26.
111. *See* Natasha Lomas, *Facebook to Exclude North American Users from Some Privacy Enhancements*, TECHCRUNCH (Apr. 4, 2018), https://techcrunch.com/2018/04/04/facebook-gdpr-wont-be-universal/; Hayley Tsukayama, *Facebook's Privacy Changes Look Different for Europeans and Americans*, WASH. POST (Apr. 20, 2018), https://www.washingtonpost.com/news/the-switch/wp/2018/04/20/facebooks-privacy-changes-look-different-for-europeans-and-americans/.
112. *See* Reddit, Privacy Policy (effective Oct. 15, 2020), https://www.redditinc.com/policies/privacy-policy.

Chapter 1

1. *Compare, e.g.,* Tim Wu, *Is the First Amendment Obsolete?, in* THE PERILOUS PUBLIC SQUARE: STRUCTURAL THREATS TO FREE EXPRESSION TODAY 28 (David E. Pozen ed., New York: Columbia University Press, 2020), *with* Mark Tushnet, *The Kids Are All Right: The Law of Free Expression and New Information Technologies* (Harvard Public Law Working Paper No. 21-09), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3714415.

2. *See, e.g.,* Geoffrey R. Stone, *Reflections on Whether the First Amendment Is Obsolete, in* THE PERILOUS PUBLIC SQUARE: STRUCTURAL THREATS TO FREE EXPRESSION TODAY 71, 72 (David E. Pozen ed., New York: Columbia University Press, 2020).
3. *See, e.g.,* Miami Herald Publ'g Co. v. Tornillo, 418 U.S. 241, 258 (1974) ("The choice of material to go into a newspaper, and the . . . treatment of public issues and public officials—whether fair or unfair—constitute the exercise of editorial control and judgment. It has yet to be demonstrated how governmental regulation of this crucial process can be exercised consistent with First Amendment guarantees of a free press as they have evolved to this time.").
4. 395 U.S. 367 (1969).
5. *Id*. at 369.
6. *See* Turner Broad. Sys., Inc. v. F.C.C., 512 U.S. 622 (1994); Miami Herald Publ'g Co. v. Tornillo, 418 U.S. 241 (1974).
7. *See* LEE C. BOLLINGER, IMAGES OF A FREE PRESS chs. 5, 6 (Chicago: University of Chicago Press, 1991).
8. *See* Syracuse Peace Council v. F.C.C, 867 F.2d 654 (D.C. Cir. 1989).
9. *See* Genevieve Lakier, *The Non-First Amendment Law of Freedom of Speech*, 134 HARV. L. REV. 2299, 2367–68 (2021)(citing examples).
10. *See, e.g.,* TIM WU, IS FILTERING CENSORSHIP? THE SECOND FREE SPEECH TRADITION (Washington, DC: Governance Studies at Brookings Institution, 2010), https://www.brookings.edu/research/is-filtering-censorship-the-second-free-speech-tradition/; Biden v. Knight First Amendment Inst. at Columbia Univ., 141 S. Ct. 1220 (2021) (Thomas, J., concurring).
11. *See, e.g.,* Hustler Magazine, Inc. v. Falwell, 485 U.S. 46, 52 (1988) ("False statements of fact are particularly valueless; they interfere with the truth-seeking function of the marketplace of ideas, and they cause damage to an individual's reputation that cannot easily be repaired by counterspeech, however persuasive or effective"); Keeton v. Hustler Magazine, Inc., 465 U.S. 770, 776 (1984) ("There is 'no constitutional value in false statements of fact'") (quoting Gertz v. Robert Welch, Inc., 418 U.S. 323, 340 (1974)).
12. 567 U.S. 709 (2012).
13. *See* Nike v. Kasky, 539 U.S. 654 (2003) (dismissing the writ of certiorari as improvidently granted); *id*. at 663–65 (Stevens, J., concurring).
14. *See, e.g,* Milkovich v. Lorain Journal Co., 497 U.S. 1, 18–23 (1990).
15. *See generally* ZEYNEP TUFECKI, TWITTER AND TEAR GAS: THE POWER AND FRAGILITY OF NETWORKED PROTEST (New Haven, CT: Yale University Press, 2017).
16. McCullen v. Coakley, 573 U.S. 464, 488 (2014) (quoting Meyer v. Grant, 486 U.S. 414, 424 (1988)).
17. *Id.* at 476.
18. *See, e.g.,* Hill v. Colorado, 530 U.S. 703, 724 (2000) (referring to "the harassment, the nuisance, the dogging, and the implied threat of physical touching" that accompany face-to-face speech).
19. *See, e.g.,* Chaplinsky v. New Hampshire, 315 U.S. 568 (1942).
20. Cohen v. California, 403 U.S. 15, 20 (1971) (citing Cantwell v. Connecticut, 310 U.S. 296, 309 (1940)).
21. *See id.* at 20 ("While the four-letter word displayed by [the protestor] in relation to the draft is not uncommonly employed in a personally provocative fashion," in this instance it was "clearly not 'directed to the person of the hearer'" and "no individual actually or likely to be present could reasonably have regarded the words . . . as a direct personal insult").
22. *See, e.g.,* Int'l Soc'y for Krishna Consciousness v. Lee, 505 U.S. 672, 675–76, 684 (1992); Heffron v. Int'l Soc'y for Krishna Consciousness, 452 U.S. 640, 643–44 (1981); *compare* In re Primus, 436 U.S. 412 (1978), *with* Ohralik v. Ohio State Bar Ass'n, 436 U.S. 447 (1978).
23. *See, e.g.,* Hill v. Colorado, 530 U.S. 703 (2000); Price v. City of Chicago, 915 F.3d 1107 (2019), *cert. denied*, 141 S. Ct. 185 (2020).
24. *See, e.g.,* People v. Shack, 86 N.Y.2d 529, 658 N.E.2d 706 (1995).
25. *See generally* Richard H. Fallon Jr., *Sexual Harassment, Content Neutrality, and the First Amendment Dog That Didn't Bark*, 1994 S. CT. REV. 1.
26. Mahanoy Area Sch. Dist. v. Levy, 141 S. Ct. 2038, 2045 (2021).

27. *See, e.g.*, Walker v. Texas Div., Sons of Confederate Veterans, Inc., 576 U.S. 200 (2015); Pleasant Grove v. Summum, 555 U.S. 460, 476 (2009) ("The Free Speech Clause restricts government regulation of private speech; it does not regulate government speech."). Sometimes governments express their views by facilitating or subsidizing the speech of private parties—roughly, speaking through independent contractors rather than employees. These selective government subsidies of private speech raise notoriously difficult questions. When the government supports speakers on one side of an issue, its actions, again, have the effect, and sometimes the purpose, of giving that side an advantage in public debate, just as suppressing the other side would—possibly to a lesser degree than suppression, but not necessarily, if the subsidy is substantial. But these problems can be left aside, because even core government speech—speech by government officials who purport to speak for the government—raises questions that become more troubling when the officials use social media.
28. *Compare* Bantam Books, Inc. v. Sullivan, 372 U.S. 58 (1963), *with* Meese v. Keene, 481 U.S. 465 (1987). *See also* Wilson v. Houston Cmty. Coll. Sys., 955 F.3d 490 (5th Cir. 2020) (holding that a government body's censure of one of its members can violate the First Amendment), *cert. granted*, 2021 WL 1602636 (Apr. 26, 2021).
29. If the complaint alleges defamatory speech that was within the scope of the official's duties, the claim will be treated as having been brought against the United States under the Federal Tort Claims Act, so the official will not be personally liable. *See, e.g.*, Haddon v. United States, 68 F.3d 1420, 1422–23 (D.C. Cir. 1995) (citing 28 U.S.C. § 2679(d)(1)). In fact, because the Federal Tort Claims Act does not waive the government's sovereign immunity from defamation actions (see 28 U.S.C. § 2680(h)), the government will not be liable, either.
30. *See, e.g.*, Council of Am. Islamic Relations v. Ballenger, 444 F.3d 359 (D.C. Cir. 2006).
31. 274 U.S. 357, 372–79 (1927).
32. *Id.* at 375.
33. *Id.* at 377.
34. *See* 567 U.S. at 729 (plurality opinion of Kennedy, J.): "A Government-created database could list Congressional Medal of Honor recipients. Were a database accessible through the Internet, it would be easy to verify and expose false claims."

Chapter 2

1. As reported by Maryland delegate James McHenry. *See* 3 MAX FARRAND, THE RECORDS OF THE FEDERAL CONVENTION OF 1787, 85 (New Haven, CT: Yale University Press, 1967).
2. In 1964, three self-published paperbacks making the case for Barry Goldwater became sensational best sellers. Nicole Hemmer, *The Three Books That Shook Conservative Media in 1964*, ATLANTIC (May 20, 2014), https://www.theatlantic.com/politics/archive/2014/05/the-three-books-that-shook-conservative-media-and-politics-in-1964/371264. As Hemmer points out, this was a one-time phenomenon that has never been repeated. *See also* NICOLE HEMMER, MESSENGERS OF THE RIGHT: CONSERVATIVE MEDIA AND THE TRANSFORMATION OF AMERICAN POLITICS (Philadelphia: University of Pennsylvania Press, 2018).
3. *See* EVGENY MOROZOV, THE NET DELUSION: THE DARK SIDE OF INTERNET FREEDOM (New York: Public Affairs, 2011); JARON LANIER, YOU ARE NOT A GADGET: A MANIFESTO (New York: Vintage, 2011).
4. As Pizzagate was a right-wing phenomenon, my examples are right-wing organizations. They have left-wing counterparts, like the Daily Kos or Talking Points Memo, though I accept the point—well documented by research—that the radical change in our media ecosystem is an asymmetric phenomenon, more widespread, more pronounced, and more extreme on the right than on the left. *See* YOCHAI BENKLER ET AL., NETWORK PROPAGANDA: MANIPULATION, DISINFORMATION, AND RADICALIZATION IN AMERICAN POLITICS (New York: Oxford Scholarship Online, 2018). That fact surely matters in thinking about a solution, but it is irrelevant to the point I am making here—which is that the problem requires fundamentally rethinking the legal norms that define the space within which any solution must be formulated. As Benkler et al. note, even if concentrated on one side of the political divide, "having a segment of the population that is systematically disengaged from objective journalism and the

ability to tell truth from partisan fiction is dangerous to any country," "makes actual governance difficult," and "is a political disaster waiting to happen." *Id.* at 16.

5. *See* Amanda Robb, *Anatomy of a Fake News Scandal*, ROLLING STONE (Nov. 16, 2017), https://www.rollingstone.com/feature/anatomy-of-a-fake-news-scandal-125877.
6. *See* ALAN BRINKLEY, VOICES OF PROTEST: HUEY LONG, FATHER COUGHLIN, & THE GREAT DEPRESSION (New York: Vintage, 1983); MARCUS SHELDON, FATHER COUGHLIN: THE TUMULTUOUS LIFE OF THE PRIEST OF THE LITTLE FLOWER (Boston: Little, Brown, 1973); John Schneider, *The Rabble-Rousers of Early Radio Broadcasting*, 42 RADIO WORLD 16–18 (2018). Our latter-day Coughlins—characters like Rush Limbaugh, Sean Hannity, Laura Ingraham, Tucker Carlson, and any number of other Fox News and talk radio personalities—are less crude and extreme in their biliousness, which gives them a semiplausible veneer of legitimacy and respectability.
7. Recall the Macedonian teenagers who notoriously created more than one hundred websites that imitated legitimate news while running hyperpartisan stories about the 2016 US elections to generate advertising revenues for themselves. *See* Craig Silverman, *This Analysis Shows How Viral Fake Election News Stories Outperformed Real News on Facebook*, BUZZFEED NEWS (Nov. 16, 2016), https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook; Craig Silverman & Lawrence Alexander, *How Teens in the Balkans Are Duping Trump Supporters with Fake News*, BUZZFEED (Nov. 3, 2016), https://www.buzzfeednews.com/article/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo.
8. *See* Elisa Shearer, *More Than Eight-in-Ten Americans Get News from Digital Devices*, PEW RES. CTR. (Jan. 12, 2021), https://www.pewresearch.org/fact-tank/2021/01/12/more-than-eight-in-ten-americans-get-news-from-digital-devices.
9. *See generally* JOHN V. PAVLIK, JOURNALISM AND NEW MEDIA (New York: Columbia University Press, 2001); JOHN V. PAVLIK, MEDIA IN THE DIGITAL AGE (New York: Columbia University Press, 2008).
10. Gregory P. Magarian, *The Internet and Social Media*, *in* THE OXFORD HANDBOOK OF FREEDOM OF SPEECH 350, 365–68 (Adrienne Stone & Frederick Schauer eds., Oxford: Oxford University Press, 2021); STUART ALLAN & EINAR THORSEN EDS., CITIZEN JOURNALISM: GLOBAL PERSPECTIVES (New York: Peter Lang, 2009).
11. Whitney v. California, 275 U.S. 357, 377 (1927) (Brandeis, J., concurring).
12. *Id.* at 375–77.
13. Peter Overby, *NRA: "Only Thing That Stops a Bad Guy with a Gun Is a Good Guy with a Gun,"* ALL THINGS CONSIDERED (Dec. 21, 2012, 3:00 PM), https://www.npr.org/2012/12/21/167824766/nra-only-thing-that-stops-a-bad-guy-with-a-gun-is-a-good-guy-with-a-gun.
14. Commentators have noted the unusual process by which an "American" identity formed among the peoples of thirteen independent colonies over the course of the Revolution and crises that followed through the adoption of the Constitution and the stabilization of politics in the early Republic. *See* YEHOSHUA ARIELI, INDIVIDUALISM AND NATIONALISM IN AMERICAN IDEOLOGY 1–13 (Cambridge, MA: Harvard University Press, 1964). Tocqueville marveled at this fact, describing to a friend the puzzle that would frame his analysis of America:

> Picture to yourself . . . if you can, a society which comprises all the nations of the world . . . people differing from one another in language, in beliefs, in opinions; in a word a society possessing no roots, no memories, no prejudices, no routine, no common ideas, no national character, yet with a happiness a hundred times greater than our own. . . . This, then, is our starting point! What is the connecting element between these so different elements? How are they welded into one people?

Quoted in Arieli, *supra*, at 17. Tocqueville had refined this question by the time he wrote *Democracy in America*, shifting his gaze from Americans' "happiness" to their unusual public spiritedness. "How is it," he asked,

> that in the United States, where the inhabitants arrived but yesterday in the land they occupy, whither they brought with them neither customs nor memories, where they

> meet for the first time without knowing each other, where, to say it in one word, the instinct of country can hardly exist—how does it come about that each man is as interested in the affairs of his township, of his canton, and of the whole state as he is in his own affairs?

ALEXIS DE TOCQUEVILLE, DEMOCRACY IN AMERICA 236 (J. P. Mayer ed., George Lawrence trans., Anchor Books, Doubleday, 1969). The answer is that Americans built a new identity on two related cornerstones: first, their commitment to an idea—the idea of republicanism that motivated the Revolution; and second, a system they devised themselves to make that idea a practical reality in day-to-day governance—the US Constitution. These became the foundation for expressing who we are as a people and what unites us as a nation: the phenomena that defined, and still today define, America and Americans uniquely in the world and in world history. Lest you doubt it, visit the National Archives, where you will see the Declaration and the Constitution laid out as on an altar, as clear a display as one could make of the one religion—albeit a civic religion—that Americans have always shared. *See generally* Larry Kramer, *"To Adjust These Clashing Interests": Negotiation and Compromise as Core Constitutional Values* (2015 Roberts Lecture, U. Pa. L. School) (manuscript on file with author). But as I argue below, that may soon need be "used to share."

15. In the United States, social deference to a leadership class persisted at least into the early years of the twentieth century. Members of the leadership class disagreed and stood on opposite sides of political issues, but they still worked within a generally shared sense of the world and managed to govern passably well until the issue of slavery proved more than they could manage. Even then, it is striking how quickly politics restabilized after the Civil War. It was this structure of politics that made tolerable the existence of an information environment dominated by partisan journals willing to spread lies and practice yellow journalism. Politics became more genuinely popular in the early twentieth century, especially in the years after World War I, but this development just happened to coincide with the rise of a new information environment dominated by a relative handful of major news organizations that, as described above, practiced serious, nonpartisan journalism with high professional standards and a centrist orientation—yielding a reasonably well-informed public operating with similar understandings of the state of the world.
16. *See* Matthew Valasik & Shannon E. Reid, *After the Insurrection: America's Far-Right Groups Get More Extreme*, CONVERSATION (Mar. 15, 2021), https://theconversation.com/after-the-insurrection-americas-far-right-groups-get-more-extreme-156463; Lee Drutman, *Theft Perception: Examining the Views of Americans Who Believe the 2020 Election Was Stolen*, Democracy Fund Voter Study Group (June 2021), https://www.voterstudygroup.org/publication/theft-perception.
17. 337 U.S. 1 (1949).
18. *Id.* at 37 (Jackson, J., dissenting).
19. *See* Frederick Schauer, *Slippery Slopes*, 99 HARV. L. REV. 361 (1985).
20. 376 U.S. 254, 280 (1964).
21. 567 U.S. 709 (2011) (slip op. at 10) (plurality op.) (quoting Brown v. Entm't Merchants Ass'n, 564 U.S. 768 (2011)).
22. *Id.*, slip op. at 4–7 (Breyer, J., concurring in the judgment).
23. *See* Paul Finkelman, *Foreign Law and American Constitutional Interpretation: A Long and Venerable Tradition*, 63 N.Y.U. ANN. SURVEY OF AM. L. 29, 37–49 (2007). The deep familiarity with and respect for lessons from other nations among the founding generation of Americans is evident from even a light perusal of the period's writings.
24. Justice Scalia once boasted, "I probably use more foreign legal materials than anyone else on the Court," before snidely adding, "Of course they are all fairly old foreign legal materials, and they are all English." As for contemporary foreign legal materials, Scalia dismissed them entirely, stating these "can *never* be relevant to an interpretation of—to the *meaning* of—the U.S. Constitution." Antonin Scalia, *Foreign Legal Authority in the Federal Courts, Keynote Address to the Annual Meeting of the American Society of International Law (March 31–April 3, 2004)*, 98 AM. SOC'Y INT'L L. PROC. 305, 306–10 (2004).

25. R. v. Oakes, 1 S.C.R. 103 (1986); *see* Grégoire Webber, *Proportionality and Limitations on Freedom of Speech*, *in* THE OXFORD HANDBOOK OF FREEDOM OF SPEECH 172, 176–180, 186 (Adrienne Stone & Frederick Schauer eds., Oxford: Oxford University Press, 2021); MARK TUSHNET, ADVANCED INTRODUCTION TO FREEDOM OF EXPRESSION 24–29 (2018).
26. *See, e.g.*, Dennis v. United States, 341 U.S. 494 (1951); Barenblatt v. United States, 360 U.S. 109 (1959); Konigsberg v. State Bar of Cal., 366 U.S. 36 (1961).
27. International Covenant on Civil and Political Rights art. 19(3), Dec. 19, 1966, 999 U.N.T.S. 171, Can. T.S. 1976 No. 47 (entered into force Mar. 23, 1976). "Ordre public" is the international law version of what in American conflict of laws is called the "public policy" doctrine and which refers to protecting "some fundamental principle of justice, some prevalent conception of good morals, some deep-rooted tradition of the common weal." *See* Loucks v. Standard Oil Co., 224 N.Y. 99, 111, 120 N.E. 198, 202 (1918).
28. International Covenant on Civil and Political Rights art. 20.
29. Charter of Fundamental Rights of the European Union art. 10 (2016) Official Journal C202, June 7, 389–405.
30. A useful overview of differences among different nations in the treatment of freedom of expression, broken out by topic and issues, can be found in THE OXFORD HANDBOOK OF FREEDOM OF SPEECH (Adrienne Stone & Frederick Shauer eds., Oxford: Oxford University Press, 2021), which compiles essays from many of the world's leading experts.
31. *Id.*, chs. 19, 25.
32. *Id.*, ch. 16.
33. *Id.*, ch. 17.
34. *Id.*, ch. 21.
35. *Id.*, ch. 27.
36. *Id.*, ch. 29.
37. *Id.* at 204.
38. *Id.*, ch. 6.
39. *See* Adam Satariano, *An Experiment to Stop Online Abuse Falls Short in Germany*, N.Y. TIMES (Sept. 23, 2021), https://www.nytimes.com/2021/09/23/technology/online-hate-speech-germany.html.
40. For typical examples, see laws in Canada, Criminal Code, RSC 1985, c C-46; and France, Loi 90-615 du 13 juillet 1990.
41. Den. Pen. Code, Straffeloven, § 266 B.
42. Strafgesetzbuch §§ 130 (1), 130(2).
43. Ice. Pen. Code, art. 233(a), L. 13/2014, 2. gr., L. 96/1973, 1. gr.
44. *See* Amanda Erickson, *Macron's Emails Got Hacked. Here's Why French Voters Won't Hear Much About Them Before Sunday's Election*, WASH. POST (May 6, 2017), https://www.washingtonpost.com/news/worldviews/wp/2017/05/06/macrons-emails-got-hacked-heres-why-french-voters-wont-hear-much-about-them-before-sundays-election.
45. Not long before the 2020 election, *Consumer Reports* prepared a guide to the varying curation policies of the largest social media platforms. *See* Kaveh Waddell, *On Social Media Only Some Lies Are Against the Rules*, CONSUMER REP. (Aug. 13, 2020), https://www.consumerreports.org/social-media/social-media-misinformation-policies.
46. 47 U.S.C. § 230.
47. *See* GARY MARCUS & ERNEST DAVIS, REBOOTING AI: BUILDING ARTIFICIAL INTELLIGENCE WE CAN TRUST (New York: Vintage, 2019).
48. Benkler et al.provide a useful, succinct summary of propaganda, as well as a compelling description of what they call the "propaganda feedback loop." *Supra* note 4, at 24–38, 75–99.
49. Roger McNamee, *How to Fix Facebook—Before It Fixes Us*, WASH. MONTHLY (Jan./Feb./Mar. 2018), https://washingtonmonthly.com/magazine/january-february-march-2018/how-to-fix-facebook-before-it-fixes-us.
50. *See* Shearer, *supra* note 8. (68 percent of Americans get their news from news websites or apps, compared to 65 percent who rely on search engines, 53 percent who rely on social media, and 22 percent who rely on podcasts).

Chapter 3

1. Vivek H. Murthy, CONFRONTING HEALTH MISINFORMATION: THE U.S. SURGEON GENERAL'S ADVISORY ON BUILDING A HEALTHY INFORMATION ENVIRONMENT (2021), https://www.hhs.gov/sites/default/files/surgeon-general-misinformation-advisory.pdf.
2. *Legislating on Hate Speech and Disinformation—Lessons from Ethiopia*, IKIGAI LAW (May 8, 2020), https://www.ikigailaw.com/legislating-on-hate-speech-and-disinformation-lessons-from-ethiopia/#acceptLicense.
3. John Stuart Mill, ON LIBERTY (London: John W. Parker & Son, 1859).
4. Tim Wu, *Is the First Amendment Obsolete?*, *in* THE PERILOUS PUBLIC SQUARE: STRUCTURAL THREATS TO FREE EXPRESSION TODAY (David E. Pozen ed., New York: Columbia University Press, 2020).
5. Casey Newton, *President Biden Goes After Facebook*, PLATFORMER (July 19, 2021), https://www.platformer.news/p/president-biden-goes-after-facebook.
6. Sheera Frenkel, *White House Dispute Exposes Facebook Blind Spot on Misinformation*, N.Y. TIMES, July 19, 2021, https://www.nytimes.com/2021/07/19/technology/facebook-misinformation-blind-spot.html.
7. Soroush Vosoughi, Deb Roy, & Sinan Aral, *The Spread of True and False News Online*, 359 SCIENCE 1146–51 (2018), https://www.science.org/doi/abs/10.1126/science.aap9559.
8. Kevin Roose, *Inside Facebook's Data Wars*, N.Y. TIMES, July 14, 2021, https://www.nytimes.com/2021/07/14/technology/facebook-data.html.
9. Newton, *supra* note 5.
10. Yochai Benkler et al., *Mail-In Voter Fraud: Anatomy of a Disinformation Campaign*, Berkman Klein Ctr. for Internet & Soc'y, Res. Pub. No. 2020-6, Oct. 2, 2020, https://cyber.harvard.edu/publication/2020/Mail-in-Voter-Fraud-Disinformation-2020.
11. Laura Hazard Owen, "*At First, Facebook Was Happy That I and Other Journalists Were Finding Its Tool Useful . . . But the Mood Shifted*," NEIMANLAB (July 14, 2021).
12. Nic Newman et al., *Reuters Institute Digital News Report 2020*, Reuters Institute for the Study of Journalism (2020), https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf.
13. Network Enforcement Act [Bearbeitungsstand] (June 12, 2017).
14. Hannah Arendt, THE ORIGINS OF TOTALITARIANISM (New York: Schocken, 1951).
15. Michael J. Bazyler, HOLOCAUST, GENOCIDE, AND THE LAW: A QUEST FOR JUSTICE IN A POST-HOLOCAUST WORLD (New York: Oxford University Press, 2017).
16. Abrams v. United States, 250 U.S. 616 (1919)
17. Brandenburg v. Ohio, 395 U.S. 444 (1969).
18. Geoffrey R. Stone, interview with the author (2021).
19. New York Times Co. v. Sullivan, 376 U.S. 254 (1964).
20. Genevieve Lakier, *The First Amendment's Real* Lochner *Problem*, 87 U. CHI. L. REV. 1241, 1311 (2020).
21. Red Lion Broadcasting Co. v. FCC, 395 U.S. 367 (1969).
22. Genevieve Lakier, interview with the author (2021).
23. Miami Herald Pub. Co. v. Tornillo, 418 U.S. 241 (1974).
24. John C. Coates, IV, *Corporate Speech & the First Amendment: History, Data, and Implications*, 30(2) CONST. COMMENT. 223–75 (2015), https://hdl.handle.net/11299/183130.
25. Citizens United v. FEC, 558 U.S. 310 (2010).
26. Catharine A. MacKinnon, *The First Amendment: An Equality Reading*, *in* THE FREE SPEECH CENTURY. 140 (Lee C. Bollinger & Geoffrey R. Stone eds., New York: Oxford University Press, 2018).
27. Janus v. Am. Fed'n of State, 138 S.Ct. 2448 (2018).
28. Robert Post, interview with the author (2020).
29. Genevieve Lakier, interview with the author (2021).
30. Ellen P. Goodman, interview with the author (2020).
31. Public Broadcasting Act of 1967, 47 U.S.C. § 396 (1967).
32. Matthew Gentzkow, interview with the author (2020).
33. Ellen P. Goodman, interview with the author (2020).

34. Adrian Croft & Geert De Clercq, *France Fights to Keep Macron Email Hack from Distorting Election*, REUTERS (May 6, 2017), https://www.reuters.com/article/us-france-election/france-fights-to-keep-macron-email-hack-from-distorting-election-idUSKBN1820BO.
35. Dominique Cardon, interview with the author (2020).
36. Arnaud Mercier, interview with the author (2020).
37. *Facebook's Algorithm: A Major Threat to Public Health*, AVAAZ (Aug. 19, 2020), https://secure.avaaz.org/campaign/en/facebook_threat_health/.
38. Fadi Quran, interview with the author (2020).
39. Ellen P. Goodman & Sanjay Jolly, *A "Full Stack" Approach to Public Media in the United States*, The German Marshall Fund of the United States (July 26, 2021), https://www.gmfus.org/news/full-stack-approach-public-media-united-states.
40. U.S. Dep't of Justice, § 230—Nurturing Innovation or Fostering Unaccountability? (June 2020). ("One of the key benefits and most frequent uses of Section 230 identified by experts and industry is the protection for online platforms from intermediary liability for defamation based on material posted by third parties on their services.").
41. Paul Barrett, *Regulating Social Media: The Fight over Section 230—and Beyond*, NYU Stern Center for Business and Human Rights (Sept. 8, 2020), https://www.stern.nyu.edu/experience-stern/faculty-research/regulating-social-media-fight-over-section-230-and-beyond; Stigler Comm. on Digital Platforms, *Final Report*, University of Chicago Booth School of Business (2019), https://www.chicagobooth.edu/-/media/research/stigler/pdfs/digital-platforms---committee-report---stigler-center.pdf.
42. Ellen P. Goodman & Karen Kornbluh, *Social Media Platforms Need to Flatten the Curve of Dangerous Misinformation*, SLATE (Aug. 21, 2020).
43. See Investigation of Competition in Digital Markets: Majority Staff Report and Recommendations: Subcom. on Antitrust, Comm. and Adminitrative Law of the Com. on Judiciary (October 2, 2020), https://judiciary.house.gov/uploadedfiles/competition_in_digital_markets.pdf?utm_campaign=4493-519
44. Cat Zakrzewski, *Lina Khan's First Big Test as FTC Chief: Defining Facebook as a Monopoly*, WASH. POST (Aug. 19, 2021), https://www.washingtonpost.com/technology/2021/08/19/ftc-facebook-lawsuit-lina-khan-deadline/.
45. H.R. 7014, Banning Microtargeted Political Ads Act, 116th Cong. (2019–20), introduced by Rep. Anna Eshoo (May 2020); and S. 4066, The Platform Accountability and Consumer Transparency (PACT) Act, 116th Cong. (2019–20), introduced by Sens. Brian Schatz and John Thune (June 2020).
46. Elizabeth Warren, *Here's How We Can Break up Big Tech*, MEDIUM (2019), https://medium.com/@teamwarren/heres-how-we-can-break-up-big-tech-9ad9e0da324c.

Chapter 4

1. *See* Schenk v. United States, 249 U.S. 47, 52 (1919).
2. *See* West Va. State Bd. of Educ. v. Barnette, 319 U.S. 624, 641 (1943).
3. *See* WILLIAM BLAKE, THE PORTABLE BLAKE (Alfred Kazin ed., New York: Viking Penguin, 1977).
4. ST. AUGUSTINE, TREATISES ON VARIOUS SUBJECTS (THE FATHERS OF THE CHURCH, VOLUME 16) 78 (Roy J. Deferrari ed., Washington, DC: Catholic University Press of America, 2010).
5. HANNAH ARENDT, BETWEEN PAST AND FUTURE: EIGHT EXERCISES IN POLITICAL THOUGHT 297 (New York: Penguin, 1968).
6. *See* Whitney v. California, 274 U.S. 357, 398 (1927) (Brandeis, J., concurring).
7. *See* Arendt, *supra* note 5, at 232.

Chapter 5

1. Marcy Gordon & Barbara Ortutay, *U.S. Senators Push Facebook Exec on Instagram Policies for Youth*, ASSOCIATED PRESS (Sept. 30, 2021, 4:43 PM), https://apnews.com/article/coronavirus-pandemic-business-technology-health-congress-25bf24d23a3ac6f08f833b8066a34000.

2. John Robertson, *"For Our Own Good": Federal Preemption of State Tort Law—Risk, Regulation, and the Goals of Environmental Protection*, 20 WM. & MARY ENVTL. L. & POL'Y REV. 143, 169 (1995).
3. Statista, market capitalization of largest companies in S&P 500 Index as of September 24, 2021.
4. Peter Eavis & Steve Lohr, *Big Tech's Domination of Business Reaches New Heights,* N.Y. TIMES (Aug. 19, 2020), https://www.nytimes.com/2020/08/19/technology/big-tech-business-domination.html.
5. Emily A. Vogels, *56% of Americans Support More Regulation of Major Technology Companies*, PEW RES. CTR. (July 20, 2021), https://www.pewresearch.org/fact-tank/2021/07/20/56-of-americans-support-more-regulation-of-major-technology-companies.
6. Gerrit De Vynck et al., *Big Tech CEOs Face Lawmakers in House Hearing on Social Media's Role in Extremism, Misinformation*, WASH. POST (Apr. 9, 2021), https://www.washingtonpost.com/technology/2021/03/25/facebook-google-twitter-house-hearing-live-updates.
7. The White House, Executive Order on Promoting Competition in the American Economy (July 9, 2021), https://www.whitehouse.gov/briefing-room/presidential-actions/2021/07/09/executive-order-on-promoting-competition-in-the-american-economy.
8. Brian Fung, *FTC Files Fresh Antitrust Complaint Seeking to Break Up Facebook*, CNN (Aug. 19, 2021), https://www.cnn.com/2021/08/19/tech/ftc-facebook-antitrust/index.html.
9. Kiran Jeevanjee et al., *All the Ways Congress Wants to Change Section 230*, SLATE (Mar. 23, 2021), https://slate.com/technology/2021/03/section-230-reform-legislative-tracker.html.
10. Dominic Rushe, *Big Tech's Big Week Raises Fears of "Blade Runner Future" of Mega-Company Rule*, GUARDIAN (Aug. 1, 2021), https://www.theguardian.com/technology/2021/aug/01/big-techs-big-week-blade-runner-future-amazon-google-apple-microsoft.
11. Brooke Auxier & Monica Anderson, *Social Media Use in 2021*, PEW RES. CTR. (Apr. 7, 2021), https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021.
12. *Id.*
13. *E.g.*, Jane Doe No. 1 v. Backpage.com, LLC, 817 F.3d 12 (1st Cir. 2016).
14. *E.g.*, Daniel v. Armslist, LLC, 2019 WI 47, 386 Wis. 2d 449, 926 N.W.2d 710, *cert. denied*, No. 19-153, 2019 WL 6257416 (U.S. Nov. 25, 2019).
15. *E.g.*, Oberdorf v. Amazon.com, Inc., 295 F. Supp. 3d 496 (M.D. Pa. 2017), *aff'd in part, vacated in part,* 930 F.3d 136 (3d Cir. 2019), *vacated en banc*, 936 F.3d 182 (3d Cir. 2019).
16. Michael L. Rustad & Thomas H. Koenig, *Taming the Tort Monster: The American Civil Justice System as a Battleground of Social Theory*, 68 BROOK. L. REV. 1, 4 (2002).
17. Michael L. Rustad & Thomas H. Koenig, *Rebooting Cybertort Law*, 80 WASH. L. REV. 335, 382 (2005).
18. Chambers v. Baltimore & Ohio R. Co., 207 U.S. 142, 148, 28 S. Ct. 34, 35, 52 L. Ed. 143 (1907).
19. Douglas A. Kysar, *The Public Life of Private Law: Tort Law as a Risk Regulation Mechanism*, 9 EUR. J. RISK REG. 48, 50 (2018).
20. Rustad & Koenig, *supra* note 17, at 385.
21. Danielle Citron & Mary Anne Franks, *Cyber Civil Rights in the Time of COVID-19*, HARV. L. REV. BLOG (May 14, 2020), https://blog.harvardlawreview.org/cyber-civil-rights-in-the-time-of-covid-19.
22. Danielle Keats Citron, *Cyber Civil Rights*, 89(61) B.U. L. REV. 62 (2009), https://www.bu.edu/law/journals-archive/bulr/volume89n1/documents/CITRON.pdf.
23. Jonathon W. Penney, *Whose Speech Is Chilled by Surveillance?*, SLATE (July 7, 2017, 7:32 AM), https://slate.com/technology/2017/07/women-young-people-experience-the-chilling-effects-of-surveillance-at-higher-rates.html; Caitlin Ring Carlson & Haley Witt, *Online Harassment of U.S. Women Journalists and Its Impact on Press Freedom*, 25(11) FIRST MONDAY (2020), https://firstmonday.org/ojs/index.php/fm/article/download/11071/9995; Lucina Di Meco & Saskia Brechenmacher, *Tackling Online Abuse and Disinformation Targeting Women in Politics*, CARNEGIE ENDOWMENT FOR INTERNATIONAL PEACE (Nov. 30, 2020), https://carnegieendowment.org/2020/11/30/tackling-online-abuse-and-disinformation-targeting-women-in-politics-pub-83331.
24. *State of the First Amendment 2019*, FREEDOM FORUM INSTITUTE (2019), https://www.freedomforuminstitute.org/first-amendment-center/state-of-the-first-amendment.

25. *See* MARY ANNE FRANKS, THE CULT OF THE CONSTITUTION (Stanford, CA: Stanford University Press, 2019).
26. Nicholas Thompson, *Instagram's Kevin Systrom Wants to Clean up the &#%$@! Internet*, WIRED (Aug. 14, 2017), https://www.wired.com/2017/08/instagram-kevin-systrom-wants-to-clean-up-the-internet.
27. John Perry Barlow, *A Declaration of the Independence of Cyberspace*, ELECTRONIC FRONTIER FOUNDATION (Feb. 8, 1996), https://www.eff.org/cyberspace-independence.
28. *Id*.
29. Marvin Ammori, *The "New"* New York Times: *Free Speech Lawyering in the Age of Google and Twitter*, 127 HARV. L. REV. 2259, 2283 (2014).
30. Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1621 (2018).
31. Adrien Chen, *Reddit CEO Speaks out on Violentacrez in Leaked Memo: "We Stand for Free Speech,"* GAWKER (Oct. 16, 2012), https://www.gawker.com/5952349/reddit-ceo-speaks-out-on-violentacrez-in-leaked-memo-we-stand-for-free-speech.
32. Patrick Howell O'Neill, 8chan, The Central Hive Of Gamergate, *Is Also An Active Pedophile Network*, Daily Dot (Nov. 17, 2014), https://www.dailydot.com/debug/8chan-pedophiles child-porn-gamergate/.
33. *Community Guidelines*, PARLER (Feb. 14, 2021), https://parler.com/documents/guidelines.pdf.
34. Emily Stewart, *Lawmakers Seem Confused About What Facebook Does—And How to Fix It*, VOX (Apr. 10, 2018), https://www.vox.com/policy-and-politics/2018/4/10/17222062/mark-zuckerberg-testimony-graham-facebook-regulations.
35. ASTRA TAYLOR, THE PEOPLE'S PLATFORM: TAKING BACK POWER AND CULTURE IN THE DIGITAL AGE 221 (New York: Henry Holt, 2014).
36. SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER 514 (New York: PublicAffairs, 2019).
37. Lucas D. Introna & Helen Nissenbaum, *Shaping the Web: Why the Politics of Search Engines Matters*, 16 INFO. SOC'Y 169, 171 (2000).
38. Moran Yemini, *The New Irony of Free Speech*, 20 COLUM. SCI. & TECH. L. REV. 119, 192 (2018).
39. Elizabeth Dwoskin & Nitasha Tiku, *How Twitter, on the Front Lines of History, Finally Decided to Ban Trump*, WASH. POST (Jan. 16, 2021, 6:50 PM), https://www.washingtonpost.com/technology/2021/01/16/how-twitter-banned-trump.
40. Travis Caldwell, *Trump's "We Love You" to Capitol Rioters Is More of the Same*, CNN (Jan. 7, 2021, 11:18 AM), https://www.cnn.com/2021/01/07/politics/trump-history-comments-trnd/index.html.
41. Adi Robertson, *Twitter Says Trump's Account Is Locked, and He's Facing a Ban*, VERGE (Jan. 6, 2021, 7:13 PM), https://www.theverge.com/2021/1/6/22217686/trump-twitter-account-locked-capitol-hill-riot-tweets-policy-violations.
42. Twitter Inc., *Permanent Suspension of @realDonaldTrump,* TWITTER BLOG (Jan. 8, 2021), https://blog.twitter.com/en_us/topics/company/2020/suspension.html.
43. Luke Mogelson, *Among the Insurrectionists*, NEW YORKER (Jan. 25, 2021), https://www.newyorker.com/magazine/2021/01/25/among-the-insurrectionists.
44. Rebecca Heilweil & Shirin Ghaffary, *How Trump's Internet Built and Broadcast the Capitol Insurrection*, VOX (Jan 8, 2021, 5:00 PM), https://www.vox.com/recode/22221285/trump-online-capitol-riot-far-right-parler-twitter-facebook.
45. Tony Romm & Elizabeth Dwoskin, *Trump Banned from Facebook Indefinitely, CEO Mark Zuckerberg Says*, WASH. POST (Jan. 7, 2021, 8:26 PM), https://www.washingtonpost.com/technology/2021/01/07/trump-twitter-ban.
46. Adi Robertson, *Twitter Bans QAnon Supporters, Including Former National Security Adviser Michael Flynn*, VERGE (Jan. 8, 2021, 5:20 PM), https://www.theverge.com/2021/1/8/22221332/twitter-ban-qanon-accounts-michael-flynn-sidney-powell-ron-watkins.
47. Jack Nicas & Davey Alba, *Amazon, Apple and Google Cut off Parler, an App That Drew Trump Supporter*, N.Y. TIMES (Jan. 9, 2021), https://www.nytimes.com/2021/01/09/technology/apple-google-parler.html.

48. Lexi Lonas, *Pompeo, Cruz and Other Trump Allies Condemn Twitter's Ban on President*, THE HILL (Jan. 9, 2021, 1:06 PM), https://thehill.com/policy/technology/533486-pompeo-cruz-and-other-trump-allies-condemn-twitters-ban-on-president.
49. Lauren Giella, *Fact Check: Did Twitter Violate President Trump's First Amendment Rights?*, NEWSWEEK (Jan. 11, 2021, 7:38 PM), https://www.newsweek.com/fact-check-did-twitter-violate-president-trumps-first-amendment-rights-1560673.
50. Abigail Covington, *Conservatives Are Upset They Are Losing Followers on Twitter*, ESQUIRE (Jan. 10, 2021), https://www.esquire.com/news-politics/a35172071/matt-gaetz-twitter-followers-purge-qanon.
51. *See* Alex Kern & Kari Paul, *Donald Trump Suspended from Facebook Indefinitely, Says Mark Zuckerberg*, GUARDIAN (Jan. 7, 2021), https://www.theguardian.com/us-news/2021/jan/07/donald-trump-twitter-ban-comes-to-end-amid-calls-for-tougher-action.
52. Corynne McSherry, *EFF's Response to Social Media Companies' Decisions to Block President Trump's Accounts*, ELECTRONIC FRONTIER FOUNDATION (Jan. 7, 2021), https://www.eff.org/deeplinks/2021/01/eff-response-social-media-companies-decision-block-president-trumps-accounts.
53. Natalie Colarossi, *ACLU Counsel Warns of "Unchecked Power" of Twitter, Facebook After Trump Suspension*, NEWSWEEK (Jan. 9, 2021, 1:21 PM), https://www.newsweek.com/aclu-counsel-warns-unchecked-power-twitter-facebook-after-trump-suspension-1560248.
54. *See* Adi Robertson, *Parler Sues Amazon for Kicking It off the Internet*, VERGE (Jan. 11, 2021), https://www.theverge.com/2021/1/11/22225072/parler-amazon-web-services-lawsuit-antitrust-twitter-capitol-riot-incitement-moderation.
55. 137 S. Ct. 1730, 1737 (2017).
56. Community Decency Act of 1996, 47 U.S.C. § 230.
57. McCullen v. Coakley, 134 S. Ct. 2518, 2529 (2014).
58. Manhattan Cmty. Access, No. 17-1702, 587 U.S. __ (2019).
59. *Id.* at 5.
60. *Id.* at 15.
61. Elizabeth Nolan Brown, *Section 230 Is the Internet's First Amendment. Now Both Republicans and Democrats Want to Take It Away*, REASON (July 29, 2019, 8:01 AM), https://reason.com/2019/07/29/section-230-is-the-internets-first-amendment-now-both-republicans-and-democrats-want-to-take-it-away.
62. Matt Laslo, *The Fight over Section 230—And the Internet as We Know It*, WIRED (Aug. 13, 2019, 3:18 PM), https://www.wired.com/story/fight-over-section-230-internet-as-we-know-it.
63. Alexis Kramer, *Armslist Online Gun Sale Case Won't Get Supreme Court Review*, BLOOMBERG LAW (Nov. 25, 2019, 9:36 AM), https://news.bloomberglaw.com/tech-and-telecom-law/armslist-mass-shooting-case-wont-get-supreme-court-review.
64. I AM JANE DOE (50 Eggs Films 2017), https://www.iamjanedoefilm.com.
65. Mary Anne Franks, *Moral Hazard on Stilts: "Zeran's" Legacy*, YAHOO!FINANCE (Nov. 10, 2017), https://finance.yahoo.com/news/moral-hazard-stilts-apos-zeran-083033447.html.
66. Kalev Leetaru, *Do Social Media Platforms Really Care About Online Abuse?*, FORBES (Jan. 12, 2017, 11:46 AM), https://www.forbes.com/sites/kalevleetaru/2017/01/12/do-social-media-platforms-really-care-about-online-abuse/#39b6354845f1.
67. Mary Anne Franks, *How the Internet Unmakes Law*, 16 OHIO ST. TECH. L. JOUR. 10–24 (2020), https://kb.osu.edu/bitstream/handle/1811/91840/OSTLJ_V16N1_010.pdf?sequence=1&isAllowed=y.

Chapter 6

1. 395 U.S. 367 (1969).
2. *See, e.g.*, Gregory P. Magarian, *How Cheap Speech Underserves and Overheats Democracy*, 54 U.C. DAVIS L. REV. 2455 (2021); Richard L. Hasen, *Cheap Speech and What It Has Done (to American Democracy)*, 16 FIRST AMEND. L. REV. 200, 201 (2017).
3. *See* Martha Minow, SAVING THE NEWS: WHY THE CONSTITUTION CALLS FOR GOVERNMENT ACTION TO PRESERVE FREEDOM OF SPEECH (New York: Oxford University Press, 2021).
4. Magarian, *supra* note 2, at 2477.

5. Eugene Volokh, *Cheap Speech and What It Will Do*, 104 YALE L.J. 1805, 1806–7 (1995).
6. 137 S. Ct. 1730 (2017).
7. *Id.* at 1733.
8. *Id.* at 1735.
9. *Id.* at 1735–36.
10. Red Lion Broadcasting Co. v. FCC, 395 U.S. 367, 375 (1969) (upholding the fairness doctrine based on broadcast spectrum scarcity).
11. *See* LORI ANDREWS, I KNOW WHO YOU ARE AND I SAW WHAT YOU DID: SOCIAL NETWORKS AND THE DEATH OF PRIVACY 121–35 (New York: Free Press, 2011).
12. DANIELLE KEATS CITRON, HATE CRIMES IN CYBERSPACE 53 (Cambridge, MA: Harvard University Press, 2014).
13. *Id.* at 35–55.
14. Marlisse Silver Sweeney, *What the Law Can (and Can't) Do About Online Harassment*, ATLANTIC (Nov. 12, 2014), https://www.theatlantic.com/technology/archive/2014/11/what-the-law-can-and-cant-do-about-online-harassment/382638.
15. *When Egypt Turned off the Internet*, AL JAZEERA (Jan. 28, 2011), https://www.aljazeera.com/news/middleeast/2011/01/2011128796164380.html.
16. Ashcroft v. ACLU, 542 U.S. 656, 667 (2004).
17. *See* ROBERT S. MUELLER, THE MUELLER REPORT: REPORT ON THE INVESTIGATION INTO RUSSIAN INTERFERENCE IN THE 2016 PRESIDENTIAL ELECTION (Washington, DC: WSBLD, 2019).
18. *See* Jones v. Dirty World Entm't Recordings LLC, 755 F.3d 398 (6th Cir. 2014) (holding that Section 230 precluded a defamation claim against a website that solicited defamatory material). *See also* Jeff Kosseff, *A User's Guide to Section 230, and a Legislator's Guide to Amending It (Or Not)*, 37 BERKELEY TECH. L.J., (forthcoming 2022), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3905347.
19. 47 U.S.C. § 230(c)(1) ("No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.").
20. 47 U.S.C. § 230(c)(2) ("No provider or user of an interactive computer service shall be held liable on account of—(A) any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected.").
21. Zeran v. Am. Online, Inc., 129 F.3d 327, 330 (4th Cir. 1997).
22. JEFF KOSSEFF, THE TWENTY-SIX WORDS THAT CREATED THE INTERNET (Ithaca, NY: Cornell University Press, 2019) (referring to the first of Section 230's two immunities: § 230(c)(1)).
23. *See* Danielle Keats Citron & Mary Anne Franks, *The Internet as a Speech Machine and Other Myths Confounding Section 230 Reform*, 2020 U. CHI. LEGAL F. 45, 46–47 (2020) ("Today, politicians across the ideological spectrum are raising concerns about the leeway provided to content platforms under Section 230. Conservatives claim that Section 230 gives tech companies a license to silence speech based on viewpoint. Liberals criticize Section 230 for giving platforms the freedom to profit from harmful speech and conduct.").
24. Jon Kyl, *Why Conservatives Don't Trust Facebook*, WALL ST. J. (Aug. 20, 2019), https://www.wsj.com/articles/why-conservatives-dont-trust-facebook-11566309603. *But see* Rebecca Heilweil, *Right-Wing Media Thrives on Facebook. Whether It Rules Is More Complicated*, VOX (Sept. 9, 2020), https://www.vox.com/recode/21419328/facebook-conservative-bias-right-wing-crowdtangle-election ("Consistently, [*New York Times* columnist Kevin] Roose found, conservative pages were beating out liberals' in making it into the day's Top 10 Facebook posts with links in the United States, based on engagement, like the number of reactions, comments, and shares the posts receive. That seems to provide evidence against the notion that Facebook censors conservatives, a complaint often trotted out by Republicans despite lacking any significant data to support their claims of systemic bias.").
25. *See, e.g.*, Protecting Constitutional Rights from Online Platform Censorship Act, H.R. 83 (117th Cong.), https://www.congress.gov/bill/117th-congress/house-bill/83/text; 21st Century Foundation for the Right to Express and Engage (FREE) in Speech Act, S. 1384 (117th Cong.),

https://www.congress.gov/bill/117th-congress/senate-bill/1384?s=9&r=2; Platform Accountability and Consumer Transparency (PACT) Act, S. 4066 (116th Cong.) (2020), https://www.congress.gov/bill/116th-congress/senate-bill/4066/text?format=xml; Safeguarding Against Fraud, Exploitation, Threats, Extremism, and Consumer Harms (SAFE TECH) Act (117th Cong.), https://www.congress.gov/bill/117th-congress/senate-bill/299/text?r=5. *See also* Kiran Jeevanje et al., *All the Ways Congress Wants to Change Section 230*, SLATE (Mar. 23, 2021), https://slate.com/technology/2021/03/section-230-reform-legislative-tracker.html.

26. Taylor Hatmaker, *Trump Vetoes Major Defense Bill, Citing Section 230*, TECH CRUNCH (Dec. 23, 2020), https://techcrunch.com/2020/12/23/trump-ndaa-veto-section-230/.
27. Sofia Andrade, *Florida's New Pro-Disney, Anti-Facebook and Twitter Law*, SLATE (May 25, 2021), https://slate.com/technology/2021/05/florida-stop-social-media-censorship-act-disney.html. *See, e.g.*, Stop the Censorship Act, H.R. 4027 (116th Cong.) (2019–2020), https://www.congress.gov/bill/116th-congress/house-bill/4027; Stop Social Media Censorship Act, H.B. 482, https://house.mo.gov/billtracking/bills211/sumpdf/HB0482I.pdf.
28. For a persuasive argument in favor of criminalizing revenge porn and an analysis of the constitutional and practical issues that might arise from doing so, *see* Danielle Keats Citron & Mary Anne Franks, *Criminalizing Revenge Porn*, 49 WAKE FOREST L. REV. 345 (2014).
29. For an excellent analysis of the deepfake problem, *see* Robert Chesney & Danielle Keats Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753 (2019).
30. NetChoice, LLC v. Moody, ___ F. Supp. 3d ___, 2021 WL 2690876, at *8 (N.D. Fla. June 30, 2021).
31. *Id.* at *24–28.
32. *See, e.g.*, Platform Censorship Act, H.R. 83 (117th Cong.); FREE Speech Act, S. 1384 (117th Cong.).
33. *See, e.g.*, United States v. Alvarez, 567 U.S. 709 (2012) (declaring unconstitutional federal law making it a crime to falsely claim receipt of military honors); R.A.V. v. City of St. Paul, 505 U.S. 377 (1992) (striking down city's hate speech ordinance).
34. *See* Manhattan Cmty. Access Corp. v. Halleck, 139 S. Ct. 1921, 1925 (2019); Prager Univ. v. Google LLC, 951 F.3d 991, 994 (9th Cir. 2020) (holding that editorial decisions made by YouTube are not subject to First Amendment scrutiny because it is a private company).
35. PACT Act, S. 4066 (116th Cong.).
36. Compare Jason Orestes, *Parler Is Not the Free Speech Utopia That Trump Allies Hope for*, WASH. EXAMINER (July 6, 2020), https://www.washingtonexaminer.com/opinion/parler-is-not-the-free-speech-utopia-that-trump-allies-hope-for (among other things, Parler prohibits lewd written descriptions of pornography).
37. Kosseff, *supra* note 18, at *34 ("If Congress changes the law to impose more potential liability for particular types of content, platforms likely will more aggressively moderate not only the content that is clearly illegal, but other user content that could even possibly fall within that category. Even if it is unclear whether the Section 230 exception would apply, or whether the platform would even face liability without Section 230 protection, the platform likely does not want to risk the cost of litigating a case on the merits.").
38. Pub. L. No. 105-304, 112 Stat. 2860 (Oct. 28, 1998).
39. Wendy Seltzer, *Free Speech Unmoored in Copyright's Safe Harbor: Chilling Effects of the DMCA on the First Amendment*, 24 HARV. J.L. & TECH. 171 (2010).
40. Mike Isaac, *Facebook Bans Anti-Vaccination Ads, Clamping Down Again*, N.Y. TIMES (Oct. 13, 2020), https://www.nytimes.com/2020/10/13/technology/facebook-bans-anti-vaccination-ads.html.
41. Elizabeth Dwoskin, *Facebook Says Post That Cast Doubt on COVID-19 Vaccine Was Most Popular on the Platform from January Through March*, WASH. POST (Aug. 21, 2021), https://www.washingtonpost.com/technology/2021/08/21/facebook-coronavirus-vaccine.
42. Elizabeth Dwoskin et al., *A Group of Moms on Facebook Built an Island of Good-Faith Vaccine Debate in a Sea of Misinformation*, WASH. POST. (Aug. 24, 2021), https://www.washingtonpost.com/technology/2021/08/23/facebook-vaccine-talk-group.

43. Jam Kotenko, *Facebook Reveals We Upload a Whopping 350 Million Photos to the Network Daily*, DIGITAL TRENDS (Sept. 18, 2013), https://www.digitaltrends.com/social-media/according-to-facebook-there-are-350-million-photos-uploaded-on-the-social-network-daily-and-thats-just-crazy.
44. *Id.*
45. Kosseff, *supra* note 18, at *34.
46. Don Reisinger, *Twitter Has Suspended 1.2 Million Terrorist Accounts Since 2015*, FORTUNE (Apr. 5, 2018), https://fortune.com/2018/04/05/twitter-terrorist-account-suspensions.
47. *See* evelyn douek, *Governing Online Speech: From "Posts-as-Trumps" to Proportionality and Probability*, 121 COLUM. L. REV. 759, 767 (2021).
48. Allow States and Victims to Fight Online Sex Trafficking Act (FOSTA), Pub. L. No. 115-164, 132 Stat. 1253 (2018).
49. Kosseff, *supra* note 18, at *34.
50. *See* Craigslist, *About: FOSTA*, https://www.craigslist.org/about/FOSTA.
51. *See* Kosseff, *supra* note 18, at *34 (explaining that the notion of what might constitute COVID misinformation changed markedly between 2020 and 2021).
52. *See infra*, section "The Problems of the Internet."
53. 376 U.S. 254 (1964).
54. *Id.* at 270.
55. *Id.* at 271.
56. *Id.* at 271–72.
57. 567 U.S. 709 (2012).
58. *Id.* at 715–16.
59. *Id.* at 715.
60. *Id.* at 725–26.
61. *Id.* at 718.
62. *Id.* at 719.
63. *See, e.g.*, Cent. Hudson Gas & Elec. Corp. v. Pub. Serv. Comm'n of N.Y., 447 U.S. 557, 566 (1980).
64. *Alvarez*, 567 U.S. at 720.
65. Hustler Magazine, Inc. v. Falwell, 485 U.S. 46, 52 (1988).
66. Brown v. Hartlage, 456 U.S. 45, 60–61 (1982).
67. Garrison v. Louisiana, 379 U.S. 64, 75 (1964).
68. 376 U.S. 254, 271–72 (1964) (quoting NAACP v. Button, 371 U.S. 415, 433 (1963)).
69. 567 U.S. at 723.
70. *See, e.g.*, Mueller, *supra* note 17, at 20.
71. *Id.*
72. Indictment, United States v. Internet Research Agency, LLC. No, 1:18-cr-00032-DLC (D.D.C. Feb. 16, 2018), 2018 WL 914777.
73. Sheera Frenkel & Katie Benner, *To Stir Discord in 2016, Russians Turned Most Often to Facebook*, N.Y. TIMES (Feb. 17, 2018), https://www.nytimes.com/2018/02/17/technology/indictment-russian-tech-facebook.html.
74. Scott Shane, *Russia Isn't the Only One Meddling in Elections. We Do It Too*, N.Y. TIMES (Feb. 17, 2018), https://www.nytimes.com/2018/02/17/sunday-review/russia-isnt-the-only-one-meddling-in-elections-we-do-it-too.html.
75. *Id.*
76. Raphael Satter et al., *Inside Story: How Russians Hacked the Democrats' Emails*, AP NEWS (Nov. 4, 2017), https://www.apnews.com/dea73efc01594839957c3c9a6c962b8a.
77. 435 U.S. 765, 767–68 (1978).
78. *Id.* at 776–77 (emphasis added).
79. 558 U.S. 310, 319 (2010).
80. *Id.* at 350, 394.
81. Pac. Gas & Elec. Co. v. Pub. Utils. Comm'n of Cal., 475 U.S. 1, 8 (1986) (plurality opinion).
82. Bluman v. F.C.C., 800 F. Supp. 2d 281, 284 (D.D.C. 2011), *aff'd*, 565 U.S. 1104 (2012).
83. *Id.* at 288.
84. 514 U.S. 334, 357 (1995).

85. *Id.* at 341–42.
86. *Id.* at 342.
87. 140 S. Ct. 2082 (2020).

Chapter 7

1. *See* Jason Murdock, *Facebook Is a Publisher and Must Be Responsible for Content, Says Founder of World's Biggest Advertising Company*, NEWSWEEK (July 21, 2020, 6:23 AM), https://www.newsweek.com/facebook-advertising-boycott-martin-sorrell-mark-zuckerberg-police-user-content-1519295 ("They all say they are tech companies. . . . They can't say they are digital engineers tightening nuts on digital pipes with digital spanners and are not responsible for digital content that flows through those pipes. They are.").
2. *See* ANDREW MARANTZ, ANTISOCIAL: ONLINE EXTREMISTS, TECHNO-UTOPIANS, AND THE HIJACKING OF THE AMERICAN CONVERSATION 48 (2020) ("Interviewed by *Time* in 2010, Facebook's founder, Mark Zuckerberg, said, 'The best stuff spreads, whether it's the best news article, or the best song, or the best product, or the best movie.' ").
3. *See* Sara Brown, *MIT Sloan Research About Social Media, Misinformation, and Elections*, MIT Sloan School of Management (Oct. 5, 2020), https://mitsloan.mit.edu/ideas-made-to-matter/mit-sloan-research-about-social-media-misinformation-and-elections.
4. *See* Amber M. Hamilton, *Silicon Valley Pretends That Algorithmic Bias Is Accidental. It's Not*, SLATE (July 7, 2021, 1:55 PM), https://slate.com/technology/2021/07/silicon-valley-algorithmic-bias-structural-racism.html.
5. *See* Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity*, 86 FORDHAM L. REV. 401 (2017) ("Omegle is not exactly a social media site for sexual predators, but it is fair to say that a social network designed for the benefit of the predator community would look a lot like Omegle.").
6. Elizabeth Dwoskin, *Facebook's Sandberg Deflected Blame for Capitol Riot, but New Evidence Shows How Platform Played Role*, WASH. POST (Jan. 13, 2021, 3:27 PM), https://www.washingtonpost.com/technology/2021/01/13/facebook-role-in-capitol-protest.
7. Sheera Frenkel & Davey Alba, *Misinformation Amplifiers Target Georgia Senate Races*, N.Y. TIMES (Dec. 22, 2020), https://www.nytimes.com/2020/12/22/technology/georgia-senate-runoff-misinformation.html.
8. Jeff Horwitz, *Facebook Algorithm Shows Gender Bias in Job Ads, Study Finds*, WALL ST. J. (Apr. 9, 2021, 1:12 PM), https://www.wsj.com/articles/facebook-shows-men-and-women-different-job-ads-study-finds-11617969600.
9. 47 U.S.C. § 230. Section 230 not only protects provider efforts to remove offending content, but also immunizes providers when they don't. Section 230(c)(2) protects "good faith" moderation efforts to "restrict access to material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable[.]" At the same time, Section 230(c)(1) shields providers that do *not* filter third-party content, providing that "no provider or user of interactive computer services shall be treated as the publisher or speaker of any information provided by an information content provider."
10. JEFF KOSSEFF, THE TWENTY-SIX WORDS THAT CREATED THE INTERNET 3 (Ithaca, NY: Cornell University Press, 2019).
11. Simon Kemp, *60% of the World Is Online—10 Big Takeaways on the State of the Internet in 2021*, NEXT WEB (Apr. 21, 2021, 8:06 AM), https://thenextweb.com/news/60-of-the-world-is-online-10-big-takeaways-on-the-state-of-the-internet-in-2021.
12. Elected representatives from all over the political spectrum agree on this point, from Senator Ted Cruz (R-Tex.) and Senator Elizabeth Warren (D-Mass.) to former president Donald Trump and President Joe Biden.
13. *Americans Going Online . . . Explosive Growth, Uncertain Destinations*, PEW RES. CTR. (Oct. 16, 1995), https://www.pewresearch.org/politics/1995/10/16/americans-going-online-explosive-growth-uncertain-destinations.
14. KOSSEFF, *supra note* 10, at 3.
15. *Id.*

16. Peter H. Lewis, *PERSONAL COMPUTERS; The Compuserve Edge: Delicate Data Balance*, N.Y. TIMES (Nov. 29, 1994), https://www.nytimes.com/1994/11/29/science/personal-computers-the-compuserve-edge-delicate-data-balance.html.
17. Alina Selyukh, *The Big Internet Brands of the '90s—Where Are They Now?*, NPR (July 25, 2016, 4:41 PM), https://www.npr.org/sections/alltechconsidered/2016/07/25/487097344/the-big-internet-brands-of-the-90s-where-are-they-now.
18. *Id.*
19. In 1995, digital advertising totaled $55 million. CHRIS TREADAWAY & MARI SMITH, FACEBOOK MARKETING: AN HOUR A DAY 3–4 (Indianapolis, IN: Wiley, 2010). As of 2021, digital advertising was expected to hit $657 billion. Georg Szalai, *Global, U.S. Ad Spending to Hit Records on COVID Rebound: Forecast*, HOLLYWOOD REP. (June 13, 2021, 8:00 AM), https://www.hollywoodreporter.com/business/business-news/advertising-record-2021-covid-rebound-1234966706.
20. *See Disinformation Online and a Country in Crisis: Hearing Before the H. Comm. on Energy and Comm.* (116th Cong.) (2021) (statement of Neil Fried at 2) ("'Primary publishers'—those responsible for approval of another's content, such as publishers of books, newspapers, and magazines—can be held culpable for libel if the content contains falsehoods that harm someone's reputation. . . . 'Distributors' (also called 'secondary publishers')—those that do not have editorial discretion over content but that are responsible for its delivery, such as libraries, bookstores, and newsvendors . . . can only be held culpable if they knew, or should have known, about the false, defamatory statements in the content they delivered.").
21. *Id.*
22. *Id.*
23. Stratton Oakmont, Inc. v. Prodigy Servs. Co., 1995 WL 323710 (N.Y. Sup. Ct. May 24, 1995); Neil Fried, *The Myth of Internet Exceptionalism: Bringing Section 230 into the Real World*, 5(2) AM. AFF. J. 179 (Summer 2021).
24. KOSSEFF, *supra* note 10, at 2.
25. Peter H. Lewis, *COMPANY NEWS; Prodigy Sees Quick Growth from Internet Web Service*, N.Y. TIMES (Jan. 31, 1995), https://www.nytimes.com/1995/01/31/business/company-news-prodigy-sees-quick-growth-from-internet-web-service.html.
26. Christopher Cox, *The Origins and Original Intent of Section 230 of the Communications Decency Act*, U. RICHMOND J.L. & TECH. BLOG (Aug. 27, 2020), https://jolt.richmond.edu/2020/08/27/the-origins-and-original-intent-of-section-230-of-the-communications-decency-act.
27. *Id.*
28. 141 CONG. REC. No. 129 H8470.
29. Jonathan A. Obar & Steven S. Wildman, *Social Media Definition and the Governance Challenge: An Introduction to the Special Issue*, 39(9) TELECOMM. POL'Y 745 (July 22, 2015).
30. The other two, Google and Wikipedia, depend on content generated from third-party users. *Id.*
31. Elisa Shearer, *More Than Eight-in-Ten Americans Get News from Digital Devices*, PEW RES. CTR. (Jan. 12, 2021), https://www.pewresearch.org/fact-tank/2021/01/12/more-than-eight-in-ten-americans-get-news-from-digital-devices.
32. *Stratton Oakmont*, 1995 WL 323710 (N.Y. Sup. Ct. May 24, 1995).
33. Maddy Osman, *Wild and Interesting Facebook Statistics and Facts (2021)*, KINSTA BLOG (Jan. 3, 2021), https://kinsta.com/blog/facebook-statistics.
34. As Justice Thomas recently noted, "[M]any courts have construed the law broadly to confer sweeping immunity on some of the largest companies in the world." Malwarebytes, Inc. v. Enigma Software Grp. USA, LLC, 141 S. Ct. 13, 208 (2020) (Thomas, J., concurring in denial of certiorari).
35. *Id.* at 329.
36. *Id.*
37. *Id.* at 328.
38. Eric Goldman, *Revenge Porn Is Bad, But It's Not GoDaddy's Fault*, FORBES (Apr. 10, 2014, 3:11 PM), https://www.forbes.com/sites/ericgoldman/2014/04/10/revenge-porn-is-bad-but-its-not-godaddys-fault/?sh=73952f313f6e.

39. *See* Neil Fried, *When It Comes to Section 230, Something Must Change*, DIGITALFRONTIERS ADVOCACY (Sept. 3, 2020), https://digitalfrontiersadvocacy.com/blogs-and-op-eds/f/when-it-comes-to-section-230-something-must-change.
40. Mark MacCarthy, *Back to the Future for Section 230 Reform*, BROOKINGS INSTITUTION (Mar. 17, 2021), https://www.brookings.edu/blog/techtank/2021/03/17/back-to-the-future-for-section-230-reform.
41. Max Fisher & Amanda Taub, *On YouTube's Digital Playground, an Open Gate for Pedophiles*, N.Y. TIMES (June 3, 2019), https://www.nytimes.com/2019/06/03/world/americas/youtube-pedophiles.html.
42. *Id.*
43. Justin Scheck et al., *Facebook Employees Flag Drug Cartels and Human Traffickers. The Company's Response Is Weak, Documents Show*, WALL ST. J. (Sept. 16, 2021), https://www.wsj.com/articles/facebook-drug-cartels-human-traffickers-response-is-weak-documents-11631812953.
44. Heather Somerville, *Airbnb's Section 230 Use Underscores Law's Reach Beyond Facebook*, WALL ST. J. (Jan. 10, 2021, 12:00 PM), https://www.wsj.com/articles/airbnbs-section-230-use-underscores-laws-reach-beyond-facebook-11610298001.
45. Peter Dizikes, *Study: On Twitter, False News Travels Faster Than True Stories*, MIT NEWS OFFICE (Mar. 8, 2018), https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308.
46. *Id.*
47. Soroush Vosoughi et al., *The Spread of True and False News Online*, 359 SCI. 1146 (2018).
48. Stephen Fowler, *How Pro-Trump Local News Sites Keep Pushing 2020 Election Misinformation*, NPR (July 2, 2021, 5:00 AM), https://www.npr.org/2021/07/02/1011916687/how-pro-trump-local-news-sites-keep-pushing-2020-election-misinformation.
49. Frenkel & Alba, *supra* note 7 (updated Jan. 5, 2021).
50. Ana Santos Rutschman, *Facebook's Latest Attempt to Address Vaccine Misinformation—And Why It's Not Enough*, HEALTH AFF. BLOG (Nov. 5, 2020), https://www.healthaffairs.org/do/10.1377/hblog20201029.23107.
51. Philip Ball, *Anti-Vaccine Movement Could Undermine Efforts to End Coronavirus Pandemic, Researchers Warn*, NATURE (May 13, 2020), https://www.nature.com/articles/d41586-020-01423-4.
52. David Lazer et al., *People Are More Anti-Vaccine If They Get Their COVID News from Facebook Than from Fox News, Data Shows*, WASH. POST (July 27, 2021), https://www.washingtonpost.com/politics/2021/07/27/people-are-more-anti-vaccine-if-they-get-their-covid-19-news-facebook-rather-than-fox-news-new-data-shows.
53. Scott Waldman, *Climate Denial Newspaper Flourishes on Facebook*, CLIMATEWIRE (Aug. 27, 2021), https://www.eenews.net/articles/climate-denial-newspaper-flourishes-on-facebook.
54. *Id.*
55. US Food and Drug Administration, *Why You Should Not Use Ivermectin to Treat or Prevent COVID-19* (Sept. 3, 2021), https://www.fda.gov/consumers/consumer-updates/why-you-should-not-use-ivermectin-treat-or-prevent-covid-19.
56. @US_FDA, Twitter (Aug. 21, 2021, 7:57 AM), https://twitter.com/us_fda/status/1429050070243192839.
57. Tim Marchman, *Facebook's Ivermectin Groups Are Unhinged and Out of Control*, VICE (Aug. 26, 2021), https://www.vice.com/en/article/bvzkjd/facebooks-ivermectin-groups-are-unhinged-and-out-of-control.
58. Letter from Senators Warren, Carper, Schatz, and Whitehouse to Mark Zuckerberg, Chairman and Chief Executive Officer, Facebook (July 15, 2020), https://www.warren.senate.gov/imo/media/doc/07.15.2020%20Letter%20from%20Sens.%20Warren,%20Carper,%20Whitehouse,%20&%20Schatz%20to%20Mr.%20Zuckerberg.pdf.
59. Kurt Wagner, *Facebook's New Target in the Misinformation War: Climate Lies*, BLOOMBERG GREEN (Apr. 13, 2021), https://www.bloomberg.com/news/articles/2021-04-13/will-facebook-stop-climate-myths-and-misinformation.

60. *Environmental Groups Call out Facebook Climate Change Disinformation,* FRIENDS OF THE EARTH (Mar. 24, 2021), https://foe.org/news/environmental-groups-call-out-facebook-climate-change-disinformation.
61. *Id.*
62. *Climate Change and Digital Advertising: Climate Science Disinformation in Facebook Advertising,* INFLUENCEMAP (Oct. 2020), https://influencemap.org/report/Climate-Change-and-Digital-Advertising-86222daed29c6f49ab2da76b0df15f76#.
63. Samuel Spies, *How Misinformation Spreads*, SOC. SCI. RES. COUNCIL (July 30, 2020), https://mediawell.ssrc.org/literature-reviews/how-misinformation-spreads/versions/1-0.
64. *Id.*
65. Yaël Eisenstat, *How to Hold Social Media Accountable for Undermining Democracy*, HARV. BUS. REV. (Jan. 11, 2021), https://hbr.org/2021/01/how-to-hold-social-media-accountable-for-undermining-democracy.
66. Elizabeth Dwoskin, *Misinformation on Facebook Got Six Times More Clicks Than Factual News During the 2020 Election, Study Says*, WASH. POST (Sept. 4, 2021), https://www.washingtonpost.com/technology/2021/09/03/facebook-misinformation-nyu-study.
67. *Id.*
68. Julia Carrie Wong, *Down the Rabbit Hole: How QAnon Conspiracies Thrive on Facebook*, GUARDIAN (June 25, 2020), https://www.theguardian.com/technology/2020/jun/25/qanon-facebook-conspiracy-theories-algorithm.
69. Keach Hagey & Jeff Horwitz, *Facebook Tried to Make Its Platform a Healthier Place. It Got Angrier Instead*, WALL ST. J. (Sept. 15, 2021), https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215.
70. *Id.*
71. *Id.*
72. Laura Edelson et al., *Far-Right News Sources on Facebook More Engaging,* MEDIUM, Mar. 3, 2021, https://medium.com/cybersecurity-for-democracy/far-right-news-sources-on-facebook-more-engaging-e04a01efae90.
73. Alex Jones, *From Memes to Infowars: How 75 Fascist Activists Were "Red-Pilled,"* BELLINGCAT (Oct. 11, 2018), https://www.bellingcat.com/news/americas/2018/10/11/memes-infowars-75-fascist-activists-red-pilled/.
74. Jonas Kaiser & Adrian Rauchfleisch, *Unite the Right? How YouTube's Recommendation Algorithm Connects the U.S. Far-Right*, MEDIUM (Apr. 11, 2018), https://medium.com/@MediaManipulation/unite-the-right-how-youtubes-recommendation-algorithm-connects-the-u-s-far-right-9f1387ccfabd.
75. Daniela Hernandez & Parmy Olson, *Isolation and Social Media Combine to Radicalize Violent Offenders*, WALL ST. J. (Aug. 5, 2019), https://www.wsj.com/articles/isolation-and-social-media-combine-to-radicalize-violent-offenders-11565041473.
76. Michael R. Gordon & Dustin Volz, *Russian Disinformation Campaign Aims to Undermine Confidence in Pfizer, Other Covid-19 Vaccines, U.S. Officials Say*, WALL ST. J. (Mar. 7, 2021), https://www.wsj.com/articles/russian-disinformation-campaign-aims-to-undermine-confidence-in-pfizer-other-covid-19-vaccines-u-s-officials-say-11615129200.
77. David Broniatowski, Amelia Jamison, SiHua Qi, Lulwah AlKulaib, Tao Chen, Adrian Benton, Sandra C. Quinn, & Mark Dredze, *Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate*, 108 AM. J. PUBLIC HEALTH, 1378–84 (2018), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6137759/.
78. Michael R. Gordon & Dustin Volz, *Russian Disinformation Campaign Aims to Undermine Confidence in Pfizer, Other Covid-19 Vaccines, U.S. Officials Say*, Wall St. J. (Mar. 7, 2021), https://www.wsj.com/articles/russian-disinformation-campaign-aims-to-undermine-confidence-in-pfizer-other-covid-19-vaccines-u-s-officials-say-11615129200.
79. Mark Kramer, *Lessons from Operation "Denver," the KGB's Massive AIDS Disinformation Campaign*, MIT PRESS READER (May 26, 2020), https://thereader.mitpress.mit.edu/operation-denver-kgb-aids-disinformation-campaign.
80. *See* Erika Kinetz, *Anatomy of a Conspiracy: With COVID, China Took Leading Role*, ASSOCIATED PRESS (Feb. 15, 2021), https://apnews.com/article/pandemics-beijing-only-on-ap-epidemics-media-122b73e134b780919cc1808f3f6f16e8.

81. Todd Spangler, *Reddit Finally Bans Hate Speech, Removes 2,000 Racist and Violent Forums Including The_Donald*, VARIETY (June 29, 2020, 10:00 AM), https://variety.com/2020/digital/news/reddit-bans-hate-speech-groups-removes-2000-subreddits-donald-trump-1234692898.
82. INSTAGRAM, *Combatting Misinformation on Instagram* (Dec. 16, 2019), https://about.instagram.com/blog/announcements/combatting-misinformation-on-instagram.
83. Associated Press, *People Who Search for White Nationalist Info on Facebook Will Soon Be Redirected to "Life After Hate,"* MARKETWATCH (Mar. 28, 2019, 8:11 AM), https://www.marketwatch.com/story/soon-people-who-search-for-white-nationalist-terms-on-facebook-will-be-redirected-to-life-after-hate-2019-03-27.
84. April Glaser, *YouTube Is Adding Fact-Check Links for Videos on Topics That Inspire Conspiracy Theories*, SLATE (Aug. 14, 2018, 10:56 AM), https://slate.com/technology/2018/08/youtube-is-adding-fact-check-links-from-wikipedia-and-encyclopedia-britannica-for-videos-on-topics-that-inspire-conspiracy-theories.html.
85. It's easy to see how social media platforms, Facing the prospect of unstructured liability, could be bullied by lawsuits into taking unpopular and counterproductive positions. Industry response to the few previous Section 230 reforms, like FOSTA/SESTA (Allow States and Victims to Fight Online Sex Trafficking Act and Stop Enabling Sex Traffickers Act), is instructive. The relatively small nonprofit site Craigslist entirely removed its "personal connections" page to eliminate the chance of illegal sex trafficking content being posted, even though the vast majority of the page's content was innocuous. While these trade-offs could be justified in cases of criminal conduct, that balance looks different in areas like defamation.
86. *Section 230 Protections*, ELECTRONIC FRONTIER FOUNDATION, https://www.eff.org/issues/bloggers/legal/liability/230.
87. Cheyenne Ubiera, *Fox Runs All-Caps Disclaimer over Trump's Lies About Stolen Election at CPAC*, DAILY BEAST (July 11, 2021, 8:20 PM), https://www.thedailybeast.com/fox-runs-all-caps-disclaimer-over-trumps-lies-about-stolen-election-at-cpac.
88. 150 A.3d 1213 (D.C. 2016).
89. *See, e.g.*, Press Release, Fed. Trade Comm'n, FTC, Nevada Obtain Order Permanently Shutting Down Revenge Porn Site MyEx (June 22, 2018), https://www.ftc.gov/news-events/press-releases/2018/06/ftc-nevada-obtain-order-permanently-shutting-down-revenge-porn. Statements couched as opinions may not be actionable as defamation or harassment.
90. *See, e.g.*, People v. Marquan M., 19 N.E.3d 480 (N.Y. 2014).
91. *See* Alison V. King, *Constitutionality of Cyberbullying Laws: Keeping the Online Playground Safe for Both Teens and Free Speech*, 63 VANDERBILT L. REV. 845 (2019).
92. *See, e.g.*, *id*. at 63–68.
93. Laurence Tribe, *First Amendment Fantasies in the Social Media Debate*, THE HILL (May 11, 2021, 8:30 AM), https://thehill.com/opinion/technology/552735-laurence-tribe-first-amendment-fantasies-in-the-social-media-debate.
94. In intentional interference with contractual relations, tortious interference, and other possible theories.
95. Caleb Ecarma, *"As Stupid as You'd Think": Trump's Social Media Lawsuit Looks Like a Mess*, VANITY FAIR (July 7, 2021), https://www.vanityfair.com/news/2021/07/trump-social-media-lawsuit.
96. Jake Lahut, *Trump Announces Lawsuits Against Facebook and Twitter, Immediately Starts Fundraising off It*, BUSINESS INSIDER (July 7, 2021), https://www.businessinsider.com/trump-facebook-twitter-lawsuit-fundraising-immediately-2021-7.
97. *Id.*
98. *See, e.g.*, Ted Cruz, *supra* note 12.
99. *See, e.g.*, Warren, *supra* note 12.
100. B. Travis Brown, *Changing the Rules of the Internet? The Uncertain Future of Platform and Intermediary Liability*, JD SUPRA (Apr. 29, 2021), https://www.jdsupra.com/legalnews/changing-the-rules-of-the-internet-the-3335391, as required by Senators Brian Schatz and John Thune's Platform Accountability and Consumer Transparency (PACT) Act, S. 4066 (116th Cong.) (2020), https://www.congress.gov/bill/116th-congress/senate-bill/4066/text?format=xml.

101. MacCarthy, *supra* note 40.
102. *Id.*
103. *See* Zack Colman & Karl Mathiesen, *Climate Scientists Take Swipe at Exxon Mobil, Industry in Leaked Report*, POLITICO (July 2, 2021, 10:00 AM), https://www.politico.com/news/2021/07/02/climate-scientists-exxon-mobile-report-497805 ("Details from the leaked report began emerging last week, days before a British television broadcaster aired a video showing a top Exxon Mobil lobbyist admitting the company worked with 'shadow groups' that engaged in disinformation campaigns around climate science.").
104. *See, e.g.*, Anna Massoglia, *"Dark Money" Groups Find New Ways to Hide Donors in 2020 Election*, OPEN SECRETS (Oct. 30, 2020, 4:26 PM), https://www.opensecrets.org/news/2020/10/dark-money-2020-new-ways-to-hide-donors.
105. Harper Neidig, *Franken Blasts Facebook for Accepting Rubles for U.S. Election Ads*, THE HILL (Oct. 31, 2017, 5:40 PM), https://thehill.com/policy/technology/358102-franken-blasts-facebook-for-accepting-rubles-for-us-election-ads.
106. *See* Anna Massoglia, *Trump 2020 Campaign Ad Payments Hidden by Layers of Shell Companies*, OPEN SECRETS (June 13, 2019, 12:42 PM), https://www.opensecrets.org/news/2019/06/trump-2020-campaign-ad-payments-hidden-by-layers-of-shell-companies.
107. 2021 National Defense Authorization Act/Corporate Transparency Act, H.R.6395 (116th Cong.) (2019–2020), Pub. L. No. 116-283 (Jan. 1, 2021).
108. DISCLOSE (Democracy Is Strengthened by Casting Light on Spending in Elections) Act of 2017, S.1585 (115th Cong.) (2017–18).
109. *See, e.g.*, Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1167–68 (9th Cir. 2008) (*en banc*) ("We believe that both the immunity for passive conduits and the exception for co-developers must be given their proper scope and, to that end, we interpret the term 'development' as referring not merely to augmenting the content generally, but to materially contributing to its alleged unlawfulness. In other words, a website helps to develop unlawful content, and thus falls within the exception to section 230, if it contributes materially to the alleged illegality of the conduct.").
110. *See, e.g.*, FTC v. Accusearch, Inc., 570 F.3d 1187, 1199 (10th Cir. 2009).
111. Horwitz, *supra* note 8.
112. *Id.*

Chapter 8

1. Esther Pan, *Iraq: Justifying the War*, FOREIGN AFF. (Council on Foreign Relations, Feb. 2, 2005, 3:27 PM), https://www.cfr.org/backgrounder/iraq-justifying-war.
2. 116 Stat. 1498, Pub. L. No. 107-243 (Oct. 16, 2002), https://uscode.house.gov/statutes/pl/107/243.pdf.
3. Judith Miller, *Aftereffects: Prohibited Weapons; Illicit Arms Kept Till Eve of War, an Iraqi Scientist Is Said to Assert*, N.Y. TIMES (Apr. 21, 2003), https://www.nytimes.com/2003/04/21/world/aftereffects-prohibited-weapons-illicit-arms-kept-till-eve-war-iraqi-scientist.html.
4. David Remnick, *Making a Case*, NEW YORKER (Feb. 3, 2003), https://www.newyorker.com/magazine/2003/02/03/making-a-case.
5. Jim Naureckas, *MSNBC's Racism Is OK, Peace Activism Is Not*, FAIR (Apr. 1, 2003), https://fair.org/extra/msnbcs-racism-is-ok-peace-activism-is-not.
6. Steve Rendall & Tara Broughel, *Amplifying Officials, Squelching Dissent*, FAIR (May 1, 2003), https://web.archive.org/web/20121209040307/http:/fair.org/extra-online-articles/amplifying-officials,-squelching-dissent.
7. Frank Newport, *Seventy-Two Percent of Americans Support War Against Iraq*, GALLUP (Mar. 24, 2003), https://news.gallup.com/poll/8038/seventytwo-percent-americans-support-war-against-iraq.aspx.
8. Dmitriy Khavin et al., *Day of Rage: How Trump Supporters Took the U.S. Capitol*, N.Y. TIMES (June 30, 2021), https://www.nytimes.com/video/us/politics/100000007606996/capitol-riot-trump-supporters.html.
9. Center for an Informed Public, Digital Forensic Research Lab, Graphika, & Stanford Internet Observatory, *The Long Fuse: Misinformation and the 2020 Election* (2021), Stanford

Digital Repository: Election Integrity Partnership. v.1.3.0, https://purl.stanford.edu/tr171zs0069; https://www.eipartnership.net/report.

10. Sheera Frenkel, *The Rise and Fall of the "Stop the Steal" Facebook Group*, N.Y. TIMES (Nov. 5, 2020), https://www.nytimes.com/2020/11/05/technology/stop-the-steal-facebook-group.html; Lena V. Groeger et al., *What Parler Saw During the Attack on the Capitol*, PARLER (Jan. 17, 2021), https://projects.propublica.org/parler-capitol-videos.
11. Renée DiResta, *It's Not Misinformation. It's Amplified Propaganda*, ATLANTIC, Oct. 9 2021, theatlantic.com/ideas/archive/2021/10/disinformation-propaganda-amplification-ampliganda/620334/.
12. Yochai Benkler, *How the "Propaganda Feedback Loop" of Right-Wing Media Keeps More Than a Quarter of Americans Siloed*, Berkman Klein Center (Nov. 7, 2018), https://cyber.harvard.edu/story/2018-11/how-propaganda-feedback-loop-right-wing-media-keeps-more-quarter-americans-siloed.
13. Catherine Kim, *Poll: 70 Percent of Republicans Don't Think the Election Was Free and Fair*, POLITICO (Nov. 9, 2020 5:00 PM), https://www.politico.com/news/2020/11/09/republicans-free-fair-elections-435488.
14. Press Release, Reuters/Ipsos, *Trump's Coattails* (Apr. 2, 2021), https://www.ipsos.com/sites/default/files/ct/news/documents/2021-04/topline_write_up_reuters_ipsos_trump_coattails_poll_-_april_02_2021.pdf.
15. Ryan Goodman & Justin Hendrix, *January 6 Clearinghouse: Congressional Hearings, Government Documents, Court Cases, Academic Research*, JUST SECURITY (Oct. 26, 2021), https://www.justsecurity.org/77022/january-6-clearinghouse.
16. Kate Starbird et al., *Disinformaton as Collaborative Work: Surfacing the Participatory Nature of Strategic Information Operations*, 3 PROCEEDINGS OF THE ACM ON HUM.-COMPUT. INTERACTION 1 (Nov. 2019); SINAN ARAL, THE HYPE MACHINE: HOW SOCIAL MEDIA DISRUPTS OUR ELECTIONS, OUR ECONOMY, AND OUR HEALTH—AND HOW WE MUST ADAPT (New York: Currency, 2020).
17. José van Dijck & Thomas Poell, *Social Media and the Transformation of Public Space*, SOC. MEDIA + SOC'Y (July–Dec. 2015), https://journals.sagepub.com/doi/full/10.1177/2056305115622482; Renée DiResta, *Mediating Consent*, RIBBONFARM: CONSTRUCTIONS IN MAGICAL THINKING (Dec. 17, 2019), https://www.ribbonfarm.com/2019/12/17/mediating-consent.
18. WALTER LIPPMANN, PUBLIC OPINION (New York: Macmillan, 1922).
19. EDWARD BERNAYS, CRYSTALLIZING PUBLIC OPINION (New Hall Press, 1923); EDWARD BERNAYS, PROPAGANDA (New York: H. Liveright, 1928).
20. MARSHALL MCLUHAN, UNDERSTANDING MEDIA: THE EXTENSIONS OF MAN (London: Taylor & Francis, 1964).
21. JACQUES ELLUL, PROPAGANDA: THE FORMATION OF MEN'S ATTITUDES (Konrad Kellen, trans., Jean Lerner, trans., New York: Vintage, 1973).
22. DANIEL J. BOORSTIN, THE IMAGE: A GUIDE TO PSEUDO-EVENTS IN AMERICA (New York: Vintage, 1992).
23. NOAM CHOMSKY, FOR REASONS OF STATE (New York: New Press, 2003).
24. Henry Farrell & Daniel W. Drezner, *The Power and Politics of Blogs*, 134 PUBLIC CHOICE 15–30 (2008), https://doi.org/10.1007/s11127-007-9198-1.
25. Douglas Quenqua, *Blogs Falling in an Empty Forest*, N.Y. TIMES (June 5, 2009), https://www.nytimes.com/2009/06/07/fashion/07blogs.html.
26. Zeynep Tufekci, *It's the (Democracy-Poisoning) Golden Age of Free Speech*, WIRED (Jan. 16, 2018, 6:00 AM), https://www.wired.com/story/free-speech-issue-tech-turmoil-new-censorship.
27. Elisa Shearer, *More Than Eight-in-Ten Americans Get News from Digital Devices*, PEW RES. CTR. (Jan. 12, 2021), https://www.pewresearch.org/fact-tank/2021/01/12/more-than-eight-in-ten-americans-get-news-from-digital-devices.
28. Sheera Frenkel et al., *How the "Plandemic" Movie and Its Falsehoods Spread Widely Online*, N.Y. TIMES (May 20, 2020), https://www.nytimes.com/2020/05/20/technology/plandemic-movie-youtube-facebook-coronavirus.html.

29. Daniel Funke, *How the Wayfair Child Sex-Trafficking Conspiracy Theory Went Viral*, POLITIFACT (July 15, 2020), https://www.politifact.com/article/2020/jul/15/how-wayfair-child-sex-trafficking-conspiracy-theor.
30. Tim Lister & Clare Sebastian, *Stoking Islamophobia and Secession in Texas—From an Office in Russia*, CNN (Oct. 6, 2017, 5:00 PM), https://www.cnn.com/2017/10/05/politics/heart-of-texas-russia-event/index.html.
31. Charles Bethea, *Confessions of a Trump Troll*, NEW YORKER (Aug. 31, 2020), https://www.newyorker.com/magazine/killing-time-dept.
32. Renée DiResta, *Online Conspiracy Groups Are a Lot Like Cults*, WIRED (Nov. 13, 2018), https://www.wired.com/story/online-conspiracy-groups-qanon-cults.
33. Alicia Wanless & Michael Berk, *Participatory Propaganda: The Engagement of Audiences in the Spread of Persuasive Communications.* Paper presented at the Conference on Social Media & Social Order, Culture Conflict 2.0, Oslo Project, Nov. 29, 2017, https://www.researchgate.net/publication/329281610_Participatory_Propaganda_The_Engagement_of_Audiences_in_the_Spread_of_Persuasive_Communications.
34. Katherine Schaeffer, *A Look at the Americans Who Believe There Is Some Truth to the Conspiracy Theory That COVID-19 Was Planned*, PEW RES. CTR. (July 24, 2020), https://www.pewresearch.org/fact-tank/2020/07/24/a-look-at-the-americans-who-believe-there-is-some-truth-to-the-conspiracy-theory-that-covid-19-was-planned.
35. Matthew A. Baum et al., *The State of the Nation: A 50-State COVID-19 Survey: Report #13: Public Trust in Institutions and Vaccine Acceptance v.1* (COVID-19 Consortium for Understanding the Public's Policy Preferences Across States, Sept. 15, 2020), https://osf.io/qzsxn/.
36. Bence Bago et al., *Fake News, Fast and Slow: Deliberation Reduces Belief in False (but Not True) News Headlines*, 149(8) J. EXP. PSYCHOL.: GEN. 1608 (2020); Gordon Pennycook & David G. Rand, *Lazy, Not Biased: Susceptibility to Partisan Fake News Is Better Explained by Lack of Reasoning Than by Motivated Reasoning*, 188 COGNITION 39–50 (2018), https://doi:10.1016/j.cognition.2018.06.011; Gordon Pennycook & David G. Rand, *Who Falls for Fake News? The Roles of Bullshit Receptivity, Overclaiming, Familiarity, and Analytic Thinking*, 88(2) J. OF PERSONALITY 185 (2019).
37. Guardian Staff and Agencies, *Most Republicans Still Believe 2020 Election Was Stolen from Trump—Poll*, GUARDIAN (May 24, 2021), https://www.theguardian.com/us-news/2021/may/24/republicans-2020-election-poll-trump-biden.
38. Nathaniel Persily, *American Elections Need Help. Here's How to Make Them Better*, WASH. POST (Jan. 22, 2014), https://www.washingtonpost.com/news/monkey-cage/wp/2014/01/22/american-elections-need-help-heres-how-to-make-them-better.
39. Jeff Horwitz & Deepa Seetharaman, *Facebook Executives Shut Down Efforts to Make the Site Less Divisive*, WALL ST. J. (May 26, 2020, 11:38 AM), https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499.
40. Renée DiResta & Tobias Rose-Stockwell, *How to Stop Misinformation Before It Gets Shared*, WIRED (Mar. 26, 2021), https://www.wired.com/story/how-to-stop-misinformation-before-it-gets-shared.
41. Ellen P. Goodman, *Digital Information Fidelity and Friction: Crafting a Systems-Level Approach to Transparency*, KNIGHT FIRST AMENDMENT INST. AT COLUMBIA UNIV., BLOG (Feb. 26, 2020), https://knightcolumbia.org/content/digital-fidelity-and-friction.
42. Renée DiResta, *There Are Bots. Look Around*, RIBBONFARM: CONSTRUCTIONS IN MAGICAL THINKING (May 23, 2017), https://www.ribbonfarm.com/2017/05/23/there-are-bots-look-around.
43. *The Long Fuse*, *supra* note 9.
44. Mark R. Warner, *White Paper: Potential Policy Proposals for Social Media and Technology Firms* (2018), https://graphics.axios.com/pdf/PlatformPolicyPaper.pdf.
45. Lee Rainie & Janna Anderson, *Theme 7: The Need Grows for Algorithmic Literacy, Transparency and Oversight,* PEW RES. CTR. (Feb. 8, 2017), https://www.pewresearch.org/internet/2017/02/08/theme-7-the-need-grows-for-algorithmic-literacy-transparency-and-oversight.

46. Joëlle Swart, *Experiencing Algorithms: How Young People Understand, Feel About, and Engage With Algorithmic News Selection on Social Media*, SOC. MEDIA + SOC'Y (Apr. 12, 2021), https://doi.org/10.1177/20563051211008828.
47. Ethan Zuckerman, *Who Filters Your News? Why We Built gobo.social,* BLOG (Nov. 16, 2017), https://ethanzuckerman.com/2017/11/16/who-filters-your-news-why-we-built-gobo-social.
48. Goodman, *supra* note 41.
49. Prosocial Design Network, https://www.prosocialdesign.org.
50. Twitter, *COVID-19 Misleading Information Policy*, https://help.twitter.com/en/rules-and-policies/medical-misinformation-zolicy.
51. Renée DiResta, *The Conspiracies Are Coming From Inside the House*, ATLANTIC (Mar. 10, 2020), https://www.theatlantic.com/ideas/archive/2020/03/internet-conspiracies-are-coming-inside-country/607645.
52. CHRIS BAIL, BREAKING THE SOCIAL MEDIA PRISM: HOW TO MAKE OUR PLATFORMS LESS POLARIZING 10 (Princeton, NJ: Princeton University Press, 2021).
53. *See* Jonathan Haidt & Tobias Rose-Stockwell, *The Dark Psychology of Social Networks*, ATLANTIC (Dec. 15, 2019), https://www.theatlantic.com/magazine/archive/2019/12/social-media-democracy/600763.
54. News Literacy Project, https://newslit.org.
55. Joan Donovan & Danah Boyd, *Stop the Presses? Moving From Strategic Silence to Strategic Amplification in a Networked Media Ecosystem*, AM. BEHAV. SCIENTIST (Sept. 29, 2019), https://doi.org/10.1177/0002764219878229.
56. BAIL, *supra* note 52, at 101; More In Common, *The Perception Gap*, https://perceptiongap.us.
57. Stanford Center for Deliberative Democracy, *America in One Room*, HELENA PROJECT, Sept. 19–22, 2019, Dallas, Texas, https://cdd.stanford.edu/2019/america-in-one-room.

Chapter 9

1. Twitter Inc., *Permanent Suspension of @realDonaldTrump*, TWITTER BLOG (Jan. 8, 2021), https://blog.twitter.com/en_us/topics/company/2020/suspension.html.
2. *Id.*
3. Twitter Inc., *World Leaders on Twitter: Principles & Approach*, TWITTER BLOG (Oct. 15, 2019), https://blog.twitter.com/en_us/topics/company/2019/worldleaders2019.html.
4. *Id.*
5. Twitter Inc., *supra* note 1.
6. Mark Zuckerberg, *Post Announcing Indefinite Suspension of Trump's Accounts*, FACEBOOK, https://www.facebook.com/zuck/posts/10112681480907401 (last visited Jan. 24, 2021).
7. Nick Clegg, *Referring Former President Trump's Suspension from Facebook to the Oversight Board*, ABOUT FACEBOOK (Jan. 21, 2021), https://about.fb.com/news/2021/01/referring-trump-suspension-to-oversight-board/; evelyn douek, *Facebook Has Referred Trump's Suspension to Its Oversight Board. Now What?*, LAWFARE, https://www.lawfareblog.com/facebook-has-referred-trumps-suspension-its-oversight-board-now-what (last visited Jan. 24, 2021).
8. evelyn douek, *Facebook's "Oversight Board:" Move Fast with Stable Infrastructure and Humility*, 21 N.C. J. L. & TECH. 1 (2019) [hereinafter douek, 2019]; evelyn douek, *What Kind of Oversight Board Have You Given Us?*, U. CHI. L. REV. ONLINE (May 11, 2020), https://lawreviewblog.uchicago.edu/2020/05/11/fb-oversight-board-edouek/ [hereinafter douek, 2020]; Kate Klonick, *The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression*, 129 YALE L.J. 2418 (2020). The Oversight Board referred the question of what to do with Trump's account going forward back to Facebook to decide. Facebook ultimately decided to suspend Trump's account for two years and only reinstate it if the risk to public safety had receded: Nick Clegg, *In Response to Oversight Board, Trump Suspended for Two Years; Will Only Be Reinstated If Conditions Permit*, FACEBOOK NEWSROOM (June 4, 2021), https://about.fb.com/news/2021/06/facebook-response-to-oversight-board-recommendations-trump/.
9. Manhattan Cmty. Access Corp. v. Halleck, 139 S. Ct. 1921, 1930 (2019) ("[W]hen a private entity provides a forum for speech, the private entity is not ordinarily constrained by the First

Amendment because the private entity is not a state actor. The private entity may thus exercise editorial discretion over the speech and speakers in the forum.").

10. Pierre-Paul Bermingham, *Merkel Among EU Leaders Questioning Twitter's Trump Ban*, POLITICO, https://www.politico.eu/article/angela-merkel-european-leaders-question-twitter-donald-trump-ban/ (last visited Jan. 24, 2021); Daniel Hurst, *Australia's Acting PM Says Capitol Attack "Unfortunate" and Condemns Twitter "Censorship" of Trump*, THE GUARDIAN, http://www.theguardian.com/australia-news/2021/jan/11/australias-acting-pm-says-capitol-attack-unfortunate-and-condemns-twitter-censorship-of-trump (last visited Jan. 24, 2021).
11. @navalny, *1. I Think That the Ban of Donald Trump on Twitter Is an Unacceptable Act of Censorship (THREAD)*, TWITTER, https://twitter.com/navalny/status/1347969772177264644 (last visited Jan. 24, 2021).
12. @jack, *Tweet Thread on the Suspension of @realDonaldTrump*, TWITTER, https://twitter.com/jack/status/1349510769268850690 (last visited Jan. 24, 2021).
13. Clegg, *supra* note 7.
14. *See* TARLETON GILLESPIE, CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION, AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA 5 (New Haven, CT: Yale University Press, 2018) (arguing that platforms have been forced to moderate content in order to maintain their continued existence by attracting advertisers and users).
15. Eric Goldman, *Of Course the First Amendment Protects Google and Facebook (and It's Not a Close Question)*, Knight First Amendment Institute's First Amendment Series (Santa Clara Univ. Legal Studies Research Paper, No. 2018-8, Feb. 26, 2020), https://knightcolumbia.org/content/course-first-amendment-protects-google-and-facebook-and-its-not-close-question [https://perma.cc/D6X4-QGUP].
16. *See, e.g.*, GILLESPIE, *supra* note 14; Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598 (2018); NICOLAS A. SUZOR, LAWLESS: THE SECRET RULES THAT GOVERN OUR DIGITAL LIVES (Cambridge: Cambridge University Press, 2019).
17. THOMAS IRWIN EMERSON, THE SYSTEM OF FREEDOM OF EXPRESSION 10 (New York: Random House, 1970).
18. Klonick, *supra* note 16, at 1618.
19. Jeffrey Rosen, *Google's Gatekeepers*, N.Y. TIMES (Nov. 28, 2008), https://www.nytimes.com/2008/11/30/magazine/30google-t.html; Marvin Ammori, *The "New"* New York Times: *Free Speech Lawyering in the Age of Google and Twitter*, 127 HARV. L. REV. 2259 (2014); Klonick, *supra* note 16.
20. Sarah Jeong, *The History of Twitter's Rules*, VICE (Jan. 14, 2016), https://www.vice.com/en_us/article/z43xw3/the-history-of-twitters-rules.
21. Catherine Buni & Soraya Chemaly, *The Secret Rules of the Internet*, THE VERGE, https://www.theverge.com/2016/4/13/11387934/internet-moderator-history-youtube-facebook-reddit-censorship-free-speech (last visited Nov. 10, 2019).
22. Samantha Murphy, *Facebook Changes Its "Move Fast and Break Things" Motto*, MASHABLE, https://mashable.com/2014/04/30/facebooks-new-mantra-move-fast-with-stability/ (last visited Apr. 29, 2021).
23. Sarah T. Roberts, *Digital Detritus: "Error" and the Logic of Opacity in Social Media Content Moderation*, 23 FIRST MONDAY (Mar. 2018).
24. *Cf.* McCulloch v. Maryland, 17 U.S. 316, 407 (1819).
25. Klonick, *supra* note 16, at 1634.
26. Frederick Schauer, *Formalism*, 97 YALE L.J. 509, 510 (1987–88)(emphasis in original).
27. Gerald E. Frug, *The Ideology of Bureaucracy in American Law*, 97 HARV. L. REV. 1276, 1282 (1984).
28. Nicholas Bagley, *The Procedure Fetish*, MICH. L. REV. 345, 372 (2019).
29. Klonick, *supra* note 16, at 1621.
30. Bagley, *supra* note 28, at 380.
31. evelyn douek, *Governing Online Speech: From "Posts-as-Trumps" to Proportionality and Probability*, 121 COLUM. L. REV. 759, 820 (2021).

32. Monika Bickert, *Defining the Boundaries of Free Speech on Social Media*, *in* THE FREE SPEECH CENTURY 254, 256 (Lee C. Bollinger & Geoffrey R. Stone eds., New York: Oxford University Press, 2019).
33. *See, e.g.*, the European Union's proposed Digital Services Act, Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020PC0825; Platform Accountability and Consumer Transparency (PACT) Act, S. 4066 (116th Cong.) (2020), https://www.congress.gov/bill/116th-congress/senate-bill/4066/text?format=xml.
34. *Santa Clara Principles on Transparency and Accountability in Content Moderation*, SANTA CLARA PRINCIPLES, https://santaclaraprinciples.org.
35. douek, 2019; douek, 2020.
36. Charter art. 2, § 1.
37. Charter art. 2, § 2.
38. *Id*.
39. *Let's Talk About Hot Tub Streams*, TWITCH BLOG (May 21, 2021), https://blog.twitch.tv/en/2021/05/21/lets-talk-about-hot-tub-streams/ (noting that "while we have guidelines about sexually suggestive content, being found to be sexy by others is not against our rules, and Twitch will not take enforcement action against women, or anyone on our service, for their perceived attractiveness").
40. *Ravelry New Policy: Do Not Post in Support of Trump or His Administration*, RAVELRY (Jan. 20, 2021), https://www.ravelry.com/content/no-trump.
41. LON L. FULLER, THE MORALITY OF LAW 39 (Rev. ed., New Haven, CT: Yale University Press, 1969).
42. J. Nathan Matias, *Preventing Harassment and Increasing Group Participation Through Social Norms in 2,190 Online Science Discussions*, PNAS 1 (2019).
43. Tom R. Tyler, *Procedural Justice, Legitimacy, and the Effective Rule of Law*, 3 CRIME AND JUSTICE 283 (2003); Facebook Data Transparency Advisory Group, *Report of the Facebook Data Transparency Advisory Group*, JUSTICE COLLABORATORY, YALE LAW SCHOOL (2019), https://law.yale.edu/system/files/area/center/justice/document/dtag_report_5.22.2019.pdf.
44. Frederick Schauer, *The Convergence of Rules and Standards*, [2003] N.Z. L. REV. 303, 315.
45. Elena Kagan, *Private Speech, Public Purpose: The Role of Governmental Motive in First Amendment Doctrine*, 63 U. CHI. L. REV. 413, 515 (1996).
46. Because when a platform decides to take down a piece of content, that decision will often be invisible to everyone except the affected user, and privacy restrictions limit what platforms can or do release about specific enforcement decisions.
47. Soraya Chemaly, *#FreeTheNipple: Facebook Changes Breastfeeding Mothers Photo Policy*, HUFFPOST, https://www.huffpost.com/entry/freethenipple-facebook-changes_b_5473467 (last visited Nov. 10, 2019).
48. *Community Standards—14. Adult Nudity and Sexual Activity*, FACEBOOK COMMUNITY STANDARDS, https://www.facebook.com/communitystandards/adult_nudity_sexual_activity (last visited Apr. 30, 2021).
49. David Gilbert, *Leaked Documents Show Facebook's Absurd "Breast Squeezing Policy,"* VICE NEWS, https://www.vice.com/en/article/7k9xnb/leaked-documents-show-facebooks-absurd-breast-squeezing-policy (last visited Apr. 30, 2021).
50. *Coordinated Harmful Activity*, TWITTER RULES, https://help.twitter.com/en/rules-and-policies/coordinated-harmful-activity (last visited Apr. 2, 2021).
51. Tom Knowles, *YouTube Bans Fake Animal Rescue Videos*, THE TIMES (Mar. 25, 2021), https://www.thetimes.co.uk/article/youtube-bans-fake-animal-rescue-videos-l5vwtnxls.
52. Tom Ginsburg, *The Judicialization of Administrative Governance*, *in* ADMINISTRATIVE LAW AND GOVERNANCE IN ASIA: COMPARATIVE PERSPECTIVES 20 (Tom Ginsburg & Albert H. Y. Chen eds., Abingdon, UK: Routledge, 2008).
53. ADRIAN VERMEULE, MECHANISMS OF DEMOCRACY: INSTITUTIONAL DESIGN WRIT SMALL, MECHANISMS OF DEMOCRACY 95 (New York: Oxford Scholarship Online, 2007).
54. SARAH T. ROBERTS, BEHIND THE SCREEN: CONTENT MODERATION IN THE SHADOWS OF SOCIAL MEDIA 172 (New Haven, CT: Yale University Press, 2019).

55. Lauren Strapagiel, *The "Proud Boys" Hashtag Has Been Taken Over by Gay Love*, BUZZFEED NEWS (Oct. 4, 2020), https://www.buzzfeednews.com/article/laurenstrapagiel/proud-boys-hashtag-reclaimed-gay-love-photos.
56. douek, *supra* note 31, at 791–92.
57. Nathaniel Persily & Joshua A. Tucker, *Conclusion: The Challenges and Opportunities for Social Media Research*, *in* SOCIAL MEDIA AND DEMOCRACY: THE STATE OF THE FIELD AND PROSPECTS 324 (Nathanel Persily & Joshua A. Tucker eds., Cambridge: Cambridge University Press, 2020).
58. ROBERTS, *supra* note 54; Casey Newton, *The Secret Lives of Facebook Moderators in America*, THE VERGE (Feb. 25, 2019), https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona.
59. Robert Gorwa et al., *Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance*, 7 BIG DATA & SOC. 1, 2 (2020); Tim Wu, *Will Artificial Intelligence Eat the Law? The Rise of Hybrid Social-Ordering Systems*, 119 COLUM. L. REV. 29 (2019).
60. douek, *supra* note 31, at 808–13.
61. *Santa Clara Principles on Transparency and Accountability*, *supra* note 34; Emma J. Llansó, *No Amount of "AI" in Content Moderation Will Solve Filtering's Prior-Restraint Problem*, 7 BIG DATA & SOC. 1, 4 (2020).
62. evelyn douek, *The Limits of International Law in Content Moderation,* 6 U.C. IRVINE J. INT'L TRAN'L & COMP. L. 37 (2021).
63. GILLESPIE, *supra* note 14, at 5.
64. *See, e.g.*, Jeff Horwitz & Deepa Seetharaman, *Facebook Executives Shut Down Efforts to Make the Site Less Divisive*, WALL ST. J., https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499 (last visited May 1, 2021); Newley Purnell & Jeff Horwitz, *Facebook's Hate-Speech Rules Collide with Indian Politics*, WALL ST. J. (Aug. 14, 2020), https://www.wsj.com/articles/facebook-hate-speech-india-politics-muslim-hindu-modi-zuckerberg-11597423346; Monika Bauerlein & Clara Jeffery, *Facebook Manipulated the News You See to Appease Republicans, Insiders Say,* MOTHER JONES (Oct. 21, 2020), https://www.motherjones.com/media/2020/10/facebook-mother-jones/.
65. This paragraph draws on evelyn douek, *Trump Is Banned. Who Is Next?*, THE ATLANTIC, https://www.theatlantic.com/ideas/archive/2021/01/trump-is-banned-who-is-next/617622/ (last visited Jan. 18, 2021).
66. https://www.c-span.org/video/?478048-1/facebook-twitter-ceos-testify-regulating-social-media-content (1:36:14 PM).
67. Ryan Mac, *Facebook Employees Are Outraged at Mark Zuckerberg's Explanations of How It Handled the Kenosha Violence*, BUZZFEED NEWS (Aug. 28, 2020), https://www.buzzfeednews.com/article/ryanmac/facebook-employees-slam-zuckerberg-kenosha-militia-shooting.
68. Robert A. Kagan, *Varieties of Bureaucratic Justice: Building on Mashaw's Typology*, in ADMINISTRATIVE LAW FROM THE INSIDE OUT: ESSAYS ON THEMES IN THE WORK OF JERRY L. MASHAW 247, 270 (Nicholas R. Parrillo ed., Cambridge: Cambridge University Press, 2017).
69. Daphna Renan, *The Law Presidents Make*, 103(5) VA. L. REV. 805, 809 (2017).
70. RICHARD H. FALLON, LAW AND LEGITIMACY IN THE SUPREME COURT 21 (Cambridge, MA: Belknap Press of Harvard University Press, 2018); Bagley, *supra* note 28, at 372.
71. FALLON, *supra* note 70, at xi.
72. Richard J. Pierce Jr., *A Comparison of the Cultures and Performance of a Modern Agency and a Nineteenth-Century Agency*, *in* ADMINISTRATIVE LAW FROM THE INSIDE OUT: ESSAYS ON THEMES IN THE WORK OF JERRY L. MASHAW 322, 342 (Nicholas R. Parrillo ed., Cambridge: Cambridge University Press, 2017).
73. Bagley, *supra* note 28, at 349, 380.
74. Renan, *supra* note 69, at 868.
75. Jeremy K. Kessler, *The Struggle for Administrative Legitimacy*, 129 HARV. L. REV. 718, 729 (2016).
76. Elizabeth Magill & Adrian Vermeule, *Allocating Power Within Agencies*, 120 YALE L.J. 1032, 1081 (2011).

77. Twitter Support, *Sharing an Article Can Spark Conversation, so You May Want to Read It Before You Tweet It. To Help Promote Informed Discussion, We're Testing a New Prompt on Android––When You Retweet an Article That You Haven't Opened on Twitter, We May Ask If You'd Like to Open It First*, @TWITTERSUPPORT (June 10, 2020), https://twitter.com/TwitterSupport/status/1270783537667551233.
78. *See, e.g.*, Jonathan Zittrain, *A Jury of Random People Can Do Wonders for Facebook*, THE ATLANTIC, https://www.theatlantic.com/ideas/archive/2019/11/let-juries-review-facebook-ads/601996/ (last visited May 3, 2021).
79. Ben Smith, *Is an Activist's Pricey House News? Facebook Alone Decides*, N.Y. TIMES (Apr. 26, 2021), https://www.nytimes.com/2021/04/25/business/facebook-nypost.html.
80. JACK GOLDSMITH & TIM WU, WHO CONTROLS THE INTERNET? ILLUSIONS OF A BORDERLESS WORLD (New York: Oxford University Press, 2008).
81. Bickert, *supra* note 32, at 260–61.
82. Dan Jerker B. Svantesson, *Internet Jurisdiction and Intermediary Liability*, *in* OXFORD HANDBOOK OF ONLINE INTERMEDIARY LIABILITY 689 (Giancarlo Frosio ed., New York: Oxford University Press, 2020).
83. douek, *supra* note 62.
84. Roscoe Pound, *The Call for a Realist Jurisprudence*, 44 HARV. L. REV. 697, 707 (1931).
85. *See, e.g.*, Jody Freeman & David B. Spence, *Old Statutes, New Problems*, 163 U. PA. L. REV. 1, 5 (2014).
86. Strapagiel, *supra* note 55.
87. evelyn douek, *The Free Speech Blind Spot: Foreign Election Interference on Social Media*, *in* DEFENDING DEMOCRACIES: COMBATING FOREIGN ELECTION INTERFERENCE IN A DIGITAL AGE 265 (Jens David Ohlin & Duncan B. Hollis eds., New York: Oxford University Press USA—OSO, 2021); Bridget Barrett, *What We Learned From Bloomberg's Online Campaign*, LAWFARE, https://www.lawfareblog.com/what-we-learned-bloombergs-online-campaign (last visited Mar. 6, 2020).
88. *See, e.g.*, James Grimmelmann, *The Platform Is the Message*, 2 GEO. L. TECH. REV. 217 (2018) (discussing the Tide pod challenge).
89. *See, e.g.*, Netzwerkdurchsetzungsgesetz [Network Enforcement Act], July 12, 2017, (Ger.) § 2(2) (NetzDG); PACT Act, S. 4066, § 5(d)(2).
90. douek, *supra* note 31, at 818–19; NIELS PETERSEN, PROPORTIONALITY AND JUDICIAL ACTIVISM: FUNDAMENTAL RIGHTS ADJUDICATION IN CANADA, GERMANY AND SOUTH AFRICA 55 (Cambridge: Cambridge University Press 2017); Vicki C. Jackson, *Constitutional Law in an Age of Proportionality*, 124 YALE L.J. 3094, 3155 (2015).
91. Charles F. Sabel & William H. Simon, *Minimalism and Experimentalism in the Administrative State*, 100 GEO. L.J. 53, 74 (2011).
92. *See, e.g.*, Schauer, *supra* note 26, at 512 (discussing *Lochner*).
93. Monica Anderson, *Americans Divided on Whether Trump Should Be Permanently Banned from Social Media*, PEW RESEARCH CENTER (May 5, 2021), https://www.pewresearch.org/fact-tank/2021/05/05/americans-divided-on-whether-trump-should-be-permanently-banned-from-social-media/.

Chapter 10

1. *See* Int'l Soc'y for Krishna Consciousness v. Lee, 505 U.S. 672, 678 (1992).
2. *See* Pleasant Grove City v. Summum, 555 U.S. 460, 469–70 (2009).
3. *See* Jack M. Balkin, *Free Speech Is a Triangle*, 118 COLUM. L. REV. 2011, 2025–28 (2018).
4. *See* INT'L COVENANT ON CIV. & POL. RTS., *opened for signature* Dec. 16, 1966, art. 19(3), S. Exec. Doc. E, 95-2 (1978), 999 U.N.T.S. 171, 178 (entered into force Mar. 23, 1976) [hereinafter ICCPR]; Human Rights Committee, General Comment No. 34: Article 19: Freedom of opinion and expression, Sept. 12, 2011, CCPR/C/GC/34 [hereinafter GC 34].
5. *See* GC 34, *supra* note 4, paras. 34–35.
6. *See, e.g.*, Henry Olsen, *Mark Zuckerberg Must Decide the Fate of Trump's Facebook Account. He Should Turn to the First Amendment*, WASH. POST (May 5, 2021), https://www.washingtonp

ost.com/opinions/2021/05/05/mark-zuckerberg-must-decide-fate-trumps-facebook-account-he-should-turn-first-amendment/.

7. *See* James Freeman, *Twitter Has a Competitor*, WALL ST. J. (Dec. 8, 2020), https://www.wsj.com/articles/twitter-has-a-competitor-11607472466.
8. *See* Knight First Amendment Inst. at Columbia Univ. v. Trump, 928 F.3d 226, 234–237 (2d Cir. 2019).
9. *See* Evelyn Mary Aswad, *The Future of Freedom of Expression Online*, 17 DUKE L. & TECH. REV. 26, 39–40 (2018); Global Network Initiative, https://globalnetworkinitiative.org/.
10. *See* REPORT OF THE SPECIAL RAPPORTEUR ON THE PROMOTION AND PROTECTION OF THE RIGHT TO FREEDOM OF OPINION AND EXPRESSION, U.N. Doc A/HRC/38/35, Apr. 6, 2018, para. 70/ [hereinafter SR REPORT]; see *generally* Aswad, *supra* note 9.
11. *See* Daphne Keller, *Who Do You Sue? State and Platform Hybrid Power Over Online Speech* 2, 11 (Hoover Inst. Aegis Series Paper No. 1902, 2019).
12. *See, e.g.*, Manhattan Cmty. Access Corp. v. Halleck, 139 S. Ct. 1921, 1926 (2019).
13. *See* Turner Broad. Sys. v. FCC, 512 U.S. 622, 629, 636 (1994); Miami Herald Publ'g Co. v. Tornillo, 418 U.S. 241, 256 (1974).
14. *See* SR REPORT, *supra* note 10, paras. 28, 44–48.
15. *See* Keller, *supra* note 11, at 16; K. Sabeel Rahman, *The New Utilities: Private Power, Social Infrastructure, and the Revival of the Public Utility Concept*, 39 CARDOZO L. REV. 1621, 1668–80.
16. Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183, 1186, 1209–10 (2016).
17. *See* Michael I. Meyerson, *Authors, Editors, and Uncommon Carriers: Identifying the "Speaker" Within the New Media*, 71 NOTRE DAME L. REV. 79, 83 (1995); Biden v. Knight First Amendment Inst. at Columbia Univ., 141 S. Ct. 1220, 1222–23 (Thomas, J., concurring).
18. Traditional public forums are those that "have immemorially been held in trust for the use of the public, and, time out of mind, have been used for purposes of assembly, communicating thoughts between citizens, and discussing public questions." Hague v. CIO, 307 U.S. 496, 515 (1939). Because this category is historically derived, it is effectively limited to public streets and parks. *See* Frederick Schauer, *Not Just About License Plates:* Walker v. Sons of Confederate Veterans, *Government Speech, and Doctrinal Overlap in the First Amendment*, 2015 S. CT. REV. 265, 283–84. Nonpublic forums include "public property which is not by tradition or designation a forum for public communication." Perry Educ. Ass'n v. Perry Local Educators, 460 U.S. 37, 46 (1983).
19. *Lee*, 505 U.S. at 678.
20. Rosenberger v. Rector & Visitors of Univ. of Va., 515 U.S. 819, 829 (1995).
21. *See Perry Educ. Ass'n*, 460 U.S. at 45.
22. *See Rosenberger*, 515 U.S. at 829.
23. *See* Lyrissa Lidsky, *Government Sponsored Social Media and Public Forum Doctrine Under the First Amendment*, 19 PUB. LAW. 2, 3 (2011) ("[A] government actor who creates a completely open, interactive Facebook page without any explicit statement of purpose probably cedes all but the most limited forms of editorial control over that forum.").
24. *See* FCC v. Pacifica Found., 438 U.S. 726 (1978).
25. *See id*. at 748–50.
26. *See* Action for Children's Television v. FCC, 58 F.3d 654, 660 (D.C. Cir. 1995); *see also* FCC v. Fox Television Stations, 556 U.S. 502, 533 (2009) (Thomas, J., concurring) ("[E]ven if this Court's disfavored treatment of broadcasters under the First Amendment could have been justified at the time of *Red Lion* and *Pacifica*, dramatic technological advances have eviscerated the factual assumptions underlying those decisions.").
27. *See Pacifica*, 438 U.S. at 748; Red Lion Broad. v. FCC, 395 U.S. 367, 389–90 (1969).
28. *See Red Lion Broad.*, 395 U.S. at 389–90.
29. Erznoznik v. City of Jacksonville, 422 U.S. 205 (1975); Cohen v. California, 403 U.S. 15 (1971); Southeastern Promotions v. Conrad, 420 U.S. 546 (1975).
30. *Erznoznik*, 422 U.S. at 209.
31. *See* United States v. Playboy Entm't Group, 529 U.S. 803, 813–15 (2000); Reno v. ACLU, 521 U.S. 844, 874–77 (1997); Sable Commc'ns of Cal. v. FCC, 492 U.S. 115, 130–31 (1989).

32. *See* Schauer, *supra* note 18, at 285; Robert C. Post, *Between Governance and Management: The History and Theory of the Public Forum*, 34 UCLA L. REV. 1713, 1801 (1987).
33. *See* Steinburg v. Chesterfield Cty. Planning Comm'n, 527 F.3d 377 (4th Cir. 2008); Lidsky, *supra* note 23, at 7.
34. This might distinguish a state-run social media platform from the comments section on a public official's tweet, blog post, or social media post, which one could plausibly argue can be limited to comments germane to the original content. *See* Lidsky, *supra* note 23, at 5–6.
35. Hateful Conduct Policy, https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy.
36. Hate Speech, https://www.facebook.com/communitystandards/hate_speech.
37. *See* R.A.V. v. City of St. Paul, 505 U.S. 377, 392 (1992); James Weinstein, *Hate Speech Bans, Democracy, and Political Legitimacy*, 32 CONST. COMMENT. 527, 545 (2017).
38. *R.A.V.*, 505 U.S. at 392.
39. Walker v. Sons of Confederate Veterans, 576 U.S. 200 (2015).
40. *Id*. at 208.
41. *See id*. at 207–8.
42. *See id.*; *Pleasant Grove City*, 555 U.S. 460; Johanns v. Livestock Mktg. Ass'n, 544 U.S. 550 (2005).
43. *Walker*, 576 U.S. at 211–13.
44. NEA v. Finley, 524 U.S. 569 (1998).
45. Rust v. Sullivan, 500 U.S. 173 (1991).
46. *See* Speiser v. Randall, 357 U.S. 513 (1958); Shapiro v. Thompson, 394 U.S. 618 (1969); Sherbert v. Verner, 374 U.S. 398 (1963).
47. *See* Perry v. Sindermann, 408 U.S. 593 (1972); Pickering v. Bd. of Educ., 391 U.S. 563 (1972).
48. *Perry*, 408 U.S. at 597.
49. *See Finley*, 524 U.S. at 583–86.
50. *See Rust*, 500 U.S. at 193.
51. *See* David Cole, *Beyond Unconstitutional Conditions: Chartering Spheres of Neutrality in Government-Funded Speech*, 67 N.Y.U. L. Rev. 675, 676 (1992); Schauer, *supra* note 18, at 291.
52. Garcetti v. Ceballos, 547 U.S. 410 (2006).
53. *See* Evelyn Aswad, *To Protect Freedom of Expression, Why Not Steal Victory from the Jaws of Defeat?*, 77 WASH. & LEE L. REV. 609, 618–19 (2020).
54. Universal Declaration of Human Rights, G.A. Res. 217 A(III), art. 19, U.N. Doc. A/810 (1948).
55. ICCPR art. 19(2).
56. ICCPR art. 19(3).
57. *See* GC 34, *supra* note 4, paras. 33–35.
58. *See* SENATE COMM. ON FOREIGN RELATIONS, REPORT ON THE INTERNATIONAL COVENANT ON CIVIL AND POLITICAL RIGHTS, S. Exec. Rep. No. 23, 1, 20–21 (102d Sess. 1992), *reprinted in* 31 I.L.M. 645 (1992).
59. GC 34, *supra* note 4, paras. 11, 28.
60. *Id*., para. 35.
61. *Id*., para. 31.
62. Coleman v. Australia, U.N. GAOR, Hum. Rts. Comm., 87th Sess., U.N. Doc. CCPR/C/87/D/1157/2003 (2006), para 7.3.
63. Zündel v. Canada, Communication No. 953/2000, U.N. Doc. CCPR/C/78/D/953/2000 (2003), para. 8.5.
64. *Id.*
65. *See* Michael O'Flaherty, *Freedom of Expression: Article 19 of the International Covenant on Civil and Political Rights and the Human Rights Committee's General Comment 34*, 12 HUM. RTS. L. REV. 630, 638 (2012).
66. *Id*. at 647–48.
67. McAuliffe v. Mayor of New Bedford, 29 N.E. 517, 517 (Mass. 1892).
68. As O'Flaherty notes, instead of deciding that Zündel had no free expression interest in holding his press conference in a government-operated space, the Committee could have

decided how his access could be regulated under the limitations language of Article 19's third section. *See* O'Flaherty, *supra* note 65, at 638.

69. *See* Locurto v. Giuliani, 447 F.3d 159 (2d Cir. 2006) (rejecting a First Amendment challenge to dismissal of New York City police officers for participating in a racist parade float while off duty).
70. Even within the category of access, the focus has been on coercive enforcement of access restrictions against individuals, as in *Coleman*, rather than the state's quasi-editorial control over particular utterances within those spaces.
71. *See* SR REPORT, *supra* note 10, paras. 2, 6.
72. *See id.*
73. *See id.*, para. 7.
74. *See id.*, para. 13.
75. *See id.*, para. 28.
76. *See id.*, para. 41.
77. *See also* Aswad, *supra* note 9, at 40–41 (arguing that the companies have no rights under international human rights law).
78. *See* New York Times Co. v. United States, 403 U.S. 713, 714 (1971); Southeastern Promotions v. Conrad, 420 U.S. 546, 552–53 (1975).
79. Even on a government platform, the government may not compel private citizens to associate themselves with its own speech. *See* Wooley v. Maynard, 430 U.S. 705, 717 (1977).
80. *See* JAMAL GREENE, HOW RIGHTS WENT WRONG (Boston: Houghton Mifflin Harcourt, 2021).
81. *Cf.* Michael Coenen, *Of Speech and Sanctions: Toward a Penalty-Sensitive Approach to the First Amendment*, 112 COLUM. L. REV. 991 (2012) (describing how little First Amendment doctrine calibrates free speech analysis to the nature of the penalty).
82. ICCPR art. 20(2).
83. Red Lion Broad. v. FCC, 395 U.S. 367, 376 (1969).
84. *See, e.g.*, Danielle Keats Citron & Jonathan W. Penney, *When Law Frees Us to Speak*, 87 FORDHAM L. REV. 2317, 2318–19 (2019) (detailing racist and sexist harassment, nonconsensual nudity, doxing, and other abusive online behavior).
85. GC 34, *supra* note 4, para. 15; Human Rights Comm., Concluding Observations regarding the Russian Federation, Dec. 1, 2003, CCPR/CO/79/RUS, para. 18; O'Flaherty, *supra* note 65, at 639.
86. *See* Sarah Joseph, "*Is Fox News a Breach of Human Rights?": The News Media's Immunity From the Guiding Principles on Business and Human Rights*, 1 BUS. & HUM. RTS. J. 229, 237 (2016).
87. The Human Rights Committee interprets rights embodied within the ICCPR as applying only to individuals. U.N. Human Rights Comm., General Comment No. 31, para. 9, U.N. Doc. CCPR/C/21/Rev.1/Add. 13 (May 26, 2004).
88. *See* Mitchell N. Berman, *Commercial Speech and the Unconstitutional Conditions Doctrine: A Second Look at "The Greater Includes the Lesser,"* 55 VAND. L. REV. 693 (2002).
89. *See* Anthony G. Amsterdam, Note, *The Void-for-Vagueness Doctrine in the Supreme Court*, 109 U. PA. L. REV. 67 (1960).
90. *Rosenberger*, 515 U.S. at 829; Cornelius v. NAACP Legal Def. & Educ. Fund, 473 U.S. 788, 804–6 (1995).
91. Schauer, *supra* note 18, at 726; Bantam Books v. Sullivan, 372 U.S. 58, 59–60, 70–71 (1963).
92. GC 34, *supra* note 4, para. 25.
93. *See* Shuttlesworth v. City of Birmingham, 394 U.S. 147, 150–51 (1969).
94. *See* Paul A. Freund, *The Supreme Court and Civil Liberties*, 4 VAND. L. REV. 533, 538 (1951).
95. *See* Schauer, *supra* note 18, at 727–28; Paris Adult Theatre I v. Slaton, 413 U.S. 49, 55 (1973).
96. *See* evelyn doeuk, *Governing Online Speech: From "Posts-As-Trumps" to Proportionality and Probability*, 121 COLUM. L. REV. 759, 763 (2021).
97. *See* Miller v. California, 413 U.S. 15, 24 (1973).
98. *See* Southeastern Promotions v. Conrad, 420 U.S. 546 (1975); *Erznoznik*, 422 U.S. 205.
99. ICCPR art. 19(3)(b).
100. *See* Müller v. Switzerland, App. No. 10737/84, 13 Eur. H.R. Rep. 212 (1991); Sarah Joseph, *Art and Human Rights Law*, *in* RESEARCH HANDBOOK ON ART AND LAW 389, 396 (Jani McCutcheon & Fiona McGaughey eds., London: Edward Elgar, 2020).

101. GC 34, *supra* note 4, para. 36; *see* O'Flaherty, *supra* note 65, at 649.
102. *See* United States v. Alvarez, 567 U.S. 709, 718–19 (2012).
103. *See* New York Times Co. v. Sullivan, 376 U.S. 254 (1964).
104. It is worth reiterating that this essay has not considered the situation of a state platform that is either exclusive or market dominant. There is good reason to think of this situation differently than a relatively open marketplace in which the state chooses to participate on equal terms with others.

Chapter 11

1. *See, e.g.*, Kelly Lewis, *Social Media Platforms Are Complicit in Censoring Palestinian Voices*, THE CONVERSATION (May 24, 2021), https://theconversation.com/social-media-platforms-are-complicit-in-censoring-palestinian-voices-161094 (describing the "mounting evidence" of the unjustified [and often unexplained] removal of pro-Palestinian content).
2. *See, e.g.*, Leah Nylen, *Social Media's "Disparate Impact" on Conservatives Unfair Trade Practice, Lee Says*, POLITICO (Oct. 28, 2020), https://www.politico.com/news/2020/10/28/conservatives-social-media-tech-hearing-433323; Rishika Pardikar, *Social Media Companies Like Instagram Are Censoring Dissent*, JACOBIN (June 1, 2021), https://www.jacobinmag.com/2021/06/censorship-facebook-instagram-twitter-india-palestine-colombia.
3. Fla. Stat. Ann. § 501.2041 ("Censor" includes any action taken by a social media platform to delete, regulate, restrict, edit, alter, inhibit the publication or republication of, suspend a right to post, remove, or post an addendum to any content or material posted by a user. The term also includes actions to inhibit the ability of a user to be viewable by or to interact with another user of the social media platform.").
4. A 2019 list of the most popular websites globally listed only one website owned and operated by a nonprofit entity—namely, Wikipedia. *Ranking the Top 100 Websites in the World*, VISUAL CAPITALIST, https://www.visualcapitalist.com/wp-content/uploads/2019/08/top-100-websites-ranking.html.
5. *See, e.g.*, Shannon C. McGregor & Daniel Kreiss, *Conservatives Say Google and Facebook Are Censoring Them. Here's the Real Background*, WASH. POST (Aug. 1, 2021), https://www.washingtonpost.com/politics/2019/08/01/are-google-facebook-censoring-conservatives-problem-is-more-widespread-than-that/.
6. Fla. Stat. Ann. §§ 106.072, 501.2041.
7. Shawn Mulcahy, *Texas Senate Approves Bill to Stop Social Media Companies From Banning Texans for Political Views*, TEX. TRIB. (Mar. 30, 2021), https://www.texastribune.org/2021/03/30/texas-soical-media-censorship. As of this writing, the law remains under consideration by the Texas legislature. Many similar bills have been proposed in other states. *See* David McCabe & Cecilia Kang, *As Congress Dithers, States Step in to Set Rules for the Internet*, N.Y. TIMES (May 14, 2021), https://www.nytimes.com/2021/05/14/technology/state-privacy-internet-laws.html.
8. American Choice and Innovation Online Act, H.R. 3816 (117th Cong.) (2021–22), https://cicilline.house.gov/sites/cicilline.house.gov/files/documents/American%20Innovation%20and%20Choice%20Online%20Act%20-%20Bill%20Text.pdf. The bill defines a "business user" to mean a person who uses the platform to sell or distribute goods or services. *Id.* This definition presumably includes not only commercial advertisers but content providers who monetize their content. *See, e.g.*, *How to Earn Money on YouTube,* https://support.google.com/youtube/answer/72857?hl=en.
9. *See* Press Release, Senator Hawley Introduces Legislation to Amend Section 230 Immunity for Big Tech Companies (June 19, 2019), https://www.hawley.senate.gov/senator-hawley-introduces-legislation-amend-section-230-immunity-big-tech-companies.
10. *See, e.g.*, Kristen Clarke & David Brody, *It's Time for an Online Civil Rights Act*, THE HILL (Aug. 3, 2018), https://thehill.com/opinion/civil-rights/400310-its-time-for-an-online-civil-rights-act (arguing for the extension of federal public accommodations laws to the social media platforms); K. Sabeel Rahman, *The New Utilities: Private Power, Social Infrastructure, and the Revival of the Public Utility Concept*, 39 CARDOZO L. REV. 1621, 1670 (2017–18)(suggesting the extension of common carrier obligations to social media platforms).

11. Tim Wu, *Is Filtering Censorship? The Second Free Speech Tradition*, BROOKINGS INST. (Dec. 27, 2010), https://www.brookings.edu/research/is-filtering-censorship-the-second-free-speech-tradition.
12. Other regulatory tools that have historically been used to protect the democratic inclusiveness of the mass public sphere include ownership controls and antitrust mechanisms such as the tool of divestiture. *See generally* C. Edwin Baker, *Media Concentration: Giving up on Democracy*, 54 FLA. L. REV. 839 (2002) (discussing the historical use of these mechanisms). In addition to these familiar regulatory tools, scholars of the social media platforms have begun to argue for new kinds of structural regulation to reshape the digital public sphere. *See, e.g.*, Ethan Zuckerman, *The Case for Digital Public Infrastructure*, KNIGHT FIRST AMENDMENT INST. (Jan 17, 2020), https://knightcolumbia.org/content/the-case-for-digital-public-infrastructure. All of these instruments, while potentially fruitful tools for cabining platform power, would require much more economic reorganization than the imposition of nondiscrimination duties on the platforms.
13. As Fred Schauer puts it:

> Private choices based on the content of speech are not, unlike racially discriminatory choices, viewed by society as inherently evil. Rather, individual decisions about speech—preferring some ideas and information to others, placing one's property at the service of some ideologies and not others—are central to the concept of a marketplace of ideas. Ideas fail or succeed according to their ability to win support in free public debate. A private person participates in that debate when he contributes the use of his property to the proponents of certain ideas; that is an act of advocacy as surely as if he were disseminating the ideas himself.

Frederick F. Schauer, Hudgens v. NLRB and the Problem of State Action in First Amendment Adjudication, 61 MINN. L. REV. 433, 448–449 (1977).

14. Tarleton Gillespie has argued, along these lines, that content "moderation is, in many ways, the commodity that platforms offer." TARLETON GILLESPIE, CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION, AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA 47 (New Haven, CT: Yale University Press, 2018). *See also* Kate Klonick, *The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression*, 129 YALE L.J. 2418, 2441 (2020) (noting that after negative public reaction to controversial content removal decisions, "[l]eaders at [Facebook] began to fully understand that content-moderation policy and enforcement were not just a 'customer service' or 'trust and safety issue'—as a communication platform, content moderation was *the product* of Facebook") [hereinafter Klonick, *Facebook Oversight Board*]. The phrase "new governors" was coined by Klonick in an earlier paper. Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598 (2018).
15. *See generally* RICHARD R. JOHN, SPREADING THE NEWS: THE AMERICAN POSTAL SYSTEM FROM FRANKLIN TO MORSE (Cambridge, MA: Harvard University Press, 1995).
16. William Jones, *The Common Carrier Concept as Applied to Telecommunications: A Historical Perspective*, CYBERTELECOM (1980), http://www.cybertelecom.org/notes/jones.htm; https://perma.cc/5AVH-CBQD.
17. *State Statutes Affecting the Associated Press*, *in* LAW OF THE ASSOCIATED PRESS 484 (1914).
18. 47 U.S.C. § 315(a).
19. 47 CFR § 76.205(a).
20. *See, e.g.*, 52 U.S.C. § 30120(b) ("No person who sells space in a newspaper or magazine to a candidate or to the agent of a candidate, for use in connection with such candidate's campaign, may charge any amount for such space which exceeds the amount charged for comparable use of such space for other purposes"). *See also* W. Va. Code § 59-3-6 (2020) (prohibiting newspapers from charging political candidates more than "private patrons").
21. Minn. Stat. § 211B.05(2) (2020) (prohibiting newspapers from charging some candidates for office more than others).
22. JOHN, *supra* note 15, at 38–39.

23. *Id.* at 31 ("To reduce the cost of securing political information for citizen-farmers, many of whom lived in the South and West, Congress increased the cost of doing business for merchants, most of whom lived in the North and East. In the broadest sense, then, this policy was not distributive, but regulatory or, more precisely, redistributive.").
24. Genevieve Lakier, *The Non-First Amendment Law of Freedom of Speech*, 134 HARV. L. REV. 2299, 2325 (2021).
25. 67 Cong. Rec. 12504 (1926) (statement of Mr. Broussard).
26. As Tarleton Gillespie notes, on social media sites "[r]ecommendation algorithms . . . not only help us find information, they also provide a means to know what there is to know and how to know it, to participate in social and political discourse, and to familiarize ourselves with the publics in which we participate [and] are now a key logic governing the flows of information on which we depend." Tarleton Gillespie, *The Relevance of Algorithms*, *in* MEDIA TECHNOLOGIES: ESSAYS ON COMMUNICATION, MATERIALITY, AND SOCIETY (Tarleton Gillespie et al. eds., Cambridge, MA: MIT Press Scholarship Online, 2014).
27. Facebook, for example, employs more than thirty thousand employees on platform "safety and security." Half of that group is responsible for content moderation. Casey Newton, *The Trauma Floor: The Secret Lives of Facebook Moderators in America*, THE VERGE (Feb. 25, 2019), https://www.theverge.com/2019/2/25/18229714. Twitter, in contrast, employs only fifteen hundred people in the work of content moderation, due to its smaller size. Jason Koebler & Joseph Cox, *How Twitter Sees Itself*, VICE (Oct. 7, 2019 9:00 AM), https://www.vice.com/en/article/a35nbj/twitter-content-moderation. The labor costs of paying content moderators are, of course, only half of the story; the platforms report expending significant money building algorithmic systems to detect unlawful or impermissible speech. *See* Hannah Bloch-Wehba, *Automation in Moderation,* 53 CORNELL INT'L L.J. 41, 47 (2020).
28. Klonick, *Facebook Oversight Board*, *supra* note 14, at 2437–38.
29. Cox v. New Hampshire, 312 U.S. 569, 574 (1941).
30. Jack Balkin has gone so far as to argue that were First Amendment standards to apply to the social media platforms, "they would quickly become ungovernable" because they "would not be able to ban . . . the sort of speech that . . . leads to trolling [or] hate speech and many other kinds of abusive speech that make the site fare less valuable for the vast majority of customers." Jack M. Balkin, *Free Speech Is a Triangle*, 118 COLUM. L. REV. 2011, 2026 (2018).
31. *See* United States Telecom Ass'n v. F.C.C., 825 F.3d 674, 743 (D.C. Cir. 2016) (explaining that the First Amendment does not apply to common carrier laws that regulate private speech intermediaries "that hold themselves out as neutral, indiscriminate conduits" of others' speech); Turner Broad. Sys., Inc. v. F.C.C., 512 U.S. 622, 642 (1994) ("[T]he First Amendment . . . does not countenance governmental control over the content of messages expressed by private individuals. . . . Laws that compel speakers to utter or distribute speech bearing a particular message are subject to the same rigorous scrutiny. In contrast, regulations that are unrelated to the content of speech are subject to an intermediate level of scrutiny.").
32. *See* Facebook Community Standards, https://www.facebook.com/communitystandards ; The Twitter Rules, https://help.twitter.com/en/rules-and-policies/twitter-rules; YouTube's Community Guidelines, https://support.google.com/youtube/answer/9288567.
33. The platforms do claim to provide nondiscriminatory access to users regardless of their identity. *See*, *e.g.*, Facebook Community Standards ("To ensure that everyone's voice is valued, we take great care to craft policies that are inclusive of different views and beliefs, in particular those of people and communities that might otherwise be overlooked or marginalized. . . . Our Community Standards apply to everyone, all around the world, and to all types of content."), https://www.facebook.com/communitystandards/introduction; Twitter Help, *Our Approach to Policy Development and Enforcement Philosophy*, https://help.twitter.com/en/rules-and-policies/enforcement-philosophy ("We have the Twitter Rules in place to help ensure everyone feels safe expressing their beliefs and we strive to enforce them with uniform consistency."). But this is quite different than proclaiming, as common carriers typically do, that they accept all speech regardless of its content.
34. 512 U.S. at 652 (concluding that the federal law requiring cable companies to carry broadcast television programming was content neutral because it did not make facial content

distinctions and was "not designed to favor or disadvantage speech of any particular content" but instead was an attempt "to protect broadcast television from what Congress determined to be unfair competition by cable systems"). *See also id.* at 645 (noting that "[s]o long as they are not a subtle means of exercising a content preference," the speaker-based distinctions written into the federal law could be considered content neutral).

35. Eugene Volokh, *Treating Social Media Platforms Like Common Carriers?*, 1 J. FREE SPEECH L. 377, 402 (2021).
36. Buckley v. Valeo, 424 U.S. 1, 17 (1976) (rejecting the claim that laws regulating election spending count as incidental regulations of speech because "[e]ven if the categorization of the expenditure of money as conduct were accepted . . . [a]lthough the Act does not focus on the ideas expressed by persons or groups subject to its regulations, it is aimed in part at equalizing the relative ability of all voters to affect electoral outcomes by placing a ceiling on expenditures for political expression by citizens and groups. . . . [I]t is beyond dispute that the interest in regulating the alleged 'conduct' of giving or spending money 'arises in some measure because the communication allegedly integral to the conduct is itself thought to be harmful.'").
37. Citizens United v. F.E.C., 558 U.S. 310, 348, 350 (2010) (denying that the government has a legitimate interest in preventing economically powerful "corporations from obtaining an unfair advantage in the political marketplace by using resources amassed in the economic marketplace"). In fact, the constitutional argument in favor of campaign finance laws is much stronger than the argument in favor of extending common carrier obligations to the platforms. This is because the primary argument for campaign finance laws historically has been that they are necessary to prevent wealthy donors from monopolizing the time and attention of political representatives, not that they are necessary to prevent corporations from exercising undue influence over public debate about elections. A desire to prevent wealthy donors from monopolizing the time and attention of elected leaders does not presume anything about the quality or character of their contribution to public debate. Instead, it vindicates the basic political equality principle of "one person one vote." *See* David A. Strauss, *Corruption, Equality, and Campaign Finance Reform*, 94 COLUM. L. REV. 1369, 1382 (1994) (making this point). A similar equality rationale cannot be mobilized, however, to justify common carrier laws for the platforms. This is because it is not necessary to require the social media companies to accept all speech to protect the right of users to *access* the platforms without discrimination on the basis of their identity. To grant a basic right of access, all that is needed is the kind of public accommodation law I discuss in the section "What Can the Legislature Do?"
38. NetChoice, LLC v. Moody, No. 4:21cv220-RH-MAF, 2021 WL 2690876, at *10 (N.D. Fla. June 30, 2021) (concluding, for this reason, that "[t]he Florida [law is] about as content-based as it gets").
39. This not a matter of interpretation. The official Sponsor's Statement of Intent included in the Committee Report on the bill declared its purpose to be preventing "social media site[s] of over 100 million users . . . from censoring a person or the content that person posts based on the person's viewpoint or on the viewpoint expressed in the post." Committee Report, Texas Senate (Mar. 15, 2021), https://capitol.texas.gov/BillLookup/Text.aspx?LegSess=87R&Bill=SB12.
40. Press Release, A Stronger Online Economy: Opportunity, Innovation, and Choice (June 11, 2021), https://cicilline.house.gov/press-release/house-lawmakers-release-anti-monopoly-agenda-stronger-online-economy-opportunity (describing the purpose of the bill to be to prohibit discriminatory conduct by dominant platforms, including a ban on self-preferencing and picking winners and losers online); American Choice and Innovation Online Act, H.R. 3816 (117th Cong.) (2021–22) (allowing the prohibited discrimination as long as the social media company can show "by clear and convincing evidence that the [discrimination] would not result in harm to the competitive process by restricting or impeding legitimate activity by business users"). The pro-competition focus of the bill is also made evident by two other provisions, which prohibit social media companies from giving preference to their own "products, services or lines of business" over those of other platform users. *Id.*
41. Turner Broad. Sys., Inc. v. F.C.C., 512 U.S. 622, 652 (1994).
42. *Id.* at 642.

43. *Id.* at 662.
44. Ward v. Rock Against Racism, 491 U.S. 781, 798 (1989).
45. *See generally* Elena Kagan, *Private Speech, Public Purpose: The Role of Governmental Motive in First Amendment Doctrine*, 63 U. CHICAGO L. REV. 413 (1996).
46. W. Va. St. Bd. of Educ. v. Barnette, 319 U.S. 624, 642 (1943).
47. *Id.* at 641.
48. Proposal for a Regulation on a Single Market for Digital Services (Digital Services Act) (DSA) (Dec. 15, 2020), https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en#documents.
49. S. 797 (117th Cong.), https://www.congress.gov/bill/117th-congress/senate-bill/797/text (requiring social media companies that remove content believed to violate their rules in response to a user's complaint to "notify the [user responsible for the content] and explain why the content was removed" and "allow [the user responsible for the content] to appeal the decision"). *See also* Sonja Hutson, *Utah Lawmaker Planning a New Version of Vetoed Social Media Free Speech Bill*, KUER 90.1 (May 12, 2021), https://www.kuer.org/politics-government/2021-05-12/utah-lawmaker-planning-a-new-version-of-vetoed-social-media-free-speech-bill (discussing legislative plans to introduce a bill guaranteeing users of social media platforms a right of appeal in Utah).
50. Turner Broad. Sys., Inc. v. F.C.C., 512 U.S. 622, 645 (1994).
51. The PACT Act merely requires that a social media company review the decision to take down or otherwise deprioritize user content that is believed to violate the platform's rules and "determine whether the content adheres to the acceptable use policy" of the platform. It provides no guidance whatsoever to the platforms on how they should do this. S. 797, *supra* note 49. In contrast, the DSA provides platform users the right to file "sufficiently precise and adequately substantiated complaints . . . electronically and free of charge" when the social media company removes their speech, suspends service, or suspends or terminates their account. DSA, *supra* note 48, at art. 17. It requires the platforms to "handle complaints submitted through their internal complaint-handling system in a timely, diligent and objective manner" and "inform complainants . . . of [their] decision"). *Id.* It also grants users a right to have their complaint adjudicated by an "out-of-court dispute settlement body" that has been certified to be "impartial and independent," to have "necessariy expertise," and be "capable of settling [the] dispute in a swift, efficient and cost-effective manner." *Id.*, art. 18. To prevent costs from disincentivizing users' willingness to take their disputes to these forums, the DSA also provides that in cases where the complaint is decided in the user's favor, "the online platform shall reimburse the recipient for any fees and other reasonable expenses that the recipient has paid or is to pay in relation to the dispute settlement." *Id.*
52. 42 U.S.C. § 2000a. Federal law also makes it unlawful to "discriminate[] against [any person] on the basis of disability in the full and equal enjoyment of the goods, services, facilities, privileges, advantages, or accommodations of any place of public accommodation." 12 U.S.C. § 12182.
53. For example, the public accommodations law of Illinois declares it unlawful to "refuse to another the full and equal enjoyment of the facilities, goods, and services of any public place of accommodation" because of their "actual or perceived: race, color, religion, national origin, ancestry, age, sex, marital status, order of protection status, disability, military status, sexual orientation, pregnancy, or unfavorable discharge from military service." 775 Ill. Comp. Stat. Ann. §§ 5/5-102, 5/5-103.
54. *See, e.g.*, Clarke & Brody, *supra* note 10; Nancy Leong & Aaron Belzer, *The New Public Accommodations: Race Discrimination in the Platform Economy*, 105 GEO. L.J. 1271 (2017).
55. DANIELLE KEATS CITRON, HATE CRIMES IN CYBERSPACE 56–72 (Cambridge, MA: Harvard University Press, 2014).

56. *See, e.g.*, Heart of Atlanta Motel, Inc. v. United States, 379 U.S. 241, 250 (1964) (the "fundamental objective" of public accommodations law is to prevent "the deprivation of personal dignity that surely accompanies denials of equal access to public establishments").
57. Hurley v. Irish-Am. Gay, Lesbian and Bisexual Grp. of Boston, Inc., 515 U.S. 557, 572–73, (1995) (noting that, while "[public accommodations laws] are well within the State's usual power to enact when a legislature has reason to believe that a given group is the target of discrimination, and . . . do not, as a general matter, violate the First or Fourteenth Amendments . . . [or] target speech or discriminate on the basis of its content," a law that prohibits parade organizers from deciding what groups may participate in the parade, under what banners, violates the organizers' First Amendment right of autonomy to choose the content of his own message and, conversely, to decide what not to say).
58. Although disparate impact is a more common standard of discrimination in the employment discrimination context, courts apply it in public accommodations cases also. *See* Leong & Belzer, *supra* note 54, at 1309.
59. *Id.*
60. Ángel díaz & laura hecht-felella, double standards in social media content moderation (New York: Brennan Center for Justice, Aug. 4, 2021) [hereinafter double standards].
61. *See* Facebook Community Standards, Twitter Rules, and YouTube's Community Guidelines, *supra* note 32.
62. Roberts v. U.S. Jaycees, 468 U.S. 609, 624 (1984).
63. double standards, *supra* note 60, at 19 ("Oftentimes the platforms will identify . . . content as violating a broad, general policy, such as 'hate speech,' without providing information on what part of the rule was violated.").

Chapter 12

1. *See* Nixon v. Herndon, 27 U.S. 536 (1927); Smith v. Allwright, 321 U.S. 649 (1944); Terry v. Adams, 345 U.S. 461 (1953).
2. *See* Marsh v. Alabama, 326 U.S. 501 (1945).
3. Mitchell v. United States, 313 U.S. 80, 97 (1941).
4. Pa. Publ'ns v. Pa. Pub. Util. Comm'n, 36 A.2d 777, 781 (Pa. 1944). *See generally* Angela J. Campbell, *Publish or Carriage: Approaches to Analyzing the First Amendment Rights of Telephone Companies*, 70 n.c. l. rev. 1071 (1992).
5. FedEx Corp. v. United States, 121 F. App'x 125, 126 (6th Cir. 2005).
6. *See* Biden v. Knight First Amendment. Inst. at Columbia Univ., 141 S. Ct. 1220 (2021) (*cert. denied*) (Thomas, J., concurring).
7. *See* Heart of Atlanta Motel, Inc. v. United States, 379 U.S. 241 (1964).
8. PruneYard Shopping Ctr. v. Robins, 447 U.S. 74, 77 (1980).
9. Rumsfeld v. F. for Acad. & Institutional Rts., Inc., 547 U.S. 47 (2006).
10. Jackson v. Metropolitan Edison Co., 419 U.S. 345 (1974); Pac. Gas & Elec. v. Pub. Utils. Comm'n of California, 475 U.S. 1 (1986).
11. Red Lion Broad. Co. v. F.C.C., 395 U.S. 367 (1969).
12. Turner Broad. Sys., Inc. v. F.C.C. (*Turner I*), 512 U.S. 622, 684 (1994); Turner Broad. Sys., Inc. v. F.C.C. (*Turner II*), 520 U.S. 180 (1994).
13. Miami Herald Pub. Co. v. Tornillo, 418 U.S. 241 (1974).
14. Citizens United v. F.E.C., 558 U.S. 310 (2010).
15. Packingham v. North Carolina, 137 S. Ct. 1730, 1732 (2017).
16. *Id.* at 1737 (quoting Reno v. ACLU, 521 U.S. 844, 870 (1997)).
17. *Twitter: Transparency and Accountability: Hearing Before the H. Comm. on Energy & Com.*, 115th Cong. 1 (2018) (testimony of Jack Dorsey, CEO, Twitter, Inc.), https://docs.house.gov/meetings/IF/IF00/20180905/108642/HHRG-115-IF00-Wstate-DorseyJ-20180905.pdf ("Twitter is used as a global town square, where people from around the world come together in an open and free exchange of ideas"), https://perma.cc/UJ89-935S; Mark Zuckerberg, *A Privacy-Focused Vision for Social Networking*, facebook (March 6, 2019), https://www.facebook.com/notes/markzuckerberg/a-privacy-focused-vision-for-soc

ial-networking/10156700570096634 ("Over the last 15 years, Facebook and Instagram have helped people connect with friends, communities, and interests in the digital equivalent of a town square."), https://perma.cc/2FS4-XS8H.

18. *See generally* Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598 (2018).
19. *See* evelyn douek (@evelyndouek), TWITTER (Sept. 30, 2020, 11:49 AM), https://twitter.com/evelyndouek/status/1311377706408435713?s=20.
20. Matthew Prince, *Why We Terminated Daily Stormer*, CLOUDFARE (Aug. 16, 2017), https://blog.cloudflare.com/why-we-terminated-daily-stormer.
21. Jackson v. Metro. Edison Co., 419. U.S. 345, 352 (1974).
22. Complaint for Injunctive and Declaratory Relief ¶ 3, at 2, Trump v. Facebook, Inc., No. 1:21-cv-22440-KMW, 2021 WL 2826779 (S.D. Fla. July 7, 2021) [hereinafter Complaint, Trump v. Facebook]; Complaint for Injunctive and Declaratory Relief ¶ 3, at 2, Trump v. YouTube Inc., No. 21-cv-22445-KMM, 2021 WL 2860054 (S.D. Fla. July 7, 2021) [hereinafter Complaint, Trump v. YouTube]; Complaint for Injunctive and Declaratory Relief ¶ 3, at 2, Trump v. Twitter, No. 1:21-cv-22441 (S.D. Fla. July 7, 2021) [hereinafter Complaint, Trump v. Twitter], https://www.wsj.com/media/TrumpvTwitter.pdf?mod=article_inline.
23. *See* Complaint, Trump v. Facebook, *supra* note 22, ¶ 26, at 6; Complaint, Trump v. Twitter, *supra* note 22, ¶ 24, at 5.
24. *See* Complaint, Trump v. Facebook, *supra* note 22, ¶¶ 167–69, at 38–39; COMPLAINT, TRUMP V. YOUTUBE, *supra* note 22, ¶¶ 147–49, at 34–35; Complaint, Trump v. Twitter, *supra* note 22, ¶¶ 131–33, at 29.
25. Complaint, Trump v. Facebook, *supra* note 22, ¶ 150, at 36; Complaint, Trump v. Twitter, *supra* note 22, ¶ 114, AT 26; COMPLAINT, TRUMP V. YOUTUBE, *supra* note 22, ¶ 130, at 32.
26. Genevieve Lakier, *Informal Government Coercion and the Problem of "Jawboning,"* LAWFARE (July 26, 2021, 3:52 PM), https://www.lawfareblog.com/informal-government-coercion-and-problem-jawboning.
27. Communications Decency Act of 1996, 47 U.S.C. § 230(c)(2)(A).
28. Note also that in the absence of the immunities Section 230 provides, the platforms would need to censor even more. They would not know, in many cases, if a given speech might be libelous, but, aware that they would be on the hook for damages if it were, they would likely need to censor borderline content that should be constitutionally protected.
29. See *infra* notes 39–41 and accompanying text.
30. *See* 2021 Fla. Laws 32, http://laws.flrules.org/2021/32.
31. NetChoice, LLC et al. v. Moody, No. 4:21CV220-RH-MAF, 2021 WL 2690876, at *1 (N.D. Fla. June 30, 2021).
32. 2021 Fla. Laws 32, at 11, http://laws.flrules.org/2021/32.
33. *Id.* at 10.
34. *NetChoice, LLC*, 2021 WL 2690876, at *8.
35. *See* Press Release, Ron DeSantis, Governor, Florida, *Governor Ron DeSantis Signs Bill to Stop the Censorship of Floridians by Big Tech* (May 24, 2021), https://www.flgov.com/2021/05/24/governor-ron-desantis-signs-bill-to-stop-the-censorship-of-floridians-by-big-tech.
36. *Knight First Amendment. Inst.* at 1224–25 (2021).
37. *See* Philip Hamburger, *The Constitution Can Crack Section 230*, WALL ST. J. (Jan. 29, 2021), https://www.wsj.com/articles/the-constitution-can-crack-section-230-11611946851; Philip Hamburger & Clare Morell, *The First Amendment Doesn't Protect Big Tech's Censorship*, WALL ST. J. (July 31, 2021), https://www.wsj.com/articles/big-tech-twitter-facebook-google-youtube-sec-230-common-carrier-11627656722.
38. Eugene Volokh, *Treating Social Media Platforms like Common Carriers?*, 1 J. FREE SPEECH L. 377 (forthcoming 2021).
39. 418 U.S. 241 (1974).
40. Craig Silverman, *Facebook Says Its Fact Checking Program Helps Reduce the Spread of a Fake Story by 80%*, BUZZFEED NEWS (OCT. 11, 2017), https://www.buzzfeednews.com/article/craigsilverman/facebook-just-shared-the-first-data-about-how-effective-its.

41. Joan E. Solsman, *Youtube's AI Is the Puppet Master Over Most of What You Watch*, CNET (Jan. 10, 2018, 10:05 AM), https://www.cnet.com/news/youtube-ces-2018-neal-mohan.
42. Renée DiResta, *Free Speech Is Not the Same as Free Reach*, WIRED (Aug. 30, 2018, 4:00 PM), https://www.wired.com/story/free-speech-is-not-the-same-as-free-reach.
43. *Political Content*, TWITTER, https://business.twitter.com/en/help/ads-policies/ads-content-policies/political-content.html (last visited Aug. 28, 2021) (prohibiting the promotion of political content in Twitter's paid advertising products).
44. *See* Mark Zuckerberg, FACEBOOK (Sept. 3, 2020), https://www.facebook.com/zuck/posts/10112270823363411; Elizabeth Dwoskin et al., *Facebook Tries to Head off Election Turmoil, Angering Both Trump and Democrats*, WASH. POST (Sept. 3, 2020, 6:50 PM), https://www.washingtonpost.com/technology/2020/09/03/facebook-political-ads.
45. *Advertising Policies*, FACEBOOK, https://www.facebook.com/policies/ads/ (last visited Aug. 14, 2021) (prohibiting, among other things, profanity and bad grammar, "before-and-after" images, promotion of payday loans, paycheck advances, or bail bonds, vaccine discouragement, and ads with nonexistent functionality like checkboxes or multiple-choice questions that do not work).
46. *Advertising Policies: Sensational Content*, FACEBOOK, https://www.facebook.com/policies/ads/prohibited_content/sensational_content (last visited Aug. 14, 2021).
47. *Advertising Policies: Low Quality or Disruptive Content*, FACEBOOK, https://www.facebook.com/policies/ads/prohibited_content/low_quality_or_disruptive_experiences (last visited Aug. 14, 2021).
48. *See* Daphne Keller, *Amplification and Its Discontents*, KNIGHT FIRST AMENDMENT. INST. AT COLUMBIA UNIV. (June 8, 2021), https://knightcolumbia.org/content/amplification-and-its-discontents.
49. *See generally Community Standards Enforcement Report: Spam*, FACEBOOK, https://transparency.fb.com/data/community-standards-enforcement/spam/facebook (last visited Aug. 28, 2021); *Community Standards Enforcement Report: Fake Accounts*, FACEBOOK, https://transparency.fb.com/data/community-standards-enforcement/fake-accounts/facebook (last visited Aug. 28, 2021).
50. *See* McIntyre v. Ohio Elections Comm'n, 514 U.S. 334, 357 (1995).
51. *See* R.A.V. v. City of St. Paul, 505 U.S. 377, 391–92 (1992).
52. *See* United States v. Alvarez, 567 U.S. 709, 722–23 (2012).
53. *See* United States v. Stevens, 559 U.S. 460 (2010); Brown v. Entm't Merchs. Ass'n, 564 U.S. 786 (2011).
54. *See* Reno v. ACLU, 521 U.S. 844 (1997); Erznoznik v. City of Jacksonville, 422 U.S. 205 (1975).
55. *Facebook Community Standards: Hate Speech*, FACEBOOK, https://www.facebook.com/communitystandards/hate_speech (last visited Aug. 14, 2021).
56. *See id.* (prohibiting exclusionary targeting of groups with protected characteristics such as race, ethnicity, national origin, disability, religious affiliation, caste, sexual orientation, sex, gender identity, and serious disease).
57. *See* Near v. Minnesota *ex rel.* Olson, 283 U.S. 697 (1931).
58. PAUL M. BARRETT, WHO MODERATES THE SOCIAL MEDIA GIANTS? A CALL TO END OUTSOURCING (New York: Stern Center for Business and Human Rights, June 2020), https://issuu.com/nyusterncenterforbusinessandhumanri/docs/nyu_content_moderation_report_final_version?fr=sZWZmZjI1NjI1Ng.
59. *See* Kate Klonick, *Inside the Team at Facebook That Dealt with the Christchurch Shooting*, NEW YORKER (Apr. 25, 2019), https://www.newyorker.com/news/news-desk/inside-the-team-at-facebook-that-dealt-with-the-christchurch-shooting.
60. *See* David Kaye (Special Rapporteur), *Promotion and Protection of the Right to Freedom of Opinion and Expression*, U.N. DOC. A/74/486 (2019), https://freedex.org/wp-content/blogs.dir/2015/files/2019/11/UNGA2019-Hate-Speech.pdf.
61. *Oversight Board Charter*, FACEBOOK (2019), https://about.fb.com/wp-content/uploads/2019/09/oversight_board_charter.pdf.
62. Case Decision 2020-004-IG-UA, OVERSIGHT BOARD (Jan. 28, 2021), https://www.oversightboard.com/decision/IG-7THR3SI1.

63. Case Decision 2020-006-FB-FBR, OVERSIGHT BOARD (Jan. 28, 2021), https://www.oversightboard.com/decision/FB-XWJQBU9A.
64. *Id.*
65. *See* evelyn douek, *Governing Online Speech: From "Posts-as-Trumps" to Proportionality and Probability*, 121 COLUM. L. REV. 759 (2021).
66. *See* Daphne Keller, *Who Do You Sue? State and Platform Hybrid Power Over Online Speech*, HOOVER INST. (Jan. 29, 2019) https://www.hoover.org/research/who-do-you-sue.
67. *See* Nate Persily, *An Act* (Oct. 5, 2021), https://tinyurl.com/persilylegislation.
68. Daphne Keller, *A Quick Take on the Final Executive Order on Platforms and CDA 230* (June 2, 2020), https://docs.google.com/document/d/1j_ZU9OUkvA5teWSC2_iQx61eB1rzbTjN55AqzmIP1c0/edit#.
69. *See* Exec. Order No. 13925, 85 Fed. Reg. 34,079 (May 28, 2020).
70. *See* evelyn douek, *The Rise of Content Cartels*, KNIGHT FIRST AMENDMENT INST. AT COLUM. UNIV. (Feb. 11, 2020), https://knightcolumbia.org/content/the-rise-of-content-cartels.
71. An important set of considerations for a platform or oversight body, therefore, should be how dominant the platform is in the information environment. In countries, like Myanmar, where Facebook *is* the internet, a different calculus might be in order.
72. *See* ANU BRADFORD, THE BRUSSELS EFFECT: HOW THE EUROPEAN UNION RULES THE WORLD (New York: Oxford Scholarship Online, 2019).
73. For a more developed version of this argument, *see* Nathaniel Persily & Joshua A. Tucker, *Conclusion: The Challenges and Opportunities for Social Media Research*, *in* SOCIAL MEDIA AND DEMOCRACY: THE STATE OF THE FIELD AND PROSPECTS FOR REFORM, 313–31 (Nathaniel Persily & Joshua A. Tucker eds., Cambridge: Cambridge University Press, 2020).

Chapter 13

1. THOMAS RID, ACTIVE MEASURES: THE SECRET HISTORY OF DISINFORMATION AND POLITICAL WARFARE (New York: Farrar, Straus and Giroux, 2020).
2. Monmouth University Polling Institute, *Majority Back Capitol Riot Commission* (West Long Branch, NJ: Monmouth University, Mar. 17, 2021), https://www.monmouth.edu/polling-institute/reports/MonmouthPoll_US_031721/.
3. *E.g.*, JONATHAN CORPUS ONG & JASON VINCENT A. CABAÑES, ARCHITECTS OF NETWORKED DISINFORMATION: BEHIND THE SCENES OF TROLL ACCOUNTS AND FAKE NEWS PRODUCTION IN THE PHILIPPINES (London: Newton Tech4Dev Network, 2018); Anupam Das & Ralph Schroeder, *Online Disinformation in the Run-Up to the Indian 2019 Election*, 24(12) INFO., COMMS. & SOC'Y 1762–78 (2020); Felipe Bonow Soares & Raquel Recuero, *Hashtag Wars: Political Disinformation and Discursive Struggles on Twitter Conversations During the 2018 Brazilian Presidential Campaign*, 7(2) SOC. MEDIA + SOC'Y (2021), https://doi.org/10.1177/20563051211009073; Samuel C. Woolley & Philip N. Howard, *Computational Propaganda Worldwide: Executive Summary* (2017), https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Casestudies-ExecutiveSummary.pdf.
4. Joseph B. Bak-Coleman et al., *Stewardship of Global Collective Behavior*, 118(27) PROC. OF NAT'L ACAD. OF SCI. e2025764118 (2021).
5. Henry Farrell & Bruce Schneier, *Common-Knowledge Attacks on Democracy* (Berkman Klein Center Research Publication 2018-7, 2018).
6. Kate Starbird et al., *Disinformation as Collaborative Work: Surfacing the Participatory Nature of Strategic Information Operations*, 3(CSCW) PROC. OF ACM ON HUMAN-COMPUTER INTERACTION 1 (2019).
7. Caroline Jack, *Lexicon of Lies: Terms for Problematic Information*, 3(22) DATA & SOC'Y 1094 (2017).
8. Claire Wardle & Hossein Derakhshan, *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making*, 27 Council of Europe Report DGI (2017)..
9. LADISLAV BITTMAN, THE KGB AND SOVIET DISINFORMATION: AN INSIDER'S VIEW (Washington, DC: Pergamon-Brassey's, 1985); RID, *supra* note 1.
10. BITTMAN, *supra* note 9, at 145.

11. SINAN ARAL, THE HYPE MACHINE: HOW SOCIAL MEDIA DISRUPTS OUR ELECTIONS, OUR ECONOMY, AND OUR HEALTH—AND HOW WE MUST ADAPT (New York: Currency, 2020).
12. ROBERT S. MUELLER, THE MUELLER REPORT: REPORT ON THE INVESTIGATION INTO RUSSIAN INTERFERENCE IN THE 2016 PRESIDENTIAL ELECTION (Washington, DC: WSBLD, 2019); U.S. SENATE. SELECT COMMITTEE ON INTELLIGENCE, RUSSIAN ACTIVE MEASURES CAMPAIGNS AND INTERFERENCE IN THE 2016 U.S. ELECTION, VOLUME 2: RUSSIA'S USE OF SOCIAL MEDIA WITH ADDITIONAL VIEWS (Washington, DC: Government Publishing Office, 2019).
13. Thomas Boghardt, *Soviet Bloc Intelligence and Its AIDS Disinformation Campaign*, 53(4) STUD. IN INTELLIGENCE 1 (2009); Douglas Selvage, *Operation "Denver": The East German Ministry of State Security and the KGB's AIDS Disinformation Campaign, 1985–1986 (Part 1)*, 21(4) J. COLD WAR STUD. 71 (2019).
14. RID, *supra* note 1.
15. BITTMAN, *supra* note 9, at 43.
16. RID, *supra* note 1.
17. *E.g.*, MUELLER, *supra* note 12; Starbird et al., *supra* note 6; *Threat Report: The State of Influence Operations 2017–2020*, FACEBOOK (May 20, 2021), https://about.fb.com/wp-content/uploads/2021/05/IO-Threat-Report-May-20-2021.pdf.
18. ONG & CABAÑES, *supra* note 3. Renée DiResta, *Dancing in the Dark: Disinformation Researchers Need More Robust Data and Partnerships*, National Endowment for Democracy, Global Insight Series 39, 40 (Jan. 2021).
19. *See, e.g.*, Shimon Kogan, Tobias J. Moskowitz, & Marina Niessner, *Social Media and Financial News Manipulation*, SSRN (Sept. 15, 2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3237763.
20. Jane Lytvynenko, *A Televangelist Peddling Silver as a Fake Cure for Coronavirus Has Been Warned by Officials to Stop*, BUZZFEED NEWS (Mar. 10, 2020), https://www.buzzfeednews.com/article/janelytvynenko/ftc-fda-scam-coronavirus-cures; Neena Satija & Lena H. Sun, *A Major Funder of the Anti-Vaccine Movement Has Made Millions Selling Natural Health Products*, WASH. POST. (Dec. 20, 2019), https://www.washingtonpost.com/investigations/2019/10/15/fdc01078-c29c-11e9-b5e4-54aa56d5b7ce_story.html.
21. ANDREW MARANTZ, ANTISOCIAL: ONLINE EXTREMISTS, TECHNO-UTOPIANS, AND THE HIJACKING OF THE AMERICAN CONVERSATION (New York: Penguin, 2020).
22. Kate Starbird, *Examining the Alternative Media Ecosystem Through the Production of Alternative Narratives of Mass Shooting Events on Twitter*, 11(1) PROC. OF INT'L AAAI CONF. ON WEB & SOC. MEDIA (2017), https://www.semanticscholar.org/paper/Examining-the-Alternative-Media-Ecosystem; Kate Starbird et al., *Ecosystem or Echo-System? Exploring Content Sharing Across Alternative Media Domains*, 12(1) PROC. OF INT'L AAAI CONF. ON WEB & SOC. MEDIA (2018), https://faculty.washington.edu/kstarbi/Starbird-et-al-ICWSM-2018-Echosystem-final.pdf.
23. BITTMAN, *supra* note 9.
24. *See, e.g.*, Sheera Frenkel, *A Freelance Writer Learns He Was Working for the Russians*, N.Y. TIMES (Sept. 2, 2020), https://www.nytimes.com/2020/09/02/technology/peacedata-writer-russian-misinformation.html; Sheera Frenkel & Julian E. Barnes, *Russians Again Targeting Americans with Disinformation, Facebook and Twitter Say*, N.Y. TIMES (Sept. 1, 2020), https://www.nytimes.com/2020/09/01/technology/facebook-russia-disinformation-election.html.
25. RID, *supra* note 1.
26. Ahmer Arif et al., *Acting the Part: Examining Information Operations Within #BlackLivesMatter Discourse*, 2(CSCW) PROC. OF ACM ON HUMAN-COMPUTER INTERACTION 1 (2018); Starbird et al., *supra* note 6; Tom Wilson & Kate Starbird, *Cross-Platform Disinformation Campaigns: Lessons Learned and Next Steps*, 1(1) HARVARD KENNEDY SCH. MISINFORMATION REV. (2020), https://misinforeview.hks.harvard.edu/article/cross-platform-disinformation-campaigns/.
27. *E.g.*, Mike Isaac & Daisuke Wakabayashi, *Russian Influence Reached 126 Million Through Facebook Alone*, N.Y. TIMES (Oct. 30, 2017), https://www.nytimes.com/2017/10/30/technology/facebook-google-russia.html; Nicholas Thompson & Issie Lapowsky, *How Russian Trolls Used Meme Warfare to Divide America*, WIRED (Dec. 17, 2018), https://www.wired.com/story/russia-ira-propaganda-senate-report/.

28. MUELLER, *supra* note 12; U.S. SENATE SELECT COMMITTEE, *supra* note 12.
29. Jen Weedon et al., *Information Operations and Facebook*, FACEBOOK (2017), https://about.fb.com/wp-content/uploads/2017/04/facebook-and-information-operations-v1.pdf; *Update on Twitter's Review of the 2016 US Election*, TWITTER BLOG (Jan. 19, 2018), https://blog.twitter.com/official/en_us/topics/company/2018/2016-election-update.html.
30. *Removing Bad Actors on Facebook*, FACEBOOK (July 31, 2018), https://about.fb.com/news/2018/07/removing-bad-actors-on-facebook/.
31. SAMANTHA BRADSHAW ET AL., THE LONG FUSE: MISINFORMATION AND THE 2020 ELECTION (Stanford, CA: Center for an Informed Public, Digital Forensic Research Lab, Graphika, & Stanford Internet Observatory, 2021).
32. *Id.*
33. Monmouth University Polling Institute, *supra* note 2.
34. Yochai Benkler et al., *Mail-In Voter Fraud: Anatomy of a Disinformation Campaign* (Berkman Center Research Publication No. 2020-6, Oct. 2, 2020), http://dx.doi.org/10.2139/ssrn.3703701.
35. BRADSHAW ET AL., *supra* note 31.
36. *Id.*
37. YOCHAI BENKLER, THE WEALTH OF NETWORKS: HOW SOCIAL PRODUCTION TRANSFORMS MARKETS AND FREEDOM (New Haven, CT: Yale University Press, 2006).
38. Joe Miller, *Zuckerberg: Facebook's Mission Is to "Connect the World,"* BBC NEWS (Feb. 24, 2014), https://www.bbc.com/news/technology-26326844.
39. Alicia Wanless & Michael Berk, *The Audience Is the Amplifier: Participatory Propaganda*, *in* THE SAGE HANDBOOK OF PROPAGANDA 85 (Paul Baines, Nicholas O'Shaughnessy, & Nancy Snow eds., London: SAGE, 2019).
40. Starbird et al., *supra* note 6.
41. Letter to Maricopa County Voters from Maricopa County, Board of Supervisors (Nov. 4, 2020), https://bloximages.newyork1.vip.townnews.com/azfamily.com/content/tncms/assets/v3/editorial/6/79/6794518e-1ef5-11eb-b43e-e75aabde9c1d/5fa3391e361b0.pdf.pdf.
42. Rosalind S. Helderman, *A Cybersecurity Expert Who Promoted Claims of Fraud in the 2020 Election Is Leading the GOP-Backed Recount of Millions of Ballots in Arizona*, WASH. POST (Apr. 16, 2021), https://www.washingtonpost.com/politics/arizona-recount-cyber-ninjas/2021/04/16/b3cbbf08-996e-11eb-962b-78c1d8228819_story.html.
43. *E.g.*, PETER POMERANTSEV & MICHAEL WEISS, THE MENACE OF UNREALITY: HOW THE KREMLIN WEAPONIZES INFORMATION, CULTURE AND MONEY. VOL. 14 (New York: Institute of Modern Russia, 2014); Spencer McKay & Chris Tenove, *Disinformation as a Threat to Deliberative Democracy*, 74(3) POL. RES. Q. 703 (2021); Farrell & Schneier, *supra* note 5.
44. Pomerantsev & Weiss, *supra* note 43.
45. Farrell & Schneier, *supra* note 5.
46. RID, *supra* note 1.
47. PETER L. BERGER & THOMAS LUCKMANN, THE SOCIAL CONSTRUCTION OF REALITY: A TREATISE IN THE SOCIOLOGY OF KNOWLEDGE (New York: Anchor, 1966).
48. RID, *supra* note 1, at 427–28.
49. YOCHAI BENKLER ET AL., NETWORK PROPAGANDA: MANIPULATION, DISINFORMATION, AND RADICALIZATION IN AMERICAN POLITICS (New York: Oxford University Press, 2018); Benkler et al., *supra* note 34.
50. *E.g.*, BRADSHAW ET AL., *supra* note 31; MUELLER, *supra* note 12.
51. ARAL, *supra* note 11; Starbird et al., *supra* note 6.
52. ANTHONY GIDDENS, THE CONSTITUTION OF SOCIETY: OUTLINE OF THE THEORY OF STRUCTURATION (Berkeley: University of California Press, 1986).
53. Gregory Berns et al., *Neural Mechanisms of the Influence of Popularity on Adolescent Ratings of Music*, 49(3) NEUROIMAGE 2687 (2010).
54. Savvas Zannettou, *"I Won the Election!": An Empirical Analysis of Soft Moderation Interventions on Twitter*, 15 PROC. OF INT'AL AAAI CONF. ON WEB & SOC. MEDIA 865 (2021).
55. Arif et al., *supra* note 26.
56. Alessandro Bessi & Emilio Ferrara, *Social Bots Distort the 2016 US Presidential Election Online Discussion*, 21(11) FIRST MONDAY (2016), https://ssrn.com/abstract=2982233;

Tuğrulcan Elmas et al., *Ephemeral Astroturfing Attacks: The Case of Fake Twitter Trends*, ARXIV PREPRINT arXiv:1910.07783 (2019); Samuel C. Woolley, *Bots and Computational Propaganda: Automation for Communication and Control, in* SOCIAL MEDIA AND DEMOCRACY: THE STATE OF THE FIELD, PROSPECTS FOR REFORM 89 (Nathaniel Persily & Joshua A. Tucker eds., Cambridge: Cambridge University Press, 2020).

57. Erin Gallagher, *Trump Trains*, MEDIUM (Sept. 15, 2019), https://erin-gallagher.medium.com/trump-trains-84bea1c3170d; Christopher Torres-Lugo et al., *The Manufacture of Political Echo Chambers by Follow Train Abuse on Twitter*, ARXIV E-PRINTS arXiv-2010 (2020).
58. Ryan Mac et al., *Facebook Stopped Employees From Reading an Internal Report About Its Role in the Insurrection. You Can Read It Here*, BUZZFEED NEWS (Apr. 26, 2021), https://www.buzzfeednews.com/article/ryanmac/full-facebook-stop-the-steal-internal-report.
59. BRADSHAW ET AL., *supra* note 31.
60. Micah Curtis, *Team Biden & CNN's Criticism of Facebook Not Censoring Trump BEFORE the Election Is Wholly Un-American*, RT.COM (June 28, 2021), https://www.rt.com/op-ed/527785-biden-cnn-facebook-trump.
61. Guy Birchall, *"We Want to Delete More!" Facebook Must Choose Freedom of Speech Over Censorship, If It Still Pretends to Be "a Platform,"* RT.COM (Aug. 4, 2020), https://www.rt.com/op-ed/497001-facebook-advertisers-hate-speech-moderators.
62. Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598 (2017).
63. Bak-Coleman et al., *supra* note 4.
64. BITTMAN, *supra* note 9, at 70.
65. *Id.* at 75.
66. Frenkel, *supra* note 24; Frenkel & Barnes, *supra* note 24.
67. DAN GILLMOR, WE THE MEDIA: GRASSROOTS JOURNALISM BY THE PEOPLE, FOR THE PEOPLE (Sebastapol, CA: O'Reilly Media, 2004).
68. BRADSHAW ET AL., *supra* note 31.
69. *The Disinformation Dozen: Why Platforms Must Act on Twelve Leading Online Anti-Vaxxers* (Center for Countering Digital Hate, 2021), https://www.counterhate.com/disinformationdozen.
70. Jeff Horwitz, *Facebook Says Its Rules Apply to All. Company Documents Reveal a Secret Elite That's Exempt*, WALL ST. J. (Sept. 13, 2021), https://www.wsj.com/articles/facebook-files-xcheck-zuckerberg-elite-rules-11631541353.
71. Renée DiResta, *Free Speech Is Not the Same as Free Reach*, WIRED.COM (Aug. 30, 2018), https://www.wired.com/story/free-speech-is-not-the-same-as-free-reach.
72. ROBERT E. KRAUT & PAUL RESNICK, BUILDING SUCCESSFUL ONLINE COMMUNITIES: EVIDENCE-BASED SOCIAL DESIGN (Cambridge, MA: MIT Press, 2012).

Chapter 14

1. *See* Danielle Keats Citron & Mary Anne Franks, *Criminalizing Revenge Porn*, 49 WAKE FOREST L. REV. 345 (2014).
2. *See, e.g.*, National Inst. of Family & Life Advocates v. Becerra, 138 S. Ct. 2361 (2018) (restricting regulation of compelled commercial speech and asserting that there are no special doctrines for the regulation of professional speech); Reed v. Town of Gilbert, 576 U.S. 155 (2015) (holding that all content-based regulations, regardless of regulatory context, are subject to strict scrutiny); Sorrell v. IMS Health Inc., 564 U.S. 552 (2011) (holding data privacy regulation unconstitutional and suggesting that content-based regulations of commercial speech are disfavored); *see also* U.S. Telecom Ass'n v. FCC, 855 F.3d 381, 418 (D.C. Cir. 2017) (Kavanaugh, J., dissenting from denial of rehearing en banc) (arguing that network neutrality rules are unconstitutional).
3. YOCHAI BENKLER, ROBERT FARIS, & HAL ROBERTS, NETWORK PROPAGANDA: MANIPULATION, DISINFORMATION, AND RADICALIZATION IN AMERICAN POLITICS (New York: Oxford Scholarship Online, 2018).
4. FCC v. Pacifica Found., 438 U.S. 726, 748 (1978) (plurality opinion).
5. *Id.* at 749–50.

6. For a list of key players and functions in basic internet services, see Matthew Prince, *Why We Terminated Daily Stormer*, CLOUDFLARE (Aug. 16, 2017), https://blog.cloudflare.com/why-we-terminated-daily-stormer [https://perma.cc/8QKY-GK8H]; see also JAMES GRIMMELMANN, INTERNET LAW: CASES & PROBLEMS 27–35 (8th ed. 2018) (describing elements of the internet "stack"); Jack M. Balkin, *Old-School/New-School Speech Regulation*, 127 HARV. L. REV. 2296, 2303–04 (2014) (listing elements of the digital infrastructure of free expression); *Free Speech: Only as Strong as the Weakest Link*, ELEC. FRONTIER FOUND., https://www.eff.org/free-speech-weak-link [https://perma.cc/X2KB-WXB5] (describing elements of digital infrastructure).
7. *See* US Telecom Ass'n v. FCC, 825 F.3d 674 (D.C. Cir. 2016) (upholding network neutrality rules).
8. Biden v. Knight First Amendment Inst. at Columbia Univ., 141 S. Ct. 1220 (2021) (Thomas, J., concurring).
9. Associated Press v. United States, 326 U.S. 1, 20 (1945).
10. Tarleton Gillespie, *Platforms Are Not Intermediaries*, 2 GEO. L. TECH. REV. 198 (2018).
11. Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1618–1630 (2018).
12. SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (New York: Public Affairs, 2019).
13. Although this essay is directed to the American context, the largest social media companies operate around the world and cause harms around the world. Among the most notorious examples is the use of Facebook's applications to promote genocide in Myanmar. See Dan Milmo, *Rohingya sue Facebook for £150bn over Myanmar genocide*, GUARDIAN ONLINE, Dec 6 2021, https://www.theguardian.com/technology/2021/dec/06/rohingya-sue-facebook-myanmar-genocide-us-uk-legal-action-social-media-violence [https://perma.cc/PDY6-L4H7].

    Facebook devotes a disproportionate amount of its resources for content moderation to America and Europe, probably because these are wealthy and mature markets for advertising and because the threat of regulation is greatest there. But "[w]ith growth largely stalled [in theUnited States] and in Europe, nearly all of Facebook's new users are coming from developing countries, where Facebook is the main online communication channel and source of news." See, e.g., Justin Scheck, Newley Purnell, and Jeff Horwitz, *Facebook Employees Flag Drug Cartels and Human Traffickers. The Company's Response Is Weak, Documents Show*, WALL STREET JOURNAL, Sept. 16, 2021, https://www.wsj.com/articles/facebook-drug-cartels-human-traffickers-response-is-weak-documents-11631812953 [https://perma.cc/2EV2-XQTD]. Facebook's content moderation services are even more lax and haphazard in the rest of the world than they are in the United States and Europe, and predictably, harms may be even greater there. See *id.* ("Activists have complained for years that Facebook does too little to protect overseas users from trouble it knows occurs on its platform. The documents show that many within Facebook agree.").
14. JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM (New York: Oxford University Press, 2019).
15. *See, e.g.,* C. Edwin Baker, Media Concentration and DEMOCRACY: WHY OWNERSHIP MATTERS (2006); FCC v. National Citizens Comm. for Broad., 436 U.S. 775, 780 (1978) ("The Commission has long acted on the theory that diversification of mass media ownership serves the public interest by promoting diversity of program and service viewpoints, as well as by preventing undue concentration of economic power."). The seminal case is National Broad. Co. v. United States, 319 U.S. 190, 206-08 (1943) (upholding FCC's power to limit media ownership in a single geographical area).
16. For an argument to this effect, *see* Kate Klonick, The Marketplace of Governance 36 (unpublished manuscript, December 12, 2021). ("Content moderation is a diseconomy of scale, and its operation is cross subsidized through complementary products. At 3 billion users worldwide the costs [for Facebook] of maintaining legibility and utility of the site have increased exponentially.")
17. Nikolas Guggenberger, Reframing the Digital Public Sphere: Industrial Policy for Digital Pluralism (2021) (unpublished manuscript).

18. Sally Hubbard, *Fake News Is a Real Antitrust Problem*, COMPETITION POL'Y INT'L (Dec. 19, 2017), https://www.competitionpolicyinternational.com/fake-news-is-a-real-antitrust-problem/.
19. evelyn douek, *The Rise of Content Cartels*, KNIGHT FIRST AMENDMENT INST. AT COLUM. UNIV. (Feb. 11, 2020), https://knightcolumbia.org/content/the-rise-of-content-cartels.
20. Guggenberger, *supra* note 17.
21. *See, e.g.*, Cory Doctorow, *Adversarial Interoperability*, EFF (Oct. 2, 2019), https://www.eff.org/deeplinks/2019/10/adversarial-interoperability [https://perma.cc/SA78-UEPN]; Mike Masnick, *Protocols, Not Platforms: A Technological Approach to Free Speech*, KNIGHT FIRST AMEND. INST. (Aug. 21, 2019), https://knightcolumbia.org/content/protocols-not-platforms-a-technological-approach-to-free-speech [https://perma.cc/4QCG-K32L].
22. Jack M. Balkin, *The Fiduciary Model of Privacy*, 134 HARV. L. REV. F. 11 (2020).
23. *Id.* at 11-13.
24. *Id.* at 13-14.
25. *See* Neil Richards & Woodrow Hartzog, *A Duty of Loyalty for Privacy Law*, 99 WASH. U. L. REV. (2020), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3642217 [https://perma.cc/4D28-8QMN].
26. *See, e.g.*, Neil M. Richards, *Reconciling Data Privacy and the First Amendment*, 52 UCLA L. REV. 1149, 1182–86, 1190–92 (2005) (noting the ubiquity of collection and use restrictions that do not violate the First Amendment).
27. Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183, 1208–9 (2016).
28. 47 U.S.C. § 230.
29. 47 U.S.C. § 230(c)(1).
30. Jack M. Balkin, *Old-School/New-School Speech Regulation*, 127 HARV. L. REV. 2296, 2309–16 (2014) (noting that withdrawing intermediary immunity predictably produces collateral censorship).
31. Eric Goldman, *Why Section 230 Is Better Than the First Amendment*, 95 NOTRE DAME L. REV. REFLECTION 33, 36–44 (2019) (noting that Section 230's substantive and procedural protections extend well beyond the protections of the First Amendment).
32. *Id.* at 36–39 (noting substantive differences between Section 230 and the First Amendment); *id.* at 39–44 (noting the many procedural benefits of Section 230).
33. Balkin, *Fiduciary Model of Privacy*, *supra* note 22, at 31–32.
34. *Id.* at 31.
35. *See supra* note 2.

Chapter 15

1. YOCHAI BENKLER ET AL., NETWORK PROPAGANDA: MANIPULATION, DISINFORMATION, AND RADICALIZATION IN AMERICAN POLITICS ch. 11 (New York: Oxford University Press, 2018); JEFFREY M. BERRY & SARAH SOBIERAJ, THE OUTRAGE INDUSTRY: POLITICAL OPINION MEDIA AND THE NEW INCIVILITY (New York: Oxford University Press, 2014); KATHLEEN HALL JAMIESON & JOSEPH N. CAPPELLA, ECHO CHAMBER: RUSH LIMBAUGH AND THE CONSERVATIVE MEDIA ESTABLISHMENT (Oxford: Oxford University Press, 2008); NICOLE HEMMER, MESSENGERS OF THE RIGHT: CONSERVATIVE MEDIA AND THE TRANSFORMATION OF AMERICAN POLITICS, POLITICS AND CULTURE IN MODERN AMERICA (Philadelphia: University of Pennsylvania Press, 2016).
2. JACOB S. HACKER & PAUL PIERSON, WINNER-TAKE-ALL POLITICS: HOW WASHINGTON MADE THE RICH RICHER—AND TURNED ITS BACK ON THE MIDDLE CLASS (1st ed., New York: Simon & Schuster, 2010); KEVIN MICHAEL KRUSE, ONE NATION UNDER GOD: HOW CORPORATE AMERICA INVENTED CHRISTIAN AMERICA (New York: Basic Books, 2015); ANDREW L. WHITEHEAD & SAMUEL L. PERRY, TAKING AMERICA BACK FOR GOD: CHRISTIAN NATIONALISM IN THE UNITED STATES (New York: Oxford University Press, 2020).
3. Richard Hofstadter, *The Paranoid Style in American Politics*, 229 (1374) HARPER'S MAG. 77 (1964).
4. C. EDWIN BAKER, MEDIA, MARKETS, AND DEMOCRACY, COMMUNICATION, SOCIETY, AND POLITICS (Cambridge: Cambridge University Press, 2002).

5. Stefano DellaVigna & Ethan Kaplan, *The Fox News Effect: Media Bias and Voting*, 122(3) Q. J. ECON. 1187 (2007).
6. Aaron Shaw & Yochai Benkler, *A Tale of Two Blogospheres: Discursive Practices on the Left and Right*, 56(4) AM. BEHAV. SCIENTIST 459 (Apr. 2012), https://doi.org/10.1177/0002764211433793.
7. BENKLER ET AL., *supra* note 1, ch. 11.
8. Nir Grinberg et al., *Fake News on Twitter During the 2016 U.S. Presidential Election*, 363(6425) SCI. 374 (Jan. 25, 2019), https://doi.org/10.1126/science.aau2706; Andrew Guess et al., *Less Than You Think: Prevalence and Predictors of Fake News Dissemination on Facebook*, 5(1) SCI. ADVANCES (Jan. 2019), https://doi.org/10.1126/Sciadv.aau4586; Jennifer Allen et al., *Evaluating the Fake News Problem at the Scale of the Information Ecosystem*, 6(14) SCI. ADVANCES (Apr. 1, 2020), https://doi.org/10.1126/Sciadv.aay3539.
9. Christopher A. Bail et al., *Assessing the Russian Internet Research Agency's Impact on the Political Attitudes and Behaviors of American Twitter Users in Late 2017*, 117(1) PROC. NAT'L ACAD. SCI. (USA) 243 (Jan. 7, 2020), https://doi.org/10.1073/pnas.1906420116.
10. *Understanding QAnon's Connection to American Politics, Religion, and Media Consumption*, PRRI BLOG, https://www.prri.org/research/qanon-conspiracy-american-politics-report/ (last visited June 10, 2021).
11. Heidi Tworek & Paddy Leerssen, AN ANALYSIS OF GERMANY'S NETZDG LAW (Transatlantic Working Group, Apr. 15, 2019), https://www.ivir.nl/publicaties/download/NetzDG_Tworek_Leerssen_April_2019.pdf.
12. Adrian Rauchfleisch & Jonas Kaiser, *Deplatforming the Far-Right: An Analysis of YouTube and BitChute*, SSRN SCHOLARLY PAPER (Rochester, NY: Social Science Research Network, June 15, 2021), https://papers.ssrn.com/abstract=3867818.
13. James Doubek, "*Conspiracy Theorist Alex Jones Apologizes for Promoting 'Pizzagate,'*" NATIONAL PUBLIC RADIO (Mar. 26, 2017), https://www.npr.org/sections/thetwo-way/2017/03/26/521545788/conspiracy-theorist-alex-jones-apologizes-for-promoting-pizzagate.
14. Philip Bump, *Analysis: Don't Blame the Seth Rich Conspiracy on Russians. Blame Americans*, WASH. POST (last visited Oct. 14, 2021), https://www.washingtonpost.com/politics/2019/07/09/dont-blame-seth-rich-conspiracy-russians-blame-americans/.
15. *Alex Jones, Infowars, and the Sandy Hook Defamation Suits*, FIRST AMEND. WATCH (last visited June 20, 2021), https://firstamendmentwatch.org/deep-dive/alex-jones-infowars-and-the-sandy-hook-defamation-suits/; Elizabeth Williamson, *Lawyers for Neo-Nazi to Defend Alex Jones in Sandy Hook Case*, N.Y. TIMES (U.S.) (July 3, 2018), https://www.nytimes.com/2018/07/02/us/politics/sandy-hook-alex-jones-lawyers.html.
16. BENKLER ET AL., *supra* note 1.
17. *Legal Updates: Learn How We Are Defending Dominion*, DOMINION VOTING SYSTEMS BLOG (last visited June 20, 2021), https://www.dominionvoting.com/legal-updates-learn-how-we-are-defending-dominion/.
18. Christine Hauser, *Alex Jones Retracts Chobani Claims to Resolve Lawsuit*, N.Y. TIMES (U.S.) (May 17, 2017), https://www.nytimes.com/2017/05/17/us/alex-jones-chobani-lawsuit.html; Rocky Barker, *Breitbart's Chobani Coverage Is a Reminder of Idaho's Experience With the Alt-Right*, IDAHO STATESMAN (Nov. 16, 2016), http://www.idahostatesman.com/news/local/news-columns-blogs/letters-from-the-west/article115283313.html.
19. BENKLER ET AL., *supra* note 1, ch. 5.
20. Ethan Zuckerman, *QAnon and the Emergence of the Unreal*, J. DESIGN & SCI. (6) (July 15, 2019), https://doi.org/10.21428/7808da6b.6b8a82b9.
21. BENKLER ET AL., *supra* note 1, ch. 3.
22. New York Times v. Sullivan, 376 U.S. 254 (1964).
23. Milkovich v. Lorain Journal Co., 497 U.S. 1, 19–21 (1990).
24. Greenbelt Coop. Pub. Ass'n v. Bressler, 398 U.S. 6, 14 (1970).
25. McDougal v. Fox News Network, 489 F. Supp. 3d 174, 182–84 (S.D.N.Y. 2020) (citing 1 RODNEY A. SMOLLA, LAW OF DEFAMATION § 6:92 (2d ed. 1986)).
26. 403 U.S. 15 (1971) (holding that wearing a jacket stating "Fuck the Draft" in a courtroom was protected speech).

27. Kathy Frankovic, *Belief in Conspiracies Largely Depends on Political Identity* (Dec. 27, 2016) (last visited Oct. 14, 2021), https://today.yougov.com/topics/politics/articles-reports/2016/12/27/belief-conspiracies-largely-depends-political-iden.
28. *Understanding QAnon's Connection*, *supra* note 10.
29. MICHAEL DIMOCK ET AL., POLITICAL POLARIZATION IN THE AMERICAN PUBLIC (Washington, DC: Pew Research Center, June 12, 2014), http://www.people-press.org/2014/06/12/political-polarization-in-the-american-public/; Mark Jurkowitz et al., *U.S. Media Polarization and the 2020 Election: A Nation Divided*, PEW RESEARCH CENTER'S JOURNALISM PROJECT BLOG (Jan. 24, 2020), https://www.journalism.org/2020/01/24/u-s-media-polarization-and-the-2020-election-a-nation-divided/.
30. Ben Smith, *Fox Settled a Lawsuit Over Its Lies. But It Insisted on One Unusual Condition*, N.Y TIMES (Bus.) (Jan. 18, 2021), https://www.nytimes.com/2021/01/17/business/media/fox-news-seth-rich-settlement.html; Michael Isikoff, *Fox Paid Seven Figures to Settle Lawsuit Over Bogus Seth Rich Conspiracy Story*, YAHOO NEWS (Nov. 24, 2020), https://www.yahoo.com/entertainment/fox-paid-seven-figures-to-settle-lawsuit-over-bogus-seth-rich-conspiracy-story-003236858.html.
31. 586 U.S. ___ (2019) (Thomas J., concurring).
32. Alyssa Rosenberg, *Review: A Sex Tape, a Conspiracy and an Empty Victory Over Gawker*, WASH. POST (last visited Oct. 14, 2021), https://www.washingtonpost.com/outlook/a-sex-tape-a-conspiracy-and-an-empty-victory-over-gawker/2018/02/23/ca60164a-069d-11e8-b48c-b07fea957bd5_story.html.
33. Deepa Seetharaman & Emily Glazer, *How Mark Zuckerberg Learned Politics*, WALL ST. J. (Tech.) (Oct. 16, 2020), https://www.wsj.com/articles/how-mark-zuckerberg-learned-politics-11602853200 (describing how Facebook "[e]ngineers redesigned their intended changes so that left-leaning sites like Mother Jones were affected more than previously planned" in response to outrage about the politically asymmetric impact of efforts to change the newsfeed algorithms to downgrade outlets that engaged in falsehoods).
34. Naomi Oreskes et al., *How American Businessmen Made Us Believe That Free Enterprise Was Indivisible From American Democracy: The National Association of Manufacturers' Propaganda Campaign 1935–1940*, *in* THE DISINFORMATION AGE, 95–119 (W. Lance Bennett & Steven Livingston eds., 1st ed., Cambridge: Cambridge University Press, 2020), https://doi.org/10.1017/9781108914628.004.
35. NAOMI ORESKES & ERIK M. CONWAY, MERCHANTS OF DOUBT: HOW A HANDFUL OF SCIENTISTS OBSCURED THE TRUTH ON ISSUES FROM TOBACCO SMOKE TO GLOBAL WARMING (New York: Bloomsbury, 2011); Anahad O'Connor, *Coca-Cola Funds Scientists Who Shift Blame for Obesity Away from Bad Diets*, WELL BLOG (Aug. 9, 2015), https://well.blogs.nytimes.com/2015/08/09/coca-cola-funds-scientists-who-shift-blame-for-obesity-away-from-bad-diets/.
36. Red Lion Broad. Co. v. FCC, 395 U.S. 367 (1969).
37. JEROME A. BARRON, FREEDOM OF THE PRESS FOR WHOM? THE RIGHT OF ACCESS TO MASS MEDIA (Bloomington: Indiana University Press, 1973).
38. Yochai Benkler, *Free As the Air to Common Use: First Amendment Constraints on Enclosure of the Public Domain*, 74(2) N.Y.U. L. REV. 366–81 (May 1999); Marvin Ammori, *Another Worthy Tradition: How the Free Speech Curriculum Ignores Electronic Media and Distorts Free Speech Doctrine*, 70 MO. L. REV. 59–123 (2005).
39. Miami Herald Publ'g Co. v. Tornillo, 418 U.S. 241 (1974).
40. On "the partner state," *see* A. Pazaitis & W. Drechsler, *Peer Production and State Theory: Envisioning a Cooperative Partner State*, *in* THE HANDBOOK OF PEER PRODUCTION 359–70 (M. O'Neil et al. eds., Hoboken, NJ: Wiley, 2021).
41. Jurkowitz et al., *supra* note 29.
42. Michal Carrington et al., *Consumption Ethics: A Review and Analysis of Future Directions for Interdisciplinary Research*, 168(2) J. BUS. ETHICS 215 (Jan. 2021), https://doi.org/10.1007/s10551-020-04425-4; Jeffery Bray et al., *An Exploratory Study Into the Factors Impeding Ethical Consumption*, 98(4) J. BUS. ETHICS 597 (2011).
43. *U.S. Annual Organic Food Sales Near $48 Billion* (last visited June 15, 2021), https://www.foodbusinessnews.net/articles/13805-us-organic-food-sales-near-48-billion; *Organic Food*

Sales Jump Nearly 13% to Record High in 2020, SUPERMARKET NEWS (May 25, 2021), https://www.supermarketnews.com/consumer-trends/organic-food-sales-jump-nearly-13-record-high-2020.

44. Yinjin Lee & Alexis Bateman, *The Competitiveness of Fair Trade and Organic Versus Conventional Coffee Based on Consumer Panel Data*, ECOLOGICAL ECONOMICS (last visited Oct. 14, 2021), https://doi.org/10.1016/j.ecolecon.2021.106986.
45. Jeffrey Kucik & Krzysztof Pelc, *Political Rhetoric as Trade Barrier: Trump's "Buy American" Campaign and US Demand for Imports*, SSRN ELEC. J. (2017), https://doi.org/10.2139/ssrn.3067518.
46. Randall Aaron Lewis & Justin M. Rao, *On the Near Impossibility of Measuring the Returns to Advertising*, SSRN ELEC. J. (2013), https://doi.org/10.2139/ssrn.2367103; Brett R. Gordon et al., *A Comparison of Approaches to Advertising Measurement: Evidence From Big Field Experiments at Facebook*, 38(2) MARKETING SCI. 193 (Mar. 2019), https://doi.org/10.1287/mksc.2018.1135.
47. *Section 4: Demographics and Political Views of News Audiences*, PEW RESEARCH CENTER—U.S. POLITICS & POLICY BLOG (Sept. 27, 2012), https://www.pewresearch.org/politics/2012/09/27/section-4-demographics-and-political-views-of-news-audiences/.

Chapter 16

1. Lois Beckett, *Trump Digital Director Says Facebook Helped Win the White House*, GUARDIAN (Oct. 8, 2017), https://perma.cc/F9SQ-FPPE.
2. Citizens United v. F.E.C., 558 U.S. 310, 392–93 (2010) (Scalia, J., concurring).
3. First Nat'l Bank of Boston v. Bellotti, 435 U.S. 765, 776 (1978).
4. Keach Hagey & Jeff Horwitz, "*Facebook Tried to Make Its Platform a Healthier Place. It Got Angrier Instead*," WALL ST. J. (Sept. 15, 2021), https://perma.cc/P28X-HET8.
5. *See, e.g.*, Yochai Benkler, *The Danger of Overstating the Impact of Information Operations*, LAWFARE BLOG (Oct. 23, 2020), https://perma.cc/PX7T-S746.
6. Tim Wu, *Machine Speech,* 161 U. PA. L. REV. 1495 (2013).
7. *See* Julia Angwin et al., *Facebook Enabled Advertisers to Reach "Jew Haters,"* PROPUBLICA, https://perma.cc/S2KU-UEGW.
8. I don't believe this description is enough to operationalize the inevitable line-drawing problem that my category creates. I hope it is enough to suggest how such operationalization might proceed. As the science of AI develops a richer conception of relative machine autonomy, the law could usefully track that conception. It's clear enough that there is no binary test that would be applied usefully. But the law will need something more substantial than intuition to indicate when the line from machines to replicants has been crossed.
9. Lawrence B. Solum, *Originalist Methodology*, 84 U. CHI. L. REV. 269, 269–70 (2017). If not, then, as Attorney General Edwin Meese remarked, the Constitution would "be like a picnic to which the framers bring the words and the judges the meaning." Address of the Hon. Edwin Meese III, Nov. 15, 1985, at 12, https://perma.cc/R5DN-YPRN.
10. Solum, *supra* note 9, at 271.
11. I agree with Richard Fallon that it is the constant obligation of courts to translate constitutional values into usable doctrine. That translation is never simple, and the connections are often not obvious. *See* Richard H. Fallon Jr., *Judicially Manageable Standards and Constitutional Meaning*, 119 HARV. L. REV. 1274 (2006).
12. United States v. Classic, 313 U.S. 299, 315–16.
13. This dynamic, between fidelity to meaning and fidelity to role, is the central thesis of LAWRENCE LESSIG, FIDELITY AND CONSTRAINT: HOW THE SUPREME COURT HAS READ THE AMERICAN CONSTITUTION (New York: Oxford University Press, 2019).
14. The Court's Fourth Amendment jurisprudence has been the most explicit in recognizing how changing technology changes the application of the Fourth Amendment, even if it doesn't change its values. That was the common understanding of the Court's decision in *Kyllo v. United States*, 533 U.S. 27 (2001). Orin Kerr smartly criticizes this understanding of the Court's role in developing Fourth Amendment jurisprudence. *See* Orin S. Kerr, *The Fourth Amendment and New Technologies: Constitutional Myths and the Case for Caution*, 102

MICH. L. REV. 801 (2004). Kerr rightly highlights the institutional constraints on the Court's capacity to evolve the constraints of the Fourth Amendment, preferring instead the work of legislatures. I agree (belatedly) with Kerr's approach (my first book, LAWRENCE LESSIG, CODE: AND OTHER LAWS OF CYBERSPACE (New York: Basic Books, 1999) took a different view), but those institutional concerns are different in the context of the First Amendment. If anything, the caution that drives Kerr's conclusion in the context of the Fourth Amendment reinforces my own conclusion in the context of the First Amendment. Radically new speech technologies should entail First Amendment caution. Rather than automatic inclusion, changing technology should raise at least the requirement of returning to first principles.

15. *See, e.g.*, Jud Campbell, *Republicanism and Natural Rights at the Founding*, 32 CONST. COMMENT 85 (2017).
16. Justice Holmes first described the idea as the "free trade in ideas" within "the competition of the market." Abrams v. United States, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting). Justice Douglas gave us the phrase "market place of ideas," United States v. Rumely, 345 U.S. 41, 56 (1953), and Justice Brennan reformed it slightly in Lamont v. Postmaster Gen., 381 U.S. 301, 308 (1965) (Brennan, J., concurring), to the more familiar "marketplace of ideas."
17. LAWRENCE LESSIG, AMERICA, COMPROMISED (Chicago: University of Chicago Press, 2018) (see the list at 142–49).
18. Andrew Kohut, *How the Watergate Crisis Eroded Public Support for Richard Nixon*, PEW RES. CTR., https://perma.cc/9RCC-TLLG.
19. *See* JANE DAILEY, WHITE FRIGHT: THE SEXUAL PANIC AT THE HEART OF AMERICA'S RACIST HISTORY 254 & n. 11 (New York: Basic Books, 2020).
20. *See* ROBERT C. POST, DEMOCRACY, EXPERTISE, AND ACADEMIC FREEDOM: A FIRST AMENDMENT JURISPRUDENCE FOR THE MODERN STATE 31 (New Haven, CT: Yale University Press, 2013); ROBERT C. POST, CONSTITUTIONAL DOMAINS: DEMOCRACY, COMMUNITY, MANAGEMENT (Cambridge, MA: Harvard University Press, 1995).
21. Bluman v. Fed. Election Comm'n, 800 F. Supp. 2d 281 (D.D.C. 2011), *aff'd*, 565 U.S. 1104 (2012).
22. Citizens United v. Fed. Election Comm'n, 558 U.S. 310 (2010).
23. *Id.* at 283.
24. *Id.* at 287 (citations omitted).
25. Miles v. City Council, 710 F.2d 1542 (11th Cir. 1983).
26. KATE DARLING, THE NEW BREED: WHAT OUR HISTORY WITH ANIMALS REVEALS ABOUT OUR FUTURE WITH ROBOTS (New York: Henry Holt, 2021).
27. *See, e.g.*, Hazelwood Sch. Dist. v. Kuhlmeier, 484 U.S. 260 (1988).
28. Or they would suggest an obvious step, but for the Court's recent tendency to draw sharp and absolute First Amendment lines, at least in some new (and old) contexts. *See, e.g.*, Reed v. Town of Gilbert, 576 U.S. 155 (2015). Suffice it to say that the First Amendment has historically been read to be much less absolute in its reach than the current Supreme Court sees it. This modern conviction may be waning, as Justice Thomas has recently signaled more interest in understanding the original reach of the First Amendment. *See* McKee v. Cosby, 139 S. Ct. 675 (2019), and more recently, Biden v. Knight First Amend. Inst. at Columbia Univ., No. 20-197, 593 U.S. (Apr. 5, 2021) (Thomas, J., concurring) ("regulations that might affect speech are valid if they would have been permissible at the time of the founding") (slip op. at 6). As all acknowledge, that original reach was far less expansive than the current Court's articulation.
29. Wu, *supra* note 6, at 1502–3.
30. *See, e.g.*, MEIR DAN-COHEN, RIGHTS, PERSONS, AND ORGANIZATIONS: A LEGAL THEORY FOR BUREAUCRATIC SOCIETY (Berkeley: University of California Press, 1986).
31. *See Bluman*.
32. 505 U.S. 377 (1992).
33. Eugene Volokh & Donald M. Falk, *First Amendment Protection for Search Engine Search Results*, 8 J.L. ECON. & POL'Y 883 (2012). Volokh has not changed his position since. *See* Eugene Volokh, *Treating Social Media Platforms Like Common Carriers?*, 1 J. OF FREE SPEECH L. 377 (2021). For related arguments, *see* Hilda Kajbaf, *The First Amendment and Modern Technology: The Free Speech Clause and Chatbot Speech*, 47 HASTINGS CONST. L.Q. 337, 339

(2020) (discussing "chat-bots"); Tony M. Massaro & Helen Norton, *Artificial Intelligence and the First Amendment*, *in* RESEARCH HANDBOOK ON THE LAW OF ARTIFICIAL INTELLIGENCE (Cheltenham, UK: Edward Elgar, 2018), https://perma.cc/HP5H-EJP3; Tony M. Massaro & Helen Norton, *Siri-ously? Free Speech Rights and Artificial Intelligence*, 110 NW. U. L. REV. 1169 (2016).

34. Volokh & Falk, *supra* note 33, at 884.
35. *See, e.g.*, Jian Zhang v. Baidu.com Inc., 10 F. Supp. 3d 433, 438 (S.D.N.Y. 2014); Langdon v. Google, Inc., 474 F. Supp. 2d 622 (D. Del. 2007); Search King, Inc. v. Google Tech., Inc., Case No. CIV-02-1457-M (W.D. Okla. May. 27, 2003).
36. *See* Volokh & Falk, *supra* note 33, at 889, citing *First Nat'l Bank of Boston*; Va. Pharmacy Bd. v. Va. Consumer Council, 425 U.S. 748 (1976).
37. Volokh, *supra* note 33.
38. Wu, *supra* note 6, at 1517–24.
39. SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER 323–24 (New York: PublicAffairs, 2019).
40. Sophie Bushwick, *What the Capitol Riot Data Download Shows About Social Media Vulnerabilities*, SCI. AM. (Jan. 27, 2021), https://perma.cc/KBW8-8D6V. Outside of the context of the internet, there is a rich literature discussing the regulability of manipulation. *See, e.g.*, Ido Kilovaty, *Legally Cognizable Manipulation*, 34 BERKELEY TECH. L.J. 449 (2019); Eric Posner, *The Law, Economics, and Psychology of Manipulation* (University of Chicago Coase-Sandor Institute for Law & Economics Research Paper No. 726), https://perma.cc/RGE9-GRYZ; Cass R. Sunstein, *Fifty Shades of Manipulation*, *in* THE ETHICS OF INFLUENCE: GOVERNMENT IN THE AGE OF BEHAVIORAL SCIENCE (Cambridge: Cambridge University Press, 2016).
41. Jean M. Twenge et al., *Increases in Depressive Symptoms, Suicide-Related Outcomes, and Suicide Rates Among U.S. Adolescents After 2010 and Links to Increased New Media Screen Time*, 6(1) CLINICAL PSYCHOL. SCI. 3 (2017), https://perma.cc/5LWS-RHE8.
42. Elroy Boers et al., *Association of Screen Time and Depression in Adolescence*, 173 JAMA PEDIATR. 853 (2019), https://jamanetwork.com/journals/jamapediatrics/fullarticle/2737909. JAMA pediatrics unhelpfully blocks tools like Perma.cc from archiving the work cited there. *See Perma: Scoping and Addressing the Problem of Link and Reference Rot in Legal Citations*, 127 HARV. L. REV. F. 176 (2014), https://perma.cc/4LZC-CE58.
43. *Algorithms and Amplification: How Social Media Platforms' Design Choices Shape Our Discourse and Our Minds: Hearing Before the U.S. Sen. Comm. on the Judiciary Subcomm. on Privacy, Technology, and the Law* (testimony of Tristan Harris) (Apr. 27, 2021).
44. *See* MICHAEL MOSS, SALT SUGAR FAT: HOW THE FOOD GIANTS HOOKED US (New York: Random House, 2013).

Chapter 17

1. On contested attribution of both quotations, see Garson O'Toole, Quote Investigator, https://quoteinvestigator.com/2014/07/13/truth/ and *id.*, https://quoteinvestigator.com/2018/11/18/know-trouble/#:~:text=It%20ain't%20so%20much%20the%20things%20that%20people%20don,%E2%80%94%20Mark%20Twain.
2. *See* Canadian Charter of Rights and Freedoms, Part I of the Constitution Act, 1982, https://web.archive.org/web/20140119183913/http://laws-lois.justice.gc.ca/eng/const/page-15.html: "2. Everyone has the following fundamental freedoms:

   (a) freedom of conscience and religion;
   (b) freedom of thought, belief, opinion, and expression, including freedom of the press and other media of communication;
   (c) freedom of peaceful assembly; and
   (d) freedom of association."

3. Government of Canada, Department of Justice, Section 2(b)—Freedom of Expression, https://www.justice.gc.ca/eng/csj-sjc/rfc-dlc/ccrf-ccdl/check/art2b.html.

4. *See, e.g.,* JAMAL GREENE, HOW RIGHTS WENT WRONG: WHY OUR OBSESSION WITH RIGHTS IS TEARING AMERICA APART (Boston: Houghton Mifflin Harcourt, 2021); Vicki C. Jackson, *Constitutional Law in an Age of Proportionality*, 124 YALE L.J. 3094, 3136 (2015).
5. Sai Teja Peddinti, Keith W. Ross, & Justin Cappos, *User Anonymity on Twitter*, InfoQ (Oct. 12, 2017), https://www.infoq.com/articles/user-anonymity-twitter/.
6. For example, https://legal.parler.com/documents/useragreement.pdf, https://www.redditinc.com/policies/user-agreement-october-15-2020.
7. In 2021, Furie was able to reclaim ownership of the image for the purposes of selling a nonfungible token (NFT), though that does not prevent its use by hate groups.
8. Paul Romer, *A Tax That Could Fix Big Tech*, N.Y. TIMES (May 6, 2019), https://www.nytimes.com/2019/05/06/opinion/tax-facebook-google.html.
9. Pruneyard Shopping Ctr. v. Robins, 447 U.S. 74 (1980).
10. ICANN Announcement, https://www.icann.org/en/announcements/details/stewardship-of-iana-functions-transitions-to-global-internet-community-as-contract-with-us-government-ends-1-10-2016-en.
11. Paul Mozur, *A Genocide Incited on Facebook, with Posts from Myanmar's Military*, N.Y. TIMES (Oct. 15, 2018), https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html.
12. *E.g.*, Ben Shapiro: https://popular.info/p/the-dirty-secret-behind-ben-shapiros.
13. Lindy West, *What Happened When I Confronted My Cruelest Troll*, GUARDIAN (Feb. 2, 2015), https://www.theguardian.com/society/2015/feb/02/what-happened-confronted-cruellest-troll-lindy-west.
14. CONGRESSIONAL RESEARCH SERVICE, SOCIAL MEDIA: MISINFORMATION AND CONTENT MODERATION ISSUES FOR CONGRESS 16 (Jan. 27, 2021), https://crsreports.congress.gov/product/pdf/R/R46662 (citing Hye Kyung Kim, Jisoo Ahn, Lucy Atkinson et al., *Effects of COVID-19 Misinformation on Information Seeking, Avoidance, and Processing: A Multicountry Comparative Study*, 1(30) SCI. COMM. (Sept. 13, 2020), https://journals.sagepub.com/doi/10.1177/1075547020959670).
15. Schenck v. United States, 249 U.S. 47, 48–49 (1919).
16. Dennis v. United States, 341 U.S. 494 (1951).
17. United States v. Alvarez, 567 U.S. 709 (2012).
18. Brandenburg v. Ohio, 395 U.S. 444 (1969).
19. *See* Cass R. Sunstein, *Does the Clear and Present Danger Test Survive Cost-Benefit Analysis?*, *in* THE FREE SPEECH CENTURY 162, 168–170, 176 (Lee C. Bollinger & Geoffrey R. Stone eds., New York: Oxford University Press, 2018).
20. *See* Peter Margulies, *The NSA in Global Perspective: Surveillance, Human Rights, and International Counterterrorism*, 82 FORDHAM L. REV. 2137 (2014), https://ir.lawnet.fordham.edu/flr/vol82/iss5/.
21. *For the 16th Year, Americans Rate PBS and Its Member Stations Most Trusted Institution* (PBS television broadcast, Aug. 17, 2020), https://www.pbs.org/about/about-pbs/blogs/news/for-the-16th-consecutive-year-americans-rate-pbs-and-its-member-stations-most-trusted-institution/; John Gramlich, *How Pew Research Center Evaluated Americans' Trust in 30 News Sources*, PEW RES. CTR. (Jan. 25, 2020), https://www.pewresearch.org/fact-tank/2020/01/24/qa-how-pew-research-center-evaluated-americans-trust-in-30-news-sources/.
22. Ethan Zuckerman, *Building a More Honest Internet*, COLUM. JOURNALISM REV. (Fall 2019), https://www.cjr.org/special_report/building-honest-internet-public-interest.php.
23. Jon Brodkin, *Making Internet Service a Utility—What's the Worst That Could Happen?*, ARSTECHNICA (May 23, 2020), https://arstechnica.com/information-technology/2020/05/worst-case-scenario-why-the-cable-lobby-is-scared-of-becoming-a-utility/.
24. Communications Act of 1934, 47 U.S.C. ch. 5 (1964), https://uscode.house.gov/view.xhtml?path=/prelim@title47/chapter5/subchapter2&edition=prelim.
25. Proposal for a Regulation on a Single Market for Digital Services (Digital Services Act) (DSA) (Dec. 15, 2020), https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package. *See* Aline Blankertz & Julian Jarusch, *How the EU Plans to Rewrite the Rules for the Internet*, BROOKINGS TECHSTREAM (Oct. 21, 2020), https://www.brookings.edu/techstream/how-the-eu-plans-to-rewrite-the-rules-for-the-internet/.

26. *See* Natasha Lomash, *Germany Tightens Online Hate Speech Rules to Make Platforms Send Reports Straight to the Feds*, TECHCRUNCH (June 19, 2020), https://techcrunch.com/2020/06/19/germany-tightens-online-hate-speech-rules-to-make-platforms-send-reports-straight-to-the-feds/ (describing the Network Enforcement Act, known as "NetzDG").
27. Alex Hern, *Online Safety Bill "A Recipe for Censorship" Say Campaigners*, GUARDIAN (May 12, 2021), https://www.theguardian.com/media/2021/may/12/uk-to-require-social-media-to-protect-democratically-important-content.
28. Kevin Paulson, *Trump's New Favorite Channel Employs Kremlin-Paid Journalist*, DAILY BEAST (Nov. 2, 2019), https://www.thedailybeast.com/oan-trumps-new-favorite-channel-employs-kremlin-paid-journalist.
29. *See* JONATHAN RAUCH, THE CONSTITUTION OF KNOWLEDGE: A DEFENSE OF TRUTH (Washington, D.C.: Brookings Institution Press, 2021).
30. This common version of his statement grew from these words in his dissenting opinion in *Terminello v. Chicago* arguing that extremism in civil liberties could destroy the nation: "The choice is not between order and liberty. It is between liberty with order and anarchy without either. There is danger that, if the court does not temper its doctrinaire logic with a little practical wisdom, it will convert the constitutional Bill of Rights into a suicide pact." 337 U.S. 1, 37 (1949).
31. President John F. Kennedy attributed this quotation to Edmund Burke, Documents on Disarmament [May 17, 1961]: Address by President Kennedy to the Canadian Parliament, J.F.K. Library (digital version), https://www.jfklibrary.org/asset-viewer/archives/JFKPOF/034/JFKPOF-034-027.
32. Washington Post v. McManus, 944 F.3d 506 (4th Cir. 2019).
33. Citizens United v. Fed. Election Comm'n, 558 U.S. 310 (2010).

Chapter 18

1. Margaret Sullivan, *Facebook Is Harming Our Society. Here's a Radical Solution for Reining It In*, WASH. POST (Oct. 5, 2021), https://www.washingtonpost.com/lifestyle/media/media-sullivan-facebook-whistleblower-haugen/2021/10/04/3461c62e-2535-11ec-8831-a31e7b3de188_story.html.
2. Cristiano Lima, *A Whistleblower's Power: Key Takeaways From the Facebook Papers*, WASH. POST (Oct. 26, 2021), https://www.washingtonpost.com/technology/2021/10/25/what-are-the-facebook-papers.
3. Morgan Keith, *Facebook's Internal Research Found Its Instagram Platform Contributes to Eating Disorders and Suicidal Thoughts in Teenage Girls, Whistleblower Says*, BUS. INSIDER (Oct. 3, 2021), https://www.businessinsider.com/facebook-knows-data-instagram-eating-disorders-suicidal-thoughts-whistleblower-2021-10.
4. NPR, *1A: Facebook Paused the Development of Instagram for Kids. Now What?*, NPR (Sept. 30, 2021), https://www.npr.org/2021/09/30/1041954649/facebook-paused-the-development-of-instagram-for-kids-now-what.
5. Deepa Seetharaman et al., *Facebook Says AI Will Clean Up the Platform. Its Own Engineers Have Doubts*, WALL ST. J. (Oct. 17, 2021), https://www.wsj.com/articles/facebook-ai-enforce-rules-engineers-doubtful-artificial-intelligence-11634338184.
6. *Digital Economy*, U.S. BUREAU OF ECONOMIC ANALYSIS, https://www.bea.gov/data/special-topics/digital-economy (last visited Dec. 10, 2021).
7. Chris Alcantara et al., *How Big Tech Got So Big: Hundreds of Acquisitions*, WASH. POST (Apr. 21, 2021), https://www.washingtonpost.com/technology/interactive/2021/amazon-apple-facebook-google-acquisitions/.
8. Clara Hendrickson, *Local Journalism in Crisis: Why America Must Revive Its Local Newsrooms*, BROOKINGS INSTITUTION (Nov. 12, 2019), https://www.brookings.edu/research/local-journalism-in-crisis-why-america-must-revive-its-local-newsrooms.
9. *COVID Data Tracker Weekly Review*, CENTERS FOR DISEASE CONTROL AND PREVENTION, https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html (last visited Dec. 10, 2021).

10. Gerrit De Vynck & Rachel Lerman, *Facebook and YouTube Spent a Year Fighting Covid Misinformation. It's Still Spreading*, WASH. POST (July 22, 2021), https://www.washingtonpost.com/technology/2021/07/22/facebook-youtube-vaccine-misinformation.
11. Craig Timberg et al., *Inside Facebook, Jan. 6 Violence Fueled Anger, Regret Over Missed Warning Signs*, WASH. POST (Oct. 22, 2021), https://www.washingtonpost.com/technology/2021/10/22/jan-6-capitol-riot-facebook.
12. Of course, there are challenging First Amendment questions on what the government can do in the content moderation arena. I do not in this paper address those issues, and I leave it to the other essays to address those questions.
13. Peyton Shelburne, *Across the Globe, Rates of Vaccine Skepticism Have Stalled*, MORNING CONSULT (Dec. 9, 2021), https://morningconsult.com/global-vaccine-tracking.
14. Drew Altman, *Vaccine Myths Persist*, Axios (May 26, 2021), https://www.axios.com/covid-vaccine-myths-5a72f75e-dd9f-4d37-a378-d7ce9e46f074.html.
15. Davey Alba & Ryan Mac, *Facebook, Fearing Public Outcry, Shelved Earlier Report on Popular Posts*, N.Y. TIMES (Oct. 25, 2021), https://www.nytimes.com/2021/08/20/technology/facebook-popular-posts.html.
16. Heather M. Scobie et al., *Monitoring Incidence of COVID-19 Cases, Hospitalizations, and Deaths, by Vaccination Status—13 U.S. Jurisdictions, April 4–July 17, 2021*, CENTERS FOR DISEASE CONTROL AND PREVENTION (Sept. 17, 2021), https://www.cdc.gov/mmwr/volumes/70/wr/pdfs/mm7037e1-H.pdf.
17. Health Misinformation Act, S. 2448, 117th Cong. (2021).
18. Lima, *supra* note 2.
19. Marcy Gordon, *WATCH: Facebook Whistleblower Testifies to Senate on Children and Social Media*, PBS NEWS HOUR (Oct. 5, 2021), https://www.pbs.org/newshour/nation/watch-live-facebook-whistleblower-testifies-to-senate-on-children-and-social-media.
20. Donie O'Sullivan et al., *Instagram Promoted Pages Glorifying Eating Disorders to Teen Accounts*, CNN (Oct. 4, 2021), https://www.cnn.com/2021/10/04/tech/instagram-facebook-eating-disorders/index.html.
21. Anna Westin Act, S. 1865, 114th Cong. (2015).
22. *Eating Disorder Statistics*, NATIONAL ASSOCIATION OF ANOREXIA NERVOSA AND ASSOCIATED DISORDERS, https://anad.org/eating-disorders-statistics (last visited Dec. 10, 2021); *Eating Disorder Statistics*, MIRASOL, https://www.mirasol.net/learning-center/eating-disorder-statistics.php (last visited Dec. 10, 2021).
23. Bobby Allyn, *Snapchat Ends 'Speed Filter' That Critics Say Encouraged Reckless Driving*, NPR (June 17, 2021), https://www.npr.org/2021/06/17/1007385955/snapchat-ends-speed-filter-that-critics-say-encouraged-reckless-driving.
24. Jamie Condliffe, *The Average American Spends 24 Hours a Week Online*, MIT Tech. Rev. (Jan. 23, 2018), https://www.technologyreview.com/2018/01/23/146069/the-average-american-spends-24-hours-a-week-online.
25. SAFE TECH Act, S. 299, 117th Cong. (2021).
26. *See, e.g.*, *Privacy Policy*, GOOGLE, https://policies.google.com/privacy?hl=en-US (last visited Dec. 10, 2021); *Data Policy*, FACEBOOK, https://www.facebook.com/policy.php (last visited: Dec. 10, 2021); *Amazon.com Privacy Notice*, AMAZON https://www.amazon.com/gp/help/customer/display.html?nodeId=GX7NJQ4ZB8MHFRNJ (last visited Dec. 10, 2021).
27. *Data Brokers: A Call for Transparency and Accountability*, FED. TRADE COMM'N (May 2014), https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf.
28. David Nield, *All the Ways Facebook Tracks You—And How to Limit It*, WIRED (Jan. 12, 2020), https://www.wired.com/story/ways-facebook-tracks-you-limit-it.
29. *Google Chrome Privacy Notice*, GOOGLE (Sept. 23, 2021), https://www.google.com/chrome/privacy.
30. Nield, *supra* note 28.
31. *See FB Earnings Presentation Q3 2021*, FACEBOOK, https://s21.q4cdn.com/399680738/files/doc_financials/2021/q3/FB-Earnings-Presentation-Q3-2021.pdf (last visited Dec. 10, 2021). Slide 4, ARPU in Q2 of 2021 in the United States and Canada.

32. Kif Leswing, *Apple's Ad Privacy Change Impact Shows the Power It Wields Over Other Industries*, CNBC (Nov. 13, 2021), https://www.cnbc.com/2021/11/13/apples-privacy-changes-show-the-power-it-holds-over-other-industries.html.
33. Consumer Online Privacy Rights Act, S. 3195, 117th Cong. (2021).
34. Children and Teens' Online Privacy Protection Act, S. 1628, 117th Cong. (2021).
35. Soofia Tariq, *Nine out of 10 Health Apps Harvest User Data, Global Study Shows*, GUARDIAN (June 16, 2021), https://www.theguardian.com/technology/2021/jun/17/nine-out-of-10-health-apps-harvest-user-data-global-study-shows.
36. Protecting Personal Health Data Act, S. 24, 117th Cong. (2021).
37. Casey Newton & Nilay Patel, *"Instagram Can Hurt Us": Mark Zuckerberg Emails Outline Plan To Neutralize Competitors*, VERGE (July 29, 2020) (emphasis added), https://www.theverge.com/2020/7/29/21345723/facebook-instagram-documents-emails-mark-zuckerberg-kevin-systrom-hearing.
38. Guildo Federico et al., *Antitrust and Innovation: Welcoming and Protecting Disruption*, 20 INNOVATION POLICY AND THE ECON. 125 (2020), at 125 https://www.journals.uchicago.edu/doi/pdf/10.1086/705642.
39. *See* Competition Issues in Labour Markets, OECD (Feb. 26, 2020), https://www.oecd.org/daf/competition/competition-concerns-in-labour-markets.htm.
40. Federico, *supra* note 38, at 125.
41. Complaint, United States v. Google LLC, No. 1:20-cv-03010 (D.D.C., Oct. 20, 2020).
42. *FTC Alleges Facebook Restored to Illegal Buy-or-Bury Scheme to Crush Competition After String of Failed Attempts to Innovate*, FED. TRADE COMM'N (Aug. 19, 2021), https://www.ftc.gov/news-events/press-releases/2021/08/ftc-alleges-facebook-resorted-illegal-buy-or-bury-scheme-crush.
43. *See, e.g.*, Diane Bartz & Paresh Dave, *"Gorilla" Google Hit With Third Lawsuit as U.S. States Sue Over Search Dominance*, REUTERS (Dec. 17, 2020), https://www.reuters.com/business/media-telecom/gorilla-google-hit-with-third-lawsuit-us-states-sue-over-search-dominance-2020-12-17; Diane Bartz et al., *U.S. States Allege Google "Unlawfully" Preserves Play Store Monopoly*, REUTERS (July 8, 2021), https://www.reuters.com/technology/dozens-us-states-sue-google-alleging-antitrust-violations-2021-07-07; Diane Bartz & Paresh Dave, *Texas, Nine U.S. States Accuse Google of Working With Facebook to Break Antitrust Law*, REUTERS (Dec. 16, 2020), https://www.reuters.com/business/texas-nine-us-states-accuse-google-working-with-facebook-break-antitrust-law-2020-12-16.
44. Competition and Antitrust Law Enforcement Reform Act, S. 225, 117th Cong. (2021).
45. *See Transforming the FTC: Legislation to Modernize Consumer Prot. Testimony Before the House Energy and Com. Comm. Subcomm. on Consumer Prot. and Com.* (July 28, 2021) (written testimony of FTC Chair Lina Khan regarding agency resources), https://energycommerce.house.gov/sites/democrats.energycommerce.house.gov/files/documents/Witness%20Testimony_FTC%20Commissioners_CPC_2021.07.28.pdf; *see also Premerger Notification Program*, FED. TRADE COMM'N, https://www.ftc.gov/enforcement/premerger-notification-program (last visited Dec. 10, 2021) for data on the increase in merger filings; Bill Baer et al., *Restoring Competition in the United States: A Vision for Antitrust Enforcement for the Next Administration and Congress*, WASHINGTON CTR. FOR EQUITABLE GROWTH 15 (Nov. 2020), https://equitablegrowth.org/wp-content/uploads/2020/11/111920-antitrust-report.pdf.
46. For example, Supreme Court decisions have made it extremely difficult to challenge unilateral refusals to deal (Verizon Commc'n Inc. v. Law Offices of Curtis V. Trinko, LLP, 540 U.S. 398 (2004)), price squeezes (Pacific Bell Telephone Co. v. linkLine Commc'ns, Inc., 555 U.S. 438, 448-51 (2009)), or predatory pricing (Brooke Group Ltd. v. Brown & Williamson Tobacco Corp., 509 U.S. 209, 223 (1993)).
47. Cat Zakrzewski, *Senators Aim to Block Tech Giants From Prioritizing Their Own Products Over Rivals*, WASH. POST (Oct. 14, 2021), https://www.washingtonpost.com/technology/2021/10/14/klobuchar-grassley-antitrust-bill.
48. Sumit Sharma, *The Case for Fair Market Rules for the Largest Online Platforms*, CONSUMER REPORTS (Oct. 14, 2021), https://digital-lab.consumerreports.org/2021/10/14/the-case-for-fair-market-rules-for-the-largest-online-platforms/ .

49. *See, e.g., Big Tech's Main Street Squeeze: Stories From Small Business Owners*, MAIN STREET AGAINST BIG TECH, https://www.mainstreetagainstbigtech.org/storybook (last visited Dec. 10, 2021).
50. Dana Mattioli, *Amazon Scooped Up Data From Its Own Sellers to Launch Competing Products*, WALL ST. J. (Apr. 23, 2020), https://www.wsj.com/articles/amazon-scooped-up-data-from-its-own-sellers-to-launch-competing-products-11587650015.
51. Aditya Kalra & Steve Stecklow, *Amazon Copied Products and Rigged Search Results to Promote Its Own Brands, Documents Show*, REUTERS (Oct. 13, 2021), https://www.reuters.com/investigates/special-report/amazon-india-rigging.
52. *Antitrust Applied: Examining Competition in App Stores: Hearing Before the Senate Comm. on the Judiciary*, 117th Cong. (April 21, 2021) (written testimony of Kirsten Daru, Chief Privacy Officer and General Counsel for Tile, Inc.), https://www.judiciary.senate.gov/download/kirsten-daru-testimony.
53. Luther Lowe, *Yelp Testifying in Google Antitrust Hearing: Self-Preferencing by Dominant Internet Platforms*, YELP BLOG (Mar. 9, 2020), https://blog.yelp.com/news/yelp-testifying-in-google-antitrust-hearing.
54. Am. Innovation and Choice Online Act, S. 2992 117th Cong. (2021).

Report of the Commission

1. *See, e.g.*, Amy Mitchell and Mason Walker, *More Americans Now Say Government Should Take Steps to Restrict False Information Online Than in 2018*, PEW RES. CTR. (Aug. 18, 2021), https://www.pewresearch.org/fact-tank/2021/08/18/more-americans-now-say-government- should-take-steps-to-restrict-false-information-online-than-in-2018/; Ariel Edwards-Levy & Jennifer Agiesta, *CNN Poll: Three Out of Four Adults Think Facebook Is Making Society Worse*, *CNN* (Nov. 10, 2021), https://www.cnn.com/2021/11/10/business/cnn-poll- facebook/index.html.
2. *See* Mark Hosenball, *Russia Used Social Media for Widespread Meddling in U.S. Politics: Reports*, REUTERS (Dec. 17, 2018), https://www.reuters.com/article/us-usa-trump-russia-socialmedia/russia-used-social-media-for-widespread-meddling-in-u-s-politics-reports- idUSKBN1OG257; *see generally* Renée DiResta et al., *The Tactics & Tropes of the Internet Research Agency*, NEW KNOWLEDGE, https://disinformationreport.blob.core.windows.net/disinformation-report/NewKnowledge-Disinformation-Report-Whitepaper-121718.pdf; Philip N. Howard et al., *The IRA, Social Media and Political Polarization in the United States, 2012–2018*, COMPUTATIONAL PROPAGANDA RES. PROJECT, https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1004&context=senatedocs; *Social Media Influencer Charged With Election Interference Stemming From Voter Disinformation Campaign*, DEP'T OF JUST. (Jan. 27, 2021), https://www.justice.gov/opa/pr/social-media-influencer-charged-election-interference-stemming-voter-disinformation-campaign.
3. *See, e.g.*, Steven Lee Myers & Paul Mozur, *China Is Waging a Disinformation War Against Hong Kong Protesters*, N.Y. Times (Aug. 13, 2019), https://www.nytimes.com/2019/08/13/world/asia/hong-kong-protests-china.html.
4. *See, e.g., The Facebook Files, Part 2: "We Make Body Image Issues Worse,"* WALL ST. J. (Sept. 14, 2021), https://www.wsj.com/podcasts/the-journal/the-facebook-files-part-2-we-make-body-image-issues-worse/c2c4d7ba-f261-4343-8d18-d4de177cf973.
5. 395 U.S. 367 (1969).
6. *See* Red Lion, 395 U.S. at 392 (observing that there is a "First Amendment goal of producing an informed public capable of conducting its own affairs").
7. *See* 47 U.S.C. § 230(c).
8. *See* 17 U.S.C. § 512.
9. Platform Accountability and Consumer Transparency Act ("PACT Act"), S.797, § 6(a), 117th Cong., 1st Sess. (2021), https://www.congress.gov/bill/117th-congress/senate-bill/797/text. The PACT Act would impose a statutory requirement on social media companies (and other interactive computer service providers) to implement such a notice-and-take-down regime. *See id.* § 5(b)-(c). Senator Sheldon Whitehouse (D-RI) and Professor Mary Anne Franks also support this approach in their essays.

10. *See* Regulation on a Single Market for Digital Services (Digital Services Act) and amending Directive 200/31/EC 5 (providing for immunity from suit for user-generated content in most circumstances unless the company has knowledge that the content is unlawful) and 14-15 (requiring social media companies to adopt a complaint mechanism for unlawful content and providing that receipt of such a complaint constitutes knowledge under Article 5).
11. Paul M. Barrett, *Who Moderates the Social Media Giants? A Call to End Outsourcing*, N.Y.U. STERN CTR. FOR BUS. AND HUM. RT., 7 (Jun. 2020).
12. Mark MacCarthy, *A Dispute Resolution Program for Social Media Companies*, BROOKINGS (Oct. 9, 2020), https://www.brookings.edu/research/a-dispute-resolution-program-for-social- media-companies/.
13. *Id.*
14. *National Advertising Division*, BBB National Programs, https://bbbprograms.org/programs/all-programs/national-advertising-division.
15. *Digital Advertising Accountability Program*, BBB National Programs, https://bbbprograms.org/programs/all-programs/daap.
16. Board Members, Board of Governors of the Federal Reserve System (Oct. 14, 2021), https://www.federalreserve.gov/aboutthefed/bios/board/default.htm.
17. David K. Li et al., *MLB Pulls All-Star Game From Atlanta in Protest of Restrictive New Voting Law*, NBC NEWS (Apr. 2, 2021), https://www.nbcnews.com/news/us-news/mlb-pulls-all-star-game-georgia-wake-restrictive-new- voting-n1262930.
18. *See* 17 U.S.C. § 512.

Concluding Statement

1. ETERNALLY VIGILANT: FREE SPEECH IN THE MODERN ERA (Lee C. Bollinger & Geoffrey R. Stone eds., Chicago: University of Chicago Press, 2001); THE FREE SPEECH CENTURY (Lee C. Bollinger & Geoffrey R. Stone eds., New York: Oxford University Press, 2018); NATIONAL SECURITY, LEAKS AND FREEDOM OF THE PRESS: THE PENTAGON PAPERS FIFTY YEARS ON (Lee C. Bollinger & Geoffrey R. Stone eds., New York: Oxford University Press, 2021); SOCIAL MEDIA, FREEDOM OF SPEECH, AND THE FUTURE OF OUR DEMOCRACY (Lee C. Bollinger & Geoffrey R. Stone eds., New York: Oxford University Press, 2022).
2. Communications Act of 1934, 48 Stat. 1064 (1934).
3. *Id.*
4. *Id.*

INDEX

For the benefit of digital users, indexed terms that span two pages (e.g., 52–53) may, on occasion, appear on only one of those pages.