NEW YORK LAW SCHOOL

DigitalCommons@NYLS

Articles & Chapters

Faculty Scholarship

Fall 2021

The Paradox of Free Speech in the Digital World: First Amendment Friendly Proposals for Promoting User Agency

Nadine Strossen

Follow this and additional works at: https://digitalcommons.nyls.edu/fac_articles_chapters

Part of the First Amendment Commons, and the Internet Law Commons

The Paradox of Free Speech in the Digital World: First Amendment Friendly Proposals for Promoting User Agency

Nadine Strossen†

The United States Supreme Court has continued a speech-protective trend dating back to the 1960s, safeguarding even the most controversial speech from government regulation, including speech that critics of this trend label with the stigmatizing terms "hate speech," "disinformation," "misinformation," "extremist speech," and "terrorist speech." In contrast, as dominant online platforms have become increasingly important forums for both individual self-expression and democratic discourse, the platforms have been issuing and enforcing increasing restrictions on their users' speech pursuant to each platform's content moderation policies. These restrictions often suppress speech that the U.S. Constitution bars government from suppressing. As private sector entities, these dominant platforms presumptively have no First Amendment obligation to host any expression or users—unless the platforms should be treated as "state actors," as multiple experts and litigants recently have argued. Moreover, platforms have their own First Amendment rights to determine which speech or speakers they wish to host. Given these platforms' outsized influence, government officials, civil society organizations, and individual experts have proposed a range of measures that would shape the platforms' exercise of their enormous power to censor ideas and speakers on their respective forums.

While many critics complain that dominant platforms are not restricting enough speech, many others lodge the opposite complaint, which is the focus of this Essay. Stressing the goal of facilitating indi-

† John Marshall Harlan II Professor of Law Emerita, New York Law School; past national President, American Civil Liberties Union (1991–2008). Professor Strossen was invited to publish this Essay in connection with the forty-third annual Foulston Siefkin Lecture, which she delivered on March 18, 2021. She thanks the following New York Law School students for their essential contributions to this Essay's footnotes: Kelly Barrett, Michael Kazamias, Kirsten Kovats, Heidi Moore, and Shukra Sabnis. She also thanks Kirsten Kovats for superb editing. Finally, she thanks the following officers and staff members of the *Washburn Law Journal*: Rachel Schwein for organizing the Foulston Siefkin Lecture, and Zach Smith, Lizzy McEntire, Ashley McCall, and Dayton Yarrow for their assistance in editing this Essay. This Essay intentionally is more lightly footnoted than the typical law journal article, but if any reader has trouble locating sources to support particular points, they should contact Professor Strossen at nadine.strossen@nyls.edu.

vidual freedom of choice, which is the ideal from a free speech perspective, this Essay discusses a range of proposed measures to constrain the dominant platforms' censorial power with the goal of promoting user agency. It outlines proposed measures that have garnered significant support, and which warrant serious evaluation, but given the complexity of the issues and the risk of unintended adverse consequences, it does not conclusively endorse implementing any proposal.

> [D]o we really want to trust a handful of chief executives with policing spaces that have become essential parts of democratic discourse? We are uncomfortable with government doing it; we are uncomfortable with Silicon Valley 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.
>
> —Emily Bazelon, journalist, 2021[1]

> While we now may be coming to the realization that the Cyber Age is a revolution of historic proportions, we cannot appreciate yet its full dimensions and vast potential to alter how we think, express ourselves, and define who we want to be. The forces and directions of the Internet are so new, so protean, and so far reaching that courts must be conscious that what they say today might be obsolete tomorrow.
>
> —U.S. Supreme Court Justice Anthony Kennedy, 2017[2]

> Alex Feerst, one of the great thinkers about Internet content moderation, has a revealing metaphor about the real-world work involved. "You might go into it thinking that online information flows are best managed by someone with the equivalent of a PhD in hydrology," he says. "But you quickly discover that what you really need are plumbers."
>
> —Daphne Keller, Director of Stanford University's Program on Platform Regulation, 2020[3]

I. INTRODUCTION

The paradox of free speech in our brave new digital world is well captured by the famous opening lines of Charles Dickens's *A Tale of Two*

---

1. Emily Bazelon, *Why Is Big Tech Policing Speech? Because the Government Isn't*, N.Y. TIMES (Jan. 26, 2021), https://www.nytimes.com/2021/01/26/magazine/free-speech-tech.html [https://perma.cc/785K-MN8Z].

2. Packingham v. North Carolina, 137 S. Ct. 1730, 1736 (2017).

3. Daphne Keller, *Broad Consequences of a Systemic Duty of Care for Platforms*, STAN. CYBER POL'Y CTR. FREEMAN SPOGLI INST. (June 1, 2020), https://cyber.fsi.stanford.edu/news/broad-consequences-systemic-duty-care-platforms [https://perma.cc/4JXE-MAUQ].

Cities: "It was the best of times; it was the worst of times."[4] Free speech has never been so strong—nor so weak. On the one hand, the United States Supreme Court has repeatedly repudiated government censorship of even the most odious, controversial speech,[5] including hate speech;[6] disinformation and misinformation;[7] and speech that advocates extremism, violence, and terrorism.[8] On the other hand, unprecedented numbers of speakers are being silenced by non-governmental censorship[9] of virtually any speech that anyone deems objectionable. This silencing is executed by dominant private platforms[10] seeking profit maximization, and by private groups pressuring these platforms to restrict controversial voices and views. Moreover, government officials across the political spectrum constantly pressure dominant platforms to limit expression in ways that the government itself could not, thereby eluding First Amendment and other constitutional restraints that would rein in direct government censorship.

---

4. CHARLES DICKENS, A TALE OF TWO CITIES 1 (1859).

5. *See, e.g.*, Brown v. Ent. Merchs. Ass'n, 564 U.S. 786, 788 (2011) (violent videogames); United States v. Stevens, 559 U.S. 460, 465 (2010) ("crush videos"); Virginia v. Black, 538 U.S. 343, 347 (2003) (cross burning); R.A.V. v. City of St. Paul, 505 U.S. 377, 393 (1992) (cross burning). It should be noted, though, that the Court also has failed to protect speech rights in important cases, including speech rights asserted by students and prisoners.

6. *See, e.g.*, Matal v. Tam, 137 S. Ct. 1744, 1764 (2017) ("[T]he proudest boast of our free speech jurisprudence is that we protect the freedom to express the thought that we hate.") (internal quotations omitted)); Snyder v. Phelps, 562 U.S. 443, 461 (2011) ("As a Nation, we have chosen . . . to protect even hurtful speech on public issues to ensure that we do not stifle public debate.").

7. *See, e.g.*, United States v. Alvarez, 567 U.S. 709, 727 (2012) ("The remedy for speech that is false is speech that is true."). "Disinformation" refers to information that the speaker knows to be false, whereas "misinformation" refers to information that is in fact false even though the speaker does not know this.

8. *See, e.g.*, NAACP v. Claiborne Hardware Co., 458 U.S. 886, 927 (1982) ("[M]ere advocacy of the use of force or violence does not remove speech from the protection of the First Amendment.") (emphasis omitted)); Brandenburg v. Ohio, 395 U.S. 444, 448 (1969) ("[T]he mere abstract teaching . . . of the moral propriety or even moral necessity for . . . resort to force and violence, is not the same as preparing a group for violent action and steeling it to such action."). *But see* Holder v. Humanitarian Law Project, 561 U.S. 1, 33 (2010) (upholding federal law that prohibited persons and organizations from providing material support to foreign terrorist groups, as applied to support that consisted of advocacy and education, even though "such support was ostensibly intended to support non-violent, non-terrorist activities").

9. "Censorship" is not a legal term of art. This Essay uses the term to refer to any speech restriction that would violate the First Amendment if the government imposed it.

10. This Essay focuses on the dominant platforms because their extraordinary size and influence, as a practical matter, force many people to use their services. *See* THE SOCIAL DILEMMA, https://www.thesocialdilemma.com/ [https://perma.cc/AU9E-B8H6] (last visited Sept. 27, 2021) (defining "the dilemma" as unprecedented and controlling because "[n]ever before have a handful of tech designers had such control over the way billions of us think, act, and live our lives"). Concerning smaller online platforms, potential users retain meaningful choices about whether to participate. *See* Sara Wilson, *The Era of Antisocial Media*, HARV. BUS. REV. (Feb. 5, 2020), https://hbr.org/2020/02/the-era-of-antisocial-social-media [https://perma.cc/Q2UY-9XLD] (describing smaller social apps as "campfires" where users are "offer[ed] a more intimate oasis"). Accordingly, such smaller platforms should (as a normative matter) have more latitude to choose what content they wish to host, and to define their social community. As this Essay explains, all private sector companies—regardless of size—have such latitude under U.S. law.

Private sector censors are not restricted by the First Amendment; to the contrary they are shielded by it.[11] When individuals and civil society organizations urge dominant platforms to restrict certain speech through various tactics such as boycotts, they are exercising their First Amendment rights of speech and association.[12] Likewise, when platforms restrict speech, they too are exercising their First Amendment rights: their rights to choose which speech and speakers they wish to host based on their editorial judgment, just as traditional media companies do.[13] Although private sector censors do not violate the First Amendment, they do reduce free speech, which adversely affects liberty, equality, and democracy. No matter how strictly courts curb government censorship, we will not enjoy meaningful freedom of speech in our digital world unless we can curb non-governmental censorship: the dominant platforms' speech restrictions and the "cancel culture" that fuels it. Therefore, free speech advocates must refocus their efforts, seeking out means that limit the private sector's speech-suppressive power—whether wielded by the dominant platforms or by social media mobs—while concomitantly respecting the First Amendment rights of these private sector actors.

Toward that end, this Essay discusses four major points. First, it elaborates on the challenge posed by the fact that dominant platforms' exercise of unprecedented censorial power over their users' speech simultaneously constitutes an exercise of the platforms' own First Amendment rights. Next, this Essay summarizes the landmark Supreme Court decision and congressional legislation issued shortly after the World Wide Web's launch into the public domain, which promised to maximize individual freedom of choice and user empowerment. Adhering to this utopian vision as the ideal toward which we should strive, this Essay then details the current dystopian situation: the dominant platforms' unprecedented censorial power and the resulting damage to free speech and other fundamental constitutional values. Finally, the Essay sketches a range of proposed means to redress this situation—namely, means to shape the dominant platforms' content moderation policies so as to maximize users' free speech protection while protec-

11. *See generally* ERWIN CHEMERINSKY, CONSTITUTIONAL LAW: PRINCIPLES AND POLICIES 553 (6th ed. 2019). Some constitutional law experts have argued that the dominant platforms' content moderation practices should satisfy either or both of the exceptions to the state action doctrine that the Supreme Court has recognized: the entanglement exception and the public function exception. These arguments are discussed *infra* Section V.A.

12. During the summer of 2020, there was such a boycott against Facebook after the release of Facebook's Civil Rights Audit report. Cat Zakrzewski & Hamza Shaban, *Facebook Met with Civil Rights Groups After Hundreds of Companies Joined Ad Boycott*, WASH. POST (July 7, 2020), https://www.washingtonpost.com/technology/2020/07/07/facebook-boycott-civil-rights-zuckerberg/ [https://perma.cc/AWC8-7UZM].

13. *See* Mia. Herald Publ'g Co. v. Tornillo, 418 U.S. 241, 255 (1974).

ting the platforms' own First Amendment rights.[14] These proposals have garnered substantial attention and support, and each warrants serious examination. Recognizing the complexity of this subject, however, the Essay does not conclusively endorse any proposal.

II. THE CHALLENGE: TO AFFORD ONLINE SPEECH THE SAME STRONG PROTECTION AS IS AFFORDED TO OFFLINE SPEECH

The Supreme Court has recurrently examined successive new communications media to assess whether—and, if so, to what extent—expression conveyed through these medias is subject to First Amendment protection. For example, the Court did not initially deem films to be entitled to any First Amendment protection at all,[15] only repudiating that position years later.[16] To this day, the Court accords over-the-air broadcast expression only limited First Amendment protection; words that the Federal Communications Commission ("FCC") deems "patently offensive" are still strictly censored on broadcast radio and television, even when they constitute essential elements of important works and even though they may be freely conveyed on all other media.[17]

As discussed below, Congress overwhelmingly passed a law in 1996—signed by then-President Bill Clinton—that would have relegated the Internet to the same second-class First Amendment status as the broadcast media.[18] To highlight the speech suppression that Congress and the Clinton Administration supported in the online context, I provide in the following paragraph one example of important speech that is still censored from broadcast media today.

In my regular on-air interviews about free speech issues, I am barred from fully describing one of the Court's most important First Amendment decisions of all time, *Cohen v. California*.[19] In *Cohen*, the Court overturned an individual's conviction for protesting the Vietnam War-era draft by

14. For my discussion of the corresponding issues concerning "cancel culture," see Nadine Strossen, *Resisting Cancel Culture: Promoting Dialogue, Debate, and Free Speech in the College Classroom*, AM. COUNCIL OF TR. AND ALUMNI (Dec. 3, 2020), https://www.goacta.org/resource/resisting-cancel-culture/ [https://perma.cc/D7PP-3E9E].

15. *See* Mut. Film Corp. v. Indus. Comm'n of Ohio, 236 U.S. 230, 244–45 (1915) *overruled in part by* Joseph Bursyn, Inc. v. Wilson, 343 U.S. 495 (1952).

16. *See Joseph Burstyn*, 343 U.S. at 502 (abandoning the position taken in *Mutual Film Corporation* and concluding that, "expression by means of motion pictures is included within the free speech and free press guaranty of the First and Fourteenth Amendments").

17. *See, e.g.*, FCC v. Pacifica Found., 438 U.S. 726, 748 (1978) ("Patently offensive, indecent material presented over the airwaves confronts the citizen, not only in public, but also in the privacy of the home, where the individual's right to be left alone plainly outweighs the First Amendment rights of an intruder.") (citing Rowan v. Post Office Dept., 397 U.S. 728 (1970)).

18. Telecommunications Act of 1996 § 502, 47 U.S.C. § 230.

19. Cohen v. California, 403 U.S. 15, 16 (1971).

wearing, inside a courthouse, a jacket proclaiming, "[f]uck the draft."[20] In contrast, to this day, that same phrase cannot be proclaimed over the airwaves—even in the context of my free-speech-focused interviews. The controversial four-letter word, which was the basis for the breach of the peace conviction that *Cohen* overturned, of course appears in the hallowed pages of the official Supreme Court Reports, because the full *Cohen* decision duly recorded the statement at issue, expletive and all. This four-letter word also appears in audio and video discussions of *Cohen* that are transmitted via film, cable, satellite, and online streaming services. But the "F-word," along with other "patently offensive" expression, is still categorically censored from broadcast media regardless of the context, even to quote language from a major Supreme Court decision. Fortunately, the Supreme Court rejected efforts by its coordinate branches of the federal government to consign Internet speech to the same reduced First Amendment protection that broadcast speech receives.[21]

III. THE LEGAL FOUNDATIONS FOR FREE SPEECH ONLINE: *RENO V. ACLU* AND SECTION 230

When the Internet first attracted the attention of the public, press, and politicians, it was greeted the same way as all new media had been throughout history: with suspicion and fear about its power to convey controversial and potentially harmful ideas and information, widely and quickly, to vast audiences encompassing children and others considered especially susceptible to negative influences. Accordingly, it is not surprising that Congress's initial legislation in response to this new medium was censorial in nature: The 1996 Communications Decency Act ("CDA") outlawed "patently offensive" or "indecent" online expression.[22] The American Civil Liberties Union ("ACLU") immediately spearheaded a First Amendment challenge to the pertinent CDA provisions, along with a diverse coalition of human rights organizations, publishers, libraries, and others engaged in disseminating information and ideas that could well be deemed "patently offensive" or "indecent" under the CDA.

In the landmark 1997 *Reno v. ACLU* decision,[23] the ACLU and its allies won a stunning, essentially unanimous,[24] Supreme Court victory. The

20. *Id.*

21. *See* Reno v. ACLU, 521 U.S. 844, 885 (1997).

22. *See id.* at 883 (discussing and ultimately holding the "indecent transmission" and "patently offensive display" provisions of the Communications Decency Act of 1996, 47 U.S.C. § 223, unconstitutional).

23. *Id.* at 849.

24. *Id.* at 886. Justice O'Connor authored a partial dissent, in which Chief Justice Rehnquist joined, but it concerned only a narrow, particular application of the statute (as applied to an online communication involving only one adult and one or more minors, such as when an adult knowingly

Court rejected the government's contention that the Internet was analogous to broadcast media and therefore should be subject to the same second-class First Amendment treatment. Specifically, the Court stressed that broadcast radio and TV afford users relatively little choice or control, and hence are "invasive" in nature, justifying top-down government regulation.[25] In contrast, the Court noted that online communications do "not 'invade' an individual's home or appear on one's computer screen unbidden."[26] Accordingly, the Court held that online communications are entitled to the same robust First Amendment protection as the print media, declaring that online end-users should be empowered to make their own choices about what material they personally wish to view, and what material they deem fit for their own young children to view.

In short, *Reno* extended to the online context the same free speech ideals that the Court had protected in the offline context for all media other than broadcast: protecting individual freedom of choice and end-user empowerment, and rejecting top-down gatekeeper control via government censorship.[27] In terms of First Amendment doctrine, this continues to mean that any restriction of online speech is subject to strict judicial scrutiny: The restriction is presumed unconstitutional, and government can only overcome that presumption by satisfying the demanding burden of proving that the restriction is necessary, and the least restrictive means, to advance a purpose or goal of compelling importance.

The *Reno* Court was willing to assume, for the sake of argument, that the CDA was designed to promote a sufficiently important goal: protecting children from expression that could potentially have a harmful impact on them. The Court concluded, however, that outlawing the targeted expression was not the least speech-restrictive alternative means to promote that goal. Rather, the Court concluded that end-user controls, implemented by parents and other adults, could protect children at least as effectively as blanket censorship—if not more effectively.[28] *Reno* also reaffirmed prior rulings that government may not restrict adults' free speech rights on the rationale that the restrictions benefited children, reasoning that precious free speech rights should not be reduced to the lowest common denominator, such that no adult may convey or receive speech that might adversely affect the most vulnerable among us, including children.[29] The Court repeatedly

sends an email to a minor); both Justices agreed with the majority's broad holdings about the law's general unconstitutionality. *Id.*

25. *Id.* at 868 (quoting ACLU v. Reno, 929 F. Supp. 824, 844 (E.D. Pa. 1996)).

26. *Id.* at 869.

27. For the sake of brevity, I will summarize these interrelated goals with the label "user agency."

28. *Reno*, 521 U.S. at 877.

29. *Id.*

endorsed these essential speech-protective principles in a series of post-*Reno* decisions concerning other congressional restrictions on online content.[30]

The preeminent touchstone for online (and offline) free speech—maximizing user agency—was also reflected, ironically, in the Communications Decency Act itself. Although *Reno* struck down the CDA's provisions outlawing "patently offensive" and "indecent" online expression, it did not strike down a remaining CDA provision that was unchallenged by the ACLU and its allies: the provision that is now widely known as "Section 230."[31] This provision, which broadly (but not completely) immunized online intermediaries from liability for either hosting or not hosting third-party content, received overwhelming bipartisan support in both houses of Congress.[32]

The ACLU and others who lobbied and litigated against the censorial CDA provisions welcomed (and continue to support) Section 230 because it fosters less speech-restrictive alternatives to government censorship. It does this by permitting online platforms that host third-party content to make their own choices about what content moderation policies to enforce (with a few limited exceptions). The platforms may adopt policies that range from permitting all constitutionally protected speech, to permitting only speech that is neither "indecent" nor "patently offensive" (just as the CDA had mandated until *Reno* invalidated the pertinent provisions), to permitting only speech that does not contain other types of controversial content. Thus, consistent with the free speech ideal of user agency, Section 230 was designed to facilitate the flourishing of myriad platforms, with multiple content moderation policies, so that end users may choose to utilize those platforms whose policies align with their personal values and concerns. This user agency goal was reflected in Section 230's title in the House of Representatives: the "Internet Freedom and Family Empowerment Act." As stated by one of its two co-sponsors, Congressman Christopher Cox (R-CA): "We want to encourage [online platforms] to do everything possible for us, the customer, to help us control, at the portals of our computer, at the front door of our house, what comes in and what our children see."[33]

30. *See* Ashcroft v. ACLU, 535 U.S. 564 (2002); Ashcroft v. ACLU, 542 U.S. 656 (2004); ACLU v. Mukasey, 534 F.3d 181 (3rd Cir. 2008), *cert. denied*, Mukasey v. ACLU, 555 U.S. 1137 (2009).

31. The number "230" was assigned to this CDA provision once it was incorporated into the United States Code, under Title 47.

32. Fair Hous. Council of San Fernando Valley v. Roommates.com, 521 F.3d 1157, 1177 (9th Cir. 2008).

33. 141 CONG. REC. H8460 (daily ed. Aug. 4, 1995) (statement of Rep. Cox).

Thanks to *Reno* and Section 230, champions of online free speech could celebrate what promised to be an unprecedented era of maximizing freedom of speech: enabling potentially everyone in the world to communicate easily, freely, and cheaply with potentially everyone else in the world, as the *Reno* decision itself predicted. In contrast with the traditional media, where powerful gatekeepers strictly limit opportunities for third parties to speak on their platforms, the Court hailed "the vast democratic forums of the Internet," where "any person with a phone line can become a town crier with a voice that resonates farther than it could from any soapbox."[34] Instead of top-down, centralized control over expressive choices, *Reno* and Section 230 vested end users with the decentralized power to choose for themselves with whom and how they would engage online.

IV. THE CURRENT DYSTOPIAN SITUATION: THE PRIVATE SECTOR'S UNPRECEDENTED CENSORIAL POWER AND ITS DAMAGE TO FREE SPEECH AND OTHER CONSTITUTIONAL IDEALS

A. "Platformization" and "Cancel Culture"

Alas, the advent of what has been called the "platformization"[35] of the online realm has eclipsed the halcyon online free speech era that dawned with *Reno v. ACLU* and Section 230. The dominant platforms continue to impose increasingly strict content moderation policies. Moreover, their takedowns of posts and speakers, in addition to other sanctions such as downranking, demonetizing, and attaching pejorative labels to content, continue to grow at an alarming rate. These speech restrictions have provoked complaints from users across the political and ideological spectrum, all of whom allege that their speech was unjustifiably restricted. Such complaints have been made by everyone from Black Lives Matter leaders[36] to Donald

---

34. *Reno*, 521 U.S. at 868, 870.

35. *See generally* Anne Helmond, *The Platformization of the Web: Making Web Data Platform Ready*, SOC. MEDIA & SOC'Y 1 (2015), https://doi.org/10.1177/2056305115603080 [https://perma.cc/N46L-38D5]. Recognizing the adverse impact on users' free speech that has resulted from the rise of dominant platforms, Mike Masnick proposed measures that would replicate the "protocol-based" early Internet era. Mike Masnick, *Protocols, Not Platforms: A Technological Approach to Free Speech*, KNIGHT FIRST AMENDMENT INST. AT COLUM. UNIV. (Aug. 21, 2019), https://knightcolumbia.org/content/protocols-not-platforms-a-technological-approach-to-free-speech [https://perma.cc/YMB7-F548]. *See infra* Section V.A.3.i.a.

36. *See, e.g.*, Complaint at 56–60, Newman et al v. Google LLC et al., No. 20-CV-04011-LHK, 2021 WL 2633423 (N.D. Cal. June 16, 2020). The Plaintiffs in *Newman* filed a class action lawsuit against YouTube, Google, and others alleging that those defendants violated their free speech rights by applying "Restricted Mode" filtering to restrict the reach of videos tagged with, among other things, the abbreviation "BLM" and "Black Lives Matter."

Trump.[37] "Cancel culture" has bolstered the speech-stifling impact of the dominant platforms' policies, with many individuals and groups pressuring these platforms to suppress ever more expression. Consequently, would-be online speakers are subject to the same top-down, centralized gatekeeping that the *Reno* Court invalidated when it was government-mandated. Now, though, that centralized control is exercised by dominant platforms, which are often pressured to act by powerful, influential individuals or groups, ranging from government officials to "Twitter mobs."

The adverse censorial impact of platformization and "cancel culture" is likely greater than that which would have resulted had the *Reno* Court allowed the speech-suppressive CDA provisions to remain intact, in at least two significant respects. First, the dominant platforms may now restrict literally *all* speech on their forums—speech with any message and from any speaker. Moreover, these platforms may completely oust a speaker, thus imposing the functional equivalent of a permanent prior restraint, which is considered the most onerous speech restriction that government could in theory impose—although the courts almost never uphold even a temporary prior restraint. For example, under Twitter's terms of service, it can remove any person from the platform "at any time for any or no reason."[38] While the CDA targeted overly broad categories of "patently offensive" and "indecent" expression, it certainly did not encompass *all* expression.

Consider, for example, "political speech"—speech about public affairs—which the Supreme Court has consistently placed at the apex of the First Amendment hierarchy because it plays an essential role in our democratic republic, where "We the People"[39] hold sovereign power. If we cannot freely exchange information and ideas with and about public officials and candidates for public office, then how can we hold them accountable? As the Court declared, "Speech concerning public affairs is more than [individual] self-expression; it is the essence of self-government."[40] Yet the dominant platforms have increasingly suppressed political speech, even when it comes from the People's most powerful elected official, and even when it involves urgent matters of public concern. Almost without exception, controversial political speech that the dominant

37. *See, e.g.*, Cristiano Lima, *How Trump's Fury at Silicon Valley Fixated on the Little-Known Section 230*, POLITICO (Dec. 2, 2020, 7:50 PM), https://www.politico.com/news/2020/12/02/trump-silicon-valley-section-230-442436 [https://perma.cc/BX5L-CFMP] (discussing then-President Donald Trump's efforts to weaken Section 230's immunity protections that are provided "for the tech platforms that fact-checked, filtered or blocked some of his messages during the months before his defeat at the polls").

38. Twitter, Inc., User Agreement, https://cdn.cms-twdigitalassets.com/content/dam/legal-twitter/site-assets/privacy-policy-new/Privacy-Policy-Terms-of-Service_EN.pdf [https://perma.cc/83UL-37NH] (effective June 18, 2020).

39. U.S. CONST. pmbl.

40. Garrison v. Louisiana, 379 U.S. 64, 74–75 (1964).

platforms have restricted as violating various content moderation policies—for example, policies restricting "hate speech" or "disinformation"—would be constitutionally protected from government censorship.

This leads to the second major respect in which the dominant platforms' current speech suppression measures pose more dangers than government censorship: The platforms' speech restrictions are not subject to any of the constitutional limits on comparable government restrictions. Under the "state action doctrine," constitutional limits restrain only action by government officials or agencies, with very narrow exceptions; consequently, there are almost no constitutional restraints on private sector actors, no matter how powerful they are.[41] Accordingly, the dominant platforms' speech-restrictive policies are entirely unconstrained by not only the First Amendment, but also the Fifth and Fourteenth Amendments' Due Process Clauses and the Fourteenth Amendment's Equal Protection Clause.

The scale of the dominant platforms' speech restrictions is so enormous as to be almost incomprehensible. For example, the most recent (third quarter 2021) "Community Standards Enforcement Report" for Meta (which owns Facebook and Instagram) reported that from July to September of 2021, it removed or otherwise "[took] action on"[42] 22.3 million Facebook posts and 6.0 million Instagram posts deemed to constitute prohibited hate speech,[43] which constitutes an average of over 314,000 posts per day. This number bears repetition and reflection: *in just one single day, just one company, enforcing just one of its many content moderation standards, removed or otherwise suppressed 314,000 communications.*

In addition to the dominant platforms exercising censorial powers that only the government had exercised in the past, the traditional relationship between government and private sector parties has also been inverted. In the past, government officials suppressed the speech of private actors, but now private actors are increasingly suppressing the speech of government officials. To cite the best-known example, after the violent, insurrectionary invasion of the U.S. Capitol on January 6, 2021, both Facebook and Twitter

41. *See infra* Section V.A.

42. "Taking action could include removing a piece of content from Facebook or Instagram, covering photos or videos that may be disturbing to some audiences with a warning, or disabling accounts." *Content Actioned*, META, https://transparency.fb.com/policies/improving/content-actioned-metric/ (last visited Nov. 14, 2021).

43. *See* Guy Rosen, *Community Standards Enforcement Report, Third Quarter 2021*, META (Nov. 9, 2021), https://about.fb.com/news/2021/11/community-standards-enforcement-report-q3-2021/; *Hate Speech*, META, https://transparency.fb.com/data/community-standards-enforcement/hate-speech/facebook/ (last visited Nov. 14, 2021) (reporting and comparing statistics over time for content flagged as hate speech between Facebook and Instagram).

removed then-President Donald Trump's accounts, either indefinitely or permanently.[44]

The platforms' suppression of speech posed a dire situation even before the COVID-19 pandemic, but the severity of that situation has since intensified. Prior to COVID-19, most of us had heard examples of problematic takedowns and restrictions on controversial speech. We had also heard complaints, at least as loud, that there had not been enough such takedowns or restrictions. Many of us had even had our own online speech subjected to seemingly arbitrary, irrational, and unjustified censorship by dominant platforms—that certainly has happened to me. I testified on these issues before a congressional committee in 2019 (along with representatives of Facebook, Google, and Twitter). I saw firsthand Democratic and Republican committee members unanimously agree on one common concern: the dominant platforms take down too much speech of some kinds, and not enough speech of other kinds. Of course, the same committee members radically disagreed as to which speech fell within each category.

Due to the impact of the COVID-19 pandemic, this speech-suppressive situation has worsened, because the dominant platforms have increased their reliance on automated tools to enforce their content moderation policies.[45] Since concepts such as "hate speech" depend on an intense evaluation of the overall context in which the speech is uttered, it is impossible for automated tools to undertake the necessary assessment. Therefore, online expression is being subjected to more, and broader, unjustified restrictions on the ground that it constitutes impermissible hate speech. History shows that steps initially taken to cope with a crisis tend to persist long after the crisis has passed. Especially given the dominant platforms' movement toward increasingly automated content moderation practices even pre-COVID-19,[46] we can anticipate growing suppression of purported hate speech, as well as other controversial speech.

---

44. *See, e.g.*, Tony Romm and 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/ [https://perma.cc/R9X8-2N4L] (explaining that Facebook indefinitely banned then-President Donald Trump because it believed that the "risks of allowing the President to continue to use [Facebook] . . . [were] simply too great"); Kate Conger and Mike Isaac, *Twitter Permanently Bans Trump, Capping Online Revolt*, N.Y. TIMES, https://www.nytimes.com/2021 /01/08/technology/twitter-trump-suspended.html?searchResultPosition=7 [https://perma.cc/LX4W-QG 9W] (Jan. 12, 2021) (discussing Twitter's ban of then-President Donald Trump "due to the risk of further incitement of violence").

45. *See e.g.*, Evelyn Douek, *COVID-19 and Social Media Content Moderation*, LAWFARE BLOG (Mar. 25, 2020, 1:10 PM), https://www.lawfareblog.com/covid-19-and-social-media-content-moderatio n [https://perma.cc/YV52-TG2E] ("Major tech companies have begun sending home human workers who review social media content and relying more heavily on artificial intelligence (AI) tools to do the job instead.").

46. *See, e.g.*, Spandana Singh, *Everything in Moderation: An Analysis of How Internet Platforms Are Using Artificial Intelligence to Moderate User-Generated Content*, NEW AM. (July 22, 2019),

B. Users Must Retain Freedom to Choose Which Controversial Speech (Not) to View

Many proponents of restricting various kinds of controversial speech ask the following question: "What is the downside of suppressing too much hate speech, or too much controversial speech of other sorts, such as dis- or misinformation?" After all, the argument goes, this speech at best provides scant, if any, social value, and at worst it causes substantial harm. Even if we accepted these propositions for the sake of argument,[47] we would nonetheless also have to recognize the inevitable dangers in entrusting to any gatekeeper—whether a government or private actor—the power to suppress these and other kinds of controversial speech.

Inevitable dangers result from the inherently vague concepts and definitions of the targeted speech. After all, hate is an emotion, consisting of subjective feelings. What one person considers hateful speech, another may well consider neutral or even loving.[48] One powerful illustration, which relates to a major Supreme Court decision,[49] happens to have ties to Topeka, Kansas–the location of Washburn University. It involved the Westboro Baptist Church ("WBC"), located in Topeka; also noteworthy is the fact that a couple of key WBC leaders graduated from the University and its law school. For many people, a paradigmatic example of hate speech is WBC's central motto, embedded in its website name: www.godhatesfags.org. Yet Megan Phelps-Roper, a devout member and leader of the WBC from early childhood until age twenty-six, has explained that WBC members sincerely, deeply believe not only that they love LGBTQ individuals, but also that they are the only people who love LGBTQ individuals. As Phelps-Roper has stressed, the Church's message is not that *it* or *its members* hate the targeted individuals, but rather that *God* hates those individuals. Those faithful to the Church believe that LGBTQ individuals are doomed to eternal damnation unless they somehow renounce their sexual orientations or gender identities. Hence, by exhorting LGBTQ individuals

---

https://www.newamerica.org/oti/reports/everything-moderation-analysis-how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content/introduction [https://perma.cc/GU7M-4AT7] (explaining that, in response to increased government and public pressure to remove offensive content, platforms that host third-party, user-generated content "have developed or adopted automated tools to enhance their content moderation practices, many of which are fueled by artificial intelligence and machine learning").

47. *See* NADINE STROSSEN, HATE: WHY WE SHOULD RESIST IT WITH FREE SPEECH, NOT CENSORSHIP 79 (2018) 121–28 (discussing evidence that constitutionally protected hate speech does not necessarily cause feared harms, including violence, discrimination, and psychic and emotional trauma).

48. *Id.* at 69–80 (discussing the ambiguities and subjectivity inherent in writing and enforcing hate speech laws).

49. Snyder v. Phelps, 562 U.S. 443 (2011) (holding that Westboro Baptist Church ("WBC") members had a First Amendment right to convey hateful messages about LGBTQ individuals and others in public places because these messages related to matters of public concern).

to do so, WBC members believe they are doing the loving work necessary to save these individuals' immortal souls.[50]

Other examples abound. For example, many people consider the Black Lives Matter movement ("BLM") to convey positive messages, including an expression of love and respect for Black people.[51] Yet, others have denounced BLM-related expression as constituting hate speech against white and non-Black people, and police officers.[52] Correspondingly, some people view the Blue Lives Matter countermovement as conveying positive, loving messages about police officers, whereas others view it as conveying hateful, negative messages against BLM activists.[53] Even the phrase "free

---

50. Terry Gross, *How Twitter Helped Change the Mind of a Westboro Baptist Church Member*, NPR (Oct. 10, 2019, 4:31 PM), https://www.npr.org/2019/10/10/768894901/how-twitter-helped-change-the-mind-of-a-westboro-baptist-church-member [https://perma.cc/DXH4-P5JV].

> Westboro would quote this passage from the book of Leviticus that, for them, shows that the definition of "love thy neighbor" is to rebuke your neighbor when you see him sinning. And if you don't do that, then you hate your neighbor in your heart. Because you are watching this person go down this bad path that is going to lead them to the curses of God in this life, and hell in the world to come. And you failed to warn them. You didn't give them the opportunity to repent, to share with them the truth of God.

Id.

51. *See, e.g.*, Leah Asmelash, *How Black Lives Matter Went from a Hashtag to a Global Rallying Cry*, CNN (July 26, 2020, 2:00 PM), https://www.cnn.com/2020/07/26/us/black-lives-matter-explainer-trnd/index.html [https://perma.cc/manage/create?folder=11648-124024] ("[O]ur work is full of love, healing and dignity . . . [a]nd we center Black people's humanity and life over our death and decimation.") (internal quotations omitted)).

52. Rudy Giuliani, *Face Facts: 'Black Lives Matter' Is All About Hate*, N.Y. POST (Sept. 24, 2020, 7:37 PM), https://nypost.com/2020/09/24/face-facts-black-lives-matter-is-all-about-hate/ [https://perma.cc/PB4P-A2B7] ("From the start, both the organization and the movement—BLM writ large—have been about hatred and violence that extends beyond police and includes all white people, all blacks who are conservative and the United States of America."); Kevin Liptak & Kristen Holmes, *Trump Calls Black Lives Matter a 'Symbol of Hate' as He Digs in on Race*, CNN (July 1, 2020, 4:32 PM), https://www.cnn.com/2020/07/01/politics/donald-trump-black-lives-matter-confederate-race/index.html [https://perma.cc/HDY6-MQZX] (explaining that then-President Donald Trump's reaction to the painting of a Black Lives Matter mural on the street in front of Trump Tower in New York City was to call "Black Lives Matter" a "symbol of hate").

53. Fionnuala O'Leary, *BLM BACKLASH What Is Blue Lives Matter and Why Do Some People Consider It Racist?*, THE SUN, https://www.the-sun.com/news/992088/blue-lives-matter-racist-flag-blm-protests/ [https://perma.cc/9Z89-RYEH] (Aug. 19, 2020, 2:31 PM) (explaining that the Blue Lives Matter flag—which "replaces the red of a traditional American flag with black and incorporates a blue band"—is considered to "[have] some racist connotations after being spotted alongside Confederate flags at the deadly white supremacist rally in Charlottesville, Virginia, in 2017") ("Despite Blue Lives Matter supporters claiming the 'Thin Blue Line' stands for professional pride, some consider the countermovement [to Black Lives Matter] to be racist because its flag has been flown by white supremacists.").

speech" has been denounced as constituting hateful, harmful speech,[54] as has the American flag[55] and the last name of the 45th U.S. President.[56]

Similar definitional problems inherently mar any top-down efforts to proscribe other categories of expression that are typically targeted for restriction, including dis- or misinformation and extremist or terrorist content. The definitional problems concerning dis- or misinformation result in large part from the fact that few statements of general interest or concern involve simply straightforward, settled factual matters. To the contrary, statements that provoke charges of dis- or misinformation almost always involve matters of interpretation, ideas, or opinion; yet, as the Supreme Court has declared, "there is no such thing as a false idea."[57] In short, any alleged dis- or misinformation that reflects an idea—no matter how disputed it might be—should not be exorcised from public debate, including on dominant platforms, based on its alleged falsity.

Likewise, the inherent subjectivity of extremist or terrorist information is well captured by the old saying that, "one person's terrorist is someone else's freedom fighter." During 2020 and 2021, the terms "extremist" and "terrorist" were hurled at all manner of protesters in the U.S., including those who protested the following: police killings of unarmed Black men, pandemic restrictions, and election policies and outcomes. In addition, as international human rights activists have pointed out, much extremist or terrorist content also usefully serves to document its creators' human rights abuses.[58] Therefore, "thanks" to the dominant platforms' (often automated) removal of extremist or terrorist content, it is far more difficult, if not impossible, to pursue justice against the abusers.[59]

---

54. *See, e.g.*, Jonathan Butcher, *College Kowtows to Students Saying 'Free Speech Harms'* (July 19, 2019), DAILY SIGNAL, https://www.dailysignal.com/2019/07/19/college-kowtows-to-students-saying-free-speech-harms/ [https://perma.cc/L3TJ-Y44G] (discussing student protests at Williams College, when it proposed to adopt the speech-protective "Chicago Principles," originally adopted by the University of Chicago; student protestors carried signs reading "free speech harms," and suggested that the Williams faculty who promoted the Chicago Principles "were trying to 'kill' the students").

55. *See* STROSSEN, *supra* note 47, at 78–79 (explaining that, in 2015, "student government leaders at the University of California, Irvine voted to ban the display of the American flag, stating that it has been flown in instances of colonialism and imperialism, and can be interpreted as hate speech") (internal quotations omitted)).

56. *See, e.g.*, Will Carless, *They Spewed Hate. Then They Punctuated It with the President's Name*, THE WORLD (Apr. 20, 2018, 11:15 AM), https://www.pri.org/stories/2018-04-20/they-spewed-hate-then-they-punctuated-it-president-s-name [https://perma.cc/7UJJ-ELE7] ("This renaissance of hate features something new: xenophobic, racist and homophobic attacks punctuated with President Donald Trump's name.").

57. Gertz v. Robert Welch, Inc., 418 U.S. 323, 339 (1974).

58. *See, e.g.*, Avi Asher-Schapiro & Ben Barkawi, *'Lost Memories': War Crimes Evidence Threatened by AI Moderation*, REUTERS (June 19, 2020, 12:43 PM), https://www.reuters.com/article/us-global-socialmedia-rights-trfn/lost-memories-war-crimes-evidence-threatened-by-ai-moderation-idUSKBN23Q2TO [https://perma.cc/4MLC-K9HE].

59. *Id.* (reporting that, since the beginning of 2020, "the rate of content takedowns of Syrian human rights documentations on YouTube roughly doubled (from 13% to 20%)").

Even expression that initially appears to convey straightforward matters of fact turns out, upon closer examination, to involve matters of interpretation. For example, assuming historians agreed on the precise number of Armenians who were killed by Turks during World War I, they still disagree about whether those killings resulted from a predetermined plan to eradicate an ethnic group, which is the definition of genocide. Therefore, to outlaw denial of "the Armenian Genocide" as a form of either hate speech or disinformation, as several countries have done, inevitably punishes even good-faith statements by world-renowned historians. For example, Princeton Professor Bernard Lewis, a noted expert on the Ottoman Empire, was subjected to a criminal prosecution and three civil suits in France for questioning whether this mass murder—which he, of course, acknowledged—should be classified as within the legal meaning of "genocide."[60]

The inevitable role of interpretation and opinion in discussing virtually all matters of public interest can also be illustrated in the context of the ongoing COVID-19 pandemic. Given scientists' evolving understanding of the novel coronavirus, what constitutes (in)accurate information concerning transmission and effective countermeasures has been the subject of vigorous disagreement among scientists and physicians, with many experts revising their own views over time. Therefore, to stifle supposed COVID-19 dis- or misinformation, as the dominant platforms have done, inevitably suppresses important—and potentially life-saving—information and discussion.

Facebook attached a stigmatizing label to COVID-19 related content in March 2021 that demonstrates the subjective—and at least potentially misleading—nature of the dominant platforms' dis- and misinformation determinations concerning the pandemic. The label branded a *Wall Street Journal* op-ed authored by John Hopkins University Professor Dr. Marty Makary as, "Missing Context. Independent fact-checkers say this information could mislead people."[61] Dr. Makary's op-ed, "We'll Have Herd Immunity by April," presented his opinion, grounded in evidence that he cited, that Americans would have sufficient immunity to sharply reduce COVID-19's spread by April 2021.[62] Evidence-based as his projection was, he offered it not as a straightforward factual matter but rather as a

60. *See* STROSSEN, *supra* note 47, at 102–03.

61. Wall Street Journal Editorial Board, *Fact-Checking Facebook's Fact Checkers*, W.S.J. (Mar. 5, 2021, 6:36 PM), https://www.wsj.com/articles/fact-checking-facebooks-fact-checkers-11614987375 [https://perma.cc/UQ2R-SVXU].

62. Marty Makary, *We'll Have Herd Immunity by April*, W.S.J. (Feb. 18, 2021, 12:35 PM), https://www.wsj.com/articles/well-have-herd-immunity-by-april-11613669731 [https://perma.cc/AJ5J-CWRR].

matter of opinion. Nonetheless, Facebook's fact-checkers mischaracterized this piece as a factual assertion, despite their own description of it as a "Misleading Wall Street Journal *opinion* piece." These fact-checkers proceeded to charge that the piece "makes the unsubstantiated claim that the U.S. will have herd immunity by April 2021." As the Wall Street Journal Editorial Board commented: "This is counter-opinion masquerading as fact checking."[63] The Editorial Board further explained: "Scientists often disagree over how to interpret evidence. Debate is how ideas are tested and arguments are refined. But Facebook's fact checkers are presenting their opinions as fact and seeking to silence other scientists whose views challenge their own."[64] In short, Facebook's label of Dr. Makary's op-ed as "misleading" was itself misleading.

Given the irreducibly vague contours of virtually any targeted controversial expression, all those who enforce limitations on such expression necessarily have broad discretionary power and will likely wield that power according to their own subjective values or those of powerful societal groups. For this reason, it is not surprising—although tragic—that hate speech policies, which are designed to protect members of traditionally disempowered minority groups, are instead disproportionately enforced against members of those groups.[65] The enforcement record from around the world, and throughout history, shows that these discretionary, malleable concepts of suppressible speech are consistently enforced to silence dissident views and minority voices.[66]

For the foregoing reasons, human rights activists and organizations worldwide have opposed restricting expression solely because of its controversial messages, whether the restrictions are enforced by government or by dominant platforms. Instead, many human rights advocates concur, speech should be restricted only if, considering its full context, it directly causes or threatens certain specific, serious, and imminent harm, which can only be averted through the restriction. This standard is incorporated into not only modern First Amendment law, but also contemporary international free speech law.[67]

---

63. Wall Street Journal Editorial Board, *supra* note 61.
64. *Id.*
65. *See* STROSSEN, *supra* note 47, at 86–91.
66. *Id.* at 81–91.
67. Nadine Strossen, *United Nations Free Speech Standards as the Global Benchmark for Online Platforms' Hate Speech Policies*, 29 MICH. ST. INT'L L. REV. 307, 334–46 (2021).

C. The Dominant Platforms' Content Moderation Policies, as well as "Cancel Culture" Pressure, Constitute Exercises of First Amendment Rights

Although the dominant platforms' content moderation policies thwart freedom of speech when they stifle user agency regarding which information and ideas to convey and receive, these policies do not violate users' First Amendment rights. Rather, these censorial actions constitute exercises of the dominant platforms' own First Amendment rights. As noted above, these private sector actors have no First Amendment obligations to anyone, but they do have First Amendment rights to be free from government regulation of their communications, including their decisions about which communications to host on their forums. Just as the government may not force newspapers to publish content that they decline to publish pursuant to their editorial judgment,[68] and just as government may not compel parade organizers to include marchers who convey messages that the organizers choose to exclude pursuant to their values,[69] so too government may not require dominant platforms to host any content that they choose to exclude pursuant to their content moderation standards. Likewise, Twitter mobs, and other individuals or groups exerting "cancel culture" pressure, also exercise First Amendment rights: the right to advocate for the removal of certain speech or speakers from platforms or other forums, and the freedom of association to join together with others in such advocacy.

There is one caveat to the foregoing generalization that individuals have a First Amendment right to advocate platform speech restrictions, which concerns individuals who are government officials. On the one hand, government officials do not forfeit the First Amendment rights they had as private citizens merely by virtue of being elected or appointed to office. To the contrary, government officials continue to enjoy the right to convey and advocate their views, including their views that platforms should not host certain content. Moreover, "We the People" have a vital First Amendment right to receive information and ideas that our elected and appointed officials convey, including on this important topic of free speech. On the other hand, government officials may not threaten platforms—even implicitly—that if they do not "voluntar[il]y" impose certain speech restrictions, the platforms will be subject to government regulations toward that end.[70]

---

68. *See* Mia. Herald Publ'g Co. v. Tornillo, 418 U.S. 241, 258 (1974).

69. *See* Hurley v. Irish-Am. Gay, Lesbian & Bisexual Grp., 515 U.S. 557, 566 (1995) (holding that the First Amendment invalidated a state law that conditioned a parade permit on the parade organizer's inclusion of marchers who expressed a message contrary to the organizer's beliefs).

70. *See, e.g.*, Bantam Books v. Sullivan, 372 U.S. 58, 68 (1963) (finding that notices that a state agency sent to booksellers, urging booksellers to suppress publications deemed "objectionable," were "reasonably understood [as orders]," and functioned as such).

In sum, when government officials advocate platform restrictions on speech, "We the People" must ensure that their advocacy does not cross the line from permissible persuasion to impermissible compulsion, which would "abridge the freedom of speech" under the First Amendment. The Supreme Court sensibly has held that determining whether particular government conduct rises to the level of "abridgment" requires a fact-specific, functional assessment. In the current context, that assessment turns on this question: whether, as a practical matter, government expression urging platforms to restrict certain user speech exerts sufficient pressure on platforms that they are, in effect, coerced into restricting speech. In particular cases, the line between permissible persuasion and impermissible coercion may be hard to draw.[71]

There is plausible evidence that many of the dominant platforms' content restrictions have in fact been adopted, at least in part, to fend off threatened government regulation.[72] Therefore, one proposal warranting analysis—which is discussed further below—is that any restrictions adopted for this reason should be subject to First Amendment limits, on either or both of two overlapping theories that the Supreme Court has recognized in factually analogous cases. The first theory asserts that government pressure on the dominant platforms constitutes a governmental "abridgment" of speech. The second theory asserts that the government's encouraging dominant platforms to adopt and enforce certain restrictions satisfies the "entanglement" exception to the state action doctrine.

Apart from potentially unconstitutional government pressure to restrict users' speech, the dominant platforms are also subject to abundant pressure to do so from private sector individuals and organizations, including many users, advertisers, and platform employees. These private sector actors are wholly within their First Amendment rights to exert whatever influence their speech may have, including speech that threatens to withdraw their business and services. Therefore, even without government threats of regulation, the dominant platforms would have substantial reasons to restrict content as a matter of prudent business judgment. That said, government pressure on dominant platforms to restrict certain speech could still plausibly violate the First Amendment, even if the platforms had other reasons for doing so. For example, the government might more overtly violate the First Amendment by explicitly ordering a speaker to cease engaging in constitutionally protected expression. In that case, the government

71. CHEMERINSKY, *supra* note 11, at 1072–73.

72. Jed Rubenfeld, *The Right Way to Regulate Big Tech*, NAT'L REV. (Aug. 26, 2019, 6:30 AM), https://www.nationalreview.com/2019/08/big-tech-regulation-right-way/ [https://perma.cc/SJG6-JJ6D] ("[I]n policing what people say online, the Big Tech platforms are also acting under pressure from, and with significant encouragement by, Congress.").

should not be absolved from punishment for such censorial measures merely because that speaker made an independent decision to halt her own expression for other reasons. Such government action can have chilling effects beyond the situation at issue.

V. EXPLORING POTENTIAL MEASURES TO SECURE FREE SPEECH ONLINE

The current online situation presents the worst of both worlds in terms of maintaining the free speech ideal of user agency: The dominant platforms exercise unprecedented censorial power, yet they have a constitutional right to exercise that power free from any constitutional limits that would enhance users' expressive opportunities. Hence, we must strive to identify the steps that could promote user agency while also protecting the dominant platforms' free speech rights.

Myriad proposals have been advanced by government officials, civil society organizations, scholars, and others. This Essay outlines nine of those proposals that have received substantial attention and that warrant examination. I stress that these proposals are worthy of *investigation*, but not necessarily *implementation*. Even if a particular strategy appears promising in its broad outlines, as the old saying puts it, "the devil is in the details." Therefore, until the details have been explored in depth and—as Daphne Keller put it in an epigram to this Essay—examined by plumbers, as well as hydrologists, I will continue to regard these as strategies meriting serious analysis, but not conclusive endorsement.

I will lay out a general overview of the proposed strategies before discussing each in a bit more detail. First, the threshold issue noted above should be explored to ascertain the extent to which the dominant platforms' speech restrictions should be viewed as government action subject to First Amendment restraints on the rationale that government officials sufficiently induced the adoption of those restrictions. Under the "entanglement" exception to the state action doctrine, government may not circumvent its First Amendment obligations by pressuring, or otherwise coopting, private sector actors to carry out its censorial agenda. To the extent that the government has inappropriately pressured the private sector to impose speech restrictions that the government itself may not, those restrictions would violate the First Amendment. Moreover, any platform-imposed speech restrictions that were sufficiently induced by the government would also have to comply with other constitutional norms, including due process and equal protection principles. For example, users whose speech was subject to restriction would be entitled to notice from the platform and an opportunity to contest its threatened enforcement.

In addition to the entanglement exception, the Supreme Court has laid out another exception to the state action doctrine, which must also be considered in the platform content moderation context: the public function exception. Although there is some recent scholarly support for the conclusion that this exception does warrant subjecting the dominant platforms' content moderation practices to constitutional limits, recent Supreme Court precedent weighs against that conclusion.

If the speech restrictions imposed by the platforms should not be viewed as government action subject to First Amendment restraints, then alternative measures should be explored to promote the user agency ideal while also respecting the platforms' First Amendment rights to engage in content moderation practices.

A. Are the Dominant Platforms' Speech Restrictions Fairly Viewed as Government Action, and Hence Subject to First Amendment (and Other Constitutional) Limitations?

Under the state action doctrine, the Constitution's guarantees of free speech, due process, equal protection, and other fundamental rights bind only government—or "state"—actors, not private sector actors.[73] However, the Supreme Court has recognized two commonsense exceptions to this doctrine. First, the "entanglement" exception imposes constitutional limits on private sector actions or policies in which the government has played a substantial role—for example, by coercing, pressuring, or collaborating with the private sector actor.[74] In such a situation, both the private sector actor and the government may be held accountable for their joint efforts. This exception is necessary to prevent the government from evading its constitutional obligations by deputizing private entities to do its "dirty work." As the Court has stated: "[A] State normally can be held responsible for a private decision [under the entanglement exception] only when it has exercised coercive power or has provided such significant encouragement, either overt or covert, that the choice must in law be deemed to be that of the State."[75] The second exception to the state action doctrine, the "public function" exception, imposes constitutional limits on a private sector actor when that actor "exercises a function 'traditionally exclusively reserved to the State.'"[76] This exception is necessary to protect

73. CHEMERINSKY, *supra* note 11, at 533.

74. *Id.* at 563.

75. Blum v. Yaretsky, 457 U.S. 991, 1004 (1982).

76. Manhattan Cmty. Access Corp v. Halleck, 139 S. Ct. 1921, 1926 (2019) (quoting Jackson v. Metro. Edison Co., 419 U.S. 345, 352 (1974)); *Jackson*, 419 U.S. at 352–53 (recognizing precedent in which the Court held that "state action [is] present in the exercise by a private entity of powers

individuals' constitutional rights from abuse by private sector actors that, in effect, stand in the government's shoes.

Many lawsuits have challenged the speech restrictions that have been increasingly enforced by the dominant platforms on First Amendment grounds, arguing that an exception to the state action doctrine should apply.[77] To the best of my knowledge, this threshold issue has been fatal to every one of these lawsuits that has finally been adjudicated; no court has yet been persuaded that either exception applies.[78] However, because these exceptions are fact-specific, courts should reexamine the applicability of both as the factually relevant situations continue to evolve. Moreover, as noted below, the discussion of the legal issues recently has intensified, with legal scholars, judges, and other experts providing more in-depth analyses of potential avenues for constraining the dominant platforms' content moderation decisions.

1. Should the Entanglement Exception to the State Action Doctrine Apply to the Dominant Platforms' Content Moderation Policies?

To date, the lower courts have not held that the entanglement exception applies to the dominant platforms' content moderation policies. Indeed, until 2021, litigants had generally not asserted this argument,[79] relying instead on the public function exception. Accordingly, neither the Supreme Court nor other courts have had occasion to address this issue. However, respected legal scholars have recently proffered several theories as to why speech restrictions imposed by the dominant platforms should be viewed as satisfying the entanglement exception, citing several Supreme Court precedents as analogous. These writings have been followed by multiple lawsuits invoking the entanglement exception. As one newsletter noted in June 2021, "[l]awsuits fighting back against Big Tech censorship at the behest of public officials are mounting."[80] The most widely discussed are Donald

---

traditionally exclusively reserved to the State," but concluding that "the supplying of utility service is not traditionally the exclusive prerogative of the State [of Pennsylvania]").

77. *See* Adi Robertson, *Social Media Bias Lawsuits Keep Failing in Court*, THE VERGE (May 27, 2020, 5:43 PM), https://www.theverge.com/2020/5/27/21272066/social-media-bias-laura-loomer-larry-klayman-twitter-google-facebook-loss [https://perma.cc/JD4P-ZTMC].

78. *Id.*; *see also* Brief for Elec. Frontier Found. as Amicus Curiae Supporting Appellees, Prager Univ. v. Google LLC, 951 F.3d 991 (9th Cir. 2020) (No. 18-15712) [hereinafter Amicus Brief for EFF].

79. The only exception to this generalization that my Research Assistant Heidi Moore (NYLS '21) had discovered prior to March 2021 is *Divino Grp. LLC v. Google LLC*, No. 19-CV-04749-VKD, 2021 WL 51715, at *5–6 (N.D. Cal. Jan. 6, 2021). The judge rejected the plaintiffs' argument that YouTube's alleged discrimination against their videos should be subject to First Amendment restraints in light of the relationship between the government and YouTube's parent company, Google. *Id.*

80. Tom Parker, *Lawsuit Against Twitter Reveals How It Works with Democrats to Censor*, RECLAIM THE NET (June 18, 2021, 2:37 PM), https://reclaimthenet.org/twitter-california-democrats-sued-cenorship-election-conversations/ [https://perma.cc/8CP9-NV4X].

Trump's class action lawsuits against Facebook, Google, and Twitter, which he filed on July 7, 2021, on behalf of not only himself, but also "[a]ll . . . platform Users . . . who have resided in the United States between June 1, 2018, and today, and had their Facebook account censored by Defendants and were damaged thereby."[81] On July 27, 2021, Trump filed an amended complaint, adding new class representatives, including Dr. Naomi Wolf, "who in the past advised campaigns of two former Democratic presidents, and was banned by Twitter for expressing her views around Covid."[82]

In a January 2021 *Wall Street Journal* op-ed, Vivek Ramaswamy and Jed Rubenfeld rejected the conventional wisdom that the dominant platforms' content moderation policies are protected by the First Amendment, not constrained by it. In their words: "Using a combination of statutory inducements and regulatory threats, Congress has co-opted Silicon Valley to do through the back door what government cannot directly accomplish under the Constitution."[83] The statutory inducement to which they refer is Section 230, which—as noted above—largely immunizes platforms from liability if they restrict speech that the government would be constitutionally barred from restricting.[84] The regulatory threats to which the op-ed refers have been issued by members of Congress, as well as the Executive branch, to pressure the dominant platforms to "voluntarily" restrict certain speech or else face direct regulations that would force them to do so. As Ramaswamy and Rubenfeld observe, the dominant platforms have implemented major new speech restrictions in response to such pressures. In an analogous case, the Supreme Court held that a private bookseller's "voluntary" decision to stop selling certain works was subject to—and violated—the First Amendment because the government had issued veiled threats to prosecute the bookseller if it did not do so.[85] As Ramaswamy and Rubenfeld

81. Complaint at 5, Trump v. Facebook (No. 1:21-cv-22440), available at https://docs.reclaimthenet.org/trump-july-7-lawsuit-facebook.pdf [https://perma.cc/KDQ6-9GW3].

82. Didi Rankovic, *Trump Big Tech Censorship Lawsuit Amended to Include 65,000 Censorship Stories from American Citizens*, RECLAIM THE NET (Aug. 4, 2021, 9:27 AM), https://reclaimthenet.org/trump-big-tech-censorship-lawsuit-amended-to-include-65000-censorship-stories-from-american-citizens/ [https://perma.cc/63TG-LELC].

83. Vivek Ramaswamy & Jed Rubenfeld, *Save the Constitution from Big Tech*, WSJ (Jan. 11, 2021, 12:45 PM), https://www.wsj.com/articles/save-the-constitution-from-big-tech-11610387105 [https://perma.cc/R3PA-KGQ6]. Ramaswamy is founder and CEO of Roivant Sciences. Rubenfeld is a professor at Yale Law School.

84. This aspect of the argument receives some support from UCLA Law Professor Eugene Volokh. *See* Eugene Volokh, *Might Federal Preemption of Speech-Protective State Laws Violate the First Amendment?*, REASON (Jan. 23, 2021, 7:02 PM), https://reason.com/volokh/2021/01/23/might-federal-preemption-of-speech-protective-state-laws-violate-the-first-amendment/ [https://perma.cc/5VE3-ZEQC].

85. Bantam Books v. Sullivan, 372 U.S. 58, 68 (1963).

concluded, "Either Section 230 or congressional pressure alone might be sufficient to create state action. The combination surely is."[86]

In an April 2021 concurring opinion, Supreme Court Justice Clarence Thomas, in dicta, indicated at least potential support for the government pressure argument, stating: "The government cannot accomplish through threats of adverse government action what the Constitution prohibits it from doing directly . . . . Under this doctrine, plaintiffs might have colorable claims against a digital platform if it took adverse action against them [e.g., removing their posts or deplatforming them] in response to government threats."[87]

2. Should the Public Function Exception to the State Action Doctrine Apply to the Dominant Platforms' Content Moderation Policies?

It would be especially challenging to find factual circumstances involving platform content moderation policies that would fit within the narrowly construed public function exception, which applies only when the government traditionally has been the exclusive body through which the function in question has been carried out. For example, the Court's factually analogous 2019 decision in *Manhattan Community Access v. Halleck*[88] concluded that this demanding public function standard was not satisfied. The Court reasoned that providing a platform for communications—in that case, via a cable company—was not a function that government traditionally has exclusively carried out. Rather, communications platforms have traditionally been provided by private sector actors, ranging from community, religious, and educational institutions, to individuals and corporations operating commercial media outlets. For this reason, *Halleck* rejected the argument that a public access cable TV station should be deemed subject to the First Amendment under the public function exception. In language that would apply to the dominant platforms, the Court explained:

> [O]peration of public access channels on a cable system is not a traditional, exclusive public function. Moreover, a private entity . . . who opens its property for speech by others is not transformed by that fact alone into a state actor. In operating the public access channels . . . a

86. Ramaswamy & Rubenfeld, *supra* note 83. If the dominant platforms' speech restrictions are fairly viewed as government action, then Section 230 would be unconstitutional to the extent that it immunizes restrictions that government could not itself impose. *See* Amicus Brief for EFF, *supra* note 78, at 8. In short, a finding that the dominant platforms' speech restrictions constitute government action would mean that those restrictions are barred by the First Amendment and not saved by Section 230. *Id.*

87. Biden v. Knight First Amend. Inst. at Colum. Univ., 141 S. Ct. 1220, 1226 (2021) (Thomas, J., concurring) (internal citations omitted).

88. Manhattan Cmty. Access Corp. v. Halleck, 139 S. Ct. 1921, 1926 (2019).

> private actor . . . is not subject to First Amendment constraints on its editorial discretion.[89]

The *Halleck* dissenters argued that the majority's public function exception is unduly narrow, but the fact remains that this narrow exception is entrenched in Supreme Court precedent. As the *Halleck* majority commented:

> Under the Court's cases, a private entity may qualify as a state actor when it exercises "powers traditionally exclusively reserved to the State." It is not enough that the federal, state, or local government exercised the function in the past, or still does. And it is not enough that the function serves the public good or the public interest in some way. Rather, to qualify as a traditional, exclusive public function within the meaning of our state-action precedents, the government must have traditionally *and* exclusively performed the function.[90]

Although the Supreme Court has not directly considered its narrow public function exception in the context of the dominant platforms' content moderation policies, many lower courts have, unanimously ruling that this exception does not subject content moderation policies to First Amendment restraints.

In a January 2021 *Wall Street Journal* op-ed,[91] Columbia University Law Professor Philip Hamburger implicitly invoked both the public function and entanglement exceptions to the state action doctrine to conclude that the dominant platforms may restrict speech only consistent with First Amendment norms. Referring to Section 230, he suggested that "the government, in working through private companies, is abridging the freedom of speech" when those private platforms restrict speech that the First Amendment bars government itself from censoring. Stressing that this argument extends only to the dominant platforms and not to "ordinary websites that moderate commentary," he reasoned that these "massive companies . . . are akin to common carriers . . . and function as public forums." He referred to the famous 1946 case *Marsh v. Alabama* in noting that "[t]he First Amendment protects Americans even in privately owned public forums, such as company towns."[92] Hamburger also analogized Section 230 immunity to past situations involving "Southern sheriffs . . . [who] used to assure Klansmen that they would face no

89. *Id.* at 1926.

90. *Id.* at 1928–29 (internal citations omitted) (emphasis in original).

91. Philip Hamburger, *The Constitution Can Crack Section 230, Opinion*, WSJ (Jan. 29, 2021, 2:00 PM), https://www.wsj.com/articles/the-constitution-can-crack-section-230-11611946851 [https://perma.cc/MM48-LYFE].

92. *Id.*; *see also* Marsh v. Alabama, 326 U.S. 501, 509 (1946) (enforcing the public function exception to bar states from "impos[ing] criminal punishment on [individuals] for undertaking to distribute religious literature in a company town").

repercussions for suppressing the speech of civil-rights marchers," to underscore the need to ensure that "government cannot immunize powerful private parties in the hope that they will voluntarily carry out" policy that the government constitutionally cannot.

3. Would Subjecting the Dominant Platforms to First Amendment Constraints Promote User Agency or Undermine It?

Aside from whether Supreme Court precedents foreclose or permit subjecting the dominant platforms' content moderation policies to First Amendment limits, some digital rights proponents have argued that requiring all online platforms to uniformly adhere to the First Amendment would undermine the ultimate goal of facilitating user agency. For example, the Electronic Frontier Foundation ("EFF") has explained that the user agency ideal would entail a diverse array of platforms with a wide range of content moderation policies, so that users would be free to choose between platforms that restrict controversial but constitutionally protected speech, such as graphic nudity or violence, and those that do not. For users who do find such speech (or other constitutionally protected speech) personally objectionable, being exposed to it could undermine their online experience, perhaps to the point of deterring them from using these platforms at all.[93] For this reason, EFF has supported the freedom of choice that platforms currently enjoy, by virtue of their private status, to impose a range of content restrictions that government constitutionally could not, because the platforms' freedom of choice also enhances users' freedom of choice. EFF explained this stance in a recent brief in which it opposed treating even the dominant platforms as government actors:

> Although it may seem counterintuitive, on balance, Internet users' rights are best served by preserving the constitutional status quo, whereby private parties who operate private speech platforms have a First Amendment right to edit and curate their sites, and thus exclude whatever other private speakers or speech they choose. To reverse the application of the First Amendment—that is, to make online platforms no longer protected by the First Amendment but instead bound by it as if they were government entities—would undermine Internet users' interests . . . [O]nline platforms would largely be prohibited from moderating content, even though content moderation can be valuable and is supported by many Internet users when carefully implemented . . . . Moderation allows online platforms to limit content in order to create affinity or niche communities dedicated to certain subject

93. *See* Amicus Brief for EFF, *supra* note 78, at 3–6.

matters or viewpoints, or to remove hateful or harassing speech that may hinder the ability of targeted users to engage with the platform.[94]

EFF also noted a further reason why the goal of user agency would be undermined by treating platforms as government actors bound to comply with the First Amendment: "[T]he emergence of new online platforms would be inhibited by the great legal uncertainty created by the imposition of [First Amendment obligations] on private platforms."[95] Since the emergence of new platforms would increase users' options, any development that operates as a barrier to entry would have a negative impact on user agency.

To be sure, EFF and other digital rights proponents also strongly critique the dominant platforms' content moderation policies, including their arbitrary and discriminatory restrictions on much important speech. Therefore, such digital rights advocates propose alternative measures to enhance user agency, as discussed below. EFF succinctly summarized its rationale for this position while rejecting the argument that platforms (including the dominant platforms) should be deemed state actors: "[T]he answer to bad content moderation isn't to empower the government to enforce moderation practices."[96]

Recognizing that the ultimate goal of user agency would be best served by multiple platforms offering varying content moderation options among which users could choose, the question is: How could we promote that result? The remainder of this Essay explores potential answers to that question.

B. How Can We Pursue the User Agency Ideal?

Below, I briefly summarize and comment on nine proposals, each of which has received substantial attention and merits investigation, and which fall into four categories: (1) *technological* features that would facilitate user agency; (2) *procedural* constraints on how the dominant platforms enforce their content moderation policies; (3) *substantive* limits on the dominant platforms' content moderation policies; and (4) *structural* alterations of the dominant platforms' size and operations.

Before laying out these proposals, I repeat the disclaimer noted above: At this point, *I endorse only further analysis of these proposals, and do not endorse any proposal on the merits.* All of these proposals are quite general

94. *Id.* at 6 (emphasis omitted).

95. *Id.* at 2.

96. David Greene, *EFF to Court: Remedy for Bad Content Moderation Isn't to Give Government More Power to Control Speech*, ELEC. FRONTIER FOUND. (Nov. 26, 2018), https://www.eff.org/deeplinks/2018/11/eff-court-remedy-bad-content-moderation-isnt-give-government-more-power-control [https://perma.cc/WD6X-GPKF].

in nature, and hence any conclusive endorsement should follow only from exhaustive examination, including analysis of implementing details and potential unintended adverse consequences. Even concerning the proposals that strike me as most promising, I withhold my endorsement, because their precise impact would depend on the specific implementing details. The prudence of such a cautious approach was underscored by Daphne Keller, Director of Stanford University's Program on Platform Regulation, in a column published shortly before this Essay went to press. Notably, Keller stressed that even a general proposal that is widely endorsed—such as increased transparency about the dominant platforms' content moderation policies—nonetheless presents complicated issues once one focuses on implementing details:

> I am a huge fan of transparency about platform content moderation. I've considered it a top policy priority for years, and written about it in detail . . . I sincerely believe that without it, we are unlikely to correctly diagnose current problems or arrive at wise legal solutions.
>
> So it pains me to admit that I don't really know what "transparency" I'm asking for. I don't think many other people do, either. . . .
>
> . . . .
>
> Researchers and civil society should assume we are operating with a limited transparency "budget," which we must spend wisely—asking for the information we can best put to use, and factoring in the cost. We need better understanding of both research needs and platform capabilities to do this cost-benefit analysis well.[97]

1. Technological Features That Would Facilitate User Agency

i. Should the Dominant Platforms Take Technological Steps that Would Facilitate User Agency?

The first proposal allows users to choose their own content moderation policies. Specifically, user agency proponents urge that the dominant platforms should be required to implement specific technological features that would provide more options to users, and thus increase user choice. For example, specifically focusing on the dominant platforms, EFF wrote, "[u]sers of social media platforms with significant market power should be empowered to choose content they want to interact with [and to filter out that which they do not] in a simple and user-friendly manner."[98] EFF

97. Daphne Keller, *Some Humility About Transparency*, CTR. FOR INTERNET & SOC'Y AT STAN. LAW SCH. (Mar. 19, 2021), http://cyberlaw.stanford.edu/blog/2021/03/some-humility-about-transparency [https://perma.cc/XST9-JP8W].

98. Christoph Schmon, *EFF Responds to EU Commission on the Digital Services Act: Put Users Back in Control*, ELEC. FRONTIER FOUND. (Sept. 4, 2020), https://kittens.eff.org/es/deeplinks/2020/09/e

elaborated, "platforms should open up their APIs [Application Programming Interfaces] to allow users to create their own filtering rules for their own algorithms. Users should also have the option to decide against algorithmically-curated recommendations altogether, or to choose other heuristics to order content."[99]

Specific technological features that would facilitate end-user content control include "interoperability"[100] and "delegability."[101] These options would permit third parties to install content moderation and curation programs that differ from the dominant platforms' current practices, so that platform users could delegate the content moderation function to third parties whose approaches better align with a particular user's preferences. As EFF has explained, the dominant platforms

> must offer possibilities for competing, not-incumbent platforms to interoperate with their key features . . . . Interoperability should also happen at the level of user interfaces. . . mak[ing] it possible for competing third parties to act on users' behalf . . . For example, if you don't like Facebook's content moderation practices, you should be able to delegate that task to another organization.[102]

In the fall of 2020, the House Judiciary Committee staff issued an antitrust report about Big Tech, which also recommended requiring interoperability as "an important complement . . . to vigorous antitrust enforcement."[103]

In a 2019 essay commissioned by the Knight First Amendment Institute at Columbia University, for "[a]n essay series reimagining the First Amendment in the digital age," Mike Masnick elaborated on the same general concepts, which he described as reinstituting the early Internet's

---

ff-responds-eu-commission-digital-services-act-put-users-back-control [https://perma.cc/PTS4-77RA].

99. Corynne McSherry, *Content Moderation and the U.S. Election: What to Ask, What to Demand*, ELEC. FRONTIER FOUND. (Oct. 26, 2021), https://www.eff.org/deeplinks/2020/10/content-moderation-and-us-election-what-ask-what-demand [https://perma.cc/D37U-CAHT]; *see also* ELI NOAM, THE CONTENT, IMPACT, AND REGULATION OF STREAMING VIDEO: THE NEXT GENERATION OF MEDIA EMERGES 388–89 (2021) (pointing out additional benefits of granting alternative information curators API-based access into dominant platforms, noting that this approach could provide not only content screening, but also privacy protection, algorithms, and recommendation engines, which would generate user choice and reduce the dominant platforms' market power).

100. "Interoperability" is "the ability to make a new product or service work with an existing product or service." Svea Windwehr & Christoph Schmon, *Our EU Policy Principles: Interoperability*, ELEC. FRONTIER FOUND. (June 18, 2020), https://www.eff.org/deeplinks/2020/06/our-eu-policy-principles-interoperability [https://perma.cc/F4CZ-8P69].

101. The notion of "delegability" puts users in control, proposing that, "[users] should be able to delegate elements of their online experience to different [competing third parties]. For example, if [users do not] like Facebook content moderation practices, [they] should be able to delegate that task to another organization." *Id.*

102. *Id.*

103. STAFF OF S. COMM. ON THE JUDICIARY, 116TH CONG., REP. ON ANTITRUST, COM. AND ADMIN. LAW 386 (Comm. Print 2020), https://judiciary.house.gov/uploadedfiles/competition_in_digital_markets.pdf [https://perma.cc/FU7T-UB6G].

"protocol-based" structure, in contrast with the current "platform-based" structure. He explained:

> Rather than relying on a single centralized platform, . . . anyone would be able to create their own set of rules—including which content do they not want to see and which content they would like to see promoted. Since most people would not wish to manually control all of their own preferences and levels, this could easily fall on any number of third parties—whether they be competing platforms, public interest organizations, or local communities. Those third parties could create whatever interfaces, with whatever rules, they wanted . . . . [M]any, many different individuals and organizations would be able to tweak the system to their own levels of comfort and share them with others—and allow the competition to happen at the implementation layer, rather than at the underlying social network level.[104]

As Masnick underscores, these technological approaches aim to restore the utopian user agency vision that was contemplated by *Reno v. ACLU* and Section 230, as discussed above:

> Moving us back toward a world where protocols are dominant over platforms could be of tremendous benefit to free speech and innovation online.
>
> Such a move has the potential to return us to the early promise of the web: to create a place where like-minded people can connect on various topics around the globe and anyone can discover useful information on a variety of different subjects without it being polluted by abuse and disinformation. Simultaneously, it could enable greater competition and innovation on the internet, while also giving end users more control over their own data and preventing giant corporations from having too much data on any particular user.[105]

2. Procedural Constraints on How the Dominant Platforms Enforce Their Content Moderation Policies

i. Should the Dominant Platforms Comply with the 2015 "Manila Principles on Intermediary Liability"?

The second proposal stems from the Manila Principles on Intermediary Liability: Best Practices Guidelines for Limiting Intermediary Liability for Content to Promote Freedom of Expression and Innovation, A Global Civil Society Initiative (the "Manila Principles" or "Principles").[106] The Manila Principles, which were adopted in 2015, have been endorsed by many civil

104. Masnick, *supra* note 35.
105. *Id.*
106. *Manila Principles on Intermediary Liability*, ELEC. FRONTIER FOUND. (Mar. 24, 2015), https://www.eff.org/files/2015/10/31/manila_principles_1.0.pdf [https://perma.cc/W8LY-JM95].

society organizations and individual experts from around the world. They reflect fundamental procedural concerns of transparency, accountability, and due process, consistent with international human rights law. Specifically, the Manila Principles lay out "baseline safeguards" to promote user agency and to protect users from censorship and other human rights abuses by governments and private "intermediaries," including the dominant platforms. Each Manila Principle is listed below, followed by my brief comments on several of them.

There are six major Principles, each of which includes subprinciples. The major Principles in some cases apply only to governments, in other cases only to intermediaries, including platforms, and in some cases to both governments and intermediaries. They read as follows:

> 1. Intermediaries should be shielded from liability for third-party content[.]
>
> 2. Content must not be required to be restricted without an order by a judicial authority[.]
>
> 3. Requests for restrictions of content must be clear, be unambiguous, and follow due process[.]
>
> 4. Laws and content restriction orders and practices must comply with the tests of necessity and proportionality[.]
>
> 5. Laws and content restriction policies and practices must respect due process[.]
>
> 6. Transparency and accountability must be built into laws and content restriction policies and practices.[107]

Below, I quote the most important subprinciples under Principles 4–6, which specifically apply to platforms, as intermediaries.

> [*Principle 4.*] [C]ontent restriction orders and practices must comply with the tests of necessity and proportionality . . . .
>
> a. Any restriction of content should be limited to the specific content at issue.
>
> b. When restricting content, the least restrictive technical means must be adopted.
>
> c. If content is restricted because it is unlawful in a particular geographical region, and if the intermediary offers a geographically variegated service, then the geographical scope of the content restriction must be so limited.
>
> d. If content is restricted owing to its unlawfulness for a limited duration, the restriction must not last beyond this duration, and the restriction order must be reviewed periodically to ensure it remains valid.[108]

107. *Id.*
108. *Id.*

> [*Principle 5*.] [C]ontent restriction policies and practices must respect due process. . . .
>
> c. Intermediaries should provide user content providers with mechanisms to review decisions to restrict content in violation of the intermediary's content restriction policies.
>
> d. In case a user content provider wins an appeal . . . against the restriction of content, intermediaries should reinstate the content.
>
> f. When drafting and enforcing their content restriction policies, intermediaries should respect human rights.[109]
>
> [*Principle 6*.] Transparency and accountability must be built into . . . content restriction policies and practices.
>
> c. Intermediaries should publish their content restriction policies online, in clear language and accessible formats, and keep them updated as they evolve, and notify users of changes when applicable.
>
> e. Intermediaries should publish transparency reports that provide specific information about all content restrictions taken by the intermediary, including actions taken on government requests, court orders, private complainant requests, and enforcement of content restriction policies.
>
> f. Where content has been restricted on a product or service of the intermediary that allows it to display a notice when an attempt to access that content is made, the intermediary must display a clear notice that explains what content has been restricted and the reason for doing so.[110]

While Principles 5 and 6 specify procedural standards, Principle 4 calls for platforms to adhere to a key substantive tenet of international human rights law governing free speech, which itself parallels a key tenet of U.S. First Amendment law: Any speech restriction must be "necessary" and "proportionate" to promote its purpose. In other words, any speech restriction must be the "least restrictive alternative"; if the restriction's purposes could be effectively promoted by an alternative measure, which is less speech-restrictive, the platform must use that alternative measure.[111] For example, if a platform determines that a particular user post violates its terms of service, the platform should only de-platform that user as a last resort, if the user continues to post messages that violate the platform's terms of service after the platform has imposed less speech-restrictive measures, such as labeling and removing noncompliant posts.

Also paralleling U.S. constitutional law, Principle 5 mandates due process in content moderation policies and practices, while its subprinciple (f) invokes international human rights law more generally.

---

109. *Id.*
110. *Id.*
111. *See, e.g.*, Strossen, *supra* note 67, at 343–44.

Since the initial issuance of the Manila Principles in 2015, the notion that the dominant platforms' content moderation policies should comply with international human rights law has garnered increasing support. In 2018, the United Nations ("U.N.") Special Rapporteur for free expression issued an influential report advocating this approach, including the necessity, proportionality, and least restrictive alternative requirements that the Manila Principles expressly invoke.[112] Other U.N. reports also have endorsed this approach, as have civil society organizations and individual experts.[113]

ii. Should the Dominant Platforms Comply with the 2018 "Santa Clara Principles on Transparency and Accountability in Content Moderation"?

The third proposal is embodied in the Santa Clara Principles, which were adopted by the EFF and other free speech and digital rights advocates from multiple countries in 2018. The Santa Clara Principles lay out a "floor" of fundamental procedural requirements for platforms to honor in their content moderation processes.[114] These constitute "initial steps . . . to provide meaningful due process to impacted speakers and better ensure that the enforcement of [the platforms'] content guidelines is fair, unbiased, proportional, and respectful of users' rights."[115] Most of the dominant platforms endorsed these principles in 2019, but according to EFF only one (Reddit) has actually adhered to them.[116]

There are three major Principles, each accompanied by subprinciples specifying more detailed standards for compliance. The three major Principles are as follows:

> 1. *Numbers*[.] Companies should publish the numbers of posts removed and accounts permanently or temporarily suspended due to violations of their content guidelines.

---

112. David Kaye (Special Rapporteur on the Promotion & Prot. of the Right to Freedom of Opinion & Expression), U.N. Doc. A/HRC/38/35 (2018) (discussing recommendations for substantive standards for content moderation based on The Guiding Principles on Business and Human Rights, as adopted by the United Nations ("U.N.") Human Rights Council in 2011).

113. *See, e.g.*, Strossen, *supra* note 67, at n.10–11, 355–57.

114. *See The Santa Clara Principles on Transparency and Accountability in Content Moderation*, https://santaclaraprinciples.org/ [https://perma.cc/E4NR-5364] (last accessed Sept. 29, 2021) [hereinafter *The Santa Clara Principles*].

115. *Id.*

116. *See, e.g.*, Jillian C. York, *What Comes Next for the Santa Clara Principles: 2020 in Review*, ELEC. FRONTIER FOUND. (Dec. 28, 2020), https://www.eff.org/deeplinks/2020/12/2020-year-review-what-comes-next-santa-clara-principles [https://perma.cc/U2WF-J7N5] ("In 2019, we succeeded in getting a dozen companies to endorse the Principles, with several companies furthering their compliance. One company, Reddit, went all the way in implementing the Principles into their platform.").

> 2. *Notice*[.] Companies should provide notice to each user whose content is taken down or account is suspended about the reason for the removal or suspension.
>
> 3. *Appeal*[.] Companies should provide a meaningful opportunity for timely appeal of any content removal or account suspension.[117]

A portion of the subprinciples specified under the "Notice" Principle states:

> [C]ompanies should provide detailed guidance . . . about what content is prohibited, including examples of permissible and impermissible content and the guidelines used by reviewers . . . When providing a user with notice about why her post has been removed or an account has been suspended, a minimum level of detail . . . includes:
>
> - [I]nformation sufficient to allow identification of the content removed.
>
> - The specific clause of the guidelines that the content was found to violate.
>
> - How the content was detected and removed (flagged by other users, governments, trusted flaggers, automated detection, or external legal or other complaint).
>
> - Explanation of the process through which the user can appeal the decision.[118]

The Santa Clara Principles note that they constitute a floor, not a ceiling. To better protect user agency, its endorsers plan to build upon that foundation by laying out more extensive and robust requirements for transparency and accountability in content moderation practices.[119] Nonetheless, even the relatively modest, process-oriented steps for which the Santa Clara Principles currently call would be a marked improvement over the status quo, where the dominant platforms are under no obligation to take any such steps, and in fact most of them do not.

We should also heed the old adage that "sunshine is the best disinfectant," recognizing that even seemingly small procedural steps can have a significant substantive impact. After all, as service-business entities, the dominant platforms have financial incentives to respond to pressures from their customers, their employees, the media, and politicians. Therefore, by providing more information about the terms and enforcement of their content moderation policies, the dominant platforms would enable these constituencies to provide critical feedback, which could in turn stimulate policy revisions. In fact, the dominant platforms whose content mod-

117. *The Santa Clara Principles*, *supra* note 114.

118. *Id.*

119. *See id.* ("These principles are meant to serve as a starting point, outlining minimum levels of transparency and accountability that we hope can serve as the basis for a more in-depth dialogue in the future.").

eration policies have been disclosed revised those policies and practices in response to public critiques. Platforms have restored expression and speakers that they had been criticized for restricting or removing;[120] platforms also have restricted or removed expression and speakers that they had been criticized for previously having failed to restrict or remove.[121]

In 2020, EFF sought comments about potential revisions to the Santa Clara Principles, announcing that it has been working with other civil society organizations to issue an updated version.[122]

iii. Should the Dominant Platforms Permit Public Interest Study of Their Architecture and Operations?

The fourth proposal seeks transparency in the public interest. In October 2020, Facebook sent a letter to two New York University ("NYU") researchers demanding that they discontinue use of a research tool that they and others had designed and considered crucial to understanding political advertising on social media platforms.[123] The researchers, Laura Edelson and Damon McCoy, are being represented in this matter by the Knight First Amendment Institute at Columbia University (the "Knight Institute"). To illuminate the issue whether and to what extent the dominant platforms should allow researchers access for purposes of conducting studies in the public interest, this section briefly outlines Edelson's and McCoy's situation and the Knight Institute's broader efforts on this front.

In 2018, the Knight Institute wrote to Facebook on behalf of a number of journalists and social scientists "who would like to pursue projects that are manifestly in the public interest but are barred by Facebook's terms of service."[124] The Knight Institute proposed that Facebook amend its terms

---

120. *See e.g.*, Sam Levin, Julia Carrie Wong & Luke Harding, *Facebook Backs Down from 'Napalm Girl' Censorship and Reinstates Photo*, THE GUARDIAN (Sep. 9, 2016, 1:44 PM) https://www.theguardian.com/technology/2016/sep/09/facebook-reinstates-napalm-girl-photo [https://perma.cc/VQ35-XZGH] (discussing Facebook's decision to reinstate the iconic Vietnam War photo featuring a naked girl after a "global outcry and accusations of abusing power") (internal quotations omitted)).

121. Sonja West & Genevieve Lakier, *The Court, the Constitution, and the Deplatforming of Trump*, SLATE (Jan. 13, 2021, 5:14 PM), https://slate.com/technology/2021/01/deplatforming-trump-constitution-big-tech-free-speech-first-amendment.html [https://perma.cc/4TX9-83QU] (discussing the implications of Facebook's and Twitter's arguably belated "recognition of the harmfulness of [Donald Trump's] speech [that] the platforms had previously largely tolerated" and their subsequent permanent and lifelong bans of Trump from their respective websites).

122. Press Release, Elec. Frontier Found., EFF Seeks Public Comment About Expanding and Improving Santa Clara Principles (Apr. 14, 2020), https://www.eff.org/press/releases/eff-seeks-public-comment-about-expanding-and-improving-santa-clara-principles [https://perma.cc/GEJ5-QETZ].

123. Press Release, Knight First Amend. Inst. at Columb. Univ., Researchers, Knight Institute Condemn Facebook Effort to Squelch Research on Disinformation (Oct. 23, 2020), https://knightcolumbia.org/content/researchers-knight-institute-condemn-facebook-effort-to-squelch-research-on-disinformation [https://perma.cc/7M9M-CZNR] [hereinafter *Facebook Effort to Squelch Research*].

124. Letter from Knight First Amendment Institute at Columbia University to Mark Zuckerberg at 1, (Aug. 6, 2018), https://int.nyt.com/data/documenthelper/135-knight-institute-letter-to-fac/3658609c9

of service "to create a safe harbor" that "would permit journalists and researchers to conduct public-interest investigations while protecting the privacy of Facebook's users and the integrity of Facebook's platform."[125] Recognizing that Facebook had recently announced certain new transparency initiatives, the Knight Institute labeled these efforts as "laudable, but . . . insufficient."[126] In 2019, in the midst of ongoing negotiations between the Knight Institute and Facebook about the Knight Institute's proposal, more than 200 "digital researchers who study the architecture and operations of social media platforms" signed an open letter in support of it.[127] They contended that "[t]here is extraordinary public interest in understanding the influence of social media platforms on society. Yet Facebook's prohibition on the use of basic digital tools obstructs much of the work that could be done to understand that influence."[128]

Edelson and McCoy had been using Ad Observer, "a browser plug-in they and others created that allows Facebook users to voluntarily share with the researchers information about the political ads shown to them by the platform."[129] This was the tool that Facebook demanded that the researchers (as well as Facebook users) cease to use. But Edelson, who is the lead researcher in the Ad Observatory project, reasoned that her team had "developed . . . the Ad Observer plug in to deliver an essential level of cybersecurity analysis . . . which makes clear who is trying to influence us and why."[130] Likewise, Alex Abdo, the Knight Institute's litigation director, put this particular controversy into a broader context that is of even greater concern: "It would be terrible for democracy if Facebook is allowed to be the gatekeeper to journalism and research about Facebook."[131]

On August 4, 2021, the Knight Institute issued a press release stating that "[a]fter months of negotiations, late yesterday evening, Facebook abruptly shut down the accounts of New York University researchers Laura Edelson and Damon McCoy, blocking their research into political ads and the spread of misinformation on the platform." Laura Edelson commented as follows:

---

f97a7f132d2/optimized/full.pdf [https://perma.cc/K3L7-F5U6].

125. *Id.* at 4.

126. *Id.* at 3.

127. Press Release, Knight First Amend. Inst. at Columb. Univ., More Than 200 Researchers Support Knight Institute Call to Facilitate Research of Facebook's Platform (June 12, 2019), https://knightcolumbia.org/content/more-than-200-researchers-support-knight-institute-call-to-facilitat e-research-of-facebooks-platform [https://perma.cc/G9XZ-29CM]. Letter to Call on Facebook to Lift Restrictions on Digital Research and Journalism on Its Platform to Mark Zuckerberg (June 12, 2019), https://docs.google.com/forms/d/e/1FAIpQLScLcoINv_uMedadkMvp4ZUSLmayCLQZPKQcbiV1cn GXtpy_4Q/viewform [https://perma.cc/Q2KJ-8H6K] [hereinafter Letter to Lift Restrictions].

128. Letter to Lift Restrictions, *supra* note 127.

129. *Facebook Effort to Squelch Research, supra* note 123.

130. *Id.*

131. *Id.*

> Facebook is silencing us because our work often calls attention to problems on its platform. Worst of all, Facebook is using user privacy, a core belief that we have always put first in our work, as a pretext for doing this. If this episode demonstrates anything it's that Facebook should not have veto power over who is allowed to study them.[132]

iv. Should the Dominant Platforms Be Barred from Surveillance, Content Curation, or Micro-Targeting Without User Notice and Consent?

The fifth proposal calls for federal legislation to enhance user agency. The essence of freedom of speech is the right of each individual to decide for herself or himself what information or ideas to convey or to receive. Therefore, we should seek means to limit the dominant platforms' policies or practices that undermine this freedom. Of greatest concern, these platforms engage in pervasive surveillance of our online communications and actions, which they then use to "micro-target" advertisements and other communications at us, and to rank and curate the content that we receive, in an effort to maintain our undivided attention.[133]

Ample evidence indicates that this micro-targeting not only shapes which communications we receive, and in what order, but also our responsive attitudes and behaviors. All this surveilling and targeting is done without even the user's awareness—much less consent. Indeed, many users are aware of neither the general problem, nor how it impacts them individually. As Professor Evelyn Aswad has forcefully argued, this nonconsensual manipulation of our communications profoundly violates the freedom of thought that is integrally intertwined with freedom of speech.[134] Also intertwined is individual privacy, which includes the right not to share information about our communications or other aspects of our personal lives without informed consent.

Government can and should play a vital legislative role in protecting the interrelated individual rights of privacy, free thought, and free speech from invasion by the dominant platforms. Such legislation has been enacted in other countries and in a number of U.S. states. National legislation had been pending in the U.S. with broad bipartisan support in the late 1990s and early 2000s, but it was derailed by the September 11, 2001, terrorist attacks,

---

132. Press Release, Knight First Amend. Inst. at Colum. Univ., Researchers, NYU, Knight Institute Condemn Facebook's Effort to Squelch Independent Research About Misinformation (Aug. 4, 2021), https://knightcolumbia.org/content/researchers-nyu-knight-institute-condemn-facebooks-effort-to-squelch-independent-research-about-misinformation [https://perma.cc/HZA7-NZ5E].

133. For a detailed discussion of the issues that this paragraph summarizes, *see* SHOSHANNA ZUBOFF, SURVEILLANCE CAPITALISM (2019); Shoshanna Zuboff, *Social Media and a Surveillance Society*, N.Y. TIMES (Feb. 6, 2021), https://www.nytimes.com/2021/02/06/opinion/letters/social-media-surveillance.html [https://perma.cc/2XXV-WRX7].

134. Evelyn Aswad, *Losing the Freedom to Be Human*, 52 COLUM. HUM. RTS. L. REV. 306 (2020).

after which privacy concerns were eclipsed by arguments for increased surveillance to deflect potential terrorist threats. In her powerful book describing the "surveillance capitalism" that is at the heart of the dominant platforms' business models, Harvard Business School Professor Emerita Shoshanna Zuboff stresses this serendipity of timing that facilitated the platforms' nonconsensual, pervasive, granular surveillance, and the resulting micro-targeting. As Zuboff explains, the dominant platforms were formed and grew in the wake of the post-9/11 lapse in concern for, and government efforts to protect, consumer privacy.

In addition to consumer-protective legislation, other legislation to curtail micro-targeting has been introduced in Congress to completely ban the micro-targeting of political ads.[135] Federal Election Commissioner Ellen Weintraub has also advocated such a ban.[136] Moreover, some user agency proponents have advocated a total ban of all content curation by the dominant platforms, so that users would receive communications in chronological order or some other neutral ranking.[137]

3. Substantive Limits on the Dominant Platforms' Content Moderation Policies

i. Should the Dominant Platforms Align Their Content Moderation Policies with International Human Rights Free Speech Norms Set Out in U.N. Treaties?

The sixth proposal seeks to align content moderation policies with the international human rights free speech norms set out in United Nations treaties. Support for this proposal has increased since its 2018 endorsement in a report by University of California, Irvine Law Professor David Kaye, in his then-capacity as the U.N. Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression.[138] Noting their enormous power over communications worldwide, Kaye urged the dominant platforms to comply with the U.N.'s Guiding Principles on Business and Human Rights[139] and, in particular, to align their content

135. Scott Nover, *Democratic Bill Seeks to Ban Microtargeting in Political Ads*, AD WEEK (May 26, 2020), https://www.adweek.com/performance-marketing/democratic-bill-seeks-to-ban-microtargeting-in-political-ads/ [https://perma.cc/WVA5-BN2R].

136. Ellen L. Weintraub, *Don't Abolish Political Ads on Social Media. Stop Microtargeting.*, WASH. POST (Nov. 1, 2019), https://www.washingtonpost.com/opinions/2019/11/01/dont-abolish-political-ads-social-media-stop-microtargeting/ [https://perma.cc/FAB8-GQB7].

137. *An Attempt at a Fundamental Rights Based Proposal*, PLATFORM REGUL., https://platformregulation.eu/ [https://perma.cc/XTL4-24S6] (last visited Sept. 29, 2021).

138. Kaye, *supra* note 112.

139. *See id.*; U.N. Human Rights Office of the High Commissioner, Guiding Principles on Business & Human Rights (2011), https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_

moderation policies with the U.N.'s international human rights standards governing free speech. These international standards are laid out in Articles 19 and 20 of the International Covenant on Civil and Political Rights ("ICCPR") ("ICCPR art. 19" and "ICCPR art. 20"),[140] to which 173 of the 195 U.N. member nations are parties. In addition, 188 nations are parties to the International Convention on the Elimination of All Forms of Racial Discrimination ("ICERD"),[141] Article 4 of which ("ICERD art. 4") concerns hate speech in particular.

Kaye's 2018 report persuasively explained that these international standards constitute appropriate guidelines for the dominant platforms, whose operations are global in scope—more so than the law of any one country or the law of regional human rights treaties and their enforcing bodies, such as the European Convention on Human Rights and the European Court of Human Rights. The dominant platforms should find it is easier to abide by a single set of norms rather than multiple sets. It is also easier for the dominant platforms to resist speech-suppressive pressures from authoritarian (and other) governments when the platforms can assert that the desired restriction would violate international norms to which most countries have acceded (likely including the countries doing the pressuring).

Kaye's 2018 report cited authoritative interpretations of the pertinent treaty provisions by U.N. officials and bodies empowered to interpret these provisions and to monitor their implementation.[142] These interpretations have strongly protected speech and strictly curtailed permissible restrictions. For example, even though ICCPR art. 20 permits governments to restrict certain hate speech, and even though ICERD art. 4 requires governments to do so, U.N. officials and bodies have construed both provisions as subject to the general speech-protective requirements in ICCPR art. 19: Any speech restriction must be narrowly formulated to promote a specific important public purpose, it must be necessary, and it must be the least speech-restrictive means to promote that purpose. These requirements closely parallel key tenets of First Amendment law, which require that speech restrictions targeting particular content or views (such as hate speech) must be narrowly tailored, necessary, and the least speech-restrictive alternative to promote a goal of compelling importance.

---

en.pdf [https://perma.cc/F9WV-HH4S].

140. International Covenant on Civil and Political Rights arts. 19–20, Dec. 16, 1966, 999 U.N.T.S. 93, https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx [https://perma.cc/GF37-3FJU].

141. International Covenant on the Elimination of All Forms of Racial Discrimination art. 4, Dec. 21, 1965, 660 U.N.T.S. 195, https://www.ohchr.org/Documents/ProfessionalInterest/cerd.pdf [https://perma.cc/W5G4-WTKP].

142. The points that this paragraph summarizes are discussed in depth in Strossen, *supra* note 67.

Kaye issued a subsequent report in 2019, specifically urging the dominant platforms to adhere to the foregoing U.N. free speech norms when enforcing content moderation policies concerning hate speech in particular.[143] Other U.N. officials and bodies have likewise endorsed the conclusion that the dominant platforms' content moderation policies should comply with U.N. free speech norms. One important example is the U.N.'s most recent, most detailed publication on point: the 2020 "United Nations Strategy and Plan of Action on Hate Speech, Detailed Guidance on Implementation for United Nations Field Presences," which was the product of sixteen U.N. entities.[144] A growing number of civil society organizations, as well as individual scholars and activists, have endorsed this international human rights law approach and begun to flesh out the details for implementing it.[145]

ii. Should Other Dominant Platforms Emulate Facebook's Oversight Board, Which Has Been Charged with Reviewing Certain Facebook Speech Restrictions to Ensure Compliance with International Human Rights Law?

The seventh proposal concerns self-regulation. Ever since Facebook announced its plans to create an Oversight Board (the "Board"), with authority to review particular content moderation decisions to which affected users object, it has repeatedly declared that the Board would be guided by international human rights principles. In March 2021, Facebook issued a new corporate policy that elaborated upon its commitment to respect human rights.[146] This commitment was reflected in the Board's governing documents, and all of its decisions to date have been based on ICCPR 19, as well as speech-protective reports issued by U.N. officials and bodies with authority to interpret and monitor compliance with ICCPR 19, including the two Special Rapporteur reports discussed in the preceding subsection. In the closely-watched decision about then-President Trump's deplatforming, the Oversight Board notably stressed the speech-protective nature of Article 19 standards, as substantially overlapping with speech-protective First

143. David Kaye (Special Rapporteur on the Promotion & Prot. of the Right to Freedom of Opinion & Expression), U.N. Doc. A/HRC/74/486 (2019) https://undocs.org/A/74/486 [https://perma.cc/E38D-RSJK].

144. United Nations Strategy and Plan of Action on Hate Speech (Sept. 2020), https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20PoA%20on%20Hate%20Speech_Guidance%20on%20Addressing%20in%20field.pdf [https://perma.cc/GX7E-PJUY].

145. *See* Strossen, *supra* note 67, at n.10–11, 20.

146. *Corporate Human Rights Policy*, FACEBOOK, https://about.fb.com/wp-content/uploads/2021/04/Facebooks-Corporate-Human-Rights-Policy.pdf [https://perma.cc/DV44-AD2G] (last visited Sept. 29, 2021); *Facebook Oversight Board Governance*, FACEBOOK OVERSIGHT BD., https://oversightboard.com/governance/ [https://perma.cc/JG7Y-6H7L] (last visited Sept. 29, 2021).

Amendment standards, stating: "[I]n many relevant respects the principles of freedom of expression reflected in the First Amendment are similar or analogous to the principles of freedom of expression in ICCPR Article 19."[147]

Between January and mid-October 2021 (when this piece went to press), the Oversight Board ruled substantively on seventeen appeals: sixteen involving Facebook's speech restrictions, and one involving Facebook's decision not to restrict the speech at issue. In the sixteen speech-restricting cases, the Oversight Board completely upheld Facebook's speech restriction in only three of them, it overturned Facebook's speech restriction in twelve of them, and in the remaining case—Donald Trump's deplatforming—the Board held that Facebook was correct in restricting Trump's access, but it also overturned Facebook's "indeterminate and standardless penalty of indefinite suspension." In the one case involving a challenge to Facebook's refusal to restrict, the Oversight Board upheld that decision. In sum, of the seventeen cases, the Oversight Board wholly affirmed the speech-protective approach in a full thirteen; it partially affirmed a speech-protective approach in one; and it wholly rejected the speech-protective approach in only three.[148]

The seventeen initial cases that the Board has decided, as of mid-October 2021, obviously constitute an infinitesimal fraction of instances in which Facebook has imposed speech restrictions; indeed, they constitute a miniscule fraction of the cases in which Board review has been sought. But these cases could have a disproportionately significant impact for at least two reasons. First, Facebook could reasonably be expected to update its content moderation policies and training materials to incorporate the Board's analyses, principles, and conclusions. Just as a single Supreme Court decision affects countless subsequent decisions by the high Court itself, as well as other courts around the U.S. (and, in fact, the world), Oversight Board rulings might also have a widespread ripple effect.

Second, in addition to resolving the particular challenges to the restricted expression, the Board's decisions also contained recommendations for revising the Facebook policies on point, including recommendations to rewrite certain policies to provide clearer guidance and more narrowly tailored standards. Although Facebook is not required to follow such recommendations, substantial business considerations weigh in favor of it doing so.

---

147. *Case Decision 2021-001-FB-FBR*, FACEBOOK OVERSIGHT BD. (May 5, 2021), https://www.oversightboard.com/sr/decision/2021/001/pdf-english [https://perma.cc/3HM8-ACQC].

148. Facebook Oversight Board Decisions, https://oversightboard.com/decision/ [https://perma.cc/7SUY-EZJL?type=image] (last visited Oct. 16, 2021).

To be sure, it is far too soon to tell whether Facebook's Oversight Board will prove to be a positive model for promoting user agency, but it is not too soon to monitor its actions and their impact with this possibility in mind.

4. Structural Alterations of the Dominant Platforms' Size and Operations

i. Should the Structure and Operations of the Dominant Platforms Be Held to Applicable Antitrust and Similar Norms to Ensure Fair Competition?

The eighth proposal would subject the dominant platforms to legal proceedings to ensure fair competition. Numerous antitrust and similar proceedings against the dominant platforms are ongoing, or being planned, including the following: lawsuits by state attorneys general, an investigation by the Federal Trade Commission, and an investigation by the U.S. Department of Justice Antitrust Division.[149] If any of these proceedings lower the entry barrier that new competitors must surmount, or reduce the dominant platforms' market power, users could find themselves with more options. This should lead to a broader array of content moderation policies among which users could choose, and thus enhance user agency.

ii. Should the Dominant Platforms Be Treated as Regulated Public Utilities or Common Carriers?

The ninth and final proposal that this Essay explores looks to the common law. A number of ideologically diverse experts have contended that the dominant platforms constitute essential infrastructure, analogous to landline telephone companies a century ago, and hence should be treated as regulated public utilities or common carriers. In January 2021, this approach received a surprising endorsement from New York University Law Professor Richard Epstein,[150] a prominent libertarian law and economics expert. Epstein invoked the common law concept of common carriers, which obligates companies to treat all users reasonably, fairly, and non-

149. Shannon Bond & Bobby Allyn, *48 AGs, FTC Sue Facebook, Alleging Illegal Power Grabs to 'Neutralize' Rivals,* NPR (Dec. 9, 2020, 2:39 PM), https://www.npr.org/2020/12/09/944073889/48-attorneys-general-sue-facebook-alleging-illegal-power-grabs-to-neutralize-riv#:~:text=Ethics-,48%20Attorneys%20General%2C%20FTC%20Sue%20Facebook%2C%20Alleging%20Illegal%20Power%20Grabs,social%20network%20an%20unfair%20advantage [https://perma.cc/Y4UE-5AKV]; Ben Brody, *The FTC's Antitrust Case Against Facebook Stakes Out New Ground,* BLOOMBERG (Dec. 16, 2020, 3:00 AM), https://www.bloomberg.com/news/articles/2020-12-16/facebook-fb-antitrust-case-has-much-different-goal-than-google-s-googl [https://perma.cc/ZSW2-78XC].

150. Tunku Varadarajan, *The 'Common Carrier' Solution to Social-Media Censorship,* WSJ (Jan. 12, 2021), https://www.wsj.com/articles/the-common-carrier-solution-to-social-media-censorship-11610732343 [https://perma.cc/HQ26-FVAJ].

discriminatorily. Likewise, in February 2021, Berkeley Law School Dean Erwin Chemerinsky—who is a prominent liberal—co-authored an op-ed with Professor Prasad Krishnamurthy urging Congress to enact a law barring "technology platforms with monopoly power" from "discriminating on the basis of political views."[151]

The most recent endorsement of exploring such an approach from a prominent member of the legal profession came from Supreme Court Justice Clarence Thomas on April 5, 2021. Concurring in the Court's decision to dismiss as moot a case that challenged former President Trump's blocking of several users from interacting with his Twitter account, Thomas's opinion (which was not joined by any other Justice) included an extended discussion of larger issues surrounding digital platforms. Although he did not conclusively endorse a particular regulatory approach, he did conclude that "[t]here is a fair argument that some digital platforms are sufficiently akin to common carriers or places of [public] accommodation" to be subject to regulations limiting their rights to exclude would-be users.[152]

VI. CONCLUSION

The unprecedented, constitutionally unconstrained power of dominant platforms to restrict free speech and user agency presents a major challenge. If "We the People" are to exercise the robust freedom of speech that is crucial to both our individual liberty and our sovereign power in our democratic republic, we must forge alternative—i.e., non-constitutional—strategies to constrain the dominant platforms' censorial power. These strategies must also respect the dominant platforms' own First Amendment rights to decide which speech and speakers to host and not to host.

Fortunately, the urgency of this topic has spurred burgeoning scholarship, discussion, and proposals by policymakers, experts in online technology, and various academic specialists. This Essay has summarized nine such proposals, which have received substantial attention, and which warrant further examination—but not yet implementation. Recognizing the

151. Prasad Krishnamurthy & Erwin Chemerinsky, *How Congress Can Prevent Big Tech from Becoming the Speech Police*, THE HILL (Feb. 18, 2021, 8:00 AM), https://thehill.com/opinion/judiciary/539341-how-congress-can-prevent-big-tech-from-becoming-the-speech-police [https://perma.cc/5RQS-GSD8]; *see* Ronald K.L. Collins, *First Amendment News 287: Bring Back the Fairness Doctrine? Destroying the Internet in Order to Save It*, FIRE (Feb. 24, 2021), https://www.thefire.org/first-amendment-news-287-bring-back-the-fairness-doctrine-destroying-the-internet-in-order-to-save-it/_[https://perma.cc/LJ2R-TMY5]; Ronald K.L. Collins, *First Amendment News 288: Krishnamurthy & Chemerinsky v. Corn-Revere—The Technology Platforms Debate Continues*, FIRE (Mar. 3, 2021), https://www.thefire.org/first-amendment-news-288-krishnamurthy-and-chemerinsky-v-corn-revere-the-technology-platforms-debate-continues/ [https://perma.cc/LUS6-T3KE].

152. Biden v. Knight First. Amend. Inst., 141 S. Ct. 1220, 1224 (2021) (Thomas, J., concurring).

novelty and complexity of these issues, we must proceed deliberately and cautiously. As the famed early twentieth-century journalist H.L. Mencken warned: "[T]here is always an easy solution to every complex problem—neat, plausible, and wrong."[153]

153. *H.L Mencken Quotes*, BRAINYQUOTE, https://www.brainyquote.com/quotes/h_l_mencken_141512 [https://perma.cc/WVM6-VK3L] (last visited Sept. 31, 2021).