University of Cincinnati Law Review

Volume 92 | Issue 2 Article 1

December 2023

Section 230 as Civil Rights Statute

Enrique Armijo
Elon University School of Law

SECTION 230 AS CIVIL RIGHTS STATUTE

Enrique Armijo[*]

CONTENTS

* Professor, Elon University School of Law; Affiliated Fellow, Yale Law School Information Society Project; Faculty Affiliate, UNC-Chapel Hill Center for Information, Technology, and Public Life; Research Fellow, George Washington University Institute for Data, Democracy & Politics. Thanks to Anupam Chander, Will Duffield, Bob Corn-Revere, Kate Ruane, Eugene Volokh, and Chris Newman for advice and suggestions, and to Morgan Ashmore, Krissy Chapman, Lyra Hanyu Jiang, Isabella Lawrence, and Raktima Roy for outstanding research assistance, as well as to the Liberty & Law Center's Free Speech, Civil Rights, and Social Progress initiative, which also provided financial support.

INTRODUCTION

Many of our most pressing and important discussions about justice, progress, and civil rights have moved online. Disaffected, isolated youth and activists advocating for social change no longer need to share physical space to connect with others who share their challenges and aspirations. But the convergence of mobility, connectivity, and technology is not the only reason why. Thanks to Section 230 of the Communications Decency Act's[1] ("Section 230") statutory immunity for social media platforms, websites and their hosts, and non-edge providers' carriage of user content, speakers can engage in speech about protest, freedom, equality, and dissent without fear of collateral censorship—threats of punishment or liability to speech distributors for the user speech they carry—from governments, authorities, and others in power who hope to silence them.

Without this immunity, underrepresented and marginalized communities would be far less able to raise awareness on issues related to their rights and survival. African Americans and other minorities would be less able to attract mainstream attention to the sometimes-deadly realities associated with interactions with the police in their communities. Women who say they have been sexually assaulted would be unable to organize online to counteract the advantages of their powerful and litigious attackers. A pseudonymous blogger would be less able to critique their local school board. Companies that provide internet infrastructure like Cloudflare and Dropbox would have to shift resources to moderating the websites and data that they host, resulting in less innovation and security, or fewer cloud-based services available to under-resourced users who wish to speak out online on issues of social change. In a world without Section 230, the new lights on these issues might never have been cast, or at least the lights would have been far less bright. And speech that much of society might find morally repugnant, such as pornography or terrorist speech, would also be suppressed—not just speech that falls *into* such categories, but also speech *about* speech in those categories. Attacking speech intermediaries is an attack on the speakers who rely on them. Conversely, protecting these intermediaries protects the speech itself.

Efforts to reform Section 230, which are currently taking place across legislative, academic, and judicial domains, are one form of intermediary attack.[2] Though the statute's immunity extends to any website or platform

1. *See* 47 U.S.C. § 230(c)(1) (stating that no "provider or user of an interactive computer service" may be "treated as the publisher or speaker of any information provided by another" person). For the sake of clarity, this Article uses the widely adopted, but largely imprecise, reference and citation to "Section 230 of the Communications Decency Act." *See* Blake E. Reid, *Section 230 of… what?*, BLAKE REID BLOG (Apr. 9, 2020), https://blakereid.org/section-230-of-what/.

2. *See, e.g.*, Press Release, Mark R. Warner, U.S. Senator from the Commonwealth of Virginia,

that hosts user speech, the immunity particularly benefits those voices from underserved, underrepresented, and resource-poor communities. By foreclosing the possibility of legal action against platforms for user speech, Section 230 limits the government and other powerful societal and corporate actors' ability to use collateral censorship to silence unpopular voices.

Though many commentators often assert that Section 230 promotes or inhibits the expression of minority viewpoints or speakers challenging the status quo,[3] this Article is the first attempt to demonstrate precisely how Section 230's distributor-based immunity protects civil rights-related speech. Before turning to the intersection of free speech, civil rights, and technology, Section I of this Article gives an overview of the common law and constitutional history of intermediary liability. This overview demonstrates how providing greater legal immunity for distributors of speech has historically resulted in more circulation of ideas that challenge the powerful and the status quo. Section II discusses how Section 230's immunity enables social media users to organize and raise awareness concerning civil rights-related issues, focusing on the online dissemination of police brutality videos, the #MeToo movement's raising of awareness around sexual harassment and assault, and the protection of anonymous digital speech. Section III examines the threats to speech in a world where such immunity does not exist. It demonstrates how a shift in the current liability rule would result in platform overmoderation of online user speech about civil rights, protest, advocacy for political change, and other controversial topics, and explains how speech-based user rights might not necessarily provide distributors a reliable basis upon which to dismiss collateral threats. Finally, Section IV addresses the claims that by granting platforms immunity for user speech, and by extension removing the possibility of platform liability for other users'

Legislation to Reform Section 230 Reintroduced in the Senate, *House* (Feb. 28, 2023), https://www.warner.senate.gov/public/index.cfm/2023/2/legislation-to-reform-section-230-reintroduced-in-the-senate-house (seeking to reform Section 230 immunity to "allow social media companies to be held accountable for enabling cyber-stalking, online harassment, and discrimination on social media platforms"); *see also* Jessica Guynn, *Trump vs. Big Tech: Everything You Need to Know About Section 230 and Why Everyone Hates It*, USA TODAY (Oct. 15, 2020), https://www.usatoday.com/story/tech/2020/10/15/trump-section-230-facebook-twitter-google-conservative-bias/3670858001/ (discussing federal and legislative efforts to revise or revoke Section 230 immunity); Jennifer Stisa Granick, *Is This the End of the Internet As We Know It?*, ACLU (Feb. 22, 2023), https://www.aclu.org/news/free-speech/section-230-is-this-the-end-of-the-internet-as-we-know-it (discussing arguments in Gonzalez v. Google and Twitter v. Taamneh, two cases seeking judicial alteration of Section 230 immunity in the U.S. Supreme Court); Adam Liptak, *Supreme Court Puts Off Considering State Laws Curbing Internet Platforms*, N.Y. TIMES (Jan. 23, 2023), https://www.nytimes.com/2023/01/23/us/scotus-internet-florida-texas-speech.html (discussing state laws in Texas and Florida seeking to compel social media platforms to carry speech).

3. *See* sources cited in *infra* Sections II-IV.

online threatening and harassment of minority speakers, Section 230 frustrates civil rights speech rather than promotes it.

I. THE LEGAL TRADITION OF PROTECTING DISTRIBUTORS TO PROTECT "CONTROVERSIAL" SPEECH

A. Why Collateral Censorship Works

Both common and constitutional law have long recognized the collateral censorship problems presented by intermediated speech. The distribution of speech usually relies on parties other than the speaker, such as a publisher, a bookseller, an internet service provider or other network infrastructure, an application provider, or an online platform. Collateral censorship occurs when the government or other powerful entity threatens to hold the distributor of speech liable for the content it distributes instead of or in addition to the speaker.[4] The threatener does so because it either dislikes or disapproves of the speech, and thus seeks to suppress it; in some cases, the speech might be about the threatener themselves. The prospect of liability results in the distributor ceasing distribution of the threatened speech because "[the distributor] has greater incentives to censor [the speaker] than [the speaker] has to self-censor."[5]

Additionally, because the distributor often distributes the speech of more than one speaker, the censor's returns on collateral censorship are compounding: by making one successful threat at the distributor level of the speech, the government or other threatener effectively silences a number of speakers that it dislikes or disfavors, as the distributor is likely to censor not just the speech that is the object of the threat to avoid direct liability, but is also likely to censor similar speech to preclude the threatener from repeating the threat.[6] Accordingly, without some legal protection for distribution, either by treating it as speech or as an independent right, the government or other powerful party will almost always choose to punish the distributor over the speaker, or at least threaten to do so, as a way to censor disfavored speech.[7]

4. J.M. Balkin, *Free Speech and Hostile Environments*, 99 COLUM. L. REV. 2295, 2296 (1999).

5. *Id.* at 2298-99.

6. *See generally* Michael I. Meyerson, *Authors, Editors, and Uncommon Carriers: Identifying the "Speaker" within the New Media*, 71 NOTRE DAME L. REV. 79, 117 (1995); Felix T. Wu, *Collateral Censorship and the Limits of Intermediary Immunity*, 87 NOTRE DAME L. REV. 293, 309 (2011).

7. For example, Christina Mulligan has argued that the First Amendment's Press Clause should be read as recognizing the threat of collateral censorship, since protecting speech's dissemination (including dissemination by technological means) is as necessary as protecting speech itself. *See* Christina Mulligan, *Technological Intermediaries and Freedom of the Press*, 66 SMU L. REV. 157, 160 (2013) ("[T]o protect the free press interests of those who endeavor to express themselves using technology of mass communication, technologists and technology must be subject to protection, for which there is no

B. Protecting Distributors from Collateral Censorship

Common law rules concerning distributor liability recognized the dangers of collateral censorship. Under the common law, a publisher of harmful speech such as a newspaper could be held liable as if it had spoken the speech itself. However, true distributors—namely those entities like bookstores, libraries, couriers, or newsstands that, unlike newspapers or broadcast stations, distribute copies of speech that have been printed or otherwise produced by others—could only be held liable for carrying speech when the distributor had knowledge of the speech's content. In such a circumstance, the distributor could be treated as the speech's publisher for liability purposes. In other words, where speech is the cause of the harm, liability for that harm should be direct, not vicarious.

The rule recognized that punishing dissemination of information could be justifiable only in those narrow circumstances where the person who "sold, rent, gave, or otherwise circulated" the speech had knowledge of the speech and its harmful nature.[8] Accordingly, the only way a bookseller or library could be liable for carrying a defamatory book is if it (or more precisely, its employees who manage book selection) knew or had reason to know of the book's defamatory content.[9] The reason for the common law rule was straightforward: If a passive distributor does not exercise editorial control over the speech it distributes, it should not be treated as if it did for liability purposes.[10] And finding secondary liability for distributors in the absence of knowledge would result in much less speech, because the economic risk of liability for speech that might be harmful would significantly outweigh the distributor's interest in carrying, selling, or otherwise making that speech available to others.

Existing First Amendment doctrine has also long recognized the dangers of collateral censorship by requiring that any potential liability for distributing speech be based on more than the fact of distribution itself.

---

direct equivalent under the Free Speech Clause.").

8. Ralph E. Helper, *Libel and Slander—Privilege of "Fair and Accurate Report" of Judicial Proceedings—Non-Liability of Vendor of Newspaper*, 37 MICH. L. REV. 1335, 1336 (1939) (alteration in original).

9. *See* RESTATEMENT (SECOND) OF TORTS § 581 (AM. L. INST. 1977) ("[O]ne who only delivers or transmits defamatory matter published by a third person is subject to liability if, but only if, he knows or has reason to know of its defamatory character.").

10. One additional common law rule applied to pure speech carriers, or (pre-internet) platforms, such as loudspeakers and typewriter and telephone companies, namely those entities that "make[] available to another equipment or facilities that [the user] may use himself for general communication purposes." *Id.* § 581 cmt. b. These carriers were not liable for carrying harmful or legally actionable speech, even if the carrier received notice that the speech it carried was actionable. *See id.*§§ 581 cmt. b, 612. This immunity was based on the principle that carriers did not review the speech they carried prior to or during carriage. *See, e.g.*, Anderson v. N.Y. Tel. Co., 320 N.E.2d 647, 649 (N.Y. 1974) (finding IBM could not be liable "were one of its leased typewriters used to publish a libel").

In *Smith v. California*,[11] a bookseller was convicted under a local ordinance that made it a crime to "have in his possession any obscene or indecent writing or book in any place of business where books are sold or kept for sale."[12] The ordinance was a strict liability statute: liability was for disseminating the speech alone, regardless of whether the bookseller had actually reviewed the offending material for sale.[13] If the bookseller stocked a book that a court later determined to be obscene, then the bookseller was criminally liable.[14] The U.S. Supreme Court held that such a rule would cause booksellers to "restrict the books [they] sell[] to those [they have] inspected,"[15] which would in turn result in the removal of books containing speech that were both covered by the statute and books whose contents were constitutionally protected. Therefore, because of the speech-adverse effects of collateral censorship, the Court concluded that strict distributor liability violated the First Amendment.

Other areas of distributor liability law support the claim that the government often targets the distributor of disfavorable content in order to suppress that content, and speech-favorable protections should thus apply at the distribution point where the government provides the pressure. In *Bantam Books, Inc. v. Sullivan*,[16] the Rhode Island Legislature established the Commission to Encourage Morality in Youth in order to "educate the public concerning any book . . . containing obscene, indecent, indecent[,] or impure language, or manifestly tending to the corruption of the youth . . . and to investigate and recommend the prosecution of all violations of said sections."[17] As described by the Supreme Court, however, the Commission's "practice" was not to advise *authors* of potential violations of these standards, but rather "to notify a distributor . . . that certain books or magazines distributed by him had been reviewed by the Commission and had been declared by a majority of its members to be objectionable for sale, distribution or display to youths under 18 years of age."[18] In other words, it was not the speaker of the objectionable material that the Commission sought to punish, but the distributors of that material. This was the approach the Commission took in the facts leading up to the *Bantam Books* case itself, sending "at least 35" findings of objectionable content material to the publishers and wholesale distributors of the books *Peyton Place* and *The Bramble Bush*,

11. 361 U.S. 147 (1959).
12. *Id.* at 148.
13. *Id.* at 152.
14. *Id.*
15. *Id.* at 153.
16. 372 U.S. 58 (1963).
17. *Id.* at 59-60.
18. *Id.* at 61.

among others.[19] These notices also reminded distributors of the Commission's "duty to recommend to the Attorney General prosecution of purveyors of obscenity."[20] And upon receiving the notices, the distributor took "steps to stop further circulation of copies of the listed publications," stopped "fill[ing] pending orders for such publications and refus[ed] new orders," and "instructed his field men to visit his retailers and pick up all unsold copies."[21] The law of obscenity reveals the same pattern: When the government wants to suppress speech that it believes to be immoral or otherwise harmful, it aims the punishments of the criminal law at the distributor level.[22]

As examined in more detail below, one primary reason for collateral censorship's effectiveness is its efficiency. Threats to distributors are by definition more effective than threats against individual speakers, because distributors usually disseminate the speech of more than one speaker. For the same reason, a distributor has little economic interest in the individual content or user that is the subject of the threat, and therefore the distributor's response incentive always cuts in favor of ceasing distribution of the threatened speech.[23] Additionally, a collateral threat need not necessarily be well-founded on an existing legal proscription or other consequence with respect to liability—a distributor has little incentive to investigate or independently analyze the merits of the claim that underlies the threat. And as also discussed below, a distributor in most cases will be reluctant to assert the rights of a speaker whose speech it carries in response to the threat, except in the clearest cases of no liability for either the speaker or distributor, or in a case where the distributor has a business-related reason for being seen as particularly user-protective.

In finding government acts of collateral censorship to violate the First Amendment, the Supreme Court has often recognized that threats and prosecutions of the distributor of objectionable speech have the practical effect of suppressing that speech. As the Court said in *Bantam Books*, for example, the "acts and practices" of the Rhode Island Commission

19. *Id.* at 61-62.

20. *Id.* at 62.

21. *Id.* at 63.

22. *See, e.g.*, Regina v. Hicklin, L.R. 3 Q.B. 360 (1868) (considering application of statute that made illegal keeping "any obscene books" for "the purpose of sale or distribution" to pamphlet reseller); Roth v. United States, 354 U.S. 476 (1957) (considering constitutionality of convictions of booksellers Roth and Alberts under statutes that made punishable "the mailing of [obscene] material" and keeping obscene material "for sale, or advertising"); Miller v. California, 413 U.S. 15 (1973) (considering application of California obscenity statute to Miller, a bookseller).

23. *See* Olivier Sylvain, *Platform Realism, Informational Inequality, and Section 230 Reform*, 131 YALE L.J. F. 475, 479 (2021) ("[S]ocial media's ability to host and distribute third-party content—and thus to connect people and build communities—is incident to its ravenous ambition to hold the attention of their consumers and collect their data for advertisers.").

"directly and designedly stopped the circulation of publications in many parts of Rhode Island," noting that "[t]he Commission's notices, phrased virtually as orders, reasonably understood to be such by the distributor, invariably followed up by police visitations, in fact stopped the circulation of the listed publications *ex proprio vigore* ['by its own force']."[24] This was "a form of effective state regulation" of the disfavored content at issue, which made actual regulation of the disfavored content itself unnecessary.[25]

***

Providing legal immunity to the distributors of controversial speech also protects speakers who discuss those topics online. The next Section applies that claim to three contemporary examples of underrepresented speech and underrepresented speakers, and shows how those speakers would be negatively affected in the absence of the immunity granted by Section 230 for those platforms, websites, and service providers who enable them to speak online.

II. HOW SECTION 230 ENABLES SPEECH ABOUT CIVIL RIGHTS

A. Protecting the Distribution of Police Brutality Videos

The internet has enabled minority and underrepresented communities to build social movements, express their views, find and connect with allies nationally and internationally, and share episodes of discrimination. Without social media, modern political movements such as the Movement for Black Lives would have been far less able to bring attention to their civil rights-based initiatives and to organize protests. The public recording and dissemination of interactions with police has become an essential component of law enforcement reform. As criminal procedure scholar Jocelyn Simonson argues, "promoting public participation in criminal justice must include facilitating the ability of civilians to observe, record, and contest police practices and constitutional norms"—particularly for "residents of neighborhoods with large concentrations of

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

25. *See also* McConnell v. FEC, 540 U.S. 93, 251 (2003) (Scalia, J., concurring in part and dissenting in part) ("To a government bent on suppressing speech, this mode of organization presents opportunities: Control any cog in the machine, and you can halt the whole apparatus. . . . Restrict the sale of books, and it matters little who prints them.").

poor people of color [who] have the most frequent contact with, but the least input into,"[26] those policies and practices.

Existing constitutional traditions have long recognized and empowered citizens to document and expose official wrongdoing, and to organize in their efforts to do something about it.[27] But social media platform technology, combined with wireless connectivity and cloud-based storage, has facilitated the recording and wide circulation of images and video, often simultaneously via real-time streaming,[28] which has brought the issue of police brutality to life for less affected communities.[29] As was the case at other pivotal points in African Americans' historical battle for civil rights, the power of the visual image has been an enormously valuable tool in "counteract[ing] widespread disbelief about the violations they face."[30] Visual proof, in short, has been one of the most essential tools in the long fight against racism in the United States. But the modern movement around police brutality has arguably been different in one notable and important respect: both the collection and dissemination of those images have primarily come from the hands of African Americans themselves.[31] Examples such as Darnella Frazier's video of George

26. Jocelyn Simonson, *Copwatching*, 104 CAL. L. REV. 391, 391-94 (2016); Seth F. Kreimer, *Pervasive Image Capture and the First Amendment: Memory, Discourse, and the Right to Record*, 159 U. PA. L. REV. 335, 350-51 (2011) ("[P]olice abuse captured by the cameras of bystanding videographers, followed by public broadcast of the footage, has become a regular feature of our public life and the underpinning of effective demands for redress.").

27. *See* Michael J. Socolow, *The Other George Floyd Story: How Media Freedom Led to Conviction in His Killer's Trial*, TECHDIRT (Apr. 23, 2021), https://www.techdirt.com/2021/04/23/other-george-floyd-story-how-media-freedom-led-to-conviction-his-killers-trial/.

28. Kreimer, *supra* note 26, at 340-41:

> Cell phone cameras . . . have radically reduced the nonmonetary cost of image capture. In modern life, cell phones constantly accompany their users. They combine effortless and immediately accessible digital photographic capability with the capacity to transmit captured images instantaneously. . . . [and] distributions channels for digital images [such as "YouTube" and "social networking sites like Facebook"] have combined with increasingly usable [] technology to enable any holder of an image to make it instantly available to the world at large.

29. *See generally* DEAN FREELON ET AL., CTR. FOR MEDIA & SOC. IMPACT, BEYOND THE HASHTAGS: #FERGUSON, #BLACKLIVESMATTER, AND THE ONLINE STRUGGLE FOR OFFLINE JUSTICE, (Feb. 29, 2016), https://cmsimpact.org/resource/beyond-hashtags-ferguson-blacklivesmatter-online-struggle-offline-justice/; *see also id.* at 8 ("[S]ocial media uniquely benefits oppressed populations" by "help[ing] level a media playing field dominated by pro-corporate, pro-government, and (in the United States) anti-Black ideologies.").

30. *See* Karen Hao, *How to Turn the Filming of Police into the End of Police Brutality*, MIT TECH. REV. (June 10, 2020), https://www.technologyreview.com/2020/06/10/1002913/how-to-end-police-brutality-filming-witnessing-legislation/ ("[T]o mobilize change against a given injustice, there must first be a majority consensus that the injustice exists," and visual proof has been historically essential for the minority to achieve that consensus, from abolitionism to anti-lynching to the images of police dogs and water attacks against marching African Americans in the 1950s.).

31. *See id.* ("[C]ompared with previous forms of witnessing, smartphones are also more accessible, more prevalent, and—most notably—controlled in many cases by the hands of Black witnesses."). African American newspapers and journalists in the Jim Crow era "built a formidable Black public sphere that

Floyd's murder; the fatal beating of Tyre Nichols; Feidin Santana's recording of Walter Scott's shooting by Charleston, South Carolina Police; and Diamond Reynolds's livestream of Philando Castille shortly after he was shot by police have brought these instances and dozens of others to life for millions on the internet.[32] Citizen-filmed and police bodycam footage of police interactions with the public have become an essential, if not *the* essential, tool in solving the issue of police accountability for misconduct.[33] Combining recording and broadcast technologies in a single device has been transformational for the police reform movement.

Technology, and its seamless "assimilation . . . into daily life," obviously play an important role in enabling this amplification.[34] But so, too, does the law. Section 230 immunizes platforms and websites from any liability for hosting user content. With a few exceptions not relevant to this Article, Section 230 prevents those seeking to have user content removed from the internet from threatening the content's host with legal action if the host fails to take the user content down. Without this immunity, actors all along the internet backbone, from service providers to websites to social media platforms, would all be much more susceptible to threats of liability and would likely respond to those threats with collateral censorship.

This is particularly true with respect to the documentation of negative interactions with police. Despite the awareness raised by these videos,

---

forever after organized and rallied dissent against African American exclusion from the project of democracy in the United States." Sid Bedingfield & Kathy Roberts Forde, *Introduction: Journalism and the World It Built*, *in* JOURNALISM & JIM CROW: WHITE SUPREMACY AND THE BLACK STRUGGLE FOR A NEW AMERICA 3, 3 (Sid Bedingfield & Kathy Roberts Forde, eds. 2021). But though the "Black journalism" of the period "became the incubator of a new tide of Black activism" and its journalists worked to "raise public consciousness to the realities of racial injustice," D'Weston Haywood, *Fight for a New America*, *in* JOURNALISM & JIM CROW: WHITE SUPREMACY AND THE BLACK STRUGGLE FOR A NEW AMERICA, *id.*, at 58, "few whites commonly read Black newspapers," though "white journalists certainly did and periodically reprinted excerpts from them," W. Fittzhugh Brundage, *The Press and Lynching*, *in* JOURNALISM & JIM CROW: WHITE SUPREMACY AND THE BLACK STRUGGLE FOR A NEW AMERICA, *id.*, at 95.

32. *See, e.g.*, Joseph Cox & Jason Koebler, *Facebook Decides Which Killings We're Allowed to See*, VICE (July 7, 2016), https://www.vice.com/en/article/8q85jb/philando-castile-facebook-live (at the time of the article, Reynolds' video of the shooting's aftermath had been viewed more than 2 million times and shared more than a quarter million times).

33. *See, e.g.*, Howard Wasserman, *Police Misconduct, Video Recording, and Procedural Barriers to Rights Enforcement*, 96 N.C. L. REV. 1313, 1316-17, 1316 nn.12-19 (2018); *see also* Mary D. Fan, *Democratizing Proof: Pooling Public and Police Body-Camera Videos*, 96 N.C. L. REV. 1639, 1643 (2018) ("Allowing community member videos [of public-police interactions] into the official record can improve police accountability; address imbalances of power in police-said, defendant-said credibility contests; and offer a fuller picture of what happened in a contested incident.").

34. Tyler Goodwin, *Breaking the Internet: Section 230 Reform and the Future of Social Movements*, THE PLUG (Oct. 21, 2021), https://tpinsights.com/breaking-the-internet-section-230-reform-and-the-future-of-social-movements/.

people in positions of power have demonstrated their intent to suppress the distribution of such speech, and displayed how they might use the law to attempt to do so. Even as the legal protections for civilian recording of police officers in public become increasingly established, in practice, officers have arrested individuals under state wiretapping statutes, obstruction with police operations and related "failure to obey"-related laws, and for disorderly conduct.[35] Many wiretapping statutes, including the federal Wiretap Act, impose civil and criminal liability on not just the recorder of the protected information, but on any party that discloses the protected information if they have "reason to know that the information was obtained" in violation of the statute.[36] This could potentially form the basis for distributor-level threats in the absence of immunity, despite potential First Amendment and knowledge-based defenses that distributors could assert in response to the threats. Similarly, in the context of police body-cam initiatives, some police officers and departments and their unions have backed away from their initial support of body cam-mandating legislation at the local, state, and federal levels, "concerned with lack of control over the decision when to record and over subsequent release and use of the resulting video, fearing officer embarrassment or worse."[37] Additionally, officers with only a rudimentary understanding of intermediary liability have attempted to put it into effect to try and frustrate distribution of videos of their interactions with the public; one Oakland police officer intentionally played Taylor Swift's music in an attempt to use copyright law to keep video of a confrontation with activists from being posted to YouTube.[38]

35. Simonson, *supra* note 26, at 427-32 (collecting examples and discussing why officers often resist being filmed even in light of law and department-level policies that protect the public's right to do so); Kreimer, *supra* note 26, at 357-58 ("[T]he spread of pervasive image capture in the last decade has been accompanied by a rich set of cases in which police have sought to prosecute critics or potential critics who capture their images[, and in] these cases, police officers and other officials have enlisted both existing statutes and creative prosecutorial discretion in the struggle to constrain inconvenient image capture."); Aidan J. Coleman & Katharine M. Janes, Comment, *Caught on Tape: Establishing the Right of Third-Party Bystanders to Secretly Record the Police*, 107 VA. L. REV. ONLINE 166, 169 & nn.13-18 (2021) (listing reports of multiple arrests of individuals for recording the police in public). Of course, these kinds of arrests are the challenged action at issue in cases establishing a First Amendment right to record the police in public. *See, e.g.*, ACLU v. Alvarez, 679 F.3d, 583, 592 (7th Cir. 2012) (listing "many recent prosecutions against individuals who recorded encounters with on-duty police officers" under Illinois wiretap statute); Sharpe v. Winterville Police Dep't, 59 F.4th, 674, 678 & n.1 (4th Cir. 2023) (appellant who attempted to livestream his own traffic stop challenging his arrest under the North Carolina "resisting, delaying, or obstructing" a police officer statute); Glik v. Cunniffe, 655 F.3d 78, 88 (1st Cir. 2011) (same, under Massachusetts wiretap statute). *See infra* Part III.B (discussing judicial affirmation of a First Amendment-derived right to record the police in public).

36. 18 U.S.C. § 2511(1)(c).

37. Wasserman, *supra* note 33, at 1317.

38. The officer's efforts, which caused the video to go viral, had the exact opposite effect. Julian Mark, *An Officer Played a Taylor Swift Song to Keep His Recording off YouTube. Instead it Went Viral.*, WASH. POST (July 2, 2021, 7:40 AM), https://www.washingtonpost.com/nation/2021/07/02/taylor-swift-

Furthermore, several state-level police officer bills of rights bar the release of officer photographs without the officer's consent, which can potentially provide officers a tool by which to attempt to frustrate the dissemination of videos in which they appear—even when those videos are distributed by the officers' own departments.[39]

Current legislative efforts, which began in the post-summer 2020 period, could potentially form additional bases for collateral threats against distributors of police brutality videos. For example, Oklahoma legislators are considering a law that criminalizes "willfully publish[ing] on a public online site or forum . . . personally identifiable information of a law enforcement officer with intent to threaten, intimidate, harass or stalk," with publishing including "a photograph or any other realistic likeness of the person" disseminated "on a public online site or forum" "by telecommunication or electronic communication."[40] In 2022, Arizona adopted a law making it a crime to record law enforcement encounters from closer than eight feet away after an officer's verbal warning to stop recording.[41] Additionally, an Indiana law went into effect on July 1, 2023 which criminalizes approaching a police officer within twenty-five feet after the officer has ordered the person to stop.[42] These efforts, which legislatures claim are "in service of purportedly neutral values such as non-interference, officer safety, public safety, or protection of officer and public privacy,"[43] can all form the basis of collateral threats against distributors of police brutality videos and livestreams—threats whose true aim is to shield official conduct from public scrutiny. According to one citizen journalist, the Indiana encroachment law has already been used to threaten him with arrest when he livestreamed police activities in public.[44] Platforms that host or livestream such videos could be threatened with aiding-and-abetting liability for these privacy-based violations.

Additionally, government attempts to address what it views as social

---

california-cop/.

39. *See, e.g.*, CAL. GOV'T CODE § 3307.5 (Deering 2023); IOWA CODE § 80F.1 (2023); LA. STAT. Ann. § 40:2532 (2023); MINN. STAT. § 626.89 (2023); NEV. REV. STAT. § 289.025 (2023).

40. H.B. 2273, 58th Leg., 1st Sess. (Okla. 2021). *See, e.g.*, Wasserman, *supra* note 33, at 1348 (describing Texas legislator's bill, introduced in 2015, that would have expanded disruption with a peace officer statute to include "filming, recording, photographing, or documenting the officer within 25 feet of the officer" (citation omitted)). The bill was later withdrawn. *See id.* at 1349.

41. ARIZ. REV. STAT. § 13-3732 (LexisNexis 2023); *see also* Lindsey Bever, *New Arizona Law Criminalizes Filming Police from Less than 8 Feet Away*, WASH. POST (July 8, 2022), https://www.washingtonpost.com/nation/2022/07/08/arizona-police-recordings-8-feet/. The law was subsequently enjoined from enforcement on First Amendment grounds. *See* Arizona Broads. Ass'n v. Brnovich, 626 F. Supp. 3d 1102 (D. Ariz. 2022).

42. H.N. 1186, 123rd Gen. Assemb., 1st Reg. Sess. (Ind. 2023).

43. *See, e.g.*, Wasserman, *supra* note 33, at 1349.

44. *ACLU of Indiana Claims New 25-Foot Encroachment Law Violates First Amendment Rights*, ACLU OF IND. (Aug. 8, 2023), https://www.aclu-in.org/en/press-releases/aclu-indiana-lawsuit-claims-new-25-foot-encroachment-law-violates-first-amendment.

media's harmful dissemination of mass shooting videos could also provide opportunities for both collateral and direct threats against distributors for hosting user content involving police brutality. For example, in May 2021, after a gunman in Buffalo used Twitch to livestream the shooting murders he committed, the Office of the New York State Attorney General proposed the creation of a "[n]ew [a]ffirmative [criminal] [l]iability" for the "creation by the perpetrator, *or someone acting in concert with the perpetrator*, of images or videos of a homicide."[45] The Attorney General also recommended a carve-out for Section 230 distributor immunity, arguing that Section 230 immunity for violent livestreams should only attach if a distributor

> [a]t a minimum, [takes] reasonable steps [to] include efforts to remove unlawful violent criminal content and content likely to incite or solicit violent crime[;] measures to prevent the platform from being used to encourage or plan acts of violence[;] and limits on live-streaming technology designed to prevent its use to further criminal acts and incite or solicit violent crimes.[46]

To be clear, as the proposal's text states, the Attorney General's recommendations are aimed at the perpetrators of violent acts, not bystanders who record police brutality and the platforms that host their videos. Elsewhere in its report, the Attorney General's Office notes that "appropriate legislation should avoid covering videos created by bystanders or passively, such as those captured by police officers' body-worn cameras."[47] The report also says any legislation should "aim to avoid imposing any penalty for videos that have educational, historical, or social benefits."[48] And current New York law grants an affirmative right to "record police activities and maintain custody and control of any such recording and of any property or instruments used in such [a] recording."[49] But any legislation adopted consistent with these recommendations would likely be unclear enough that it could form the basis for effective distributor-level threats. For example, revoking distributor immunity on the grounds that the distributor failed to "prevent the platform from being used to encourage or plan acts of violence" could create possible liability for a wide range of social media uses, from posting public-police interactions to planning public protests that might in part turn violent. It is also unclear whether a "bystander" exception

45. LETITIA JAMES, OFF. OF THE N.Y. STATE ATT'Y GEN., INVESTIGATIVE REPORT ON THE ROLE OF ONLINE PLATFORMS IN THE TRAGIC MASS SHOOTING IN BUFFALO ON MAY 14, 2022, at 41 (Oct. 18, 2022), https://ag.ny.gov/sites/default/files/buffaloshooting-onlineplatformsreport.pdf.

46. *Id.* at 44.

47. *Id.* at 41.

48. *Id.* at 42.

49. *See* N.Y. Admin. Code § 14-189(b) (2023).

would immunize videomaking and posting, given the fact that some brutality videos are made by the victims of the brutality themselves. In sum, attempts to guard against live-streamed violence that rely on increasing liability for the distributor of the livestream run the risk of overinclusivity, and one cost of that overinclusivity is likely to be civil rights-related speech. The current liability rule, which prevents distributors from being held liable for videos posted by users, prevents this censorship from happening.

As discussed in Section III.B. below, a growing number of appellate courts have recognized a First Amendment-based right to record the police in public.[50] And a constitutional protection from prosecution for recording logically extends to protection for distribution—clearly so in the case of the user-as-distributor themselves, while less clear but also likely for third-party distributors. But before even reaching the First Amendment issue, under the current immunity regime, threats to distributors for the posting of police brutality videos are largely ineffective. Police reform-related speech thus directly benefits from distributor-level protection.

B. #MeToo: Section 230 as Shield Against Defamation-based Distributor Threats

Platform immunity also enabled another recent social media-based societal movement, the #MeToo campaign, which seeks to bring awareness and accountability around issues of sexual harassment and abuse. Starting in 2017, the use of the #MeToo hashtag on social media permitted thousands of sexual assault victims to speak out, often against assaulters who are powerful men.[51] The movement focuses on specific cases of assault, "allows women and other survivors to acknowledge their own trauma and suffering while showing themselves, each other, and the world that every person who has suffered is not alone," and seeks to spur discussion around social norms related to workplace harassment and discrimination.[52] The movement has led to legal reform as well; just a

50. *See infra* Part III.B and accompanying notes (discussing Eleventh, Ninth, First, Seventh, Fifth, and Third Circuit opinions).

51. *See Leading with Empathy: Tarana Burke and the Making of the MeToo Movement*, HARV. KENNEDY SCH. CASE #2197.0 (Nov. 16, 2020), https://case.hks.harvard.edu/leading-with-empathy-tarana-burke-and-the-making-of-the-me-too-movement/.

52. JoAnne Sweeny, *The #MeToo Movement in Comparative Perspective*, 29 AM. U. J. GENDER, SOC. POL'Y, & L. 33, 34-35 (2020); *see also* Shana L. Maier et al., *Rape Victim Advocates' Perceptions of the #MeToo Movement: Opportunities, Challenges, and Sustainability*, 38 J. INTERPERSONAL VIOLENCE 336, 337-39 (2022) ("The movement's efforts to raise awareness about the frequency of sexual victimization can also improve the understanding that it is a societal issue, not a personal one," leading to "a greater understanding of sexual violence.").

year after #MeToo went viral on social media, the Equal Employment Opportunity Commission reported a 13% increase in workplace harassment claims, "a jump [the agency] attributed to the #MeToo movement."[53]

Every social movement that encourages public accusations of wrongful conduct is inevitably going to include some accusations that turn out to be false. In the context of reports to police, where it is more difficult for victims to remain anonymous and there are legal consequences to making false reports, existing empirical evidence indicates that the instances of false claims of sexual assault or harassment are low.[54] No analogous comprehensive study of #MeToo-related false reporting has yet been undertaken; one might assume that in the lower-stakes context of anonymous speech on the internet, there may be a greater prevalence of false reports of sexual assault.[55] In any case, however, the tort of defamation permits any named subject of an accusation to both claim falsity and threaten the accuser with legal consequences.[56] Since the common law of defamation treats the dissemination of an allegedly defamatory statement the same as its initial utterance, the subject can threaten the disseminator with legal liability as well, assuming other elements of the tort are met.[57]

Section 230 immunity, however, prevents this from happening to

53. Ellen McCarthy, *#MeToo Raised Awareness About Sexual Misconduct. Has It Curbed Bad Behavior?*, WASH. POST (Aug. 14, 2021), https://www.washingtonpost.com/lifestyle/style/andrew-cuomo-me-too/2021/08/13/1ae95048-fbed-11eb-8a67-f14cd1d28e47_story.html (noting that almost twenty states enacted new sexual harassment protections in the same period); Sweeny, *supra* note 52, at 42 (discussing state-level banning of nondisclosure agreements in sexual misconduct cases).

54. A 2015 meta-study of seven false reporting studies calculated a 5% false report rate of sexual assaults to police. Claire E. Ferguson & John M. Malouff, *Assessing Police Classifications of Sexual Assault Reports: A Meta-Analysis of False Reporting Rates*, 45 ARCHIVES SEXUAL BEHAV. 1, 18-19 (2016); *see also* Sweeny, *supra* note 52, at 46 (citing *False Reporting*, NAT'L SEXUAL VIOLENCE RES. CTR. (2012), https://www.nsvrc.org/publications/false-reporting-overview). Another study reviewing ten years of reported cases to a university police department found that 5.9% of the reports were false. David Lisak et al., *False Allegations of Sexual Assault: An Analysis of Ten Years of Reported Cases*, 16 VIOLENCE AGAINST WOMEN 1318, 1329 (2010).

55. False complainants to police can be prosecuted and convicted, so the stakes associated with a false complaint to police are obviously higher for the accuser. *See* Andre W.E.A. DeZutter et al., *Motives for Filing a False Allegation of Rape*, 47 ARCHIVES SEXUAL BEHAV. 457, 457 (2018). False accusations in the Title IX context also have consequences for accusers that can range from academic sanction to civil suit by the accused to false-report-based criminal charges. *See, e.g.*, Julianne McShane, *Stanford University Employee Charged with Making 2 False Sexual Assault Allegations*, NBC NEWS (March 16, 2023), https://www.nbcnews.com/news/crime-courts/stanford-university-employee-charged-making-2-false-sexual-assault-all-rcna75264.

56. *See, e.g.*, Madison Pauly, *She Said, He Sued*, MOTHER JONES (Mar./Apr. 2020), https://www.motherjones.com/crime-justice/2020/02/metoo-me-too-defamation-libel-accuser-sexual-assault/ (discussing defamation claims brought against #MeToo accusers, noting that "at least 100 defamation lawsuits have been filed against accusers since 2014," and that "the threat of being sued, and the expense of mounting a legal defense, has deterred many survivors who seek to speak out").

57. RESTATEMENT (SECOND) OF TORTS § 578 (AM. L. INST. 1977).

online speech. Though a person falsely accused of sexual harassment can sue their accuser for posting an accusation, they cannot sue Instagram, Facebook, or Twitter for failing to take the accusation down. As discussed in Section I above, the distributor of speech has little if any incentive to continue carrying user speech for which it might be liable when that speech has even a *de minimis* prospect of subjecting the distributor to liability; indeed, the distributor has little incentive to even investigate the veracity of the accuser's falsity claim before deciding to take the accusation down. So, without Section 230's protections, #MeToo would not have had nearly the impact on society that it has had, because alleged harassers—again, most of whom have been more powerful and better resourced than their accusers—could silence accusations about them by threatening platforms with liability for distributing statements that the harasser claims are defamatory and false.

To be sure, the well-established, speech-protective law of public figure defamation, particularly the actual malice rule from *New York Times v. Sullivan*,[58] would apply to #MeToo-related, distributor-level threats in the absence of Section 230 immunity. Once a defamation plaintiff is deemed to be a public official or public figure, the requirement that they show that the false and defamatory statement about them was made with actual malice would apply to the distributor as well as the speaker should they seek to hold the distributor liable. This would be even more difficult for a plaintiff to establish than in an ordinary defamation claim against the speaker alone because of the absence of proof with respect to the distributor's state of mind. For example, the question whether clear and convincing evidence exists of a platform's knowledge of the falsity of a user's posted statement raises a host of notice-related issues, such as whether the threatener provided notice that the statement was false, whether that notice was received and sufficiently reliable to provide knowledge to the distributor of likely falsity, and whether the platform continued to distribute the statement despite that notice. However, a distributor who receives notice of a possibly defamatory statement pursuant to a collateral litigation threat "[has] a natural incentive simply to remove [the statement] upon notification,"[59] and then investigate issues with respect to the statement's falsity or its own state of mind with respect to that falsity later, if at all. And the possibility of distributor-level defamation liability would likely cause platforms, websites, and servers to decline to host certain kinds of user content that would expose them to potential liability.[60] Accordingly, Section 230's immunity ensures that the

58. 376 U.S. 254, 279-80 (1964).

59. Zeran v. Am. Online, Inc., 129 F.3d 327, 333 (4th Cir. 1997).

60. *See* Note, *Section 230 as First Amendment Rule*, 131 HARV. L. REV. 2127, 2038 (2018) ("[P]olitically controversial speech or business and product reviews may be more likely to lead to

party best suited to litigate the presence or absence of actual malice—i.e., the speaker—is the one doing so.

C. Protecting User Speech About Abortion

In the wake of the Supreme Court's *Dobbs v. Jackson Women's Health Organization*'s[61] 2022 decision finding that abortion is not a fundamental right, it is no surprise that many states have criminalized the procedure. However, several of those states have taken, or are considering, the additional step of criminalizing or imposing civil liability for *speech* about abortion, on the longstanding theory that speech which is essential to the commission of a criminal act is not protected by the First Amendment. For example, four days after the *Dobbs* decision, South Carolina legislators proposed a law that would make it "unlawful to aid, abet, or conspire with someone to procure an abortion"; "aiding-and-abetting" in this context means providing information on abortion services to "a pregnant woman, by telephone, internet, or other mode of communication."[62] Oklahoma enacted a law in May 2022 that allows a civil action against any person who "knowingly engages in conduct that aids or abets the performance or inducement of an abortion."[63] As a general matter, the First Amendment is not a defense to aiding-and-abetting criminal liability.[64] Accordingly, under such laws, a social media platform or website could in theory be threatened with aiding-and-abetting-based liability for "providing" access to user posts that provided information on obtaining an abortion—even a legal abortion in a state that has not banned the procedure—on the theory that the platform or website was accessible in the banning state.

The First Amendment's application in the abortion context is even less

---

defamation actions than more mundane content[,] or bloggers might decline to include a comment section.").

61. 142 S. Ct. 2228 (2022).

62. S.1373, 124th Gen. Assemb., Reg. Sess. (S.C. 2022), https://www.scstatehouse.gov/sess124_2021-2022/bills/1373.htm; Dessie Otachliska, *Free Speech Post-*Dobbs: *The Constitutionality of State and Federal Restrictions on the Dissemination of Abortion-Related Information*, N.Y.U. J. LEGIS. & PUB. POL'Y QUORUM (2023), https://nyujlpp.org/quorum/otachliska-free-speech-post-dobbs/. The bill in South Carolina was based on model legislation from the National Right to Life Committee. *See* Memorandum from The Bopp Law Firm, PC on NRLC Post-*Roe* Model Abortion Law (June 15, 2020) (on file with the National Right to Life Committee), https://www.nrlc.org/wp-content/uploads/NRLC-Post-Roe-Model-Abortion-Law-FINAL-1.pdf.

63. Carmen Forman, *Oklahoma Gov. Kevin Stitt Signs Nation's Strictest Abortion Ban. It Starts Immediately*, OKLAHOMAN (May 25, 2022), https://www.oklahoman.com/story/news/2022/05/25/oklahoma-implements-nations-most-restrictive-abortion-ban-kevin-stitt-roe/9891816002/. The Oklahoma law's private enforcement mechanism mirrors that of Texas's, a law which was the subject of pre-*Dobbs* abortion-related litigation at the Supreme Court. *See id.*

64. *See generally, e.g.*, Rice v. Paladin Enters., Inc., 128 F.3d 233 (4th Cir. 1997).

clear than in cases involving speech about police brutality. In the 1975 case of *Bigelow v. Virginia,* the Supreme Court declared that the state of Virginia could not criminalize a newspaper advertisement regarding available abortion services in the state of New York under a law that outlawed "the sale or circulation of any publication" that "encourage[s] the procuring of abortion."[65] The ad at issue, however, provided information not only "to readers possibly in need" of an abortion themselves, but also those with a "genuine interest in[] the subject matter or the law of another State."[66] In other words, the ad did not simply provide instructions to Virginians seeking to elude the state's abortion ban about how to obtain an abortion in a state that permitted them; it also shared information that merited a high level of First Amendment protection, namely information about the law of another state. Accordingly, the Court said that a state "may not, under the guise of exercising internal police powers, bar a citizen from another State from disseminating information about an activity that is legal in that State."[67] And in the latter case of *United States v. Edge Broadcasting*,[68] the Court held that it did not violate the First Amendment when Congress banned lottery advertising in states where lotteries were illegal, but permitted such advertising in states that ran their own lotteries under the more lenient test for commercial speech set out in *Central Hudson Gas & Elec. Corp. v. Public Service Commission*.[69] Congress had a "substantial interest" in "supporting" the states' bans on lotteries by banning their advertising.[70] Under this logic, states' efforts to ban abortion advertising on the theory that such ads "provide information" concerning abortions may be upheld.

Additionally, as the history of collateral censorship shows, when parties can point to laws to threaten legal liability, they usually do not hesitate to rely on them to threaten distributors. If a platform, website, or edge provider is threatened with aiding-or-abetting prosecution for hosting instructional speech about abortion, or (for example) advertising the availability of abortion pills, the distributor is likely to remove the speech. Even a broad interpretation of the protections afforded by *Bigelow* could be deemed to not apply to speech online posted from a state where abortion is legal that would be accessible anywhere. Law enforcement

65. 421 U.S 809, 812-13, 825-26 (1975).

66. *Id.* at 822.

67. *Id.* at 824-25; *see also* John Villasenor, *Can a State Block Access to Online Information About Abortion Services?*, BROOKINGS (July 27, 2022), https://www.brookings.edu/articles/can-a-state-block-access-to-online-information-about-abortion-services/.

68. *See generally* 509 U.S. 418 (1993).

69. 447 U.S. 557, 564 (1980).

70. 509 U.S. at 426.

and legislators seeking to enforce their home-state bans on abortion speech would argue that such speech is intended to aid-and-abet abortions not just in states where the procedure is legal, but in states where it is not.[71] The possibility of such threats results in less speech about that topic—even in states in which abortion is protected by the law.

D. Distributor Immunity as Protective of Anonymous User Speech

As noted above, many of the accusations that fueled the #MeToo movement were made anonymously. The history of anonymous speech in the United States is closely tied to the expressive components of social movements. Minorities and other oppressed groups using anonymous speech to speak out against their unfair treatment by those in power has a history going back to the Founding.[72] African Americans took advantage of anonymity to speak and associate around civil rights activism in the 1960s American South.[73] This is all equally true at present and online, where anonymous speech is particularly important for minorities and queer youth seeking allies and speaking out against unfair treatment.[74] Anonymity and pseudonymity give speakers breathing room to advocate for social change without fear of retribution from the powerful.[75] Accordingly, to the extent those communities both use the internet for civil rights-related speech and disproportionately use anonymity online, reducing protections for anonymous speech would also have a disproportionate effect on marginalized communities and speakers.

Distributor immunity protects user anonymity. In the absence of liability for carriage of distributor anonymous speech, the party claiming to be harmed must shift their efforts to user unmasking instead, and unmasking is often not an easy task. First, the ability to identify

71. *See* Note, Samantha Mitchell, *First Amendment Speech Protections in a Post-*Dobbs *World: Providing Instruction on Instructional Speech*, 91 FORDHAM L. REV. 1521, 1524 (2023).

72. JEFF KOSSEFF, THE UNITED STATES OF ANONYMOUS: HOW THE FIRST AMENDMENT SHAPED ONLINE SPEECH 16-21 (2022).

73. *See, e.g.*, *id.* at 25-31 (discussing NAACP v. Alabama, 357 U.S. 449 (1958), and the U.S. Supreme Court's recognition therein that "the First Amendment protects not only the right to association, but also the right to exercise it anonymously").

74. *See, e.g.*, Elliot Harmon, *In Debate over Internet Speech Law, Pay Attention to Whose Voices Are Ignored*, THE HILL (Aug. 21, 2019), https://thehill.com/opinion/technology/458227-in-debate-over-internet-speech-law-pay-attention-to-whose-voices-are/; Jillian C. York & Dia Kayyali, *Facebook's "Real Name" Policy Can Cause Real-World Harm for the LBGTQ Community*, ELEC. FRONTIER FOUND. (Sept. 16, 2014), https://www.eff.org/deeplinks/2014/09/facebooks-real-name-policy-can-cause-real-world-harm-lgbtq-community.

75. *See, e.g.*, McIntyre v. Ohio Elec. Comm'n, 514 U.S. 334, 357 (1995) ("Anonymity is a shield from the tyranny of the majority. . . . [it] protect[s] unpopular individuals from retaliation . . .at the hand of an intolerant society.").

anonymous users will often depend on the data retention or user anonymity policies of ISPs, websites, and platforms, such as the existence of a real name or IP address retention policy and the like. These can operate as de facto statutes of limitations on claims against users.[76] Furthermore, there are robust First-Amendment-informed protections around user anonymity that courts consider when a plaintiff seeks to force a website, platform, or ISP to identify its users.[77] Distributors will often assert these defenses in response to those attempts—particularly where the distributor's business model either relies directly on anonymous user-generated, opinion-based content, such as the employer review site Glassdoor,[78] or where the distributor values being seen as anonymity-protective in the interest of earning corporate goodwill for privacy-based competitive perception advantages or other reasons.[79]

Whether distributors would be as aggressive in defending efforts to unmask users in the absence of their own immunity for the anonymous user's speech remains unanswered. Under the existing regime, when the

76. Some commentators have argued that the solution to these problems is to require user anonymity to be more traceable where an anonymous user has caused a cognizable harm. *See* Mark Lemley, *Rationalizing Internet Safe Harbors*, 6 J. TELECOMM. & HIGH TECH. L. 101, 117 (2007); Danielle Keats Citron, *Cyber Civil Rights*, 89 B.U. L. REV. 61, 123 (2009); A. Michael Froomkin, *Anonymity and its Enmities*, 1. J. ONLINE L. art. 4, ¶¶ 14-18 (1995).

77. KOSSEFF, *supra* note 72, at 113-32 (discussing judicial approaches in assessing subpoenas aimed at websites that seek to unmask anonymous posters).

78. *See Glassdoor Employee Free Speech/User Anonymity Fact Sheet: Courts Regularly Back Glassdoor's Fight To Protect User Anonymity and Employee Free Speech*, GLASSDOOR (updated July 2021), https://about-content.glassdoor.com/app/uploads/sites/2/2021/07/Glassdoor-Legal-Fact-Sheet_July-2021.pdf [https://web.archive.org/web/20211030183042/https://about-content.glassdoor.com/app/uploads/sites/2/2021/07/Glassdoor-Legal-Fact-Sheet_July-2021.pdf] (noting that Glassdoor has "succeeded in protecting the anonymity of our users leaving reviews on [its] site in over 100 cases," and discussing several such victories in state and federal courts).

79. *See, e.g.*, Eriq Gardner, *Google to Fight Subpoena Demands Over "Shitty Media Men" Spreadsheet*, HOLLYWOOD REP. (Oct. 12, 2018), https://www.hollywoodreporter.com/business/business-news/google-fight-subpoena-demands-shitty-media-men-spreadsheet-1151932/ (discussing a situation where a defamation plaintiff suing over allegations made in the Google spreadsheet "Shitty Media Men List" sought to unmask anonymous collaborators on and viewers of the document, and the Google spokesperson was quoted as stating the company will "oppose any attempt . . . to obtain information about this document from us"); Claire Lampen, *Google Won't Disclose Names of 'Sh*tty Men in Media' List Contributors*, NEW YORK (Jan. 25, 2019), https://www.thecut.com/2019/01/google-declines-to-disclose-sh-tty-men-in-media-metadata.html. None of the anonymous contributors named as Jane Doe defendants in the defamation litigation were ever identified in court prior to the suit's settlement in 2023. *See* Jessica Testa, *'Media Men' Lawsuit Ends in a Settlement*, N.Y. TIMES (Mar. 6, 2023), https://www.nytimes.com/2023/03/06/style/media-men-list-settlement-moira-donegan.html#:~:text=rape%20or%20harassment.%E2%80%9D-,Ms.,expenses%20raised%20more%20than%20%24116%2C000. For a platform-wide survey, *see Twitter Transparency Reports: Information Requests*, TWITTER (last updated July 28, 2022), https://transparency.twitter.com/en/reports/information-requests.html#2021-jul-dec [https://web.archive.org/web/20220728161807/https://transparency.twitter.com/en/reports/information-requests.html#2021-jul-dec] (noting that in the second half of 2021, Twitter "objected to 29 U.S. civil requests for account information that sought to unmask the identities of anonymous speakers," and that "[n]o [user] information was produced on 93% of all unmasking requests").

distributor has certainty with respect to the prospect of no liability for the anonymous user's conduct, the threatener is unable to threaten to "unmask or else." On the other hand, if collateral censorship threats were made more effective by increasing prospective distributor liability for anonymous user speech, the "unmask or else" threat is back on the table of options, and the dynamics shift away from less powerful speakers and toward those who seek to suppress their speech. Distributors are more likely to defend anonymity when they have less to lose; a loss under the current liability regime only means providing the requested user data, assuming it remains available. But if an unmasking demand could be coupled with a meaningful threat of legal liability for distribution of the anonymous content, the risk of liability may cause platforms to comply more often with requests to disclose user identities to make the risk of liability go away. Furthermore, when the threatener has the option to threaten the distributor with liability for the substance of the anonymous content, the threat would allow plaintiffs to "get the relief they are most likely to want—removal of the [anonymously posted] content."[80]

III. SHIFTING THE IMMUNITY RULE: CONSEQUENCES

A. Disproportionate Effects of Overmoderation

In an online world where the current immunity rule is shifted to permit more potential exposure to legal risk for platforms and websites' distribution of user content, those entities will overmoderate that content to avoid liability and preclude repeat threats. Overmoderation is by definition overinclusive, and results in platforms' "err[ing] on the side of caution and tak[ing] . . . down" any content that might give rise to liability.[81] The net result of overinclusive moderation is thus less speech online. More importantly for present purposes, existing evidence shows that the burden of overmoderation falls disproportionately on marginalized voices discussing controversial topics.

The current content moderation regime already demonstrates that "the viewpoints of communities of color, women, LGBTQ+ communities, and religious minorities are at risk of over-enforcement"[82]—especially, and

80. KOSSEFF, *supra* note 72, at 137.

81. Daphne Keller, *Toward a Clearer Conversation About Platform Liability*, KNIGHT FIRST AMEND. INST. (Apr. 6, 2018), https://knightcolumbia.org/content/toward-clearer-conversation-about-platform-liability; *see also* Enrique Armijo, *Reasonableness as Censorship: Section 230 Reform, Content Moderation, and the First Amendment*, 73 FLA. L. REV. 1199, 1217-18 (2021) ("In the absence of a mechanism by which all third-party content is screened prior to its posting (a virtual impossibility for Facebook, YouTube, or Twitter, at least), platforms will err on the side of removing any third-party speech that might be the basis . . . for legal liability.").

82. Angel Diaz & Laura Hecht-Felella, *Double Standards in Social Media Content Moderation*,

most troublingly, when members of those groups are discussing issues of race, gender, and discrimination.[83] As Daphne Keller has argued, "a growing body of evidence suggests" overinclusive removals "disproportionately harm [speech of] vulnerable or disfavored groups."[84] Additionally, Ari Ezra Waldman notes that removal systems that rely on flagging by other users at the screening level can over-censor "queer content by subjecting it to the heteronormative judgments" of majority groups, which also results in disproportionate moderation effects.[85] This uneven burden operates through the platforms' content moderation policies themselves as well as in the enforcement actions that platforms take in implementing those policies.[86] And automated content moderation systems, embedded with implicit biases or inadequate data sets in some of the countries in which the largest websites and platforms operate, "can put speech from marginalized communities at [further] risk of over-removal;"[87] evidence with respect to these systems already indicates that algorithmic content moderation "can have higher false positive rates for under-represented speakers than for white ones."[88] For example, a user sharing an instance of discrimination against them can (and often is) treated as a discriminator for purposes of moderating their speech.

Another reason overmoderation falls disproportionately on underrepresented groups is rooted in those groups' reasons for being disproportionately heavy users of new communication technologies in the

---

BRENNAN CTR. FOR JUST. (2021), https://www.brennancenter.org/our-work/research-reports/double-standards-social-media-content-moderation.

83. *See, e.g.*, Oliver L. Haimson et al., *Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas*, 5 PROC. ASS'N FOR COMPUTING MACH. ON HUM.-COMPUT. INTERACTION 1 (2021); Jessica Guynn, *Facebook While Black: Users Call it Getting "Zucked," Say Talking About Racism Is Censored Speech*, USA TODAY (Apr. 24, 2019), https://www.usatoday.com/story/news/2019/04/24/facebook-while-black-zucked-users-say-they-get-blocked-racism-discussion/2859593002.

84. Keller, *supra* note 81, at n.13 (citing several studies and examples).

85. Ari Ezra Waldman, *Disorderly Content*, 97 WASH. L. REV. 907, 909 (2022).

86. *See* Diaz & Hecht-Felella, *supra* note 82, at 5 (discussing how policies barring terrorism-related content "disproportionately target speech from Muslim and Arabic-speaking communities"); *id.* at 10 (detailing how Black activists posting examples of attacks targeting them have been subjected to Facebook enforcement actions); *id.* at 11 ("[R]esearchers found that models for automatic hate speech detection were 1.5 times more likely to flag tweets written by self-identified Black people as offensive or hateful, with tweets written using African American English 'more than twice as likely to be labeled as "offensive" or "abusive."'") (citing Maarten Sap et al., *The Risk of Racial Bias in Hate Speech Detection*, PROC. 57TH ANN. MEETING OF THE ASS'N FOR COMPUTATIONAL LINGUISTICS 1668, 1668-1678 (2019)).

87. *Id.* at 11; *see also* Matthew Ingram, *An Election in Cambodia Exposes Facebook's Shortcomings (Again)*, COLUM. JOURNALISM REV. (July 27, 2023), https://www.cjr.org/the_media_today/cambodia_hun_sen_facebook.php (arguing that Facebook's content moderation team in Cambodia "fail[ed] to take action to prevent violent threats").

88. Enrique Armijo, *Speech Regulation by Algorithm*, 30 WM. & MARY BILL RTS. J. 245, 257 & n.52 (2021); *see also generally* Anupam Chander, *The Racist Algorithm?*, 115 MICH. L. REV. 1023 (2017).

first place. As discussed in Section I above, minorities and underrepresented communities have turned to social media to "assume[] the role of news creators and distributors"[89] of information about their communities because they believe that the mainstream media has failed to fairly or adequately do so.[90] To the extent underrepresentation and bias occurs in the mainstream information ecosystem, it falls to members of the underrepresented groups to "set their own agendas" themselves.[91] It should therefore be unsurprising that both minority speakers and minority views, including speakers and views related to civil rights and societal injustice, are overrepresented in the user-generated content space of social media.

A recent categorical narrowing of Section 230's immunity demonstrates how distributors respond to an increase in immunity risk. In 2018, Congress jointly passed and President Trump signed into law the Allow States and Victims to Fight Online Sex Trafficking Act ("FOSTA"), which expanded existing federal sex trafficking crimes to include "knowingly assisting, supporting, or facilitating [sex trafficking]."[92] FOSTA also excluded state criminal prosecutions for sex trafficking and civil claims against websites from Section 230 under the "knowingly assisting, supporting, or facilitating" provision. Even before the bill was signed into law, the prospect of liability caused platforms and websites to begin overmoderating user content. Craigslist.com removed its "Personals" section entirely. In a message to its users, Craigslist said

89. DEAN FREELON ET AL., KNIGHT FOUND., HOW BLACK TWITTER AND OTHER SOCIAL MEDIA COMMUNITIES INTERACT WITH MAINSTREAM NEWS7 (Feb. 27, 2018), https://kf-site-production.s3.amazonaws.com/media_elements/files/000/000/136/original/TwitterMedia-final.pdf. "Black Twitter" is a catch-all term for a "collective of active, primarily African American Twitter users who have created a virtual community that participates in continuous real-time conversations" around "sociopolitical changes" as well as humor, popular culture-related commentary, and other aspects of Black identity. Feminista Jones, *Is Twitter the Underground Railroad of Activism?*, SALON (July 17, 2013), https://www.salon.com/2013/07/17/how_twitter_fuels_black_activism/.

90. And this perception is obviously driven by what minorities and underrepresented communities view as a lack of representation of their communities in the mainstream media itself. *See* FREELON ET AL., *supra* note 89, at 81-82 ("News outlets will never be able to tell the complete story about issues that matter to marginalized people without substantial representation on the masthead."). *See generally* ALLISSA V. RICHARDSON, BEARING WITNESS WHILE BLACK: AFRICAN AMERICANS, SMARTPHONES, AND THE NEW PROTEST #JOURNALISM (2020) (discussing African American use of social media and mobile technology to raise awareness about police brutality within the larger context of media witnessing theory, citizen journalism, and the documenting of anti-Black racism).

91. FREELON ET AL., *supra* note 89, at 9; *see also id.* at 13 ("Participants [in study of Black Twitter participants] use Twitter to circulate and raise awareness about these concerns on their own terms without waiting for professional journalists to take interest."); *id.* at 38 ("[Study respondents] described Black Twitter's development as a space in which Black people discuss issues of concern to themselves and their communities—issues they say either are not covered by mainstream media, or are not covered with the appropriate cultural context.").

92. 18 U.S.C. § 1591(e)(4); *see generally* Eric Goldman, *The Complicated Story of FOSTA and Section 230*, 17 FIRST AMEND. L. REV. 279 (2019).

that because FOSTA "subject[s] websites to criminal and civil liability when third parties (users) misuse online personals unlawfully," the law "jeopardize[ed]" all of the website's "other services," and so the Personals section was taken "offline."[93] Website infrastructure provider Cloudflare removed its services from Switter, a sex work-friendly social media network with nearly a half-million users.[94] The burden of overmoderation in response to FOSTA's shifting of the liability rule has also been borne most on sex workers, many of whom are of minority or intersectional identities, and who have been driven offline and into more physically dangerous methods of communication and commerce.[95] In the case of FOSTA, a shifting of the distributor liability rule has resulted in not just less speech for underrepresented groups, but less safety as well.[96]

The fact that speakers expressing minority viewpoints are both overrepresented and overmoderated in online speech spaces compounds the problems associated with a change in the existing liability rule that would permit greater distributor liability. Returning to the issue of police violence, consider, for example, an algorithm that a platform adopts that scans for streamed videos of gun violence in light of the prospect of increased liability for hosting such content.[97] Whether the algorithm took such videos down either *ex ante*, immediately after they were posted, or prevented their distribution as soon after their posting as possible, it is unlikely the automated process could distinguish between a massacre like the Christchurch murders and the police shootings of Philando Castile and Sean Reed.[98] In the absence of intermediary immunity for permitting user posts of those videos and in the presence of a possible collateral

93. *about > FOSTA*, CRAIGSLIST, https://www.craigslist.org/about/FOSTA; Aja Romano, *A New Law Intended to Curb Sex Trafficking Threatens the Future of the Internet as We Know It*, VOX (July 2, 2018), https://www.vox.com/culture/2018/4/13/17172762/fosta-sesta-backpage-230-internet-freedom.

94. Quinta Jurecic, *The Politics of Section 230 Reform: Learning from FOSTA's Mistakes*, BROOKINGS (Mar. 1, 2022), https://www.brookings.edu/articles/the-politics-of-section-230-reform-learning-from-fostas-mistakes/.

95. *See, e.g.*, Samantha Cole, *Trump Just Signed SESTA/FOSTA, a Law Sex Workers Say Will Literally Kill Them*, VICE (Apr. 11, 2018), https://www.vice.com/en/article/qvxeyq/trump-signed-fosta-sesta-into-law-sex-work; Charles Matula, *Any Safe Harbor in a Storm: SESTA-FOSTA and the Future of § 230 of the Communications Decency Act*, 18 DUKE L. & TECH. REV. 353, 358-59 (2020) (citing study on how electronic clearinghouse made female sex workers safer).

96. *See* Kendra Albert et al., *FOSTA in Legal Context*, 52 COLUM. HUM. RTS. L. REV. 1084, 1090 (2021) (noting that FOSTA's effects on speech and safety have been "most impactful on sex workers facing multiple forms of marginalization, including Black, [B]rown, and Indigenous workers, trans workers, and workers from lower socio-economic classes").

97. For a high-level description of algorithmic content moderation, see generally Renkai Ma & Yubo Kou, *How Advertising-Friendly Is My Video?: YouTuber's Socioeconomic Interactions with Algorithmic Content Moderation*, 5 PACM HUM.-COMPUT. INTERACTION 1, 5-6 (2021).

98. *See, e.g.*, Diaz & Hecht-Felella, *supra* note 82, at 11 ("[Automated review systems] may miss situations in which the image or video is being shared for purposes such as criticism, artistic expression, satire, or news reporting").

censorship threat for the content of the post, there is no reason for the hosting platform or website to develop a more tailored approach. This would result in a speech space much more hostile to marginalized groups' attempts to attract attention to underreported issues and minority views.

Prescreening and overmoderation in the absence of immunity would take place at the topic level as well. And more platform censorship would result in less speech on those topics overall. For example, take speech about conspiracy theories such as QAnon or Pizzagate, or reporting about terrorism. Mainstream news reports often rely on extremist groups' own online statements to understand and report threats to the public.[99] And academics with legitimate research interests in conspiracy theorists might be less able to find postings that would help explain or expose those phenomena to broader public view. If platforms or websites were subjected to effective legal threats for carrying such statements, the net result of overmoderation would be less speech about topics that are already underdiscussed because of their status outside the mainstream. These problems would be exacerbated by algorithmically-based overmoderation as well; automated capture and removal of potentially offending content would have even more trouble in distinguishing between, for example, content by terrorists and content *about* terrorism[100]—a problem that has already arisen even under the current immunity rule.[101]

99. *See* Brief of the Reporters Committee for Freedom of the Press and the Media Law Resource Center as Amici Curiae Supporting Respondents at 17, Gonzalez v. Google, 143 S. Ct. 1191 (2023) (No. 21-1333) [hereinafter Brief of Reporters Committee]. The information-based costs of overmoderation would be particularly felt with respect to breaking news on these topics, where "journalists routinely rely on first-person accounts of unfolding events to understand what is happening on the ground and document the first few minutes before reporters can arrive on the scene." *See id.* at 11.

100. *See, e.g.*, Diaz & Hecht-Felella, *supra* note 82, at 5 (discussing how platforms' takedown reports under the Global Internet Forum to Counter Terrorism relies on "expansive categories like 'support' and 'glorification,'" which results in "a wide range of political speech and human rights documentation [] inevitably [being] swept up in a removal dragnet").

101. *See, e.g.*, Martin Belam, *Twitter Under Fire After Suspending Egyptian Journalist Wael Abbas*, GUARDIAN (Dec. 18, 2017), https://www.theguardian.com/media/2017/dec/18/twitter-faces-backlash-after-suspending-egyptian-journalist-wael-abbas (suspending account of prominent Egyptian journalist who posted civil rights-related abuses by the Egyptian government); Kevin Anderson, *YouTube Suspends Egyptian Blog Activist's Account*, GUARDIAN (Nov. 28, 2007), https://www.theguardian.com/news/blog/2007/nov/28/youtubesuspendsegyptianblog (suspending account of same activist after his posting of videos showing police beatings and sexual assaults); *see also* Brief of Reporters Committee, *supra* note 99, at 9 ("It is already difficult for platforms to distinguish between a terrorist recruiting video and a news report on that video when moderating content at scale, a dynamic that leads to the removal of essential journalism, . . . [and] undermining 230's protections would exacerbate the dynamic.").

B. Why User-based Speech Protections are not Enough

The effectiveness of possible collateral censorship threats against online distributors in the absence of Section 230 immunity must also take into account the body of speech-protective First Amendment and other law that would shield many, if not most, of the civil rights-related user speech discussed here.

For example, the constitutional right to record police activity in public—now recognized by multiple federal courts of appeal[102]—would be available to distributors to assert on users' behalves in response to collateral threats. And several of these decisions explicitly rely on a First Amendment-derived right to *receive* information about the police's furtherance of their duties; to the extent they do, their rationales would provide support for a distributor's stand-alone right to disseminate or livestream user recordings as well, since it cannot be the case that the right to disseminate is only derivative of the user's right to record and thus belongs to them alone.[103] Distributor-based liability under wiretap statutes only extends to information that it is unlawful to collect in the first instance;[104] given most statutes' requirement that the recorded person expected their communications to not be recorded, a user's recording of an officer in public therefore likely could not form the basis for a distribution-based claim against a website or platform. In other words, if a user's recording has not violated a wiretap statute, then a distributor's posting of that video cannot either.[105] Here, too, the First Amendment also

102. *See, e.g.*, Sharpe v. Winterville Police Dep't 59 F.4th 674, 680-81 (4th Cir. 2023); Ness v. City of Bloomington, 11 F.4th 914, 923 (8th Cir. 2021); Fields v. City of Philadelphia, 862 F.3d 353, 359 (3d Cir. 2017); Turner v. Lieutenant Driver, 848 F.3d 678, 688 (5th Cir. 2017); ACLU v. Alvarez, 679 F.3d 583, 595 (7th Cir. 2012); Glik v. Cunniffe, 655 F.3d 78, 88 (1st Cir. 2011); Smith v. City of Cumming, 212 F.3d 1332, 1333 (11th Cir. 2000); Fordyce v. City of Seattle, 55 F.3d 436, 439 (9th Cir. 1995). Lower courts have been consistent in holding the same. *See, e.g.*, Arizona Broads. Ass'n v. Brnovich, 626 F. Supp. 3d 1102 (D. Ariz. 2022).

103. *See Ness*, 11 F.4th at 923 ("The acts of taking photographs and recording videos are entitled to First Amendment protection because they are an important stage of the speech process that ends with the dissemination of information about a public controversy."); *Alvarez*, 679 F.3d at 595 ("The act of *making* an audio or audiovisual recording is necessarily included within the First Amendment's guarantee of speech and press rights as a corollary of the right to disseminate the resulting recording.") (emphasis in original); *Glik*, 655 F.3d at 82 ("Gathering information about government officials in a form that can readily be disseminated to others serves a cardinal First Amendment interest in protecting and promoting 'the free discussion of governmental affairs.'"); *Fields*, 862 F.3d at 358-59 ("[W]e are now in an age where the public can record our public officials' conduct and easily distribute that recording widely. . . . [r]ecordings also facilitate discussion because of the ease in which they can be widely distributed via different forms of media.").

104. *See, e.g.*, 18 U.S.C. § 2510(2) (defining "oral communication" under the Wiretap Act as "any oral communication uttered by a person exhibiting an expectation that such communication is not subject to interception under circumstances justifying such expectation").

105. For federal law purposes, a party seeking to hold a distributor liable would have to show that

limits the statute's application to distributors' liability where the information is lawfully acquired and the information in the recording is of public concern.[106] Other speech-protective decisions would also support the constitutional right to make and disseminate recordings of official activity, such as the Supreme Court's recognition in *Snyder v. Phelps* that the First Amendment precludes state law tort claims for speech on public streets regarding issues of public concern.[107]

The same is true for other categories of protected user speech. As discussed above, a threatener of a defamation suit who is a public figure would have to show actual malice by clear and convincing evidence not just on the part of the speaker, but on that of the distributor as well. Punishing platforms and other hosts under existing incitement-to-riot statutes for users' organizing online protest activity would run headlong into imminence-related challenges.[108] And the Supreme Court's recent rejection of aiding-and-abetting liability for Twitter under the Antiterrorism Act shows that merely hosting harmful user content is likely insufficient assistance for a platform to be held civilly or criminally liable for any harms that the content might cause.[109]

But the fact that these substantive defenses are likely available, and would likely defeat any legal process brought by a party following through on a collateral threat, does not necessarily mean that a distributor would assert them in the absence of immunity. In situations where a user's First Amendment right is clearly established, a distributor might do so, but in the face of ambiguity, the distributor is likely to remove the content. And in the analogous context of qualified immunity, even the issue of whether a First Amendment right to record is clearly established can be contentious, and a process in which the threatener's incentive is to define the right as narrowly as possible to demonstrate it is not. Take the example of *Sharpe v. Winterville Police Department*, the most recent federal court of appeals case to find a First Amendment right to record.[110] There, the Fourth Circuit defined the right at issue for both First Amendment and qualified immunity purposes as not a broad right to record the police in

the distributor knew "sufficient facts concerning the circumstances of the interception such that the defendant [distributor] could, with presumed knowledge of the law, determine that the interception was prohibited in light of Title III [of the Act]." Thompson v. Dulaney, 970 F.2d 744, 749 (10th Cir. 1992).

106. Bartnicki v. Vopper, 532 U.S. 514, 533-34 (2001).

107. 562 U.S. 443, 459 (2011).

108. Armijo, *supra* note 88, at 1233-34 (explaining that First Amendment analysis of imminence's incitement requirement changes "when the speech the government seeks to punish or proscribe is not heard by a gathered mob but read on a screen by individuals making up a geographically diffuse audience").

109. Twitter, Inc. v. Taamneh, 143 S. Ct. 1206, 1230-31 (2023).

110. 59 F.4th 674, 681 (4th Cir. 2023).

public, but rather to "livestream their own traffic stop."[111] Other cases the plaintiff used to attempt to clearly establish such a right were unavailing, according to the court, because they were about "video recordings, not livestreams" of police, and "the people doing the recording tend[ed] to be bystanders, not the subjects of the stop itself."[112] These distinctions would bear on not just whether the right demonstrated by the plaintiff was clearly established, but also the constitutionality of the police defendants' actions, since according to the court "[a] different balance is struck when an officer prevents a bystander from recording someone else's traffic stop than when the officer prevents a passenger from livestreaming their own stop."[113] And despite the emergence of a First Amendment right to record, state courts have upheld wiretap-based prosecutions of citizens who have recorded police officers during traffic stops where the officers claimed to not know they were being recorded.[114] Where the contours of the user's right are uncertain, the distributor is unlikely to take the lead in litigating them.[115]

In a real sense, however, discussing the merits of possible claims against distributors belies the point of their immunity in the first place. The threat to the distributor is likely sufficient to overcome potential litigation over the possible applicability of these rights because the distributor lacks incentives to assert any First Amendment or other speech-based rights enjoyed by the user on the user's behalf in any instance where the validity of the user's defense is uncertain. The same is true for any First Amendment or other protective right to distribute. Civil rights or protest-related speech is generally not a remunerative category of speech for the platform, such that the platform will incur costs to defend keeping it up in the face of a threat, absent some business goodwill-related interest that allows it to be seen as a defender of controversial user content—and an internet company's calculations on those issues can turn on a dime.[116] So the kinds of speech that are most

111. *Id.* at 683.

112. *Id.* at 683-84.

113. *Id.* at 684. The court noted the police department's arguments that a right to livestream one's own traffic stop might survive constitutional scrutiny since "livestreaming a traffic stop endangers officers because viewers can locate the officers and intervene in the encounter," and that "violence against police officers has been increasing—including planned violence that uses new technologies." *Id.* at 682.

114. *See* Commonwealth v. Hyde, 750 N.E.2d 963, 966-68 (Mass. 2001).

115. Eric Goldman, *Why Section 230 Is Better Than the First Amendment*, 95 NOTRE DAME L. REV. REFLECTION 33, 42 (2019) ("Unlike Section 230, constitutional litigation is rarely quick or cheap," and unlike immunity defenses, which are usually litigated on motions to dismiss, "constitutional doctrines often raise sufficient factual questions that courts wait until summary judgment (or later) before disposing of an unmeritorious case.").

116. *Id.* at 41; *see, e.g.*, Anne D'Innocenzio, *Target Becomes Latest Company to Suffer Backlash for LGBTQ+ Support, Pulls Some Pride Month Clothing*, ASSOCIATED PRESS (May 24, 2023), https://apnews.com/article/target-pride-lgbtq-4bc9de6339f86748bcb8a453d7b9acf0; Lydia Polgreen,

likely to be targeted and cause the most public controversy also may be the kinds that the distributors would be least likely to defend in the absence of immunity.

IV. SECTION 230 AS FACILITATOR OF CIVIL RIGHTS SPEECH SUPPRESSION

The above Sections make the affirmative case that Section 230's immunity for distributors results in more speech on issues of public concern than would be the case in a world where the current rule was shifted to permit greater potential liability. However, some commentators and lawmakers have made the opposite claim: that by protecting distributors from liability for user speech that harasses underrepresented speakers, and for their moderation decisions with respect to user speech, the statute actually results in less speech online. Revoking either or both of these immunities, so goes this argument, would result in more online speech about civil rights. But these arguments fail to consider two important results that would arise from immunity revocation. First, making distributors liable for online abuse would necessarily make collateral threats of all types more effective, not just threats for failure to protect users from harassment by other users. Second, forcing distributors to carry user speech with which they disagree violates another critical speech-based civil right, namely, the associational rights of the distributors themselves.

A. "Good" Collateral Censorship: Reducing Harassment of Speakers Expressing Minority Views

While Section 230 reduces the effectiveness of collateral censorship-based threats against underrepresented speakers and by extension underrepresented speech, an argument is also made that it enables the silencing of those views as well. Mobs and individuals, usually acting anonymously, often engage in systemic and harmful abuse and harassment online, including stalking, extortion, threats of physical violence, and sexual harassment.[117] Danielle Citron and Mary Anne

The Puny Power of 'Woke Capitalism,' N.Y. TIMES (July 12, 2023), https://www.nytimes.com/2023/07/12/opinion/woke-capitalism-diversity-lgbt-companies-pride.html ("For those on the left who take comfort in seeing big companies take bold stands on issues they care about, I'm here to tell you that those companies care much more about their bottom line than your beloved issue[.]").

117.

> Online, bigots can aggregate their efforts even when they have insufficient numbers in any one location to form a conventional hate group. They can aggregate their offline identities from their online presence, escaping social opprobrium and legal liability for destructive

Franks have shown that women of color and sexual minorities are the most frequent targets of online abuse.[118] This abuse results in emotional distress and other real offline harms, but in the context of speech it often results in exit: victims "shut down social media profiles and email accounts and withdraw from public discourse," including on "controversial topics."[119] Other parts of the online communications ecosystem demonstrate how underrepresented voices are often silenced through manipulation of existing processes; in the analogous case of notice-and-takedown requests under the Digital Millennium Copyright Act, there is some evidence of disproportionate takedown requests filed against minority creators and content entrepreneurs.[120]

According to these commentators, Section 230's immunity plays a significant, and negative, role in enabling this conduct. By immunizing harassment, Section 230 empowers threats toward minority and underrepresented speakers by other users. This activity results in less speech because the threatened speakers choose to exit. In the absence of the liability rule, collateral censorship would occur, but it would, on balance, be better for civil rights-related speech online, because much of the speech the platforms would censor in the presence of potential liability would be that of users who threaten and harass underrepresented speakers such as minorities, women, or others who express minority points of view. Reducing that harassment via collateral platform threats would result in more speech. The immunity, claim Citron and Franks, "takes away the leverage" that the harassed user might have to force the platform to act to reduce harassment.[121] One example where this loss of leverage is felt is with respect to victims' efforts to take down "revenge pornography," or the nonconsensual online posting of sexually graphic images or videos; in the absence of Section 230 immunity, a website or platform could be effectively threatened with liability for hosting revenge porn, or for failing

> acts. Both of these qualities are crucial for the growth of anonymous online mobs that attack women, people of color, religious minorities, gays, and lesbians.
>
> . . . .
>
> Some victims respond by . . . going offline. [This harassment] work[s] to suppress the speech [of members of those groups.]

See Citron, *supra* note 76, at 63-67.

118. Danielle Keats Citron & Mary Anne Franks, *The Internet as Speech Machine and Other Myths Confounding Section 230 Reform*, 2020 U. CHI. LEGAL F. 45, 54 (citing, *inter alia*, Maeve Duggan, Online Harassment 2017 Study, PEW RES. CTR. (July 11, 2017), https://www.pewresearch.org/internet/2017/07/11/online-harassment-2017/).

119. *Id.* at 55.

120. *See, e.g.*, Jonathon Penney, *Privacy and Legal Automation: The DMCA as a Case Study*, 22 STAN. TECH. L. REV. 412, 470-71 (2019) (showing disproportionate chilling effects against women in the DMCA context).

121. Citron & Franks, *supra* note 118, at 53.

to take such content down after having received notice from the victim.[122]

To be sure, in theory all the aforementioned overmoderation incentives that the absence of immunity would exacerbate apply with equal force to harassment claims. In the face of a threat of prospective liability for failing to take down or remove harassing content, the platform will comply with the threatener's wishes. But in the anti-Section 230 case, where the party threatening the platform is the subject of harassment, the user speech being removed from circulation by the distributor as a result of the threat is less deserving of protection. Reducing harassing content results in more speech from speakers who might otherwise be driven off the platform. Since, as discussed above, those speakers who are now more likely to remain are often from underrepresented communities and expressing minority or anti-status quo points of view, the net result of this kind of collateral censorship is more civil rights-related speech online. Section 230 thus does not protect speech about civil rights, but rather permits that speech's suppression. This argument, however, only considers one kind of collateral threat.

In both the pro- and anti-immunity cases, in the absence of immunity a platform is more likely to take down user speech in response to a claim that the speech offends the threatener's rights or otherwise causes the threatener harm. But when predicting the effects of a potential threat, the status of the threatener matters. It is likely that in a distributor liability world where collateral threats are possible, platforms and other distributors will be more responsive to collateral threats made by better-resourced parties, on the grounds they are more likely to follow through on their threats with legal process. The history of anti-SLAPP statutes, which arose in the context of real estate developers threatening community residents who were expressing opposition to development projects with frivolous litigation, supports this prediction.[123] In the absence of immunity, a government official or other powerful actor's takedown demand is likely to be more successful than a user's harassment-based claim. If this is true, then preserving the immunity likely results in more user speech than removing the immunity, even if

122. Since Section 230 does not preempt federal criminal law, Congress could effectively pass a law criminalizing revenge pornography distribution without carving out Section 230 immunity. For the reasons discussed in this Article, doing so would require narrow definitions of what constitutes revenge porn and a notice requirement to platforms or websites before imposing distributor-based liability. For an example of such an effort whose notice-and-takedown requirements are modeled after the Digital Millenium Copyright Act, see Allison Tungate, *Bare Necessities: The Argument for a 'Revenge Porn' Exception in Section 230 Immunity*, 23 INFO. & COMMC'NS TECH. L. 172 (2014).

123. The acronym "SLAPP" stands for "strategic lawsuits against public participation." *See, e.g.*, Paul D. Wilson, *Of Sexy Phone Calls and Well-Aimed Golf Balls: Anti-SLAPP Statutes in Recent Land-Use Damages Litigation*, 36 URB. LAW. 375, 375 (2004) (responding to "real estate developers, upset about position to their projects, [] cowing project opponents into submission by filing frivolous lawsuits against them," state legislatures enacted anti-SLAAP statutes).

the victims of online harassment are disproportionately from underrepresented communities. This is so because there is no way to permit the harassment-based threats, which Citron and Franks predict will have normatively positive consequences for speech because of those who are more likely to suffer harassment, without also enabling threats from those subjects of speech that are more powerful than the user. Defining "harassing speech" precisely enough in order to carve that speech out from Section 230 immunity would raise significant vagueness and overmoderation concerns.[124] And if immunity has been revoked altogether over the concerned voices of harassed speakers, speakers who might have remained due to a reduction of harassment would still be subjected to the same potential collateral censorship discussed in Sections II and III above—including threats to the distributor from the harassers themselves.

B. Content Moderation as Civil Rights Speech-Related Censorship

A sub-part of Section 230, 230(c)(2), grants immunity to distributors for moderating user content so long as those decisions are made in good faith.[125] This immunity permits platforms to take down or deprioritize user speech, including speech that platforms do not like, and much of which would be considered civil rights-related speech by those engaged in it, without fear of creating legal liability when doing so; some recent COVID-19-related examples include speakers opposed to government-imposed lockdowns or claims of vaccine inefficacy.[126] For those who believe platforms exercise too much discretion in deciding what user content to remove, this provision is additional evidence that the statute diminishes rather than promotes civil rights-related speech online—namely by giving platforms the unlimited right to block speech with which they disagree. The Section 230-as-anti-civil-rights claim here is that the statute immunizes distributors from engaging in viewpoint discrimination, which makes nonconformist speech a target for content

124. *See, e.g.*, People v. Golb, 15 N.E.3d 805 (N.Y. 2014) (finding New York's aggravated harassment statute, which criminalized harassment for "communicat[ing] with a person . . . by telephone, by telegraph, or by mail, or by transmitting or [by] delivering any other form of written communication, in a manner likely to cause annoyance or alarm," unconstitutionally vague and overbroad).

125. *See* 47 U.S.C. § 230(c)(2) (stating that there is no liability for "any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected").

126. *See* Citron & Franks, *supra* note 118, at 46-47 ("Conservatives claim that Section 230 gives tech companies a license to silence speech based on viewpoint." (citation omitted)); Ethan Yang, *Reddit's Censorship of the Great Barrington Declaration*, AM. INST. FOR ECON. RSCH. (Oct. 8, 2020), https://www.aier.org/article/reddits-censorship-of-the-great-barrington-declaration/.

moderation. And all "nonconformist" means in this context is the speech that the platform or other distributor takes down because it does not like it, as contrasted with the "conformist" speech that the distributor leaves up because it does.

In the absence of federal-level consensus on how Section 230 should be reformed, some states have used this theory as an opportunity to step in. In 2021, Texas passed House Bill 20, which addressed "censorship of or certain other interference with digital expression, including expression on social media platforms." The law aims to prevent platforms from "censor[ing]"—i.e., "block[ing], ban[ning], remov[ing], deplatform[ing], restrict[ing] . . . [or] deny[ing] equal access or visibility to" "a user, a user's expression, or a user's ability to receive the expression of another person based on the viewpoint of the user . . . [or] the viewpoint represented in the user's expression."[127] Similarly, the same year, Florida passed Senate Bill 702, which states that a social media platform may not "delete, regulate, restrict, edit, alter, inhibit the publication or republication of, suspend a right to post, remove, or post an addendum to any content or material posted by a user."[128] To the supporting lawmakers, these laws, not Section 230(c)(2), support civil rights-related speech—they require platforms to carry underrepresented speakers and speech that the platform disagrees with. On appeal, the Texas statute was upheld, the Florida statute was invalidated, and the U.S. Supreme Court has granted certiorari to resolve the circuit split.[129]

As an initial matter, it is likely that Section 230(c)(2) will preempt the content moderation-related provisions of both state laws. As the district court reviewing the Florida law found, Section 230's federal preemption provision directs that "no liability may be imposed under any State law or local law that is inconsistent with this section," and any part of the Florida statute "applicable to a social media platform's restriction of access to posted material" would be inconsistent with Section 230(c)(2)'s grant of immunity for those same decisions.[130] More directly, however, the First Amendment's compelled speech doctrine protects private distributors' rights to disassociate with speakers and speech with which they disagree. Forcing such associations violates the civil rights of the distributor. As the Court explained most recently in *303 Creative LLC v. Elenis*, "no

127. TEX. CIV. PRAC. & REM. CODE § 143A.001(1), .002 (West 2023).

128. FLA. STAT. § 501.2041(1)(b) (2023).

129. NetChoice, LLC v. Moody, 546 F. Supp. 3d 1082, *cert. granted*, No. 22-277 (Sept. 29, 2023).

130. NetChoice, LLC v. Moody, 546 F. Supp. 3d 1082, 1089-90 (N.D. Fla. 2021) (citing 47 U.S.C. § 230(e)(3)).

government may affect a 'speaker's own message' by 'forc[ing]' her to 'accommodate' views she does not hold."[131]

Elenis confirms that the First Amendment's civil rights-based guarantee of freedom from government-compelled speech trumps a state government's view of the types of speech that should and should not be disseminated. To be sure, there are arguments to be made as to whether distributors of online speech should not discriminate against ideas with which they disagree. And platforms themselves for the most part adamantly deny that they do.[132] But importantly, the First Amendment prevents governments from doing anything about it if they decide to do so.

CONCLUSION

In Section 230's statement of purpose, Congress stated that the statute's immunity rules were intended to help protect the internet as "a forum for a true diversity of political discourse."[133] Understanding collateral censorship—and in particular how it would operate with respect to minority points of view on the internet in the absence of immunity—demonstrates how the statute achieves this goal.

131. 143 S. Ct. 2298, 2319 (2023) (citing Rumsfeld v. F. for Acad. & Institutional Rts., Inc., 547 U.S. 47, 61-64 (2006)).

132. *See, e.g.*, Adam Gabbatt, *Claims of Anti-Conservative Bias by Social Media Firms Is Baseless, Report Finds*, GUARDIAN (Feb. 1, 2021), https://www.theguardian.com/media/2021/feb/01/facebook-youtube-twitter-anti-conservative-claims-baseless-report-finds.

133. 47 U.S.C. § 230(a)(3).