EMPOWERING SPEECH BY MODERATING IT

Danielle Keats Citron[*] *& Jonathan Penney*[**]

Content moderation is typically viewed as an affront to free expression. When companies remove online abuse, they face accusations of censorship. Lost in the discussion is the fact that victims of intimate privacy violations and cyberstalking typically—and regrettably—withdraw from on- and offline activities. Online assaults chase targeted individuals offline; they silence victims. Content moderation can secure opportunities for people to speak. Legal and corporate prohibitions against intimate privacy violations and cyberstalking can help provide the reassurance that victims need to stay online. They can endow individuals with a sense of trust so they continue to use networked technologies to express themselves. Those prohibitions are consonant with First Amendment doctrine and free

---

[*] Danielle Keats Citron, a Member of the American Academy of Arts and Sciences since 2023, is the Jefferson Scholars Foundation Schenck Distinguished Professor in Law at the University of Virginia. She has been a member of Facebook's Non-Consensual Intimate Imagery Task Force since 2011, served as a member of Twitter's Trust and Safety Task Force from 2016 to 2021, and has been a member of Spotify's Trust and Safety Council since its inception in 2021. Citron serves as the Vice President of the Cyber Civil Rights Initiative, an organization devoted to fighting for civil rights and liberties in the digital age. She is the author of *The Fight for Privacy: Protecting Dignity, Identity and Love in the Digital Age* (2022) and *Hate Crimes in Cyberspace* (2014). She was named a MacArthur Fellow in 2019. This article was originally published in August 2024 in Volume 153, Issue 3 of *Daedalus*, the Journal of the American Academy of Arts and Sciences. The citation format has been largely preserved from that publication.

[**] Jonathon Penney is a legal scholar and social scientist based at Osgoode Hall Law School at York University in Toronto, a Faculty Associate of Harvard's Berkman Klein Center for Internet & Society, and a Research Fellow at the Citizen Lab based at the University of Toronto's Munk School of Global Affairs and Public Policy. His research on law, technology, and human rights has been published in leading U.S. law reviews and social science journals and has received international press coverage. He and Danielle Keats Citron are working together to explore how laws and other measures taken to battle online abuse can empower the online speech and engagement of victims.

speech values. Combating online abuse isn't a zero-sum game with free speech as the loser. Rather, it can free us to speak by changing the culture that rewards abuse and encourages self-censorship.

Introduction

A myth of epic proportion has gained traction: that any effort to moderate online speech is a zero-sum game, with free expression as the loser. When social media companies remove destructive posts that violate terms of service, people cry, "Censorship!" Alex Jones, founder of the far-right conspiracy news site Infowars, accused YouTube of "killing the First Amendment" after the company blocked videos that revealed maps of the homes of Sandy Hook families.[1] This isn't just an extremist view: the Pew Research Center has found that a majority of people believe that companies are engaged in "political censorship" when they moderate content.[2] Some legislators have made this view a cornerstone of their political philosophy. At a House Oversight and Accountability Committee hearing in February 2023, Representative Lauren Boebert denounced Twitter as a "speech overlord." To the company's former head of Trust and Safety, Yoel Roth, she angrily admonished, "How dare you" shadow-ban my posts (even though no evidence supported the claim and former Twitter executives denied it). Representative Marjorie Taylor Greene stated that Big Tech was silencing Americans.[3] The censorship narrative has gained traction in state legislatures as well. Underlying this view is the assumption that content moderation has no upside for free expression.

The outcry is similarly strident at the suggestion that law should curtail online abuse. Online assaults that include doxing, intimate privacy violations, and threats are dismissed as weak attempts to "blow off steam." Any effort to address them is viewed as a threat to free speech. The ACLU, for instance, has adamantly opposed

---

[1] Matt Taibbi, "Beware the Slippery Slope of Facebook Censorship," *Rolling Stone*, August 2, 2018, https://www.rollingstone.com/politics/politics-features/facebook-censor-alex-jones-705766.

[2] Jessica Guynn, "'They Want to Take Your Speech Away,' Censorship Cry Unites Trump Supporters and Extremists after Capitol Attack," *USA Today*, January 15, 2021, https://www.usatoday.com/story/tech/2021/01/15/censorship-trump-extremists-facebook-twitter-social-media-capitol-riot/4178737001.

[3] David Edwards, "Lauren Boebert Furious over Twitter Censoring Her Account," *Salon*, February 9, 2023, https://www.salon.com/2023/02/09/lauren-boebert-furious-over-twitter-censoring-her-account_partner.

the passage of laws penalizing the nonconsensual disclosure of intimate images. These laws risk chilling legitimate expression, the ACLU has argued, even though the laws made clear that they would not cover matters of legitimate public interest. Under law's blighting stare, free expression is impossible.[4]

For more than a decade, we have been interrogating these claims. Rather than vanquishing free expression, combating online abuse frees people to speak. In the face of online assaults that amount to cyberstalking or intimate privacy violations, targeted individuals stop expressing themselves. They close their social media accounts, lest perpetrators exploit those accounts to attack them. They withdraw from family and friends. If their loved ones try to "talk back" to abusers, they face terrifying online assaults themselves. Victims and their loved ones are silenced and terrorized. Research makes clear that online abuse exacts significant costs to free expression.

As our research suggests, legal and industry interventions against such abuse make space for more expression rather than less. Such interventions enable victims to speak their truths. Rather than silencing speech that deserves normative protection, law and corporate policies enable victims to trust companies enabling communications so they can reveal themselves and share their truths.

I.

Legislators aren't just talking about the "censorship" of social media companies—they are doing something about it. Florida has prohibited big tech companies from removing, filtering, or downgrading journalists' speech, while Texas has barred them from moderating any user-generated content based on viewpoint, with some narrow exceptions. Under the Texas law, a "social media platform may not censor a user, a user's expression, or a user's ability to receive the expression of another person based on: (1) the viewpoint of the user or another person; (2) the viewpoint represented in the user's expression or another person's expression; or (3) a user's geographic location in this state or any part of this state."[5]

---

[4] Danielle Keats Citron, *Hate Crimes in Cyberspace* (Cambridge, Mass.: Harvard University Press, 2014), 190–193; and Mary Anne Franks, *The Cult of the Constitution* (Redwood City, Calif.: Stanford University Press, 2019), 160–198.

[5] Texas Civil Practice and Remedies Code § 143A.002(a).

The Fifth Circuit upheld the Texas law, finding that social media companies are public utilities and must take all comers. The court vacated a preliminary injunction of the bill, enabling it to go into effect, on the grounds that the law does not chill speech but rather chills censorship. The court underscored that social media companies failed to "mount any challenge under the original public meaning of the First Amendment."[6]

The Fifth Circuit baldly and incorrectly asserted that content platforms "exercise no editorial control or judgment." Having worked with social media companies for more than a decade, reviewing their internal speech rules, we have learned that these companies actively moderate online content, banning, filtering, highlighting, and prioritizing all sorts of speech, including proscribable speech like cyberstalking, terroristic threats, and nonconsensual intimate images, as well as protected expression like hate speech, misinformation, and disinformation. Social media companies are unlike telephone companies and telephone providers, which perform no role in deciding who may use their services. Social media companies are more analogous to newspapers, bookstores, or entertainment companies that enjoy First Amendment protections as speakers in their own right.

The Florida law met a decidedly different fate: the Eleventh Circuit upheld the preliminary injunction, finding that the Florida law was not likely to survive First Amendment review.[7] The court held that the Florida law's restrictions on a social media company's ability to moderate content triggered First Amendment scrutiny. The court highlighted decisions protecting the editorial discretion of publishers and media companies, noting that when social media companies remove or de-prioritize user-generated posts, they are making a judgment about the value of such content. The court found that the statute was unlikely to survive "intermediate—let alone strict—scrutiny" because a state has no legitimate interest in counteracting private speech decisions "by tilting the public debate in a preferred direction."[8]

---

[6] *NetChoice, LLC v. Paxton*, 49 F.4th 439, 445 (5th Cir. 2022). The panel defined "censor" to mean "to block, ban, remove, deplatform, demonetize, de-boost, restrict, deny equal access or visibility to, or otherwise discriminate against expression." Ibid., 446 (citing Texas Civil Practice and Remedies Code § 143A.001[1]).

[7] *NetChoice, LLC v. Attorney General, State of Florida*, 34 F.4th 1196, 1203 (11th Cir. 2022).

[8] 34 F.4th 1196, 1209, 1216 (11th Cir. 2022).

In *Moody v. Netchoice*, the Supreme Court endorsed the notion that a social media company's content-moderation decisions constitute speech that implicates the First Amendment. While vacating the Fifth and Eleventh Circuit decisions on grounds unrelated to the First Amendment merits, the Court provided guidance on the First Amendment question. The Court explained that deciding whether third-party speech will be included or excluded, pursuant to a social media company's terms of service, amounts to editorial choices protected by the First Amendment and that "[h]owever imperfect the private marketplace of ideas," it is far worse to have the government decide when speech is imbalanced and "coerc[e] speakers to provide more of some views or less that others."[9]

That strikes us as right. A private party's ability to block or filter someone else's constitutionally protected speech is part of the First Amendment tradition. Under that tradition, unlike the government, whose laws should not favor certain ideas or speakers over others, private parties are permitted, even expected, to shape norms around speech activity.[10] Generally speaking, the "government can't tell a private party or entity what to say or how to say it."[11] The government should not be in the business of telling social media companies what kinds of speech it must affiliate with (or not affiliate with).

Beyond the doctrinal point, the larger normative point remains: social media sites should be allowed to make choices about online content. They should be free to moderate their users' activities to match their priorities. They should be permitted to ban cyberstalking, threats, doxing, and nonconsensual pornography. The good of free expression, in fact, depends on their doing so.

II.

Every day, people—more often, marginalized people—face online abuse that makes it impossible for them to speak.[12] Online abuse may involve cyberstalking:

---

[9] *Moody v. Netchoice, LLC*, 603 U.S. ___ (2024).

[10] Frederick F. Schauer, Hudgens v. NLRB *and the Problem of State Action in First Amendment Adjudication*, 61 MINN. L. REV. 433, 448–49 (1977); and Frederick F. Schauer, "The Ontology of Censorship," in *Censorship and Silencing: Practices of Cultural Regulation*, ed. Robert C. Post (Los Angeles: Getty Research Institute, 1998), 147, 160–164.

[11] *NetChoice, LLC v. Attorney General*, 34 F.4th at 1203.

[12] Danielle Keats Citron, *Sexual Privacy*, 128 YALE L.J. 1870, 1905–21 (2019); and Asia A. Eaton, Holly R. Jacobs, and Yanet Ruvalcaba, *2017 Nationwide Online Study of Nonconsensual Porn Victimization and Perpetration* (Miami: Cyber Civil Rights Initiative, 2017). According to a 2017 survey

repeated targeting of specific individuals with defamatory lies, threats, and privacy violations. Lies accuse victims of being prostitutes or having sexually transmitted infections; threats invoke sexual violence; privacy invasions include doxing. When victims appear to be people of color or LGBTQIA+, the abuse is suffused with racist, homophobic, and transphobic invective. Online abuse also includes intimate privacy violations, such as the nonconsensual recording and sharing of someone's intimate images.[13]

Consider the cyberstalking campaign faced by Nina Jankowicz, a researcher specializing in state-sponsored disinformation. In April 2021, the Biden administration tapped Jankowicz to lead a new group in the Department of Homeland Security (DHS) called the Disinformation Governance Board. The board would coordinate DHS efforts to highlight trustworthy information about high-stakes issues like COVID-19 response measures and cybersecurity events. Within twenty-four hours of the board's announcement, prominent far-right media outlets and influencers attacked Jankowicz as a threat to democracy whose work would inevitably distort the truth and censor free speech.

Jankowicz faced ferocious, threatening, and destructive online abuse. Posters accused her of spreading disinformation, rather than combating it (which she had done throughout her career and would have continued to do at DHS). Videos were doctored to make it seem that she thought certain people should be able to edit others' tweets, which she had never said. Detractors began circulating her contact information online. Jankowicz received frightening emails, texts, voicemails, and letters that threatened rape and death. At the time, Jankowicz was nine months pregnant with her first child.

---

conducted by Australia's e-Safety Commissioner, women were twice as likely to be victims of nonconsensual disclosure of intimate images, and Indigenous Australians were twice as likely to have experienced the abuse of their intimate images than non-Indigenous Australians. Nicola Henry, Clare McGlynn, Asher Flynn, et al., *Image-Based Sexual Abuse: A Study on the Causes and Consequences of Non-Consensual Nude or Sexual Imagery* (London: Routledge, 2020), 35–36. Of the fifteen thousand deepfake videos posted online in 2019, about 95 percent inserted women's faces into porn. Henry Ajder, Giorgio Patriani, Francesco Cavalli, and Lauren Cullen, *The State of Deepfakes: Landscape, Threats, and Impact* (Amsterdam: Deeptrace, 2019).

[13] Danielle Keats Citron, *The Fight for Privacy: Protecting Dignity, Identity, and Love in the Digital Age* (New York: Random House, 2022); and Citron, *Hate Crimes in Cyberspace*.

The Biden administration shut the board down, and Jankowicz resigned. Security consultants advised Jankowicz and her husband to relocate, an unrealistic suggestion given that Jankowicz was due to give birth. Fox News television guests remarked with glee that their "side" had emerged victorious and "got her bounced." Jankowicz retreated into silence for months. She stopped using social media. She shut down her Twitter account. She felt unsafe to leave her home.[14]

High-profile individuals like Jankowicz aren't the only ones facing online assaults that chase them offline. "Joan," a recent law school graduate, stayed in a hotel while traveling for work. When she returned home, she received an email from a stranger. The email included a video of her showering and urinating in the hotel bathroom, a video that she never knew existed, let alone gave anyone permission to take. The emailer, presumably a hotel employee, threatened to post the video on adult sites and to send it to Joan's LinkedIn contacts unless she sent additional nude photos and videos of herself. After Joan refused, the emailer made good on the threats. The emailer sent the video to Joan's graduate school classmates and her work colleagues (who the emailer presumably found via her LinkedIn profile). The emailer posted the video (with her name in the title of the video) on adult sites, including PornHub. The video appeared on dating sites next to the suggestion that Joan was available for sex.

Joan did everything that she could to get the videos and posts taken down, but she was met with a brick wall of silence. Most adult sites ignored her requests to remove the video. PornHub, the most popular adult site in the world, initially took down the videos in response to Joan's complaints. Unfortunately, the privacy invader kept reposting the video. After a while, PornHub stopped responding to Joan's requests for help. Despite Joan's best efforts, the video appeared on adult sites and many of the postings had thousands of views.

For Joan, as for so many people facing such abuse, privacy violations are never-ending. No matter what Joan did, the video remained online. For months and months, Joan searched for new postings every day and found more and more sites where the video had been posted. Joan felt scared and alone. No space seemed safe—not a public restroom, gym locker, or fitting room. If a hotel employee could

---

[14] Shannon Bond, "She Joined DHS to Halt Disinformation. She Says She Was Halted . . . by Disinformation," NPR, May 21, 2022, https://www.npr.org/2022/05/21/1100438703/dhs-disinformation-board-nina-jankowicz.

hide a camera in her room, so could those with access to other places in her life where she expected and deserved privacy.

Joan shuttered her social media accounts. Retreating from online engagement seemed necessary, but it wasn't what she wanted. The privacy invader seemingly identified her friends and coworkers from her social media accounts, so Joan closed her Facebook account, even though it was how she kept in touch with friends from college and high school. She took down her LinkedIn profile, even though she knew that she needed to be on the site if she ever wanted to change jobs.

Telling her boss about what had happened was a nightmare. Although her boss conveyed support, Joan could not help but think that her employer and coworkers now saw her as a nude body on the toilet and in the shower. She was humiliated. Joan suffered severe anxiety and depression. She lost a significant amount of weight; it was a way for her to regain control over her body and make it difficult for people to recognize her from the video. She worked out every day in the hope that gaining strength would enable her to fend off attackers. Joan worried that someone might respond to the fake ads and accost her offline.

The experience fundamentally changed the arc of Joan's life. Joan was engaged at the time of the initial privacy violation. Her fiancé was kind and supportive in ways large and small. He helped Joan contact adult sites and request the removal of the videos. When it became unbearable for Joan to check the sites, he monitored Google for new postings of the video. Joan and her fiancé delayed their wedding. As Joan explained to Danielle Keats Citron, how could she get married when she felt afraid to leave her house? (They eloped two years later.)

Long after the initial emails and posts, Joan felt watched and unsafe. Any time her laptop or phone seemed to slow down or have issues, she immediately thought that her tormentor had hacked her devices. Joan's sense of ease—her preternatural optimism—was gone, thanks to the violation of her intimate privacy.

Young women, sexual and gender minorities, and people of color suffer a disproportionate amount of cyberstalking and intimate privacy violations. The self-censorship that Joan and Jankowicz experienced is typical. Researchers have found that cyber gender harassment results in victims' withdrawal from online discourse, friendships, family, and romantic relationships.

As Jonathon Penney has found, women are statistically more chilled in their speech and engagement when targeted with online abuse.[15] A report issued in 2016 explained that "younger women are most likely to self-censor to avoid potential online harassment: 41% of women ages 15 to 29 self-censor, compared with 33% of men of the same age group and 24% of internet users ages 30 and older (men and women)."[16]

Studies show that online abuse imperils female politicians' expression. A NATO study released in 2020 found that female Finnish cabinet ministers received a disproportionate number of abusive tweets containing sexually explicit and racist abuse and demeaning gendered expletives like "slut" and "whore."[17] A 2019 study found that 28 percent of Finnish female municipal officials targeted with misogynistic hate speech reported being less willing than they would have been otherwise to make decisions that might unleash online abuse.[18] Iiris Suomela, a member of Finland's ruling coalition, has explained that her fear of misogynistic online abuse has changed the way that she talks about and addresses issues. The country's first Black woman member of Parliament, Bella Forsgrén, echoed her colleague's sentiments in saying that she must think twice about the discussions that she participates in and how she talks about the issues, lest she face online backlash.

Intimate privacy violations have a similar silencing effect. In the face of the nonconsensual taking, use, and sharing of intimate images, women are inclined to

---

[15] Jonathon Penney, "Internet Surveillance, Regulation, and Chilling Effects Online: A Comparative Case Study," *Internet Policy Review* 6 (2) (2017): 1, 19.

[16] Amanda Lenhart, Michele Ybarra, Kathryn Zickuhr, and Myeshia Price-Feeney, *Online Harassment, Digital Abuse, and Cyberstalking in America* (New York and San Clemente, Calif.: Data & Society Research Institute and Center for Innovative Public Health Research, 2016), 4, https://www.datasociety.net/pubs/oh/Online_Harassment_2016.pdf.

[17] Kristina Van Sant, Rolf Fredheim, and Gundars Bergmanis-Korāts, *Abuse of Power: Coordinated Online Harassment of Finnish Government Ministers* (Riga: NATO Strategic Communications Centre of Excellence, 2020), 50–51.

[18] Leonie Cater, "Finland's Women-Led Government Targeted by Online Harassment," Politico, March 17, 2021, https://www.politico.eu/article/sanna-marin-finland-online-harassment-women-government-targeted.

self-censor and to connect with fewer individuals.[19] They are more likely to withdraw from online activities, including shutting down their accounts.[20]

Victims of intimate privacy violations often isolate themselves. They disconnect from loved ones and from online connections. As sociolegal scholar Nicola Henry and her coauthors explain, such isolation is "due to a profound breach of trust, not only in relation to the abuser, but from family, friends, and the world around them."[21] Victims feel like they can no longer "trust anyone" or "anything."[22] Developing or sustaining close relationships can be difficult in the aftermath of intimate privacy violations. Victims feel alienated from loved ones who find it difficult to understand what happened.[23]

In writing her book *The Fight for Privacy*, Citron interviewed more than sixty people whose intimate privacy had been violated. They hailed from the United States, the United Kingdom, India, and Iceland. Most of those individuals were women, sexual and gender minorities, and people of color, who often had intersecting marginalized identities. Nearly every single person experienced a blow to their willingness to express themselves. They shut down their social media accounts. They stopped emailing and texting friends. They stopped dating. They deleted their online dating apps. They feared new relationships, including friendships. They lost trust in the world around them and in their ability to safely express themselves online and off.

---

[19] Henry, McGlynn, Flynn, et al., *Image-Based Sexual Abuse*, 59.

[20] The CCRI study found that 26 percent of survey respondents closed Facebook accounts, 11 percent closed Twitter accounts, and 8 percent closed LinkedIn accounts. "End Revenge Porn: A Campaign of the Cyber Civil Rights Initiative, Inc.," Cyber Civil Rights Initiative, https://www.cybercivilrights.org/wp-content/uploads/2014/12/RPStatistics.pdf (accessed June 25, 2024).

[21] Henry, McGlynn, Flynn, et al., *Image-Based Sexual Abuse*, 58.

[22] Ibid., 59. This accords with a 2019 study that found that the nonconsensual taking, sharing, or use of intimate images engenders an "intense shift" toward a position of lack of trust. Mollie C. DiTullio and Mackenzie M. Sullivan, "A Feminist-Informed Narrative Approach: Treating Clients Who Have Experienced Image-Based Abuse," *Journal of Feminist Family Therapy* 31 (2–3) (2019): 100, 104.

[23] Ibid.

III.

Law and industry practices can provide meaningful protection for intimate privacy.[24] We can and should bring law to bear to combat intimate privacy violations. Rules governing the nonconsensual filming, recording, or otherwise collecting of intimate images or information raise few, if any, First Amendment concerns because they separate the public sphere from the private. Trespass laws, the intrusion-on-seclusion tort, and video voyeurism laws have withstood constitutional challenge. Computer hackers and peeping toms cannot avoid criminal penalties by insisting that they were only trying to discover information that the public would benefit from knowing.[25]

What about the argument that the disclosure of intimate images involves the discloser's speech so it cannot be the basis of civil remedies or criminal penalties? When the government regulates speech based on the content of that speech, it usually must satisfy what is called "strict scrutiny" review. Strict scrutiny is a difficult standard to satisfy because government should not be in the business of favoring some ideas and disfavoring others. But laws can satisfy that tough standard if those laws serve a compelling interest that cannot be promoted through less restrictive means. Criminal laws banning nonconsensual pornography, crafted with the help of the Cyber Civil Rights Initiative, have faced constitutional challenge and survived the crucible of strict scrutiny review.[26] The supreme courts of Illinois, Indiana, Minnesota, and Vermont have upheld their states' nonconsensual intimate imagery statutes on the grounds that their statutes were justified by the compelling governmental interest in preventing the "permanent and severe" harms posed by nonconsensual intimate images and because the statutes were narrowly tailored to serve that interest.[27] The First Amendment would preclude specific legal actions if the public would have a legitimate interest in seeing nonconsensual intimate images. The fact that the public is interested in someone's intimate images does not, however, turn those images into matters of public interest. This is the case both for

---

[24] We leave the details for readers of Citron's *The Fight for Privacy*.

[25] Ibid., 144.

[26] Both of us have positions within the Cyber Civil Rights Initiative. Citron serves as the Vice President and Penney as an adviser.

[27] *State v. Katz*, 179 N.E.3d 431 (Ind. 2022); *State v. Casillas*, 952 N.W.2d 629 (Minn. 2020); *People v. Austin*, 155 N.E.3d 439 (Ill. 2019), *cert. denied*, 141 S. Ct. 233 (2020); *State v. VanBuren*, 214 A.3d 791 (Vt. 2019).

private people whose lives are not under public inspection and for celebrities whose intimate lives are public obsessions.

Law and industry also can and should curtail cyberstalking. Although cyberstalking often involves communications, it targets specific individuals with harassing speech that can be regulated. Courts have upheld cyberstalking convictions because the harassing speech either fell within recognized First Amendment exceptions or involved speech that has enjoyed less rigorous protection, such as true threats, defamation of private individuals, and the nonconsensual disclosure of private communications on purely private matters.[28] The Supreme Court recently ruled in *Counterman v. Colorado* that the First Amendment requires proof that a defendant was reckless about the terrorizing nature of a threat to criminally punish a "true threat." The law can regulate true threats, but there must be proof that the defendant consciously disregarded the risk that their speech activity would be viewed as threatening in order to prevent the chilling of protected speech.[29]

As we wait for law to protect intimate privacy as vigorously and comprehensively as it should, content platforms should protect people from intimate privacy violations. If and when law and market measures move in that direction, the expressive impact will be profound. Not only would law and corporate speech policies deter and reduce online abuse, mitigating its chilling impacts, but law and corporate speech policies also would say to intimate privacy victims that they matter, that they can express themselves knowing that companies and the law can help them if their intimate privacy is violated.

This is known as the expressive function of law: how law shapes behavioral norms by changing the social meaning of behavior.[30] When a law is passed, it provides a powerful symbolic or "informational" signal as to wider popular attitudes

---

[28] Citron, *Hate Crimes in Cyberspace*, 199–217.

[29] *Counterman v. Colorado*, 600 U.S. ___ (2023); and Danielle Keats Citron, *From Bad to Worse: Stalking, Threats, and Chilling Effects*, 2023 SUP. CT. REV. 175. Mary Anne Franks authored an amicus brief in the *Counterman* case; Citron signed that brief along with Erwin Chemerinsky, Michael Dorf, Eric Segall, and Cristina Tilley. The brief argued that the First Amendment does not require a specific-intent requirement for stalking and threats.

[30] Alex C. Geisinger & Michael Ashley Stein, *Expressive Law and the Americans with Disabilities Act*, 114 MICH. L. REV. 1061, 1061–62 (2016); and Richard H. McAdams, *The Expressive Powers of Law: Theories and Limits* (Cambridge, Mass.: Harvard University Press, 2015).

about social behavior—about what behaviors warrant legal penalty.[31] This is especially so in democracies, where laws tend to reflect the broader electorate's norms and values. When a law is passed to protect intimate privacy, it signals popular support for the protection of victims and recognizes the value of their autonomy and dignity, including their expressive engagement. Protective measures adopted by social media companies also have an expressive function: they say that victims' speech and ongoing presence and engagement are corporate priorities, that they are important to the social media community itself and worthy of protection.

In addition to enunciating attitudes and values, law provides signals about the risks associated with certain behavior: namely, that perpetrators of online abuse will be prosecuted, securing space for victims to speak and engage openly free from fear.[32] Through this informational and signaling function, the law has expressive impact that affects behavior—both in the near term as people respond to the law's messages—like victims speaking out more—and over time, as people internalize the attitudes and norms expressed by the law.[33]

A growing body of behavioral research explores how laws that restrict and curtail forms of online abuse have these expressive impacts. In 2019, we wrote about the expressive impact of cyber harassment laws.[34] We drew on Penney's empirical evidence that cyber harassment laws have a salutary impact on people's online speech and engagement, particularly for women.[35] Penney administered an original online survey to 1,296 adults based in the United States, which described to participants a series of hypotheticals.[36] One scenario concerned participants being made aware that the government had enacted a new law with tough civil and criminal penalties for cyber harassment. Responses offered a range of insights. They suggested that a cyber harassment law would have few chilling effects on regular

---

[31] Geisinger & Stein, *supra* note 30, at 1062; and McAdams, *The Expressive Powers of Law*, 139–141.

[32] McAdams, *The Expressive Powers of Law*, 152–155, 162–165.

[33] Ibid., 140–141.

[34] Danielle Keats Citron & Jonathon W. Penney, *When Law Frees Us to Speak*, 87 FORDHAM L. REV. 2317 (2019).

[35] Penney, "Internet Surveillance, Regulation, and Chilling Effects Online."

[36] Citron & Penney, *supra* note 34, at 2330.

speech.[37] Of the participants, 87 percent indicated that a cyber harassment law would have no impact or would make it more likely for them to speak and write online.[38]

Most states have cyberstalking laws on the books, but that is regrettably where they remain. Police rarely enforce those laws because they are misdemeanors (and thus are not worth their time and resources) and because law enforcement often dismisses the attacks as just "boys being boys."[39] We need state lawmakers to reform those laws, treating them as felonies, and to spend resources training law enforcement to investigate reports, rather than turn victims away.

Doing so would have great value. Penney's empirical research has shown that cyber harassment laws actually encourage online expression, particularly for women, rather than suppress online expression, as it is widely assumed (or at least assumed by advocacy groups like the ACLU).[40] Penney's analysis reveals a gender effect in response to the law: female participants in the survey were statistically more likely to engage online in response to the cyber harassment law in a variety of ways.[41] Female survey participants reported being more likely to share content online and more likely to engage on social network sites in response to the government enacting cyber harassment laws. We have argued elsewhere that cyber harassment laws would have that salutary impact given law's expressive value.[42] Those laws would tell victims that their safety and online engagement are valued, that they will be protected, and that they matter.[43]

In 2021, we teamed up again, along with media studies scholar Alexis Shore, to conduct empirical research on the potential impact of both legal and industry efforts to protect intimate privacy (with a special focus on the responsibilities of

---

[37] Ibid.

[38] Ibid., 2331–2332.

[39] Citron, *Hate Crimes in Cyberspace*, 83–88.

[40] Ibid.

[41] Ibid.

[42] Ibid.

[43] Jonathon Penney, "Online Abuse, Chilling Effects, and Human Rights," in *Citizenship in a Connected Canada: A Research and Policy Agenda*, ed. Elizabeth Dubois and Florian Martin-Bariteau (Ottawa: University of Ottawa Press, 2020), 207.

online platforms).[44] Our findings suggest that both legal protections and industry measures would engender trust in companies and the legal system such that individuals would be more inclined to engage in self-expression online.

In one experimental study, participants were exposed to different protective sexual privacy interventions. We found that participants who had previously experienced forms of online abuse—including intimate privacy violations—were more inclined to disclose and express intimate information after becoming aware of measures enacted to protect intimate privacy. That finding held across all conditions—for interventions involving both legal and platform-based measures—though participants presented with *platform*-based measures were even more likely to be willing to engage in intimate expression.

In another experimental study with a pre/post-longitudinal design, our results found that both legal and platform-based intimate privacy measures had a positive impact on trust among participants, especially for participants from marginalized populations. After participants were made aware of both legal and platform-based intimate privacy measures, trust became a stronger predictor of intimate expression online and offline, and that predictive relationship was even stronger among women, especially those who had previously experienced online abuse. We also found that both legal and platform measures increased trust in partners, such that they would be inclined to share and disclose intimate information to them, among participants from various marginalized groups—Latinos, African Americans, Asian Americans, Pacific Islanders—who are most often the targets of online abuse and intimate privacy violations.

These findings suggest that legal and platform-based intimate privacy measures can promote trust, leading to greater intimate expression and sharing over the long term. Both studies suggest that individuals will feel more inclined to engage in intimate expression with partners if they know that platforms have legal incentives to protect them from illegality online and that they are engaging efforts pursuant to those requirements.

This is a crucial point: our ongoing research with Shore suggests that legal measures that incentivize social media companies to address intimate privacy vio-

---

[44] The Knight Foundation supported our empirical research project with a $75,000 grant and we are grateful for their support.

lations can result in even more speech, not less. For instance, as Citron has proposed elsewhere, the law that currently shields social media companies from liability even if their platforms encourage or solicit intimate privacy violations should be reformed.[45] Congress surely never meant to provide a free pass to sites whose purpose is intimate privacy violations and online assaults. Sites that deliberately or purposefully solicit, encourage, or leave up material that they know (or have reason to know) constitutes stalking, harassment, or intimate privacy violations should not enjoy immunity from liability. This would not mean that content platforms would be strictly liable for intimate images or cyberstalking posted by users. Individuals whose intimate images appear on the sites without consent would have to bring legally cognizable claims against those sites. They should have a chance to do so.[46] And reform to that federal law would have salutary effects on all of us. People might be more likely to engage in intimate expression online and offline if they know that their intimate privacy enjoys protection—this is especially true for women. We might hear more women's voices, a win for civil rights and civil liberties.

We are at a tipping point. Our intimate privacy is being violated when we most need it. We need to protect intimate privacy for the good of free expression. In short, our findings suggest that protecting intimate privacy can help provide the reassurance that victims need to express themselves, rather than retreating into silence. Law and self-governance aimed to protect intimate privacy can indeed free us to speak.

---

[45] Danielle Keats Citron, *How to Fix Section 230*, 103 BOS. U. L. REV. 713 (2023).

[46] Of course, those lawsuits would have to press claims that can be squared with the First Amendment. One can imagine sites like Hidden Camera or MrDeepFakes, which traffic in intimate privacy violations, could face tort claims for enabling crime.