ARTICLE

HOW TO FIX SECTION 230

DANIELLE KEATS CITRON*

ABSTRACT

Section 230 is finally getting the clear-eyed attention that it deserves. No longer is it naïve to suggest that we revisit the law that shields online platforms from liability for enabling illegality. The harm wrought is now undeniable for victims of online assaults and intimate privacy violations. The market has not fixed this problem. Content platforms lack sufficient incentive to combat online abuse because they generate significant profits from our likes, clicks, and shares. Victims can't sue sites that earn advertising fees from their suffering. The status quo is particularly costly for women, children, and minorities who lose their ability to speak, work, and love. Inaction signals our society's indifference to vulnerable people enduring online abuse that robs them of their civil rights and civil liberties.

We need to fix § 230. Reform must be approached with humility and care, lest it spur platforms to over—or under—moderate in ways that do more harm than good. The legislative solutions offered here grow out of a decade of experience working with tech companies, online abuse victims, and legislative staff. While the over-filtering provision, § 230(c)(2), should be preserved, the under-filtering provision, § 230(c)(1), should be revised. Sites that deliberately encourage, solicit, or maintain intimate privacy violations, cyber stalking, or cyber

* Jefferson Scholars Foundation Schenck Distinguished Professor in Law, Chapman & Caddell Professor of Law, LawTech Center Director, University of Virginia School of Law; Vice President, Cyber Civil Rights Initiative; 2019 MacArthur Fellow. Ken Abraham went above and beyond with extensive feedback; Jeff Kosseff has been a constant source of wisdom; Colin Anderson, Katie Fallow, Mary Anne Franks, and David Stoppler were crucial sounding boards. Jackson Barnett, Rachel Bayefsky, Ryan Calo, Julie Cohen, Gian Marco Caletti, Stacey Dogan, Gilad Edelman, Julie Grant, David Greene, Quinta Jurecic, Peter Kaplan, Won-Mo Lee, Nick Nugent, Jonathon Penney, Alexis Shore, Marshall Van Alstyne, Ari Waldman, and Jason Walta provided generous help. Deep gratitude to Carrie Goldberg for all that she does to change the status quo—if there is any lawyer who I want with me to fight battles, it is her. Thanks also to Stacey Dogan's Boston University School of Law platform regulation class for their insightful comments on this piece. I am grateful to the *Boston University Law Review* for continuing our tradition of publication (seventh time!). Special thanks to Sydney Sullivan and the editorial staff for their comments. I appreciate the support and advice of Dean Risa Goluboff, librarians Kathryn Boudoris and Leslie Ashbrook, and research assistants Bao Chao, Julia Schur, Rebecca Weitzel, and Shinae Yoon.

harassment should not enjoy immunity from liability. Beyond carving out those bad actors, the under-filtering provision should be conditioned on a duty of care when claims involve intimate privacy violations, cyber stalking, or cyber harassment. Rather than an unguided duty of care, lawmakers should specify the obligations involved, drawing on key lessons from the trust and safety field. Under my proposal, companies would have to show that they took steps to address abuse that inhibits self-expression and ruins livelihoods.

CONTENTS

INTRODUCTION

Section 230 of the Communications Decency Act is all the rage.[1] If you had told me a decade ago that my call for § 230 reform[2] would be taken seriously, I would not have believed you. Then, any criticism of § 230 was viewed as heretical. I first pitched the notion of reform at the inaugural Privacy Law Scholars Conference in June 2008.[3] After the session on my then-draft article, *Cyber Civil Rights*,[4] an established law and tech scholar introduced himself to me (I was a newcomer to the legal academy).[5] With a stern look on his face, he asked me if I intended to "jail communists."[6] "Your challenging Section 230 is like stabbing the First Amendment in the heart."[7]

That conversation made clear to me that it would be difficult to convince people that § 230 reform could be good for both privacy and free speech. The prevailing view was that any change would be a zero-sum game.[8] Section 230's

[1] *See* 47 U.S.C. § 230; *see also* Gilad Edelman, *Everything You've Heard About Section 230 Is Wrong*, WIRED (May 6, 2021, 7:00 AM), https://www.wired.com/story/section-230-internet-sacred-law-false-idol/ (describing intensifying debate over § 230 reform).

[2] *See generally* Danielle Keats Citron, *Cyber Civil Rights*, 89 B.U. L. REV. 61 (2009) [hereinafter Citron, *Cyber Civil Rights*].

[3] The inaugural Privacy Law Scholars Conference was held at the George Washington University Law School, thanks to the leadership of Daniel J. Solove and Chris Hoofnagle. *See* BERKELEY L. SCH. & GEO. WASH. UNIV. L. SCH., PRIVACY LAW SCHOLARS CONFERENCE (2008), http://sites.law.berkeley.edu/privacylaw/wp-content/uploads/sites/8/2013/02/Privacy-Law-Scholars-Conference-2008.pdf [https://perma.cc/JGX8-WQTQ].

[4] Citron, *Cyber Civil Rights*, *supra* note 2, at 121-25 (proposing to condition § 230 immunity for under filtering illegality on "duty of care" standard, which would include requiring traceable anonymity so perpetrators could be caught and sued). In that piece, I offered the approach but did not flesh out the precise details. *See id.* In my book, *Hate Crimes in Cyberspace*, I proposed a carveout for Bad Samaritans that solicited and peddled nonconsensual pornography and online abuse amounting to cyber stalking. *See* DANIELLE KEATS CITRON, HATE CRIMES IN CYBERSPACE 177-81 (2014) [hereinafter CITRON, HATE CRIMES IN CYBERSPACE]. With Benjamin Wittes, in 2017, I returned to my original idea of a duty of care with potential statutory language. *See* Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity*, 86 FORDHAM L. REV. 401, 418-19 (2017). We laid out why reasonableness was the right approach but did not expand upon how it might be operationalized on the ground. *See id.* at 419. In an essay for the *University of Chicago Legal Forum*, Mary Anne Franks and I considered a variety of potential reform proposals. *See* Danielle Keats Citron & Mary Anne Franks, *The Internet as a Speech Machine and Other Myths Confounding Section 230 Reform*, 2020 U. CHI. LEGAL F. 45, 69-74. This Article aims to bring together my ideas, refine them, and analyze unexplored issues and weaknesses. My intended audience for this Article is federal lawmakers, staff, scholars, and advocacy groups in the hopes that it serves as a blueprint for reform.

[5] *See* Edelman, *supra* note 1.

[6] *Id.*

[7] *Id.*

[8] *See, e.g.*, JEFF KOSSEFF, THE TWENTY-SIX WORDS THAT CREATED THE INTERNET 208 (2019) [hereinafter KOSSEFF, TWENTY-SIX WORDS] ("Section 230 has become so intertwined with our fundamental conceptions of the Internet that any wholesale reductions to the

legal shield could not be reformed to protect privacy without risking the free flow of ideas.[9]

Since its passage, § 230's legal shield has been an article of faith, hailed as nothing less than the law that "created the internet."[10] But the impact of that law is a mixed bag. Yes, § 230 has enabled expression by protecting platforms from liability related to publishing user-generated content.[11] Yes, it has made space for the development of a vast array of online services, from search engines to social networks.[12] In the mid-1990s, the commercial internet was in its infancy.[13] Early U.S. internet service providers like AOL and Prodigy had twelve million subscribers *in total*.[14] The absence of liability meant that search engines could link to sites, blogs, and other online activity without fear that they would be liable for defamatory comments.[15] Social media companies could welcome all subscribers without worrying about facing lawsuits for having published subscribers' posts.[16] Supporters of § 230 argue that § 230 is why the internet is now full of commerce, noisy discourse, and political dissent—including the Arab Spring, Black Lives Matter, and #MeToo movements.[17]

But § 230 has a destructive side—a fact that supporters usually either ignore or insist must be tolerated (by someone else, of course).[18] What the law's enthusiasts think should be overlooked or endured demands attention. Section 230's legal shield has enabled online abuse that has destroyed people's lives.[19] Under the judiciary's sweeping interpretation of § 230's legal shield, sites bear no responsibility for destructive uses of their services, even when they deliberately solicit or encourage those uses.[20] It has given a free pass to sites that profit from intimate privacy violations, harassment, and stalking. Women and

immunity could irreparably destroy the free speech that has shaped our society in the twenty-first century.").

[9] *See id.*

[10] *See id.* at 3-5; Bryan Pietsch, Isobel Asher Hamilton & Katie Canales, *The Facebook Whistleblower Told Congress It Should Amend Section 230, the Internet Law Hated by Both Biden and Trump. Here's How the Law Works.*, INSIDER (Oct. 6, 2021, 11:39 AM), https://www.businessinsider.com/what-is-section-230-internet-law-communications-decency-act-explained-2020-5 [https://perma.cc/6SN5-57H2].

[11] *See* Citron & Wittes, *supra* note 4, at 406-08.

[12] For an indispensable history of § 230's adoption and judicial interpretation over the years, see KOSSEFF, TWENTY-SIX WORDS, *supra* note 8, at 77-144.

[13] *See* Citron & Wittes, *supra* note 4, at 411.

[14] *See id.* (citing Reno v. ACLU, 521 U.S. 844, 850-51 (1997)).

[15] *See id.*

[16] *See id.*

[17] *See* DANIELLE KEATS CITRON, THE FIGHT FOR PRIVACY: PROTECTING DIGNITY, IDENTITY, AND LOVE IN THE DIGITAL AGE 84-85 (2022) [hereinafter CITRON, THE FIGHT FOR PRIVACY].

[18] This resonates with gun rights activists' approach to the Second Amendment, something that Mary Anne Franks tackles with clarity and brilliance in her book, *The Cult of the Constitution*. *See* MARY ANNE FRANKS, THE CULT OF THE CONSTITUTION 11 (2019).

[19] *See* CITRON, HATE CRIMES IN CYBERSPACE, *supra* note 4, at 35-50.

[20] *See infra* Section I.B.

people in marginalized groups have shouldered much of the abuse, which has chilled victims' expression.[21]

Section 230 is why the United States is a haven for sites trafficking in intimate privacy violations.[22] U.S.-based revenge-porn sites operate with impunity, thanks to § 230.[23] When it comes to fabricated nude imagery like deepfake sex videos that amount to defamation, the law's immunity is ironclad. The SPEECH Act of 2010 prevents victims from obtaining defamation judgments outside the U.S. and enforcing them in the U.S. if those judgments would not comport with the First Amendment or § 230.[24]

In short, § 230 is not the win for free expression that boosters claim. Just as the law's broad interpretation frees platforms to publish accounts of current events, it gives them license to encourage online abuse that silences victims. It provides social media platforms a free pass to host posts by civil rights protestors *and* sexual predators. It lets sites solicit political commentary *and* hidden-camera feeds. It enables companies to design their platforms to enhance the visibility of user-generated art *and* deepfake sex videos. Section 230 is why the internet is filled with war footage *and* death threats, encyclopedia entries *and* rape videos, restaurant reviews *and* nonconsensual pornography.[25]

Section 230 has not worked as its drafters intended. In 1996, Representatives Chris Cox and Ron Wyden called upon content platforms to act as "Good Samaritans" in blocking and screening offensive speech.[26] At that time, they could have hardly imagined the array of businesses and activities that would emerge online. Cox or Wyden surely did not mean to insulate from liability sites that deliberately solicited intimate privacy violations or that continuously allowed predators to remain on services that matched children with strangers.[27]

It is no longer heretical to suggest that § 230 should be reformed.[28] A consensus is emerging that shielding online platforms from liability without preconditions wasn't the best idea after all.[29] Section 230 has cut off important

[21] *See infra* notes 148-49 and accompanying text.

[22] *See infra* notes 116-17 and accompanying text.

[23] *See infra* notes 116-17 and accompanying text.

[24] *See* Securing the Protection of our Enduring and Established Constitutional Heritage (SPEECH) Act, Pub. L. No. 111-223, 124 Stat. 2380 (2010) (codified as amended at 28 U.S.C. §§ 4101-4105).

[25] *See* CITRON, THE FIGHT FOR PRIVACY, *supra* note 17, at 84-85.

[26] *See* 47 U.S.C. § 230(c) ("Protection for 'Good Samaritan' blocking and screening of offensive material . . . .").

[27] *See, e.g.*, M.H. *ex rel.* C.H. v. Omegle.com, LLC, No. 21-cv-814, 2022 WL 93575, at *3 (M.D. Fla. Jan. 10, 2022) (granting § 230 immunity to Omegle for alleged sex crime perpetrated by Omegle user against child plaintiff).

[28] *See* Citron & Franks, *supra* note 4, at 46-48 (discussing broad support across political spectrum for § 230 reform).

[29] *See, e.g.*, Press Release, U.S. Dep't of Just., The Justice Department Unveils Proposed Section 230 Legislation (Sept. 23, 2020), https://www.justice.gov/opa/pr/justice-department-unveils-proposed-section-230-legislation [https://perma.cc/E3SU-MMKD]; Ashley Johnson & Daniel Castro, *Proposals To Reform Section 230*, INFO. TECH. & INNOVATION FOUND. (Feb.

legal pathways, preventing victims from seeking relief and policymakers from pursuing reforms. Section 230 has made it impossible for the law to develop a response to deliberate profiteering from destructive harassment, stalking, and intimate privacy violations.

Time and practice have made clear that tech companies "don't have strong enough incentives to protect their brands by policing their platforms."[30] Thousands of sites earn advertising fees from intimate privacy violations that they have solicited.[31] These sites externalize harm from online assaults that they do not have to internalize.

We must not lose sight of the fact that although platforms don't shoulder responsibility for online abuse that they enable, someone *is* paying.[32] Victims of online abuse experience never-ending privacy invasions, emotional suffering, and reputational damage.[33] They have difficulty getting and keeping jobs.[34] Their family, friends, and colleagues abandon them when they most need support.[35] They stop engaging online and offline; their self-expression is chilled; their voices are silenced. Section 230 ensures that victims cannot sue the entities that have solicited their suffering.

This Article offers a legislative solution borne out of on-the-ground experience working with tech companies and victims of cyber stalking and intimate privacy violations. Part I begins with a short overview of the concerns that led Congress to adopt § 230. Part II highlights the costs to civil rights and civil liberties under the status quo. It shows that legal change is worth the candle but must be calibrated with care, so we don't inadvertently undermine civil rights and civil liberties in the name of protecting them. Part III lays out my proposal. I start by underscoring the parts of the law that should not be altered. Then, I turn to the focus of my proposal: excluding bad actors and conditioning the under-filtering provision on a particularized duty of care. I conclude by exploring synergies with other nations' efforts.

22, 2021), https://itif.org/publications/2021/02/22/proposals-reform-section-230 [https://perma.cc/6WX2-WHVQ]; Cameron F. Kerry, *Section 230 Reform Deserves Careful and Focused Consideration*, BROOKINGS: TECHTANK (May 14, 2021), https://www.brookings.edu/blog/techtank/2021/05/14/section-230-reform-deserves-careful-and-focused-consideration/ [https://perma.cc/PAS2-CYMD]; Michael D. Smith & Marshall Van Alstyne, *It's Time To Update Section 230*, HARV. BUS. REV. (Aug. 12, 2021), https://hbr.org/2021/08/its-time-to-update-section-230 [https://perma.cc/WHM7-C8CE].

[30] Smith & Van Alstyne, *supra* note 29.

[31] For a detailed view of the monetization of intimate privacy violations, see CITRON, THE FIGHT FOR PRIVACY, *supra* note 17, at 72-76.

[32] Or, as my colleague Ken Abraham has said to me, strict liability is the coin of the realm when it appears that no liability rules the day. Someone always bears the costs—a no liability rule means that victims bear the costs alone.

[33] *See* Danielle Keats Citron & Daniel J. Solove, *Privacy Harms*, 102 B.U. L. REV. 793, 830-31 (2022).

[34] *See* CITRON, HATE CRIMES IN CYBERSPACE, *supra* note 4, at 9.

[35] *See* CITRON, THE FIGHT FOR PRIVACY, *supra* note 17, at 41-43.

I. HOW DID WE GET HERE?

As I explain in my book, *The Fight for Privacy*,

> A brick-and-mortar business that makes it easy for third parties to stalk and invade the privacy of victims faces tort liability for enabling the abuse. A hard-copy magazine that published user-submitted nonconsensual porn encounters a blizzard of privacy lawsuits. But when those activities happen online, companies are shielded from liability. We have Section 230[] to thank for that.[36]

To understand how we got here, we need to look at the law's history. This Part provides a brief overview.

A. *Back to* Prodigy

In the mid-1990s, the growing internet posed a challenge to federal lawmakers.[37] While lawmakers wanted to encourage the growth of the internet by keeping it open and free, they also understood that these characteristics allowed for the posting of illegal and "offensive" material.[38] They recognized that tech companies would need to play a role in moderating content, as federal agencies could not deal with all "noxious material" on their own.[39]

In 1995, a New York trial court rendered a decision suggesting that any and all efforts by platforms to moderate online content was a legally risky endeavor.[40] *Stratton Oakmont, Inc. v. Prodigy Services Co.*[41] involved the alleged defamation of Stratton Oakmont, a company run by Jordan Belfort, known to many as the "Wolf of Wall Street."[42] Someone accused Stratton Oakmont of fraud on a message board hosted by internet service provider

---

[36] *Id.* at 84.

[37] *Id.*

[38] *See* CITRON, HATE CRIMES IN CYBERSPACE, *supra* note 4, at 170; Citron, *Cyber Civil Rights*, *supra* note 2, at 116 n.377; Citron & Wittes, *supra* note 4, at 404-06.

[39] Citron & Wittes, *supra* note 4, at 403.

[40] *See* Stratton Oakmont, Inc. v. Prodigy Servs. Co., No. 31063/94, 1995 WL 323710, at *4-5 (N.Y. Sup. Ct. May 24, 1995).

[41] *Prodigy*, 1995 WL 323710.

[42] *See* Matthew Partridge, *Great Frauds in History: Jordan Belfort and Stratton Oakmont*, MONEYWEEK (Aug. 7, 2019), https://moneyweek.com/512249/great-frauds-in-history-jordan-belfort-and-stratton-oakmont [https://perma.cc/B6ZB-VEXJ]. Belfort eventually faced federal prosecution and pleaded guilty for running a boiler room that defrauded investors with pump and dump stock sales. *Id.* He served twenty-two months in prison and was ordered to pay back $110.4 million to the people he defrauded. Stefania Bianchi & Mahmoud Habboush, *Wolf of Wall Street Belfort Is Aiming for $100 Million Pay*, BLOOMBERG (May 19, 2014, 9:05 AM), https://www.bloomberg.com/news/articles/2014-05-19/wolf-of-wall-street-belfort-sees-pay-top-100-million-this-year. He wrote a book about his crimes called *The Wolf of Wall Street*, which was made into a movie starring Leonardo DiCaprio. *Id.* At a speech that Belfort gave in 2014, he said of his criminal career: "I got greedy . . . . Greed is not good." *Id.*

Prodigy.[43] The firm brought defamation claims against Prodigy.[44] Prodigy argued that it should not be liable for allegedly defamatory posts that it did not know about because it was not a publisher and thus not strictly liable for the content.[45] The court rejected Prodigy's argument, explaining that because Prodigy had used software to filter out profanity (in an effort to be a "family-oriented" platform), it had assumed the role of a publisher.[46] The court found that Prodigy's efforts to moderate content increased its liability for alleged defamation.[47]

The *Prodigy* decision caused a stir, even though—as every law student and lawyer knows—a state's trial court does not bind the rulings of any other court in that state, let alone courts in other states or the federal courts.[48] By contrast, rulings by a state's highest court or the U.S. Supreme Court would serve as precedent that lower state or federal courts must follow.[49]

Although no other court had adopted the *Prodigy* approach, the decision caught the attention of Chris Cox and Ron Wyden, both Congressmen at the time.[50] Representatives Cox and Wyden feared that nascent internet companies would heed the case's message. Under the logic of that decision, if tech companies proactively removed or blocked "noxious material," then they would be treated as publishers of any material that they had not removed or filtered.[51] Companies might refrain from moderating user-generated content if doing so would make it more likely that they would bear legal responsibility for defamatory posts. That set Representatives Cox and Wyden on the path to pass legislation that would nullify the ruling in *Prodigy*.

The law that Cox and Wyden drafted was included in a statute dedicated to ridding the internet of pornography—the Communications *Decency* Act ("CDA") was certainly true to its name.[52] The CDA imposed criminal penalties on anyone who knowingly used the internet to display "patently offensive . . . sexual or excretory" activities.[53] Simply put, it criminalized porn.

---

[43] *See Prodigy*, 1995 WL 323710, at *1.

[44] *See id.* at *1.

[45] *See id.* at *2-3.

[46] *See id.* at *4.

[47] *See id.* at *5.

[48] *See Legal Research: An Overview: Mandatory v. Persuasive Authority*, UCLA SCH. OF L. HUGH & HAZEL DARLING L. LIBR., https://libguides.law.ucla.edu/c.php?g=686105&p=5160745 [https://perma.cc/6GKP-YFJM] (last updated Feb. 14, 2023, 5:03 PM).

[49] *Id.*

[50] *See* KOSSEFF, TWENTY-SIX WORDS, *supra* note 8, at 2. Other Congressmen echoed their concern. *See* 141 CONG. REC. 22046 (1995) (statement of Rep. Bob Goodlatte) (arguing Prodigy should not be responsible for editing out information posted by users on its bulletin board).

[51] *See Prodigy*, 1995 WL 323710, at *4 (finding Prodigy was "publisher" in part because it controlled content posted by users).

[52] *See* Communications Decency Act of 1996, Pub. L. No. 104-104, 110 Stat. 133.

[53] *Id.* § 502, 110 Stat. at 134.

In 1997, the Supreme Court struck down most of the law on constitutional grounds, except for Cox and Wyden's handiwork in § 230.[54]

The two-part shield from liability that Cox and Wyden devised appeared in § 230(c), its title reading: "Protection for 'Good Samaritan' blocking and screening of offensive material."[55] The provision was meant to incentivize private efforts to combat "offensive" material.[56] Section 230(c)(1)—which my colleague and Cyber Civil Rights Initiative ("CCRI") President Mary Anne Franks has described as the "'leave up' provision"—addresses the under removal of content.[57] It states that "[n]o provider or user of an interactive computer service *shall be treated as the publisher or speaker* of any information provided by another information content provider."[58] Section 230(c)(2), conversely, concerns the over removal of content, or as Franks calls it, the "'take down' provision."[59] Under the subtitle "Civil liability," the provision declares that providers or users of interactive computer services will not be held liable 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."[60]

Section 230(c)'s legal shield has a few exemptions. Excluded from the legal shield's provisions are violations of federal criminal law, intellectual property claims, and the Electronic Communications Privacy Act.[61] In 2018, the Fighting Online Sex Trafficking Act ("FOSTA")[62] updated the list of exemptions to include certain civil and state criminal laws addressing the knowing facilitation of sex trafficking.[63]

In 1996, Representatives Cox and Wyden could not have imagined what the internet would mean to public and private life in the decades to come.[64] Then, it was not obvious that the internet would become integral to daily life, that it

[54] *See* Reno v. ACLU, 521 U.S. 844, 870-79 (1997) (holding most of CDA was unconstitutional because it was impermissibly vague and imposed restrictions on free speech).

[55] 47 U.S.C. § 230(c).

[56] *See id.* § 230(b) (describing § 230's policy goals).

[57] *See Hearing on Holding Big Tech Accountable: Targeted Reforms to Tech's Legal Immunity Before Subcomm. on Commc'ns & Tech. of the H. Comm. on Energy & Com.*, 117th Cong. 5 (2021) [hereinafter Franks Testimony] (written testimony of Mary Anne Franks, Professor, University of Miami), https://www.congress.gov/117/meeting/house/114268/witnesses/HHRG-117-IF16-Wstate-FranksM-20211201-U1.pdf [https://perma.cc/FR47-SGXJ].

[58] 47 U.S.C. § 230(c)(1) (emphasis added).

[59] Franks Testimony, *supra* note 57, at 5.

[60] 47 U.S.C. § 230(c)(2).

[61] *Id.* § 230(e).

[62] *See* Fighting Online Sex Trafficking Act, Pub. L. No. 115-164, 132 Stat. 1253 (2018).

[63] 47 U.S.C. § 230(e).

[64] *See* Citron & Wittes, *supra* note 4, at 411.

would enhance, replicate, and change crucial spaces and activities.[65] As we know now, networked technologies are indispensable to work, education, close relationships, intellectual discoveries, reading, sex, dating, and health services.[66] As I explained in my book, *Hate Crimes in Cyberspace*, the internet is embedded in everything we do and everywhere we go.[67] And, as became clear during the COVID pandemic, platforms like Zoom serve varied purposes.[68]

Cox and Wyden were prescient. According to Cox, if this "amazing new thing" was "going to blossom," companies should not be "punished for *trying* to keep things clean."[69] Cox told National Public Radio in 2018 that "[t]he original purpose of [§ 230] was to help clean up the Internet, not to facilitate people doing bad things on the Internet."[70] Wyden agreed, noting that the key to § 230 "was making sure that companies in return for that protection—that they wouldn't be sued indiscriminately—were being *responsible* in terms of policing their platforms."[71]

B. *Broad Judicial Interpretation*

The judiciary's broad interpretation of § 230, however, has departed from this original vision.[72] Instead of treating § 230 as a legal shield for Good Samaritans attempting to filter and block illegality and "offensive" content (the statute's words, not mine), courts have extended the provision far beyond its text and originally intended purpose.[73]

Instead of providing legal cover to Good Samaritans, "Bad Samaritans have been immunized from liability."[74] Sites that intentionally solicited privacy violations have enjoyed immunity from liability.[75] This was true for The Dirty, whose site operator curated and posted "scoops" about people (including nude

[65] *See* CITRON, HATE CRIMES IN CYBERSPACE, *supra* note 4, at 101.

[66] *Id.*

[67] *Id.*

[68] To people's great dismay, Zoom enabled work life and home life to inadvertently collapse on each other. Journalist Jeffrey Toobin's case comes to mind. *See* Laura Wagner, *New Yorker Suspends Jeffrey Toobin for Masturbating on Zoom Call*, VICE: MOTHERBOARD (Oct. 19, 2020, 2:07 PM), https://www.vice.com/en/article/epdgm4/new-yorker-suspends-jeffrey-toobin-for-zoom-dick-incident [https://perma.cc/W8BB-HKXN].

[69] Alina Selyukh, *Section 230: A Key Legal Shield for Facebook, Google Is About To Change*, NPR (Mar. 21, 2018, 5:11 AM), https://www.npr.org/sections/alltechconsidered/2018/03/21/591622450/section-230-a-key-legal-shield-for-facebook-google-is-about-to-change [https://perma.cc/C66X-HDMK].

[70] *Id.*

[71] *Id.* (emphasis added).

[72] CITRON, THE FIGHT FOR PRIVACY, *supra* note 17, at 87.

[73] *See* Citron & Wittes, *supra* note 4, at 414-15.

[74] CITRON, THE FIGHT FOR PRIVACY, *supra* note 17, at 87.

[75] *Id.*

images),[76] and nonconsensual intimate imagery site Texxxan.[77] Sites that deliberately enhanced the visibility of illegality while ensuring that perpetrators could not be identified have also been shielded from liability.[78] This was true for Backpage's promotion of ads that trafficked minors for sex work.[79]

The statute's legal shield also has been interpreted to negate any remedy, even ones that are not only easy and inexpensive to administer but also would significantly improve victims' lives.[80] For instance, the California Supreme Court ruled that § 230 excused the review site Yelp from complying with a court order to remove defamatory content posted by a user.[81]

Courts have attributed this broad interpretive approach to the fact that § 230's adoption was driven by "First Amendment values."[82] However, Congress's goals also included "the development of technologies which maximize user control over what information is received" and the "*vigorous enforcement of Federal criminal laws to deter and punish trafficking in obscenity, stalking, and harassment by means of computer*."[83] As Mary Anne Franks has wisely explained, "[T]he law [was] intended to promote and protect the values of

[76] *See* Jones v. Dirty World Ent. Recordings LLC, 755 F.3d 398, 402-03 (6th Cir. 2014).

[77] *See* GoDaddy.com, LLC v. Toups, 429 S.W.3d 752, 753 (Tex. App. 2014); *see also* Andrew McDiarmid, *Decisive Section 230 Victory for GoDaddy in Revenge Porn Case*, CTR. FOR DEMOCRACY & TECH. (Apr. 15, 2014), https://cdt.org/insights/decisive-section-230-victory-for-godaddy-in-revenge-porn-case/ [https://perma.cc/5WEN-UUN6].

[78] *See* Doe No. 1 v. Backpage.com, LLC, 817 F.3d 12, 29 (1st Cir. 2016).

[79] *See id.* This occurred before Congress amended § 230 in 2018 to exempt sites that knowingly facilitate sex trafficking. *See* Fighting Online Sex Trafficking Act, Pub. L. No. 115-164, 132 Stat. 1253 (2018).

[80] *See, e.g.*, Hassell v. Bird, 420 P.3d 776, 779 (Cal. 2018). I have explored how § 230 has prevented victims from obtaining injunctive relief in my scholarship. *See generally* Danielle Keats Citron, *Privacy Injunctions*, 71 EMORY L.J. 955 (2022) [hereinafter Citron, *Privacy Injunctions*].

[81] *See Hassell*, 420 P.3d at 779. The court order was entered on a default judgment. *Id.* at 781. The plaintiffs in the case, a lawyer and a law firm, sued Ava Bird for defamation, false light, and intentional infliction of emotional distress in connection with her review of the firm on Yelp. *Id.* at 779-80. After the defendant failed to appear in court, the plaintiffs moved for a default judgment. *Id.* at 780. The trial court held an evidentiary hearing and then ruled in favor of the plaintiffs and ordered Bird to remove the defamatory reviews. *Id.* at 780-81. After Bird failed to remove the posts, plaintiffs served a copy of the default judgment on Yelp, leading to Yelp's motion to set aside the judgment on the grounds of § 230, which was denied. *Id.* at 781. The California Court of Appeals found that the trial court properly extended the order for injunctive relief to reach Yelp even though Yelp was not a party in the case. *Id.* at 782-83. It ruled that the trial court had the authority to require Yelp to remove the statements deemed defamatory because the injunction prohibiting Bird from repeating those statements was issued after a hearing and the issuance of a default judgment. *Id.* As noted above, the California Supreme Court found that the trial court lacked any authority over Yelp due to § 230. *Id.* at 779.

[82] *See, e.g.*, *Backpage.com*, 817 F.3d at 29 ("If the evils that the appellants have identified are deemed to outweigh the First Amendment values that drive the CDA, the remedy is through legislation, not through litigation.").

[83] 47 U.S.C. § 230(b) (emphasis added).

privacy, security and liberty alongside the values of open discourse."[84] And in some cases, the claims at issue have little to do with the publication of speech.[85]

Consider Matthew Herrick's awful encounter with Grindr, an LGBTQ dating app.[86] Herrick sought the help of Grindr after his ex-boyfriend started using the dating app to stalk him.[87] Fake profiles appeared on the app, with nude photos of Herrick next to invitations to his apartment to play out "rape fantasies."[88] In numerous emails to Grindr, Herrick explained that the fake profiles put him in serious danger.[89] As many as twenty-three men came to his apartment on a daily basis expecting sex, having been told to view his resistance as part of the "fantasy."[90] Over ten months, 1,400 men confronted him.[91] Because Herrick's police precinct had done nothing to help him, Grindr was his last option.[92] While Grindr's security team could and should have played an important role in minimizing the damage and the danger to Herrick, the company ignored his messages.[93] Herrick's only response from Grindr was an automatically generated email: "Thank you for your report."[94]

Citing § 230 immunity, a trial and an appellate court dismissed Herrick's case, and the Supreme Court refused to hear an appeal.[95] Herrick's lawsuit alleged

[84] Mary Anne Franks, *The Lawless Internet? Myths and Misconceptions About CDA Section 230*, HUFFPOST (Feb. 17, 2014), https://www.huffpost.com/entry/section-230-the-lawless-internet_b_4455090 [https://perma.cc/XVD4-XU87].

[85] *See, e.g.*, Gibson v. Craigslist, Inc., No. 08 Civ. 7735, 2009 WL 1704355, at *1-2 (S.D.N.Y. June 15, 2009) (dismissing, on § 230 grounds, claim alleging that Craigslist owed duty of care to shooting victim for enabling sale of gun on site).

[86] *See* Herrick v. Grindr, LLC, 306 F. Supp. 3d 579, 584 (S.D.N.Y. 2018) (finding Grindr immune from liability under § 230 even after failing to prevent abuse of platform to target and harass defendant despite defendant's repeated pleas for help), *aff'd*, 765 F. App'x 586 (2d Cir. 2019). I discussed Herrick's case in my book review of *Sabrina*. *See* Danielle Keats Citron, *Cyber Mobs, Disinformation, and Death Videos: The Internet as It Is (and as It Should Be)*, 118 MICH. L. REV. 1073, 1089-91 (2020) [hereinafter Citron, *Cyber Mobs*] (reviewing NICK DRNASO, SABRINA (2018)).

[87] *Herrick*, 306 F. Supp. 3d at 584-85.

[88] *Id.* at 585.

[89] *Id.* at 586.

[90] Carrie Goldberg, Herrick v. Grindr*: Why Section 230 of the Communications Decency Act Must Be Fixed*, LAWFARE (Aug. 14, 2019, 8:00 AM), https://www.lawfareblog.com/herrick-v-grindr-why-section-230-communications-decency-act-must-be-fixed [https://perma.cc/ZXF7-XF24].

[91] *Id.*

[92] *See* CARRIE GOLDBERG, NOBODY'S VICTIM: FIGHTING PSYCHOS, STALKERS, PERVS, AND TROLLS 36 (2019). Eventually, law enforcement arrested Herrick's ex, but it was after months and months of men confronting him at home and at work. *See id.* at 37.

[93] *See id.* at 37-38 (noting Grindr responded to none of fifty complaints made by Herrick and others on his behalf).

[94] *Id.* at 38.

[95] *See* Herrick v. Grindr, LLC, 306 F. Supp. 3d 579, 601 (S.D.N.Y. 2018), *aff'd*, 765 F. App'x 586 (2d Cir. 2019), *cert. denied*, 140 S.Ct. 221 (2019); *see also* Alexis Kramer, *Grindr Harassment Case Won't Get Supreme Court Review*, BLOOMBERG L. (Oct. 7, 2019, 9:51 AM),

that Grindr had defectively designed its product—the app was not designed to identify and block abusive users based on IP address, geolocation, or other industry-standard techniques, which made it difficult for the app to stop abuse from continuing.[96] As Herrick's counsel argued, Grindr's design choice concerned the company's actions, not user-generated content.[97] Even still, the trial and appellate courts found § 230 barred Herrick's claims.[98]

Even though Grindr could minimize the harm by adding capabilities to identify and block users based on IP addresses or geolocation (a common capability for online platforms, including dating apps), even though the company knew that people were using the app to invade others' intimate privacy and ignored repeated pleas for help, even though the company profited from the data that the fake profiles generated, the courts absolved Grindr of any responsibility. Grindr did not have to internalize the costs that it had externalized onto Herrick because § 230 barred the suit. Reflecting on Herrick's terrible predicament, his attorney, Carrie Goldberg, told me that she has "lost hope in there being a judicial fix to Section 230."[99] With regret, she explained, "It used to be that for the cost of an index number [to file a lawsuit], the poorest person in the world could hold the most powerful corporation accountable for the harms they caused. Those days are gone."[100]

Even social networks that have served as hunting grounds for child predation have enjoyed § 230's legal shield. Omegle is a social media site that lets users talk to strangers. As Benjamin Wittes and I wrote in 2017, Omegle "seems to understand" that it operates as a social media site for sexual predators.[101]

> The opening paragraph—the same one in which the site proclaims itself a great way to meet new friends—warns that "[p]redators have been known to use Omegle, so please be careful." The site's legal disclaimer, also on its front page, specifically warns: "Understand that human behavior is fundamentally uncontrollable, that the people you encounter on Omegle may not behave appropriately, and that they are solely responsible for their own behavior. Use Omegle at your own peril."[102]

On March 20, 2020, C.H., an eleven-year-old girl, encountered a child predator and privacy invader on the site.[103] The man, who called himself "John

https://news.bloomberglaw.com/tech-and-telecom-law/grindr-harassment-case-wont-get-supreme-court-review.

[96] *Herrick*, 306 F. Supp. 3d at 585.

[97] *Id.*

[98] *Id.* at 584.

[99] E-mail from Carrie Goldberg, Founder, C.A. Goldberg, PLLC, to Danielle Keats Citron, Professor of L., Univ. of Virginia Sch. of L. (Aug. 2, 2019) (on file with author).

[100] *Id.*

[101] Citron & Wittes, *supra* note 4, at 401.

[102] *Id.* (alteration in original) (emphasis removed) (footnote omitted).

[103] M.H. *ex rel.* C.H. v. Omegle.com, LLC, No. 21-cv-814, 2022 WL 93575, at *1 (M.D. Fla. Jan. 10, 2022).

Doe," told C.H. that he knew where she lived and that he would hack her family's electronic devices if she did not disrobe for him on camera.[104] Out of fear, C.H. complied with John Doe's demands.[105] C.H.'s parents sued Omegle on her behalf, alleging that it had negligently enabled John Doe's criminal scheme of sextortion.[106] Here again, the district court dismissed the claims on § 230 grounds.[107] The court held that FOSTA—the 2018 sex trafficking carveout to § 230's legal shield—was inapplicable because while the complaint alleged that Omegle generally knew that the site was used by child predators to commit violations of intimate privacy like sextortion, it lacked allegations that the site had specific knowledge about the abuse in C.H.'s case.[108]

The common thread weaving through these cases is that the courts have sapped § 230's Good Samaritan concept of its meaning. Sites have no liability-based incentive to take down illicit material, especially if that material gets them extra clicks, while victims have no legal leverage to insist otherwise.[109] Platforms have no legal incentive to take steps to identify and remove predators before they threaten and extort children like C.H. into revealing their bodies on camera.[110]

In October 2022, the Supreme Court agreed to hear a case involving the reach of § 230(c)(1).[111] Whether that decision will impact the rulings that I have underscored is unclear. The Court is considering whether § 230(c)(1)'s immunity extends to a company's algorithmic recommendations of terrorism-

[104] *Id.* This is a common approach of intimate privacy invaders known as "sextortionists." As I explore in my book, sextortion—extortionate threats often demanding nude images or sex acts online—commonly impacts women, girls, and boys. CITRON, THE FIGHT FOR PRIVACY, *supra* note 17, at 31-33.

[105] *M.H.*, 2022 WL 93575, at *1.

[106] *Id.* at *2.

[107] *Id.* at *6.

[108] *Id.* at *6-7.

[109] *See* Citron, *Cyber Mobs*, *supra* note 86, at 1085-91. A few federal lower courts have been convinced by arguments made initially by Carrie Goldberg in the *Grindr* case yet rejected by the Second Circuit—that claims for defective design fall outside of § 230's legal shield because they concern what a content platform did rather than what their users said or published. *See, e.g.*, Lemmon v. Snap, Inc., 995 F.3d 1085, 1094 (9th Cir. 2021); A.M. v. Omegle.com, LLC, No. 21-cv-01674, 2022 WL 2713721, at *5 (D. Or. July 13, 2022).

[110] *See M.H.*, 2022 WL 93575, at *6.

[111] Two cases were consolidated and granted certiorari to address whether algorithmic recommendation systems that amplify terrorism-related content enjoy immunity from liability under § 230(c)(1) of the CDA. *See generally* Gonzalez v. Google LLC, 2 F.4th 871 (9th Cir. 2021), *cert. granted*, 143 S. Ct. 80 (2022), *and cert. granted sub nom.* Twitter, Inc. v. Taamneh, 143 S. Ct. 81 (2022); Taamneh v. Twitter, Inc., 343 F. Supp. 3d 904 (N.D. Cal. 2018), *rev'd and remanded sub nom. Gonzalez*, 2 F.4th 871. The Court could interpret § 230(c)(1) as strictly a mechanism to overrule *Prodigy*. The Court could look at the text and history and find that it provides a legal shield for sites that leave up defamation, eliminating strict publisher liability for defamation and defamation-adjacent claims. Predicting how the Court will rule is a tall task, one that I won't even begin to try.

related content, not whether the immunity covers a site's deliberate solicitation and toleration of harmful abuse.[112]

Digital platforms wielding enormous power bear few responsibilities or obligations, thanks to the judiciary's broad interpretation of § 230(c)(1).[113] As Mary Anne Franks persuasively explained in congressional testimony:

> Rather than encouraging the innovation and development of measures to fight online abuse and harassment, (c)(1) removes incentives for online intermediaries to deter or address harmful practices no matter how easily they could do so. It effectively grants powerful corporations a super-immunity, encouraging them to pursue profit without internalizing any costs of that this [sic] pursuit. It eliminates real incentives for tech corporations to design safer platforms or more secure products. Section 230(c)(1)'s preemptive immunity ensures that no duty of care ever emerges in a vast range of online scenarios and eliminates the incentives for the best positioned party to develop responses to avoid foreseeable risks of harm.[114]

This free pass creates a "'moral hazard,' ensuring that the multibillion-dollar corporations that exert near-monopoly control of the Internet are protected from the costs of their risky ventures even as they reap the benefits."[115]

Crucially, it isn't just the dominant tech companies that benefit. Small operations also profit from deliberately hosting illegality that destroys lives. Section 230 is why nonconsensual intimate imagery sites are hosted—and thriving—in the United States (and other countries where the risk of liability is low). After notorious nonconsensual pornography site Anon-IB was taken down by Dutch authorities in 2018, [116] it reappeared in early 2020, hosted on a server pinging from Chicago and San Francisco, a fact discovered by my then-research

---

[112] *See Gonzalez*, 2 F.4th at 913 ("There is no question § 230(c)(1) shelters more activity than Congress envisioned it would. Whether social media companies should continue to enjoy immunity for the third-party content they publish, and whether their use of algorithms ought to be regulated, are pressing questions that Congress should address."); *Taamneh*, 343 F. Supp. 3d at 908 ("Plaintiffs' claims are based not upon the content of ISIS's social media postings, but upon Defendants[']) *provision of the infrastructure* which provides material support to ISIS." (alteration in original)).

[113] *See* Rebecca Tushnet, *Power Without Responsibility: Intermediaries and the First Amendment*, 76 GEO. WASH. L. REV. 986, 987-88 (2008).

[114] Franks Testimony, *supra* note 57, at 5.

[115] *Id.* at 7.

[116] *See* Alexa Liautaud, *Revenge Porn Site Anon-IB Just Got Shut Down by Dutch Police*, VICE: NEWS (Apr. 26, 2018, 10:51 AM), https://www.vice.com/en/article/pax5bv/revenge-porn-site-anon-ib-just-got-shut-down-by-dutch-police [https://perma.cc/Z7DU-SZZ7]. For stories about the reappearance of the site and the damage wrought on people living in the United States, see Bethan Kapur, *An Army of Women Are Waging War on the Web's Most Notorious Revenge Porn Site*, MEL MAG., https://melmagazine.com/en-us/story/anon-ib-revenge-porn-badass-army [https://perma.cc/Q3HT-TKH8] (last visited Mar. 17, 2023).

assistant Rebecca Weitzel. In January 2022, the Guardian reported that the site was being hosted in Russia.[117]

Let me give you a sense of the number of sites devoted to intimate privacy violations. In 2013, my research assistants and I found forty sites devoted to nonconsensual intimate imagery.[118] By 2020, that number had grown to more than *9,500* sites.[119] These "[s]ites earn money by charging subscribers monthly fees, collecting ad revenue from people's clicks, or amassing personal data, which they can sell. (Even if people are paying subscription fees, their personal data is likely being collected and shared so the sites make money both ways.)"[120]

The profits are considerable. For instance, one online forum devoted to sharing "upskirt" or "creepshot" images requires users to pay to access certain content, "with 'premium tier' access for $10 or a 'superior tier' pass costing $20."[121] The Candid Forum, a similar site, had over 220,000 members as of 2018.[122] In 2017, another site, The Candid Board, had a similar fee structure and 180,000 members.[123] As of 2015, an upskirt photo site, which averaged 70 million daily page views, had an estimated worth of 100 million dollars.[124]

Section 230 has an "outsized impact" due to "the lack of geographic borders online."[125] Accordingly, "[w]hen non-US sites remove nonconsensual intimate images, perpetrators do the next best thing—they post the images on US sites."[126] Attorneys in Italy and law enforcers in South Korea have told me that they have no leverage to pressure sites to take down their clients' or citizens' images; courts in their countries have no power to order U.S.-hosted sites to do anything because they lack jurisdiction over them.[127] In the hopes of voluntary cooperation of online platforms, foreign law enforcement officers have asked

---

[117] Anna Moore, '*I Have Moments of Shame I Can't Control': The Lives Ruined by Explicit 'Collector Culture*,' GUARDIAN (Jan. 6, 2022, 1:00 AM), https://www.theguardian.com/world/2022/jan/06/i-have-moments-of-shame-i-cant-control-the-lives-ruined-by-explicit-collector-culture [https://perma.cc/JZ45-496F].

[118] CITRON, HATE CRIMES IN CYBERSPACE, *supra* note 4, at 51.

[119] CITRON, THE FIGHT FOR PRIVACY, *supra* note 17, at 71.

[120] *Id.* at 71-72.

[121] Joseph Cox, *Inside the Private Forums Where Men Illegally Trade Upskirt Photos*, VICE: MOTHERBOARD (May 8, 2018, 1:50 PM), https://www.vice.com/en/article/gykxvm/upskirt-creepshot-site-the-candid-forum [https://perma.cc/G8C2-MTD6].

[122] *Id.*

[123] *See* James Rodger, *Massive Hack Sees 180,000 The Candid Board Account Details Leaked*, BIRMINGHAM LIVE (Jan. 26, 2017, 8:07 AM), https://www.birminghammail.co.uk/news/midlands-news/massive-hack-sees-180000-candid-12506027 [https://perma.cc/MT22-V588].

[124] CITRON, THE FIGHT FOR PRIVACY, *supra* note 17, at 72.

[125] *Id.* at 89.

[126] *Id.*

[127] *See* Zoom Interview with Won-Mo Lee, S. Korean Dir. Gen. of Digit. Sex Crimes Info. Bureau Rev. Bd. (Nov. 23, 2020) (notes on file with author); Zoom Interview with Gian Marco Caletti, Rsch. Fellow, Free Univ. of Bozen-Bolzano (Dec. 7, 2020) (notes on file with author).

me to connect them with U.S. tech companies, but those efforts have not been particularly successful.[128] The images will likely remain up, no matter how much victims complain. Accordingly, "perpetrators can always torment victims on sites hosted in the United States, and victims' home countries can't do anything about it."[129]

Section 230 has been extended to other countries via trade agreements. The United States exported § 230 to Canada and Mexico through the United States-Mexico-Canada Trade Agreement, which went into effect on July 1, 2020.[130] Under the agreement, signatories must refrain from taking measures that would render interactive computer services liable for content created by others.[131] Tech industry lobbying groups pressed hard for this development, illustrating how companies like Alphabet (Google's parent company) and Meta (Facebook's and Instagram's parent company) think the immunity is integral to their continued profitability.[132] A federal law has further extended the immunity's reach by making foreign libel judgments unenforceable in the United States unless the judgment would comport with the First Amendment *and* § 230.[133]

---

[128] CITRON, THE FIGHT FOR PRIVACY, *supra* note 17, at 89; *see also* Zoom Interview with Won-Mo Lee, *supra* note 127; Zoom Interview with Julie Grant, Australian e-Safety Comm'r (Nov. 20, 2020) (notes on file with author).

[129] CITRON, THE FIGHT FOR PRIVACY, *supra* note 17, at 89. In writing my book, *The Fight for Privacy*, I interviewed more than sixty victims of intimate privacy violations. Fifteen of those victims hailed from Iceland, India, the United Kingdom, and Australia; they described struggling with getting their intimate images taken down from sites hosted in the United States and in other domains like Russia.

[130] *See* Agreement Between the United States of America, the United Mexican States, and Canada art. 19.17, July 1, 2020, Off. of the U.S. Trade Representative.

[131] *See id.* This provision does not apply to Mexico until July 1, 2023. *See id.* art. 19.17, annex A-1.

[132] *See* Han-Wei Liu, *Exporting the First Amendment Through Trade: The Global "Constitutional Moment" for Online Platform Liability*, 53 GEO. J. INT'L L. 1, 40 (2021) ("[T]ech companies have lobbied strongly to include [Article 19.17] immunity language in trade pacts . . . ."). It is true that Meta's Chief Executive Officer, Mark Zuckerberg, has appeared in advertisements saying that his company acknowledges the need for legislation that would make them responsible for user-generated content. *See* Mike Isaac, *Mark Zuckerberg's Call To Regulate Facebook, Explained*, N.Y. TIMES (Mar. 30, 2019), https://www.nytimes.com/2019/03/30/technology/mark-zuckerberg-facebook-regulation-explained.html. I would like to think that my pressing the company's safety officials to adopt my proposals (which I have done on several occasions, including once to Zuckerberg himself) moved the needle. But I don't think that I—or any advocates—had much to do with this approach. It seems more like a smokescreen than a commitment to reform. In advertisements and interviews, Meta has expressed a desire for regulation along the lines I have suggested while, at the same time, in the halls of Congress pressing for the exportation of § 230's legal shield. I am not buying its public relations campaign.

[133] *See* Securing the Protection of Our Enduring and Established Constitutional Heritage Act, Pub. L. No. 111-223, 124 Stat. 2380 (2010) (codified as amended at 28 U.S.C. §§ 4101-4105).

II. UNDERSTANDING THE STAKES FOR CIVIL RIGHTS AND CIVIL LIBERTIES

In response to efforts to reform § 230, advocates and scholars insist that intermediary liability is unnecessary because victims can sue their attackers. The Electronic Frontier Foundation ("EFF"), an esteemed civil liberties organization, argues that "Section 230 means that if you break the law online, you should be the only one held responsible."[134] Or, as Jason Kelley wrote for the organization, "Section 230 makes only the speaker themselves liable for their speech, rather than the intermediaries through which that speech reaches its audiences."[135] Practice, however, does not match theory.

As Herrick, C.H., and countless other people have experienced, victims of online abuse can't sue perpetrators because too few lawyers offer pro bono or low-cost services and because attackers usually lack funds from which the victim can recover (and thus contingency representation is unlikely).[136] They get no help from law enforcement officers often due to dismissive attitudes, victim blaming, or questioning the credibility of victims (usually a combination of those factors).[137] Victims have no means to stop the abuse, which impairs their ability to speak, work, network, and love.[138]

Thanks to § 230(c)(1), the law cannot reach platforms facilitating abuse even though the law would otherwise reach them if their operations occurred in physical space. It cannot be used to force platforms to internalize costs that they externalize and enable. Section 230(c)(1) ensures that parties best situated to minimize the damage—content platforms—have no legal reason to help victims and considerable reasons to ignore complaints because the abuse attracts attention and generates income.

This Part highlights the price that women, children, and minorities have paid, and will continue to pay, because the law has prevented them from seeking redress from the parties best situated to prevent or minimize the damage. It then discusses the consequences of poorly drafted reform, which has undermined civil rights and civil liberties while failing to make women and children safer.

---

[134] Jason Kelley, *Section 230 Is Good, Actually*, ELEC. FRONTIER FOUND. (Dec. 3, 2020), https://www.eff.org/deeplinks/2020/12/section-230-good-actually [https://perma.cc/5BQ6-W73M].

[135] *Id.*

[136] *See* CITRON, HATE CRIMES IN CYBERSPACE, *supra* note 4, at 122; CITRON, THE FIGHT FOR PRIVACY, *supra* note 17, at 90-93; Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753, 1792-93 (2019); Danielle Keats Citron & Mary Anne Franks, *Criminalizing Revenge Porn*, 49 WAKE FOREST L. REV. 345, 358 (2014); Citron, *Cyber Civil Rights*, *supra* note 2, at 90; Citron, *Privacy Injunctions*, *supra* note 80, at 971; Danielle Keats Citron, *Sexual Privacy*, 128 YALE L.J. 1870, 1929-30 (2019) [hereinafter Citron, *Sexual Privacy*].

[137] CITRON, THE FIGHT FOR PRIVACY, *supra* note 17, at 76-80.

[138] *Id.* at 80-81.

A. *Speech and Other Crucial Opportunities at Stake*

Politicians have called for the wholesale revocation of § 230.[139] On the campaign trail in 2019, President Joseph Biden took that position.[140] On the one hand, if we got rid of § 230's legal shield, platforms would face a range of potential claims for illegality hosted on their services, including claims related to intimate privacy violations, cyber stalking, and cyber harassment. Platforms wouldn't be strictly liable for user-generated content in most instances (except for the publishers of defamation and other claims with strict publisher liability).[141] Plaintiffs would have to prove their cases. The common law would develop in a wide range of areas, including claims related to the negligent or deliberate enablement of intimate privacy violations.[142]

Nonetheless, *Prodigy* taught us important lessons. Without the legal shield, platforms might curtail their moderation efforts to avoid any species of publisher (strict) or distributor (knowledge) liability.[143] On the flip side, they might err on the side of caution and take down all sorts of online activity, especially if people complained about it. They might "filter, block, or remove posts if their continued display" gained them little and risked much litigation.[144] There might be a rise of the "heckler's veto," where "people complain about speech because they dislike the speakers or object to their views, not because they have suffered actual harm."[145]

All of this is hypothetical. We don't know precisely how much online speech would be deterred or removed if § 230's legal shield disappeared. We can guess (after all, it is a counterfactual). Colleagues at EFF predict that public discourse

[139] *See* Adi Robertson, *Lots of Politicians Hate Section 230—But They Can't Agree on Why*, VERGE: POLICY (June 24, 2020, 10:28 AM), https://www.theverge.com/21294198/section-230-tech-congress-justice-department-white-house-trump-biden [https://perma.cc/2PRY-XMA8].

[140] *See* Cristiano Lima, *Biden: Tech's Liability Shield 'Should Be Revoked' Immediately*, POLITICO (Jan. 17, 2020, 10:56 AM), https://www.politico.com/news/2020/01/17/joe-biden-tech-liability-shield-revoked-facebook-100443 [https://perma.cc/G22R-HDPD].

[141] *See* David Greene, *Publisher or Platform? It Doesn't Matter.*, ELEC. FRONTIER FOUND. (Dec. 8, 2020), https://www.eff.org/deeplinks/2020/12/publisher-or-platform-it-doesnt-matter [https://perma.cc/ZHC7-YENY] (explaining liability for republishing someone else's defamatory statements).

[142] Danielle Keats Citron, *Mainstreaming Privacy Torts*, 98 CALIF. L. REV. 1805, 1836-43 (2011) (exploring potential for negligent enablement claims that might be brought against content platforms in world where § 230(c)(1) was not interpreted to bar them).

[143] *See* Greene, *supra* note 141 (describing difference between "distributor liability," which applies to entities like booksellers and newsstands, and "publisher liability," which applies "to those who engage[] with the other person's speech in some way, whether by editing it, modifying it, [or] affirmatively endorsing it").

[144] CITRON, HATE CRIMES IN CYBERSPACE, *supra* note 4, at 179.

[145] *Id.*

would be chilled.[146] We might see the over removal of speech about reproductive health, politics, and so on. We might, or we might not.

But as research and scholarship have shown, some things are *certain*. If we do nothing to fix § 230 then setbacks to civil rights and civil liberties will continue. Online abuse has destroyed victims' crucial life opportunities, including their ability to express themselves.[147] Young women, sexual and gender minorities, and nonwhite people are disproportionately the targets of intimate privacy violations or cyber harassment, which is often a perfect storm of threats, defamation, and privacy violations.[148] Intimate privacy violations—including the nonconsensual taking, extorting, manufacturing, or distributing of people's intimate images—inhibit victims' self-expression.[149] Researchers have found that victims of intimate privacy violations withdraw from online discourse, friendships, family, and romantic relationships.[150]

Cyber gender harassment inflicts similar damage.[151] As legal scholar and social scientist Jonathon Penney has found, women are statistically more chilled

[146] *See* Kelley, *supra* note 134. I am grateful to EFF's Civil Liberties Director David Greene and his team for engaging with my proposals and offering helpful feedback and criticism.

[147] ASIA A. EATON, HOLLY JACOBS & YANET RUVALCABA, CYBER C.R. INITIATIVE, 2017 NATIONWIDE ONLINE STUDY OF NONCONSENSUAL PORN VICTIMIZATION AND PERPETRATION 23-24 (2017), https://www.cybercivilrights.org/wp-content/uploads/2017/06/CCRI-2017-Research-Report.pdf [https://perma.cc/SAW2-7QNH] (demonstrating heightened negative mental health outcomes and higher level of psychological problems as result of nonconsensual porn); *see also* CITRON, HATE CRIMES IN CYBERSPACE, *supra* note 4, at 26 (describing how cyber harassment deprives victims of essential life activities, including self-expression).

[148] *See* CITRON, THE FIGHT FOR PRIVACY, *supra* note 17, at 39-40. Many victims have more than one marginalized identity. *Id.*

[149] *See* Citron, *Sexual Privacy*, *supra* note 136, at 1899; Mary Anne Franks, *"Revenge Porn" Reform: A View from the Front Lines*, 69 FLA. L. REV. 1251, 1293-94 (2017).

[150] *See, e.g.*, Nicola Henry & Anastasia Powell, *Beyond the 'Sext': Technology-Facilitated Sexual Violence and Harassment Against Adult Women*, 48 AUSTRALIAN & N.Z. J. CRIMINOLOGY 104, 114 (2015).

[151] *See* NICOLA HENRY, CLARE MCGLYNN, ASHER FLYNN, KELLY JOHNSON, ANASTASIA POWELL & ADRIAN J. SCOTT, IMAGE-BASED SEXUAL ABUSE: A STUDY ON THE CAUSES AND CONSEQUENCES OF NON-CONSENSUAL NUDE OR SEXUAL IMAGERY 6-12 (2020); *see also* CITRON, HATE CRIMES IN CYBERSPACE, *supra* note 4, at 196 (documenting silencing impact of cyber stalking); ASHER FLYNN, NICOLA HENRY & ANASTASIA POWELL, MORE THAN REVENGE: ADDRESSING THE HARMS OF REVENGE PORNOGRAPHY 5 (2016) (explaining harms caused by nonconsensual distribution of images, including stalking, humiliation, and loss of employment); Citron, *Cyber Civil Rights*, *supra* note 2, at 71-81 (providing examples of silencing impact of online abuse); Mary Anne Franks, *Unwilling Avatars: Idealism and Discrimination in Cyberspace*, 20 COLUM. J. GENDER & L. 224, 229 (2011) (noting women "shut down their blogs" and "take down social networking profiles" in response to sexism in cyberspace); Kira Allmann, Max Harris & Laura Hilly, *Old Problems, New Media: Revenge Porn and the Law*, OXFORD HUM. RTS. HUB: RIGHTSUP, at 32:00 (Oct. 11, 2015), https://ohrh.law.ox.ac.uk/old-problems-new-media-revenge-porn-and-the-law/ (interviewing Mary Anne Franks as she explains that nonconsensual publication of intimate images "has become the way to shut a woman up").

in their speech and engagement when targeted with online abuse.[152] A report issued by Data and Society 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)."[153]

Gendered harassment impedes women across society from expressing themselves. 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."[154] A 2019 study found that 28% of Finnish female municipal officials targeted with misogynistic hate speech reported being less willing than they would be otherwise to make decisions that might unleash online abuse.[155] 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.[156] 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.[157]

Cyber harassment has deterred women from considering political careers.[158] A 2017 study found that, of the Australian women surveyed, 80% of women over the age thirty-one reported that the media's mistreatment of female

[152] *See* Jonathon W. Penney, *Internet Surveillance, Regulation, and Chilling Effects Online: A Comparative Case Study*, INTERNET POL'Y REV., May 2017, at 1, 19.

[153] AMANDA LENHART, MICHELE YBARRA, KATHRYN ZICKUHR & MYESHIA PRICE-FEENEY, ONLINE HARASSMENT, DIGITAL ABUSE, AND CYBERSTALKING IN AMERICA 4 (2016), https://www.datasociety.net/pubs/oh/Online_Harassment_2016.pdf [https://perma.cc/XYF7-EBQG].

[154] *See* KRISTINA VAN SANT, ROLF FREDHEIM & GUNDARS BERGMANIS-KORĀTS, NATO STRATEGIC COMMC'NS CTR. OF EXCELLENCE, ABUSE OF POWER: COORDINATED ONLINE HARASSMENT OF FINNISH GOVERNMENT MINISTERS 50-51 (2021), https://stratcomcoe.org/publications/abuse-of-power-coordinated-online-harassment-of-finnish-government-ministers/5 [https://perma.cc/3XSS-N8HD]. Regrettably, the news keeps providing examples. After videos featuring the Finnish Prime Minister Sanna Marin enjoying herself at a party went viral, cyber gender harassment soon followed, even as male Australian Prime Minister Anthony Albanese was lauded for chugging a beer at a concert. Sravasti Dasgupta, *Australian PM Cheered for Chugging Beer at Concert as Finland's Sanna Marin Forced To Defend Partying*, INDEPENDENT (Aug. 24, 2022, 12:43 PM), https://www.independent.co.uk/news/world/australasia/anthony-albanese-beer-concert-sanna-marin-b2151604.html.

[155] *See* Leonie Cater, *Finland's Women-Led Government Targeted by Online Harassment*, POLITICO (Mar. 17, 2021, 2:00 PM), https://www.politico.eu/article/sanna-marin-finland-online-harassment-women-government-targeted/ [https://perma.cc/HUZ7-H99B].

[156] *Id.*

[157] *Id.*

[158] *See* CAROLINE CRIADO PEREZ, INVISIBLE WOMEN: DATA BIAS IN A WORLD DESIGNED FOR MEN 280-81 (2019).

politicians made it less likely that they would go into politics.[159] Intimate privacy violations have taken future careers in politics entirely off the table. This is especially true for sexual and gender minorities.[160] When "Ben" was working for a U.S. Senator as a legislative aide, he learned that his name and nude images had been posted on a gay revenge porn site.[161] He told me that he has always wanted to go into politics but thinks it is now impossible because "[t]he photos will always hang over [his] head, ready to be leveraged against [him]."[162]

Intimate privacy violations and cyber harassment make it impossible for people to feel safe, to get and keep jobs, and to take advantage of life's other key opportunities on equal terms.[163] When victims' nude images appear online or are emailed to colleagues and family, or they face rape and death threats and their home address is posted online next to the suggestion that they fantasize about rape, their lives are plagued with fear, worry, and pain. The emotional harm victims face is significant and long-lasting.[164] In a survey conducted by CCRI in 2012 and 2013, 93% of victims facing nonconsensual pornography reported having suffered "significant emotional distress."[165] Minor victims are especially vulnerable to depression and suicide.[166]

As I have maintained: "intimate privacy violations undermine equality and, ultimately, democracy" because victims experiencing privacy violations and cyber harassment "may internalize the invidious messages that society sends to women and minorities about their bodies and sexuality."[167]

> Intimate privacy violations reinforce destructive bigoted and gendered stereotypes. When a woman's nude photo appears in the search results of her name, she will be thought of as a damaged slut. If she is trans or queer, then her naked body may be viewed as disgusting, even degenerate. If she

[159] *Id.* at 281.

[160] *See* CITRON, THE FIGHT FOR PRIVACY, *supra* note 17, at 46.

[161] *Id.*

[162] *Id.*

[163] *See* CITRON, HATE CRIMES IN CYBERSPACE, *supra* note 4, at 19-20.

[164] CITRON, THE FIGHT FOR PRIVACY, *supra* note 17, at 41.

[165] CYBER C.R. INITIATIVE, INC., END REVENGE PORN: REVENGE PORN STATISTICS 1 (2014), https://www.cybercivilrights.org/wp-content/uploads/2014/12/RPStatistics.pdf [https://perma.cc/4TF5-RPPA].

[166] CITRON, THE FIGHT FOR PRIVACY, *supra* note 17, at 42. The following appears on the r/SuicideWatch subreddit:

> I already battle with being suicidal and having bipolar 2 disorder, but now my underage nude photo from high school has shown up on the site anon-ib. every time I email the website owner that it is an underage photo it gets taken down, and then someone else posts it again with my full name, town, school, and personal information. authorities can't do anything because it's based in another country. I can't believe I have to battle this and it heavily contributes to my suicide ideations on a daily basis . . . .

u/dogmama5, REDDIT (Nov. 11, 2020, 1:53 AM), https://www.reddit.com/r/SuicideWatch/comments/js3cm9/revenge_porn_site_anonib_is_ruining_my_life/ [https://perma.cc/M4A3-C39B].

[167] CITRON, THE FIGHT FOR PRIVACY, *supra* note 17, at 43.

> is Black, Asian, or Latina, then gender and racial stereotypes ferment further into a toxic brew, casting her as unworthy of privacy because her hypersexuality got her into the mess.[168]

These invidious attitudes have economic consequences, as employers have internalized these same messages.[169] According to a study conducted by Cross-Tab Marketing Services, nearly 80% of employers consult search engines to collect intelligence on job applicants, and 55% of those employers decline to interview or hire people because their search results featured "unsuitable photos."[170] Victims of nonconsensual pornography sometimes lose their jobs and are not able to find new ones.[171] Why would employers keep or hire people whose online reputations might reflect poorly on their businesses? This was Annie Seifullah's experience. The New York Department of Education dismissed Seifullah from her job as high school principal after her intimate images were published online because she had brought "widespread negative publicity, ridicule and notoriety" to the school system she worked for.[172]

B. *FOSTA's Cautionary Tale*

Before considering the particulars, we need to make sure that reform proposals do not come at the expense of the life opportunities of vulnerable people. Lessons from misguided changes to § 230 can help us avoid such problems. Recent reforms to § 230 have undermined civil rights and civil liberties while failing to secure greater safety for the most vulnerable among us. That is the story of FOSTA and its aftermath.[173]

In 2018, Congress sought to tackle the scourge of sex trafficking.[174] Lawmakers condemned classified advertising sites like Backpage.com that helped sex traffickers accomplish their crimes while profiting handsomely.[175] But the law Congress passed in response to this problem—FOSTA—has done little to curtail sex trafficking.[176] Instead, FOSTA has made life more difficult

---

[168] *Id.*

[169] *See id.* at 44.

[170] CROSS-TAB, ONLINE REPUTATION IN A CONNECTED WORLD 9-10 (2010), https://www.job-hunt.org/guides/DPD_Online-Reputation-Research_overview.pdf [https://perma.cc/W6DB-LDCQ].

[171] CITRON, THE FIGHT FOR PRIVACY, *supra* note 17, at 44.

[172] Annie Seifullah, *Revenge Porn Took My Career. The Law Couldn't Get It Back*, JEZEBEL (July 18, 2018), https://jezebel.com/revenge-porn-took-my-career-the-law-couldnt-get-it-bac-1827572768 [https://perma.cc/P3JJ-KC2G].

[173] *See* Fighting Online Sex Trafficking Act, Pub. L. No. 115-164, 132 Stat. 1253 (2018).

[174] *See* Quinta Jurecic, *The Politics of Section 230 Reform: Learning from FOSTA's Mistakes*, BROOKINGS (Mar. 1, 2022), https://www.brookings.edu/research/the-politics-of-section-230-reform-learning-from-fostas-mistakes/ [https://perma.cc/96LK-PBA6].

[175] *See id.*

[176] *See id.*

for prosecutors to pursue cases against sex traffickers and more dangerous for people engaged in consensual sex work.[177]

The situation leading up to FOSTA's enactment was stark. Thanks to § 230's legal shield, state and local prosecutors were barred from bringing charges against sites under state criminal laws for aiding and abetting sex trafficking.[178] State attorneys general could not seek civil penalties from sites that enabled sex trafficking.[179] Victims could not sue sites for tortiously facilitating their sexual exploitation.[180] Victims could not seek redress from the parties—advertising platforms like Backpage.com—in the best position to minimize their harm.

Congress got involved after members learned that Backpage.com had been shielded from liability thanks to § 230, even though the site's operators had *helped* sex traffickers get around prohibitions on sex ads posted by sellers under eighteen, which had resulted in the rape of women and children.[181] At the time, Backpage.com was the largest and most profitable outlet for posting sex ads in the United States.[182] Adults selling sex consensually often relied on the site to find clients.[183] But the site also hosted countless ads at the behest of sex traffickers.[184] Backpage.com assisted sex traffickers by regularly editing ads to remove keywords that would signal illegality such as "teenage" or "little girl."[185]

Members of Congress set out to tackle the problem of online sex trafficking and the role of platforms in enabling sexual exploitation. They did so even though new information had emerged showing Backpage.com helped sex traffickers write advertisements, which arguably would have jeopardized the website's ability to use § 230 as a shield in the future.[186]

Regrettably, legislative negotiations became messy. I worked with Republican and Democratic senators on the effort.[187] At every stage, advisers to lawmakers sought to ensure that FOSTA would tackle the problem in a narrow and effective way.[188] However, the bill continuously became more bloated and confusing. The result was a disappointment, to say the least.

As Quinta Jurecic explained in a report for the Brookings Institution, FOSTA is "a hodgepodge of a law with a number of moving pieces—few of which are

[177] *See id.*
[178] *Id.*
[179] *Id.*
[180] *Id.*
[181] *Id.*
[182] *Id.*
[183] *Id.*
[184] *Id.*
[185] *Id.*
[186] *Id.*
[187] I advised Senator Robert Portman's staff and worked closely with then-Senator Kamala Harris's tech adviser, Jonathan Mayer. Mayer and I offered lawmakers different approaches; none were adopted.
[188] *See* Jurecic, *supra* note 174.

clearly defined."[189] FOSTA has four main parts. First, it created a new federal crime of "own[ing], manag[ing], or operat[ing]" a platform "with the intent to promote or facilitate the prostitution of another person," codified at 18 U.S.C. § 2421A.[190] Under this provision, more severe penalties are available if the defendant either "promotes or facilitates the prostitution of 5 or more persons" or "acts in reckless disregard of the fact that such conduct contributed to sex trafficking."[191] Second, FOSTA expanded existing federal sex trafficking law to encompass "knowingly assisting, supporting, or facilitating" sex trafficking, codified at 18 U.S.C. § 1591.[192] Third, it created a new exception in § 230 for federal civil claims under 18 U.S.C. § 1595, which establishes a federal civil remedy for violations of 18 U.S.C. § 1591.[193] FOSTA also excluded from § 230(c)(1)'s legal shield state criminal prosecutions of conduct that would violate 18 U.S.C. § 1591 and conduct violating 18 U.S.C. § 2421A.[194] Accordingly, state prosecutors can bring criminal charges under coextensive state law.[195] Lastly, FOSTA allowed state attorneys general to bring federal civil claims under 18 U.S.C. § 1595 on behalf of state residents harmed by conduct violating 18 U.S.C. § 1591.[196]

Taken together, FOSTA was an attempt to increase the number of enforcers (state prosecutors, state attorneys general, and private litigants) and the available criminal and civil tools (with the addition of new exemptions to § 230) for fighting sex trafficking. More potential litigants and prosecutors, and more potential avenues for criminal and civil liability, what could go wrong? Unfortunately, quite a bit.

After FOSTA's passage, Quinta Jurecic and I teamed up to write about our concerns.[197] In a report for the Hoover Institution, we underscored the law's confusing language. We argued that "FOSTA's unclear '*knowingly* facilitating' language could perversely push platforms" to either engage in no moderation at all or, alternatively, to "engage in over-the-top moderation to prove their anti-sex-trafficking bona fides and to strengthen their argument that they did not

[189] *Id.*

[190] *Id.* (citing 18 U.S.C. § 2421A(a)).

[191] *Id.* (citing 18 U.S.C. § 2421A(b)).

[192] *Id.* (citing 18 U.S.C. § 1591(e)(4)).

[193] *Id.* (referencing 28 U.S.C. § 230(e)(5)(A)).

[194] *Id.* (referencing 28 U.S.C. § 230(e)(5)(B)-(C)).

[195] *Id.*

[196] *Id.*

[197] *See* DANIELLE CITRON & QUINTA JURECIC, PLATFORM JUSTICE: CONTENT MODERATION AT AN INFLECTION POINT 7 (2018) [hereinafter CITRON & JURECIC, PLATFORM JUSTICE], https://www.hoover.org/sites/default/files/research/docs/citron-jurecic_webreadypdf.pdf [https://perma.cc/A8DQ-HT55]. We have joined forces again to reflect upon the aftermath of FOSTA with the benefit of several years of hindsight. *See generally* Danielle Keats Citron & Quinta Jurecic, *FOSTA's Mess*, VA. J.L. & TECH. (forthcoming 2023) [hereinafter Citron & Jurecic, *FOSTA's Mess*] (on file with author).

knowingly facilitate such activity in any given case."[198] As we said then, "Overly aggressive moderation would likely involve shutting down hubs devoted to sex advertising or even websites that are known to host such advertising, even if the majority of users turn to the platform for other purposes."[199] We raised the specter of the total removal of sexual expression online—no matter if there were no connections to sex trafficking.[200] Sites might look to algorithmic filtering to solve the problem, which would result in the blocking or removal of "*anything* that relates to sex, including activities that have nothing to do with illegal sex trafficking."[201] Our concern was that "aggressive over-removal [of sexual expression] seem[ed] the most likely danger."[202] We were far from the only people raising this issue—sex workers and advocates had voiced these concerns from the beginning in lobbying against the legislation.[203]

We were, unfortunately, right to worry. Online platforms reacted as we feared—and worse. As sex worker advocate Kate D'Adamo explained in a Brookings Institution webcast, the internet was once a crucial space for sex workers to find safety, clients, and support.[204] Now, with FOSTA, sex workers have been denied online outlets that let them share tips with each other and to connect with clients in ways that made their lives and livelihoods better.[205] In the wake of FOSTA's passage, websites hosted in the United States soon began shuttering classified ads sections.[206] Craigslist explained that the legal risk was too great for it to maintain that corner of the site, on which users often posted solicitations for sex: "Any tool or service can be misused. We can't take such a risk without jeopardizing all our other services."[207]

Sex workers found themselves shut out of online spaces. Sites removed sexual content that might bear any relationship to sex work, including consensual sex work. As Kendra Albert and their coauthors have explained, platforms "frequently train their employees to 'identify' or profile people in the sex trades, and exclude those people from their services."[208] Indeed, after FOSTA's passage, "male escorting sites shut down or sharply limited access."[209] One sex worker said that his website provider cut off its services with no explanation,

---

[198] CITRON & JURECIC, PLATFORM JUSTICE, *supra* note 197, at 7 (emphasis added).

[199] *Id.*

[200] *Id.*

[201] *Id.*

[202] *Id.*

[203] *Id.* at 5-6.

[204] *See The Future of Section 230 Reform*, BROOKINGS, at 30:30 (Mar. 14, 2022, 10:00 AM), https://www.brookings.edu/events/the-future-of-section-230-reform/.

[205] *See* Jurecic, *supra* note 174.

[206] *See* CITRON & JURECIC, PLATFORM JUSTICE, *supra* note 197, at 6.

[207] *Id.*

[208] Kendra Albert, Emily Armbruster, Elizabeth Brundige, Elizabeth Denning, Kimberly Kim, Lorelei Lee, Lindsey Ruff, Korica Simon & Yueyu Yang, *FOSTA in Legal Context*, 52 COLUM. HUM. RTS. L. REV. 1084, 1151 (2021) (footnote omitted).

[209] Citron & Jurecic, *FOSTA's Mess*, *supra* note 197, (manuscript at 9).

causing him to lose the personal website and email he used to advertise to and communicate with clients.[210] Cloudflare, a web infrastructure company, pulled its services from Switter, a social network that had served as an "online refuge" for sex workers whose content had been removed from other parts of the web after FOSTA's passage.[211] After terminating its service to Switter, Cloudflare's general counsel Douglas Kramer explained that the termination was "related to our attempts to understand FOSTA, which is a very bad law and a very dangerous precedent."[212]

What about the potential upsides in combating sex trafficking? There have been no state prosecutions since FOSTA's passage.[213] Perhaps the reason is FOSTA's confusing language. In 2018, Jurecic and I wondered if the law's unclear wording "might discourage state and local prosecutors from expending scarce resources on enforcement actions."[214] It is not clear if state attorneys general have brought claims on behalf of residents for civil penalties.[215]

At the time of FOSTA's passage, the U.S. Department of Justice ("DOJ") voiced concerns about the law's lack of clarity.[216] In June 2021, the Government Accountability Office ("GAO") issued a report about FOSTA's impact. According to the GAO, federal prosecutors have not found much use for the criminal statute that FOSTA created, 18 U.S.C. § 2421A.[217] The DOJ has brought only two cases under this section, including one in which the defendant successfully moved for an acquittal on that count.[218] In the GAO's estimation, federal prosecutors may not have needed the additional criminal law given that federal prosecutors successfully brought federal racketeering and money laundering charges against online platforms.[219]

Is it possible that the absence of prosecutions demonstrates that FOSTA has appreciably deterred online sex trafficking? Unfortunately, it is unlikely. No data in the GAO report suggests a decrease in rates of online trafficking or sales in consensual sex.[220] Rather, the report "described an online sex trade that hasn't shrunken since April 2018, but instead fragmented across a number of platforms and apps, some of which moved overseas."[221] Immediately after U.S. law enforcement shut down Backpage.com, other websites, such as

[210] *Id.*
[211] *Id.*
[212] *Id.*
[213] *Id.*
[214] *Id.* (manuscript at 10).
[215] *See* Albert et al., *supra* note 208, at 1131.
[216] Citron & Jurecic, *FOSTA's Mess*, *supra* note 197, (manuscript at 10).
[217] *Id.*
[218] *Id.*
[219] *Id.* (citing Backpage.com as example of using existing federal law rather than FOSTA).
[220] *Id.*
[221] *Id.*

OneBackpage.com and Backpage.ly, hosted in Poland, emerged.[222] These sites bore the disclaimer, "FOSTA-SESTA—No Operator of this site reviews content or otherwise screens the content of this site."[223] Sites located outside the United States are beyond the reach of state or federal prosecutors.[224] These overseas platforms have refused to cooperate with U.S. federal agents.[225] Accordingly, the FBI has encountered greater difficulty in finding both sex trafficking victims and traffickers themselves.[226]

Have civil suits been successful? This aspect of FOSTA could be a helpful way to force sites enabling sex trafficking to internalize the costs that they have externalized onto victims.[227] However, similar to criminal cases, this has not been the case. While plaintiffs have brought claims seeking civil penalties against sites enabling their abuse, the statute's subpar drafting has left courts confused over a number of issues, "from what specific conduct prohibited by federal sex trafficking law is included in FOSTA's Section 230 carveout; to whether FOSTA also permits state civil claims concerning sex trafficking; to what exactly constitutes 'participation in a venture.'"[228] Courts disagree over whether FOSTA implicitly exempted state civil claims from § 230's legal shield.[229] Courts also disagree over whether plaintiffs are required to demonstrate the required elements of a criminal claim under 18 U.S.C. § 1591 to trigger the exemption from § 230's legal shield.[230]

Because these ambiguities have not been resolved, platforms will remain cautious, erring on the side of removing sexual expression. Kendra Albert and co-authors rightly argue that "though the exact legal applicability of FOSTA is speculative . . . even the threat of an expansive reading of these amendments has

[222] *Id.*

[223] *Id.* (manuscript at 10-11).

[224] *Id.* (manuscript at 11).

[225] *Id.*

[226] *Id.*

[227] *Id.*

[228] *Id.*

[229] *Id. Compare* Doe v. Reddit, Inc., No. SACV 21-768, 2021 WL 4348731, at *7 (C.D. Cal. July 12, 2021) (dismissing state law claims), *with In re* Facebook, Inc., 625 S.W.3d 80, 100 (Tex. 2021) (allowing state law claims to proceed because they were functionally equivalent to 18 U.S.C. § 1595).

[230] Citron & Jurecic, *FOSTA's Mess*, *supra* note 197, (manuscript at 11). *Compare* Doe v. Kik Interactive, Inc., 482 F. Supp. 3d 1242, 1251 (S.D. Fla. 2020) (dismissing plaintiff's claim against message service for facilitating sex trafficking because private rights of action for violations of 18 U.S.C. § 1591 are only allowed upon allegation that defendant subjectively knew that it had participated in sex trafficking venture), *and* J.B. v. G6 Hosp., LLC, No. 19-CV-07848, 2021 WL 4079207, at *15-16 (N.D. Cal. Sept. 8, 2021) (same), *with* Doe v. Twitter, Inc., 555 F. Supp. 3d 889, 925-26 (N.D. Cal. 2021) (refusing to dismiss plaintiffs' claims against Twitter on grounds that it was sufficient to allege that Twitter knew or should have known that plaintiffs were victims of sex trafficking and did not need to allege heightened knowledge), *abrogated by* Does 1-6 v. Reddit, Inc., 51 F.4th 1137 (9th Cir. 2022).

had a chilling effect on free speech."[231] Those costs are paired with the lack of any real benefits for sex trafficking survivors. This is a path to be avoided.

C. *Value of Reform*

With FOSTA in the backdrop, some believe that § 230 reform efforts are doomed to failure. They point to FOSTA's failure as a warning that additional reform may make matters worse. Without question, FOSTA's shortcomings serve as a roadmap of what *not* to do. But these shortcomings aren't a reason to ignore the hefty costs to civil rights and civil liberties in the here and now. They aren't a reason to be resigned to the status quo, especially considering the important expressive and practical potential of legal reform.

We should not ignore the suffering of victims of nonconsensual intimate imagery and cyber harassment. Victims contact us at CCRI and share their terrible predicaments with our hotline. They email Franks and I, explaining that law enforcement won't help, that they can't afford legal counsel, and that platforms hosting the abuse won't take down their images. In some cases, sites double down on the cruelty. In connection with my article, *Privacy Injunctions*, I interviewed a woman who faced nonconsensual pornography at the hands of an ex-partner who asked a site to take down their images.[232] Not only did the site refuse to take down the images, but it also posted her request and mocked her.[233]

If crafted well, § 230 reform and subsequent industry measures could send the clear message to victims that platforms and law protect intimate privacy. These reforms could signal to platforms that encouraging or failing to tackle intimate privacy violations and other online abuse is wrong.[234] Over the past two years, Jonathon Penney and I have been exploring how law can free victims to speak, and in 2019, we wrote about the expressive impact of cyber harassment laws.[235] We drew on Penney's important empirical evidence that cyber harassment laws have a salutary impact on people's online speech and engagement, particularly women.[236] Penney administered to 1,296 U.S.-based adults an original online survey that described to participants a series of hypotheticals.[237] One scenario concerned participants being made aware that the government had enacted a new law with tough civil and criminal penalties for

---

[231] Albert et al., *supra* note 208, at 1157.

[232] Citron, *Privacy Injunctions*, *supra* note 80, at 956-58.

[233] *Id.* at 957.

[234] Throughout my work, I have been guided by the powerful scholarship of John Goldberg and Benjamin Zipursky, both mentors, who show the importance of law's recognition of wrongs. *See, e.g.*, JOHN C.P. GOLDBERG & BENJAMIN C. ZIPURSKY, RECOGNIZING WRONGS 3-6 (2020).

[235] *See* Danielle Keats Citron & Jonathon W. Penney, *When Law Frees Us To Speak*, 87 FORDHAM L. REV. 2317, 2320 (2019) (describing evidence that cybersecurity protections encourage victims to "stay engaged online rather than retreating into silence").

[236] *See* Penney, *supra* note 152, at 19-20.

[237] Citron & Penney, *supra* note 235, at 2329-30.

cyber harassment.[238] Responses offered a range of insights.[239] Of the participants, 87% indicated that a cyber harassment law would have no impact or would make it more likely for them to speak and write online.[240]

Crucially, Penney's empirical research showed that a cyber harassment law might actually *encourage* online expression, particularly for women.[241] Penney's analysis revealed a gender disparity in response to the law: Female respondents were statistically more likely to engage online in response to the cyber harassment law in a variety of ways, including being more likely to share content online and more likely to engage on social network sites.[242] Penney and I joined together to argue that cyber harassment laws would have that salutary impact given law's expressive value.[243] Those laws would tell victims that their safety and online engagement are valued, that they will be protected, and that they matter.[244]

In 2021, Penney and I teamed up again 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 online platforms.[245] Researcher Alexis Shore joined us in that effort. Our preliminary findings suggest that both legal protections and market measures would engender trust in companies and the legal system such that individuals would be more inclined to engage in self-expression online.

---

[238] *Id.* at 2329.

[239] *Id.* at 2330.

[240] *Id.*

[241] *See id.* at 2331-32.

[242] *See id.* at 2331.

[243] In 2009, I wrote an article arguing that laws combating cyber gender harassment would have a crucial expressive value in telling victims that they were protected and that their life opportunities and suffering mattered. *See* Danielle Keats Citron, *Law's Expressive Value in Combating Cyber Gender Harassment*, 108 MICH. L. REV. 373, 407-14 (2009). I contended that those laws would free victims to speak. *See id.* at 413. At the time, I had no empirical evidence to support my argument—that is, until Jonathon Penney conducted his crucial studies. This is true of Penney's work more generally. His empirical research and insights on law's expressive value have been invaluable to information privacy scholars. *See, e.g.*, Ivan Manokha, *Surveillance, Panopticism, and Self-Discipline in the Digital Age*, 16 SURVEILLANCE & SOC'Y 219, 229 (2018) (citing Penney's empirical work on chilling effect of internet surveillance). What privacy scholars have long argued—including that government surveillance chills self-expression—Penney has so proven.

[244] *See* Jonathon W. Penney, *Online Abuse, Chilling Effects, and Human Rights*, *in* CITIZENSHIP IN A CONNECTED CANADA: A RESEARCH AND POLICY AGENDA 207, 213-14 (Elizabeth Dubois & Florian Martin-Bariteau eds., 2020).

[245] *See generally* Danielle Keats Citron, Jonathan Penney & Alexis Shore, Platforms, Privacy, and Power (unpublished manuscript) (on file with author). The Knight Foundation supported our empirical research project with a $75,000 grant. *See Knight Foundation Announces a Wide Range of New Grants To Support Research on Internet Governance*, KNIGHT FOUND. (June 26, 2020), https://knightfoundation.org/press/releases/knight-foundation-announces-a-wide-range-of-new-grants-to-support-research-on-internet-governance/ [https://perma.cc/4JNB-QUSZ].

In one experiment, participants were exposed to different protective privacy interventions. We found that participants who previously experienced forms of online abuse—including intimate privacy violations—were more inclined to engage in intimate expression after becoming aware of measures enacted to protect intimate privacy. That finding held when we told participants about legal interventions. It also held when we separately presented platform-based measures: Participants indicated that they would be even more likely to engage in intimate expression if *platforms* took measures to protect against intimate privacy violations.

In another experiment, our preliminary results show that both legal and platform-based intimate privacy measures have a positive impact on the trust that participants had vis-à-vis partners. After participants were told about legal and platform-based intimate privacy measures, trust became a strong predictor of intimate expression, and that predictive relationship was even stronger among participants who previously experienced online abuse. We found that both legal and platform measures increased trust in partners, especially among members of minority groups that disproportionately face online abuse and intimate privacy violations.

These findings suggest that legal- and platform-based intimate privacy measures have potential to promote trust in partners, law, and platforms. Although these findings are only preliminary, both studies suggested that people were more likely to engage in intimate expression if they knew that their intimate privacy enjoyed protection: This effect was even stronger among female victims of online abuse.

Section 230 reform has important potential, but it must be done right. The FOSTA experience teaches us that vague, poorly defined carveouts to § 230 can spur platforms to over moderate, with potentially disastrous effects for vulnerable people. Just as FOSTA has cut sex workers off from key resources needed to work safely, research indicates that queer content is often the first to vanish when platforms attempt to moderate more stringently.[246] Reforms must proceed carefully to avoid inadvertently pushing platforms to shut down entire swaths of speech.

III. A ROADMAP FOR FEDERAL LAWMAKERS

Humility is essential in reform efforts, lest they do more harm than good.[247] This admonition seems especially important in this fraught moment. In 2022, intimate privacy and free speech faced serious headwinds. In *Dobbs v. Jackson Women's Health Organization*,[248] the Supreme Court overturned decades of

[246] *See* Ari Ezra Waldman, Disorderly Content 33 (Aug. 16, 2021) (unpublished manuscript), https://ssrn.com/abstract=3906001 [https://perma.cc/TT7N-EAJ5].

[247] I have always tried to write in this fashion, but recent events and decisions suggest even greater caution.

[248] 142 S. Ct. 2228 (2022).

precedent in holding that individuals' reproductive autonomy and privacy enjoy no protection under the U.S. Constitution's Due Process Clause.[249]

In the shadow of *Dobbs*, state legislatures have not only criminalized early-term abortions but also have considered banning *speech* related to abortion rights.[250] Even if such speech bans (should they materialize) were struck down on First Amendment grounds,[251] platforms might err on the side of caution and filter or remove information related to reproductive health. Such chilling would be even more likely if Congress amended § 230 in a clumsy manner, as was the case with FOSTA.[252]

---

[249] *Id.* at 2258.

[250] *See* Hayley Tsukayama, *A Proposed Antiabortion Law Infringes on Free Speech*, SCI. AM. (Aug. 3, 2022), https://www.scientificamerican.com/article/a-proposed-antiabortion-law-infringes-on-free-speech/ [https://perma.cc/W74Z-25T6]. A South Carolina bill makes it a crime to "aid, abet, or conspire" with someone to procure an abortion, mirroring a National Right to Life Committee blueprint bill, which is designed to be copied by state lawmakers across the nation. *Id.* The South Carolina bill "allows the prosecution of any person who provides information regarding self-administered abortions or the means to obtain an abortion to a 'pregnant woman' or someone acting on 'behalf of a pregnant woman.'" *Id.* If Justice Clarence Thomas has his way, as his concurrence in *Dobbs* suggests, we might see efforts to criminalize speech related to contraception, parental decisions, or gay intimacy. *See generally* Danielle Keats Citron & Peter Kaplan, Data Handmaidens (Nov. 8, 2022) (unpublished manuscript) (on file with author).

[251] *See* Eisenstadt v. Baird, 405 U.S. 438, 456, 458-59 (1972) (Douglas, J., concurring) (agreeing with overturning of conviction of lecturer at Boston University for violating statute criminalizing sale of contraception because conviction was fundamentally based on defendant's First Amendment right to talk to faculty and students about his views on contraception).

[252] Supreme Court justices also have called for the reexamination of bedrock First Amendment protection for public figures being sued for defamation. *See* Adam Liptak, *Two Justices Say Supreme Court Should Reconsider Landmark Libel Decision*, N.Y. TIMES (July 2, 2021), https://www.nytimes.com/2021/07/02/us/supreme-court-libel.html (explaining Justices Thomas and Neil Gorsuch have expressed support for view that *N.Y. Times Co. v. Sullivan*, 376 U.S. 254, 269-71 (1964), and its progeny "warranted a reassessment"). By my lights, actual malice is a crucial safeguard in cases involving speech about matters of legitimate public importance. Dissenting from a denial of certiorari in a case involving the son of the Albanian President, Justice Gorsuch expressed concern that the actual malice standard has been stretched too far, that it immunizes people who spread vicious reputation-harming lies about private individuals' private matters because, as he noted, "ignorance is bliss." Berisha v. Lawson, 141 S. Ct. 2424, 2425-30 (2021) (Gorsuch, J., dissenting). He asserted that private individuals are prevented from bringing valid defamation claims because everyone is a public figure in an internet age. *Id.* at 2429. Gorsuch noted the decline of defamation suits and that one out of five favorable jury verdicts have their awards overturned in posttrial motion practice. *Id.* at 2428. He cited two cases to support the notion that dismissals stem from the application of the actual malice standard in cases where plaintiffs are purely private individuals suing for falsehoods about their private lives. *Id.* at 2429. But the dearth of cases may have far less to do with the misapplication of the actual malice standard—which is what Gorsuch is addressing—than the fact that private individuals can't bring defamation claims because they can't afford counsel, attackers have no money to

Section 230 can and should be reformed to protect intimate privacy and free expression, but we should proceed with care and caution. We need to rule out approaches that might inadvertently make things worse rather than better. In other words, we should not pursue reform if it would harm intimate privacy and self-expression more than it would help. In that spirit, Part III explores aspects of the legal shield that should remain and then proposes narrow reforms that target the problems highlighted in Part II.

A. *Preserving § 230(c)(2)*

First, I turn to the part of the legal shield that should be preserved: Section 230(c)(2), which provides an immunity for the "good faith" removal or filtering of user-generated content. As Representatives Cox and Wyden intended, this section leaves companies free to moderate activity because they are best situated to minimize harm.[253]

Some commentators and state lawmakers want to eliminate this provision and replace it with something radically different. In their view, law shouldn't encourage self-monitoring with a legal shield. To the contrary, it should ban content platforms from moderating user-generated content. Florida and Texas have taken steps in that direction.[254] Florida has prohibited big companies from removing, filtering, or downgrading journalists' speech[255] while Texas has

recover, or attackers cannot be identified. *See* CITRON, HATE CRIMES IN CYBERSPACE, *supra* note 4, at 164-65; Citron, *Cyber Civil Rights*, *supra* note 2, at 62-66.

[253] As Ryan Calo and I are exploring in a draft article, "good faith" provides broad discretion to "interactive computer services" to filter or remove harassing, obscene, or "otherwise objectionable" material. *See generally* Ryan Calo & Danielle Keats Citron, The End of Content Moderation (Nov. 8, 2022) (unpublished manuscript) (on file with author). The title of § 230(c) is, after all, "Protection for Good Samaritan blocking and screening of offensive material." 47 U.S.C. § 230(c). As the Ninth Circuit has noted, "good faith" involves a "subjective standard whereby internet users and software providers decide what online material is objectionable." *See, e.g.*, Enigma Software Grp. USA, LLC v. Malwarebytes, Inc., 946 F.3d 1040, 1044 (9th Cir. 2019) (citing Zango, Inc. v. Kaspersky Lab, Inc., 568 F.3d 1169, 1173 (9th Cir. 2009)). In the rare case that litigants challenge the immunity on the grounds that the removal of content was not in good faith, those efforts have been largely unsuccessful except in cases where the defendant's actions have been tantamount to anticompetitive behavior. *See id.* at 1052.

[254] The Florida law, entitled the Stop Social Media Censorship Act, was signed into law in May 2021 and was preliminarily enjoined by the Eleventh Circuit. *See* NetChoice, LLC v. Att'y Gen., Fla., 34 F.4th 1196, 1231 (11th Cir. 2022). Although a federal district court preliminarily enjoined Texas House Bill 20, the Fifth Circuit stayed the injunction, finding that the law comports with the First Amendment, which I will discuss below. *See infra* notes 268-71 and accompanying text.

[255] The Florida law defines "censor" to include "any action taken by a social media platform to delete, regulate, restrict, edit, alter, inhibit the publication or republication of, suspend a right to post, remove, or post an addendum to any content or material posted by a user." FLA. STAT. ANN. § 501.2041(1)(b) (West 2022). The law specifically prohibits a social media platform from "tak[ing] any action to censor, deplatform, or shadow ban a journalistic enterprise based on the content of its publication or broadcast." *Id.* § 501.2041(2)(j).

barred them from moderating content based on its viewpoint with some narrow exceptions.[256]

Requiring social media and other tech companies to host certain speakers or speech raises considerable First Amendment concerns. The First Amendment "constrains governmental actors and protects private actors."[257] A private party's ability to block or filter someone else's constitutionally protected speech is a part of First Amendment tradition. Under that tradition, unlike the government whose laws should not favor certain ideas or speakers over others, private parties are expected to shape norms around speech activity.[258]

Generally speaking, "the government can't tell a private person or entity what to say or how to say it."[259] In *Miami Herald Publishing Co. v. Tornillo*,[260] the Court struck down a statute requiring newspapers to publish op-eds by political candidates who wanted to respond to a newspaper editor's criticism.[261] The Court held that the plaintiff newspaper could not be forced to publish speech because they deserved to have editorial control over what appeared in their pages.[262]

Viewed through the lens of *Tornillo*, § 230(c)(2) safeguards a private party's ability to speak or affiliate (or not to speak or affiliate) with expression as it wishes: it "protects the rights of non-government actors to restrict, ignore, or

---

[256] The Texas law provides:

> (a) 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.

TEX. CIV. PRAC. & REM. CODE ANN. § 143A.002(a) (West 2021).

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

[258] As Frederick Schauer explains:

> [I]ndividual decisions about speech—preferring some ideas and information to others, placing one's property at the service of some ideologies and not others—are central to the concept of a marketplace of ideas. Ideas fail or succeed according to their ability to win support in free public debate. A private person participates in that debate when he contributes the use of his property to the proponents of certain ideas; that is an act of advocacy as surely as if he were disseminating the ideas himself.

Frederick F. Schauer, Hudgens v. NLRB *and the Problem of State Action in First Amendment Adjudication*, 61 MINN. L. REV. 433, 448-49 (1977) [hereinafter Schauer, *Problem of State Action*]; *see also* Frederick Schauer, *The Ontology of Censorship*, *in* CENSORSHIP AND SILENCING: PRACTICES OF CULTURAL REGULATION 147, 160-64 (Robert C. Post ed., 1998) [hereinafter Schauer, *The Ontology of Censorship*].

[259] *NetChoice, LLC*, 34 F.4th at 1203.

[260] 418 U.S. 241 (1974).

[261] *Id.* at 255.

[262] *Id.* at 258 (emphasizing First Amendment's protection of editorial control, including content decisions for newspapers).

refuse to associate with other people's speech."[263] When social media companies engage in content moderation, they make choices about the kinds of speech welcome on their services.[264] Since their inception, social media companies have been moderating subscribers' activities to match their corporate priorities.[265] At the same time, social networks and other content platforms aren't exactly like newspapers that investigate events and stories to enable discourse and self-governance. They perform some functions of the press, but their business model is data surveillance.[266]

Some scholars contend that because platforms wield enormous power over online speech, they are akin to common carriers that can be and are required to take all comers.[267] Indeed, the Fifth Circuit has embraced this argument in reversing the preliminary injunction of Texas House Bill 20, which bans large social media companies from removing user-generated speech based on a speaker's viewpoint or geography.[268] The court vacated the preliminary injunction, enabling the law to go into effect, on the grounds that the law does

[263] Matthew Ingram, *Talking About Section 230 with Mary Anne Franks*, GALLEY BY COLUM. JOURNALISM REV., https://galley.cjr.org/public/conversations/-MfdcaH845oRYp43dV3y [https://perma.cc/96EB-WYBB] (last visited Mar. 17, 2023).

[264] *See* Genevieve Lakier & Nelson Tebbe, *After the "Great Deplatforming": Reconsidering the Shape of the First Amendment*, LPE PROJECT (Mar. 1, 2021), https://lpeproject.org/blog/after-the-great-deplatforming-reconsidering-the-shape-of-the-first-amendment [https://perma.cc/D567-ADV7].

[265] For a brilliant argument about the nuance required in thinking through the analogies that ought to govern the question, see the Knight Institute's amicus brief in the challenge to Florida's social media law that required companies of a certain size to carry all content of politicians and media outlets. Brief of Amicus Curiae The Knight First Amendment Institute at Columbia University in Support of Plaintiffs-Appellees at 3, NetChoice, LLC v. Att'y Gen., Fla., 34 F.4th 1196 (11th Cir. 2022) (No. 21-12355). I am grateful to Katie Fallow for coming to talk to my free speech seminar about the litigation and the Knight Center's strategy and thinking.

[266] For a sample of the debate on what analogies best suit content platforms, see Genevieve Lakier, *The Problem Isn't the Use of Analogies but the Analogies Courts Use*, KNIGHT FIRST AMEND. INST. (Feb. 26, 2018), https://knightcolumbia.org/content/problem-isnt-use-analogies-analogies-courts-use [https://perma.cc/ZGE8-WP5J]; and Heather Whitney, *Search Engines, Social Media, and the Editorial Analogy*, KNIGHT FIRST AMEND. INST. (Feb. 27, 2018), https://knightcolumbia.org/content/search-engines-social-media-and-editorial-analogy [https://perma.cc/2XQJ-5WCE].

[267] *See, e.g.*, Eugene Volokh, *Treating Social Media Platforms Like Common Carriers?*, 1 J. FREE SPEECH L. 377, 382 (2021).

[268] 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." *Id.* at 446 (citing TEX. CIV. PRAC. & REM. CODE ANN. § 143A.001(1) (West 2021)). As my colleague Fred Schauer has noted, the term "censor" carries a negative connotation—it should not be used to cover decisions to refrain from speaking or removing information if doing so accords with expert opinion, as it does for bookstores and newspapers or for other ways that private parties express themselves by affiliating or disaffiliating themselves with certain ideas. *See* Schauer, *The Ontology of Censorship*, *supra* note 258, at 147.

not chill speech but rather chills censorship.[269] The court underscored that the platforms failed to "mount any challenge under the original public meaning of the First Amendment."[270] The court found that content platforms are more akin to telephone companies that must carry the content of all comers than military interviews on law school campuses whose conduct, not speech, was at issue when schools refused to host them on campus.[271]

The Fifth Circuit's panel decision sits on tenuous doctrinal ground. It may be cast into doubt or reversed if the Tenth Circuit sitting en banc or the Supreme Court weighs in. Social media companies and other content platforms are more akin to newspapers, bookstores, or entertainment companies than telephone companies. The Fifth Circuit baldly and incorrectly asserted that platforms "exercise virtually no editorial control or judgment."[272] Having worked with companies for more than a decade, reviewing their internal speech rules and community guidelines, they actively moderate online content, banning all sorts of speech, including protected expression like hate speech and misinformation.[273] They are far from telephone companies like AT&T that perform no role in deciding who may pay for their services. Crucially, the Fifth Circuit did not consider whether § 230(c)(2) preempted the statute, finding that the platforms had waived their right to raise the issue because they did not address it before the district court.[274]

In the end, the question isn't whether social networks can be regulated, but rather whether regulation would survive judicial scrutiny. The Florida statute was subjected to strict scrutiny review and has been temporarily enjoined.[275]

---

[269] *Paxton*, 49 F.4th at 494 ("We reject the Platforms' attempt to extract a freewheeling censorship right from the Constitution's free speech guarantee. The Platforms are not newspapers. Their censorship is not speech. They're not entitled to pre-enforcement facial relief. And HB 20 is constitutional because it neither compels nor obstructs the Platforms' own speech in any way.").

[270] *Id.* at 454. The Fifth Circuit's "originalist" approach has scant resonance with the Supreme Court's modern understanding of the First Amendment. *See, e.g.*, N.Y. Times Co. v. Sullivan, 376 U.S. 254, 269-71 (1964). For an exploration of debates around founding-era understandings of the First Amendment, see generally Jud Campbell, *Natural Rights and the First Amendment*, 127 YALE L.J. 246 (2017).

[271] *See Paxton*, 49 F.4th at 458-59, 461-62. The court took an overly narrow and formalistic view of what constitutes editorial discretion, finding that it involves the selection of content before it is hosted, not ex post removal of content. *See id.* at 464-65.

[272] *Id.* at 459.

[273] *See* CITRON, THE FIGHT FOR PRIVACY, *supra* note 17, at 151-52; CITRON, HATE CRIMES IN CYBERSPACE, *supra* note 4, at 168; Danielle Keats Citron & Helen Norton, *Intermediaries and Hate Speech: Fostering Digital Citizenship for Our Information Age*, 91 B.U. L. REV. 1435, 1462-82 (2011).

[274] *Paxton*, 49 F.4th at 468 n.24.

[275] The Eleventh Circuit held that the Florida law's prohibition of certain social media companies from deplatforming political candidates, prioritizing or deprioritizing posts "by or about" candidates, or removing posts by a "journalistic enterprise" triggered strict scrutiny review because its requirements interfered with private actors' speech. *See* NetChoice, LLC v. Att'y Gen., Fla., 34 F.4th 1196, 1206, 1209 (11th Cir. 2022). Judge Kevin Newsom, writing

Although the Fifth Circuit subjected the Texas law to intermediate scrutiny and struck down the preliminary injunction, that ruling likely will not stand the test of time and litigation.[276]

Beyond the doctrinal assessment is the normative one. By my lights, we are better off allowing social networks and websites to moderate users' activities, rather than requiring them to carry all content. We should remember the lesson, underscored by my colleague Frederick Schauer, that the First Amendment "was not intended to force such neutrality on private persons. Indeed, the ideals of free public debate and a marketplace of ideas presume that there will be partisanship and preference for some ideas over others."[277] If social media sites were not allowed to make choices about online content and had to host nonconsensual pornography, harassment, hate speech, and spam, few people would use them. (I certainly would not.) Cox and Wyden were right to make clear that companies should be shielded from liability if they filtered, blocked, or removed harassing and otherwise "offensive" speech.

B. *Excluding Certain Bad Actors from § 230(c)(1)*

Any effort to reform the under-filtering provision should begin by excluding sites that purposefully or deliberately solicit, encourage, or keep up intimate privacy violations, stalking, or harassment. Online abuse makes it impossible for victims to speak, work, and engage with the world around them. Congress should carve out sites and other content platforms whose raison d'être is online abuse that destroys people's ability to speak, work, and love.

Congress surely never meant to provide a free pass to sites whose purpose is intimate privacy violations and online assaults. Nonconsensual intimate imagery sites aren't trying to help strangers, as Good Samaritans are; they are the antithesis of responsible actors imagined by Cox and Wyden.[278] The purpose of § 230 was to incentivize responsible content moderation, not to shield from liability sites that make a mockery of the concept.[279]

Congress could add the following sentence at the end of § 230(c)(1): *Nothing in this section shall be construed to immunize from liability sites and other content platforms that purposefully or deliberately solicit, encourage, or keep up material that they know or have reason to believe constitutes stalking, harassment, or intimate privacy violations.*[280]

---

for a three-judge panel, found it substantially likely that the law's content-moderation restrictions would not survive strict scrutiny review. *Id.* at 1209.

[276] *See Paxton*, 49 F.4th at 494.

[277] Schauer, *Problem of State Action*, *supra* note 258, at 450.

[278] *See* CITRON, HATE CRIMES IN CYBERSPACE, *supra* note 4, at 173.

[279] *See supra* Section I.A (discussing passage of and motivations surrounding § 230).

[280] I have urged Congress to amend § 230 to exclude sites that "encourage cyber stalking or nonconsensual pornography and make money from its removal *or* that principally host cyber stalking or nonconsensual pornography." CITRON, HATE CRIMES IN CYBERSPACE, *supra* note 4, at 177.

Intimate privacy violations would be defined as the "nonconsensual photographing, filming, recording, digital fabrication, exploitation, extortion, sending, or disclosure of intimate images" (nude or partially nude images of breasts, buttocks, and genitals or images of sexual activities).[281] Cyber stalking and cyber harassment would be defined as a persistent course of conduct targeted at a specific individual that causes serious emotional distress or the fear of physical harm.[282]

This carveout is narrow. It only applies to sites that deliberately or purposefully solicit, encourage, or keep up online abuse that chills speech and ruins lives. It accords with § 230's policy goal to combat cyber stalking and harassment, which often include intimate privacy invasions as part of the abuse.[283]

The DOJ has proposed a broader carveout that would exclude from 230(c)(1)'s legal shield any site that "purposefully facilitates federal crimes."[284] Platforms might have difficulty assessing how the carveout would operate because federal criminal law covers a wide range of activity. My proposed abuse-specific exemption instead identifies the precise illegality exempted from the legal shield. It also covers intimate privacy violations that destroy lives and are not covered by federal criminal law.

Critics might argue that federal criminal law falls outside § 230's legal shield so the DOJ's proposed exemption would be duplicative. However, no federal prosecutor has ever brought a case against a website operator for aiding and

[281] CITRON, THE FIGHT FOR PRIVACY, *supra* note 17, at 142; *see also* Citron, *Sexual Privacy*, *supra* note 136, at 1947-48. The definition could also extend to the nonconsensual posting of people's health and medical information, as I fear that we might soon see sites devoted to outing women and girls who obtained abortion care. *See* Danielle Keats Citron, *The End of* Roe *Means We Need a New Civil Right to Intimate Privacy*, SLATE: FUTURE TENSE (June 27, 2022, 11:36 AM), https://slate.com/technology/2022/06/end-roe-civil-right-intimate-privacy-data.html.

[282] *See* CITRON, HATE CRIMES IN CYBERSPACE, *supra* note 4, at 143-45.

[283] Federal criminal law does prohibit both cyber stalking and cyber harassment, though it has yet to address intimate privacy violations in a comprehensive manner. *See* CITRON, THE FIGHT FOR PRIVACY, *supra* note 17, at 131-47. The federal Video Voyeurism Act only covers intimate privacy violations occurring on federal lands and parks. 18 U.S.C. § 1801 (limiting criminal liability to privacy violations committed "in the special maritime and territorial jurisdiction of the United States"). Thanks to the advocacy and drafting skills of Mary Anne Franks, Congress has passed a bipartisan bill to criminalize the nonconsensual disclosure of intimate images, but it does not expressly cover deepfake sex videos and other manufactured intimate images. *See* Tal Kopan & Megan Cassidy, *Should Revenge Porn Be a Federal Crime? Here's What the 190 Bay Area Cases Reveal*, S.F. CHRON. (May 7, 2022, 1:45 PM), https://www.sfchronicle.com/bayarea/article/revenge-porn-law-17143216.php.

[284] *See* U.S. DEPT. OF JUST., SECTION 230—NURTURING INNOVATION OR FOSTERING UNACCOUNTABILITY? (2020), https://www.justice.gov/file/1286331/download [https://perma.cc/F5JF-AV5S]. I worked closely with the DOJ attorneys who drafted this report.

abetting cyber stalking and the DOJ has had plenty of time and opportunity to bring such cases.[285]

Let's return to the Candid Forum for an illustration of a site that would fall outside the legal shield under my proposal.[286]

> Running atop every page on the Candid Forum is a cartoon scene of two men wearing camouflage and using binoculars to watch bikini-clad women on the beach. The site's front page says that "[s]exy up-skirts have never been easier to capture thanks to cell phone cameras, so we're getting more submissions than ever."[287]

As *Vice*'s Joseph Cox found, thousands of threads included "upskirt images of potentially underage girls wearing school uniforms, women shopping, and women on public transport, with the photographer often stalking the target for extended periods of time."[288] The Candid Forum would have difficulty suggesting that it did not deliberately encourage or solicit users to take and share intimate privacy violations. The carveout that I propose would allow victims to sue bad actor sites like the Candid Forum.

Of course, a narrow carveout for bad actors like the Candid Forum would not mean that it would be strictly liable for intimate images posted by users. Individuals whose intimate images appear on the site without their permission would have to bring legally cognizable claims against the site. They would have the burden of proving those claims by a preponderance of the evidence. They should be given a chance to do so.

Might bad actors move overseas where U.S. plaintiffs could not sue them and where the host countries might tolerate them? Perhaps, but moving overseas might undermine their operational viability. As Nicholas Nugent explains, foreign sites that use no local intermediaries are at a disadvantage.[289] The farther data must travel to reach users, the longer it takes to access that information, and the less people will be inclined to use a slow-loading site.[290] Moving offshore also subjects sites to foreign regulation, which can be even more onerous than U.S. law. Foreign states are less protective of free speech. Indeed, as Nugent noted to me, "Russia might be a great place to operate revengeporn.ru until the FSB demands access to the operator's servers and customer lists."[291] In short,

---

[285] *See* Citron, *Sexual Privacy*, *supra* note 136, at 1929-30 (discussing law enforcement's unwillingness "to expend scarce resources on combating sexual-privacy invasions").

[286] *See supra* note 122 and accompanying text.

[287] CITRON, THE FIGHT FOR PRIVACY, *supra* note 17, at 72.

[288] Cox, *supra* note 121.

[289] *See* Nicholas J. Nugent, *The Five Internet Rights*, 98 WASH. L. REV. (forthcoming 2023) (manuscript at 44, 50-52), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4292196 [https://perma.cc/HZ63-49M9].

[290] *See id.*

[291] E-mail from Nicholas Nugent, Program Dir., Karsh Ctr. for L. & Democracy, to Danielle Keats Citron, Professor of L., Univ. of Virginia Sch. of L. (Aug. 8, 2022) (on file with author).

excluding bad actors from the legal shield would likely have an impact on sites that deliberately host abuse for profit—they would have to internalize some of the costs that they externalize, a profoundly important step forward.

C. *Setting a Particularized Duty of Care*

Clarifying the boundaries of § 230 to exclude bad actors is necessary but not sufficient for meaningful reform. As I will explore, I have revised my thinking over the past five years. I have fine-tuned my approach to reflect lessons from FOSTA and mainstream hostility to the liberty and speech rights of the vulnerable.

Here is an overview of my proposal: Congress should reform § 230(c)(1) to make the legal shield conditional in certain circumstances. The "leave up" provision should be conditioned on a duty of care in matters involving intimate privacy violations, cyber stalking, and cyber harassment. In such cases, if platforms want to assert a § 230(c)(1) defense, they must prove that they took reasonable steps to address intimate privacy violations, cyber stalking, or cyber harassment, even if their efforts failed to address the abuse in a particular case. Congress should lay out the specific obligations involved in "reasonable steps," which I specify below.

Let me explain the evolution of my thinking. In 2017, I teamed up with national security expert and *Lawfare* founder Benjamin Wittes to write statutory language that would begin to capture a duty of care for content platforms invoking § 230(c)(1)'s legal shield.[292] As we proposed, § 230(c)(1)'s legal shield should be conditioned on a showing that a content platform had taken "*reasonable steps to prevent or address unlawful uses of its services*."[293] Under our proposal, "unlawful uses" referred to activities that would violate existing law—criminal and civil law, whether statutes, regulations, or the common law.[294] Our proposal did not specify what "reasonable steps" entailed.[295]

Our proposal, if adopted, would change the way that the defense is currently adjudicated. Under current law, defendants can move to dismiss claims early in the litigation on the ground that plaintiff's claims aren't legally cognizable given the legal shield afforded by § 230(c). In assessing motions to dismiss, courts accept the plaintiff's facts as true. As we have seen in Part I, courts routinely dismiss claims against platforms, reasoning that even if everything plaintiffs say is true, defendant platforms cannot be sued for content posted by users.[296]

By contrast, the "reasonable steps" approach that Wittes and I proposed would preclude early dismissals of lawsuits because plaintiffs would surely dispute the defendant's affirmative defense that it took reasonable steps to address the

[292] *See* Citron & Wittes, *supra* note 4, at 418-19.

[293] *Id.* at 419; *see also* Citron & Franks, *supra* note 4, at 71.

[294] *See* Citron & Wittes, *supra* note 4, at 419.

[295] *Id.*

[296] *See supra* Part I.

illegality at issue.[297] The dispute could not be resolved at the motion-to-dismiss stage because the "reasonable steps" inquiry would involve a mixed question of law and fact. In discovery, facts would be unearthed about what steps the defendant took to address the illegality at issue (facts) and whether experts thought those steps satisfied the reasonableness inquiry (legal). Juries would be tasked with assessing both questions: first, what activity the defendant took to address the illegality at issue, and second, whether those actions amounted to reasonable steps.

Our proposal would introduce new challenges, responsibilities, and costs. Content platforms would have to take steps to address illegality if they wanted to invoke and earn the legal shield's protection.[298] Under that proposal, there would be no predetermined roadmap for what would constitute "reasonable steps," but platforms would have some guidance. For the last decade, the tech industry has developed intricate rules and processes around content that they consider off limits, including illegality like harassment, stalking, and nonconsensual pornography. A professional organization—the Trust and Safety Professional Association—is devoted to providing best practices on content moderation.[299] In my individual capacity and on behalf of CCRI (alongside our president, Mary Anne Franks), I have worked with tech companies to develop those best practices.[300]

---

[297] If plaintiff's counsel had a nonfrivolous basis to make that argument.

[298] Or as Gilad Edelman has put it, companies faced with a reasonable steps approach would have to compete on safety. Edelman, *supra* note 1.

[299] *See* TR. & SAFETY PRO. ASS'N, https://www.tspa.org/ [https://perma.cc/3GMZ-VMMX] (last visited Mar. 17, 2023). The Trust and Safety Professional Association's Executive Director, Charlotte Willner, joined the organization after working in online safety for fifteen years, first at Facebook and then at Pinterest. *Our Team*, TR. & SAFETY PRO. ASS'N, https://www.tspa.org/about-tspa/team/ [https://perma.cc/6HMM-3SW7] (last visited Mar. 17, 2023).

[300] I began working with Twitter in 2009. For a brief overview of my work with Twitter and the company's evolving content moderation policies before the company's takeover by Elon Musk, see Danielle Keats Citron & Hany Farid, *This is the Worst Time for Donald Trump To Return to Twitter*, SLATE: FUTURE TENSE (Nov. 20, 2022, 7:59 PM), https://slate.com/technology/2022/11/trump-returning-to-twitter-elon-musk.html [https://perma.cc/2G47-BLWQ]. In 2012, I also began working with Facebook and Microsoft, followed by work with Spotify, Bumble, TikTok, Snapchat, and Twitch. I have also served on working groups spearheaded by the Anti-Defamation League ("ADL") to devise best practices related to hate speech and stalking, and I have offered testimony in the United States and the United Kingdom on the topic. *See Best Practices for Responding to Cyberhate*, ANTI-DEFAMATION LEAGUE, https://www.adl.org/best-practices-for-responding-to-cyberhate [https://perma.cc/YMH2-VM96] (last visited Mar. 17, 2023) (describing recommendations of ADL's Working Group on Cyberhate for providers and internet community to combat cyberhate); CITRON & JURECIC, PLATFORM JUSTICE, *supra* note 197, at 4-16 (discussing Hoover Institution's Working Group on National Security, Technology, and Law's legislative solutions for abuse of website platforms' services); DANIELLE KEATS CITRON, WRITTEN TESTIMONY OF DANIELLE KEATS CITRON: "MISOGYNISTIC CYBER HATE SPEECH" 1 (2011), https://digitalcommons.law.umaryland.edu/cgi/viewcontent.cgi?article=2143&context=fac_pubs [https://perma.cc

Our proposal would incentivize firms to act responsibly, as Cox and Wyden wanted, but it also would entail considerable uncertainty. Firms would not have a playbook of what to do because the proposal lacked the steps involved in a duty of care. As my colleague Kenneth Abraham has highlighted, when statutes don't lay out the obligations entailed in reasonable care, juries must create them.[301] Under our proposal, firms wouldn't have been able to predict with certainty whether they were taking the appropriate precautions against different types of illegality. Over time, firms would learn from experiences with litigation, but they would not know for sure if their actions fell inside or outside the safe harbor.

Wittes and I did not grapple with this uncertainty, and FOSTA provides important lessons in moving forward. FOSTA illustrates the fallout of legislative reform that creates uncertainty and re-introduces the moderator's dilemma.[302] On the one hand, and this is powerfully important, we might see strong protections against illegality. On the other, and this is equally concerning in a post-*Dobbs* world, we might see overly aggressive steps that do more harm than good.

Given the vulnerability of civil rights and civil liberties in the wake of *Dobbs* and the risk of overreaction that could compound those vulnerabilities, I have refined my thinking. Rather than a condition that would apply to all (unspecified) illegality, a duty of care requirement should only apply to claims involving intimate privacy violations, cyber stalking, or cyber harassment. We know that those forms of online abuse are exacting steep costs to civil rights and liberties. That way, platforms will have a legal incentive to design content moderation policies and practices to address those problems.

Rather than an unguided duty of care, lawmakers should specify the obligations required, drawing on key lessons from the trust and safety field. In its definition of "reasonable steps," Congress should lay out the following duties. First, platforms should give individuals a way to report intimate privacy violations, cyber stalking, or cyber harassment.[303] Second, they should have

/U2DF-HZHM] (testifying before U.K. Inter-Parliamentary Committee on Antisemitism for Task Force on Online Hate about misogynistic cyber hate).

[301] Kenneth S. Abraham, *The Trouble with Negligence*, 54 VAND. L. REV. 1187, 1194 (2001). As Abraham explains:

> [C]ivil juries that create norms in unbounded negligence cases have greater formal legitimacy as sources of norms than industry customs and professional standards, but no greater, and arguably less, legitimacy than statutes. In contrast, custom and professional standards have a measure of market accountability that civil juries lack. In these respects, the legitimacy comparison between bounded and unbounded cases is inconclusive.

Id. at 1208. On the other hand, where there are preexisting norms, "[o]utcomes are likely to be more predictable" and "recurring cases are more likely to be treated alike." *Id.* at 1209.

[302] *See supra* Section II.B (discussing FOSTA and its limitations).

[303] *See, e.g.*, Copia Inst., *Newsletter Platform Substack Lets Users Make Most of the Moderation Calls (2020)*, TR. & SAFETY FOUND. (Apr. 2021), https://trustandsafetyfoundation.org/blog/newsletter-platform-substack-lets-users-make-most-of-the-moderation-calls-2020/ [https://perma.cc/A6RY-ABWQ].

processes that enable them to address those reports.[304] Third, they should endeavor to prevent intimate privacy violations, cyber stalking, or cyber harassment from recurring on their services.[305] Fourth, platforms should be subject to certain minimum logging requirements so that individuals who sue users for online abuse can get access to the information needed to identify their abusers and prove their case in court.[306] Fifth, platforms should remove, delete, or otherwise make unavailable intimate images, real or fake, that have been posted or shared without the subject's consent.[307] This would take a page from the notice-and-takedown requirement in the Digital Millennium Copyright Act.[308] Last, platforms should remove or block content that courts have adjudicated as unlawful.[309]

[304] *See* Citron & Norton, *supra* note 273, at 1468-76.

[305] *See* CITRON & JURECIC, PLATFORM JUSTICE, *supra* note 197, at 9; *see also* Copia Inst., *Chatroulette Leverages New AI To Combat Unwanted Nudity (2020)*, TR. & SAFETY FOUND. (Jan. 2021), https://trustandsafetyfoundation.org/blog/chatroulette-leverages-new-ai-to-combat-unwanted-nudity-2020/ [https://perma.cc/3VLV-RH92]; Copia Inst., *Discord Adds AI Moderation To Help Fight Abusive Content (2021)*, TR. & SAFETY FOUND. (Oct. 2021), https://trustandsafetyfoundation.org/blog/discord-adds-ai-moderation-to-help-fight-abusive-content-2021/ [https://perma.cc/U3DH-M6M8]. Another example of firms taking steps to prevent online abuse from recurring was Facebook's policies to combat nonconsensual pornography. Facebook took hashes of material that in its view constituted nonconsensual pornography (a violation of its terms of service) so that it could prevent the material from reappearing on Facebook or Instagram. The hash approach uses the PhotoDNA model created by Hany Farid and Microsoft. I highlighted such industry developments in my article *Sexual Privacy*. *See* Citron, *Sexual Privacy*, *supra* note 136, at 1955 & n.559, 1956-58. In a crucial step forward, TikTok and Bumble have joined the effort begun by Meta (formerly known as Facebook) to share access to a database containing hashes of nonconsensual intimate imagery so that it will not appear on their platforms. *See* Olivia Solon, *TikTok and Bumble Join Fight To Stop Spread of 'Revenge Porn*,*'* BLOOMBERG (Dec. 1, 2022, 6:45 AM), https://www.bloomberg.com/news/newsletters/2022-12-01/tiktok-and-bumble-join-fight-to-stop-spread-of-revenge-porn. The U.K.-based organization StopNCII.org runs the database on behalf of Facebook and Instagram and now TikTok and Bumble. *See id.*

[306] Citron, *Cyber Civil Rights*, *supra* note 2, at 123-24. In his important new book, Jeffrey Kosseff explores the role that traceability anonymity can and should play while protecting the commitment to anonymous speech. JEFFREY KOSSEFF, THE UNITED STATES OF ANONYMOUS: HOW THE FIRST AMENDMENT SHAPED ONLINE SPEECH 5, 233-72 (2022) [hereinafter KOSSEFF, THE UNITED STATES OF ANONYMOUS] ("[T]o continue the US tradition of anonymity, lawmakers should supplement the First Amendment rights and anonymity technology with robust privacy laws that restrict the ability of private parties and the government to collect, use, and share identifying information.").

[307] *See* Citron, *Privacy Injunctions*, *supra* note 80, at 968-69.

[308] *See* Derek E. Bambauer, *Exposed*, 98 MINN. L. REV. 2025, 2055 (2014).

[309] This would solve the problem raised in *Hassell*, though that case involved defamation. *See supra* note 80 and accompanying text. Jeffrey Kosseff has called for a "modest amendment to Section 230 to clarify that it does not block an order to take down material that has been adjudicated to be defamatory." KOSSEFF, THE UNITED STATES OF ANONYMOUS, *supra* note 306, at 137.

To enable the duty of care to evolve to address emerging problems and solutions, Congress should authorize the Federal Trade Commission ("FTC") or another expert agency to engage in rulemakings to flesh out the duty of care requirement. An expert agency could ensure that as technology and content moderation practices evolve, so, too, will commitments to taking reasonable steps to address online abuse. As new kinds of privacy violations inevitably emerge, so will new strategies for tackling them.

Let's return to the lawsuit that Herrick brought against Grindr.[310] We don't have a factual record about Grindr's practices, but we do know a few things from plaintiff's pleadings. Grindr had notice (more than 100 emails) that someone was using the dating app to impersonate Herrick and to post his nude images without consent yet still it did not change its design to enable it to track and remove such individuals. Grindr seemingly did nothing to prevent the abuse from recurring—it could have, but refused to, block IP addresses. If my reform proposal were in place, Herrick would likely have argued that Grindr's § 230(c)(1) defense was unavailing; the parties would have engaged in discovery about whether Grindr's took the prescribed statutory steps; the jury would have assessed whether Grindr warranted the legal shield. But none of this happened: the court dismissed Herrick's lawsuit and Grindr never had to respond to Herrick's claims.

What about Omegle?[311] Recall that the site was sued by C.H. and her parents after the site connected C.H. with a predator who violated her intimate privacy by extorting nude images from her. The lawsuit concerned the site's enablement of intimate privacy violations (sextortion), but of course there was no discovery on the question of whether the website had any processes to address intimate privacy violations and cyber harassment. There is much that we don't know because the lawsuit was dismissed before any facts could be unearthed about the site's activities. That would not be the case if Congress reformed § 230 along the lines that I have laid out.

A duty of care would require discovery and fact finding on the question of whether platforms like Grindr and Omegle followed the prescribed reasonable steps.[312] That would entail litigation costs. Platforms would not be able to obtain

[310] *See supra* notes 86-100 and accompanying text.

[311] *See supra* notes 101-08 and accompanying text.

[312] Might Congress be able to curtail protracted litigation by instructing courts to treat a conditional § 230(c)(1) defense like a personal jurisdiction defense? For personal jurisdiction challenges, courts, not juries, resolve the question. Motions to dismiss on personal jurisdiction grounds involve limited discovery on factual questions like whether the defendant purposefully availed itself of a forum, as required by the Fourteenth Amendment's Due Process analysis, or engaged in substantial commerce with the state, as required by long-arm statutes. *See, e.g.*, Danielle Keats Citron, *Minimum Contacts in a Borderless World: Voice over Internet Protocol and the Coming Implosion of Personal Jurisdiction Theory*, 39 U.C. DAVIS L. REV. 1481, 1501-29 (2006). Congress might be able to give courts the role of determining if platforms earned the legal shield in cases involving stalking, harassment, or intimate privacy violations (so long as doing so would not violate the Seventh Amendment's guarantee of a jury trial). That would be possible if that determination could be understood as

early dismissals via motions to dismiss if claims involved intimate privacy violations, cyber stalking, or cyber harassment. In cases involving such claims, companies might settle some claims, even if they are weak, to avoid litigation costs. They also might settle claims with victims who have been wronged, who have suffered grave harm, and who deserve compensation. (All other cases would operate as they do now with platforms moving to dismiss on § 230 grounds with no need to prove to a jury that they engaged in reasonable steps to address the wrongful activity alleged in the suit.)

In the shadow of potential litigation, some platforms might automatically take down content identified as intimate privacy violations, cyber stalking, or cyber harassment. Some of that content might constitute life-damaging, speech-undermining online abuse; its removal would minimize the damage suffered by victims. In other instances, reported content might not constitute online abuse. In such a case, the heckler's veto would be at play.[313] Speech might be removed not because it constitutes destructive abuse but because someone does not like it. The heckler's veto is a risk worth taking to minimize online abuse that silences victims' speech and destroys their life opportunities.

This is a crucial point: research that Penney, Shore, and I are doing suggest that a duty of care might result in even more speech.[314] 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.

Some argue that the reasonable steps approach favors the dominant platforms because they already have content moderation policies, processes, and staff in place and can afford to take the reasonable steps articulated above, whereas startups would struggle to pay for any of it.[315] It is a fair point. But it is worth underscoring that startups can enable destructive abuse just as the dominant platforms can. We should not act as if smallness means less destruction. *It does not*. Anon-IB likely has few employees (it's hard to know anything about its operation), but it packs a punch, tormenting thousands upon thousands of women and girls across the globe.[316]

jurisdictional—that is, aimed at figuring out whether the court should be involved at all, rather than answering questions on the merits. *See* Scott Dodson, *Jurisdiction and Its Effects*, 105 GEO. L.J. 619, 623-33 (2017). If that were the case (and I am not sure if it is), courts could order limited discovery and resolve the issue early in the litigation. Judicial findings could serve as precedent on the factual question of whether a platform's practices satisfied the duty of care. I so appreciate my colleague Rachel Bayefsky for talking with me about this issue.

[313] *See supra* notes 143-45 and accompanying text.

[314] *See supra* Section II.C.

[315] *See, e.g.*, Aaron Mackey, *Two Different Proposals To Amend Section 230 Share a Similar Goal: Damage Online Users' Speech*, ELEC. FRONTIER FOUND. (June 18, 2020), https://www.eff.org/deeplinks/2020/06/two-different-proposals-amend-section-230-share-similar-goal-damage-online-users [https://perma.cc/7NE9-JBD2] (arguing DOJ's proposed changes to § 230 would "impose onerous obligations that would make it incredibly difficult for any new platform to compete with the handful of dominant platforms that exist today").

[316] *See supra* notes 116-17 and accompanying text.

Tech companies will oppose this approach, no question about it. They will face suits that would require them to defend their policies and practices as consistent with the duty of care. In some cases, their practices will be vindicated and shown as reasonable but all the while they will have paid lawyers' fees. On the other hand, their practices may not constitute reasonable steps. None of this will be cheap. But neither is the status quo with lives and careers ruined and expression silenced.

Another possibility, in lieu of a narrow duty of care, is for § 230 to add to the small list of laws that fall outside the legal shield. I have worked on proposals that would carve out from the immunity civil rights legislation and laws related to cyber stalking. A pressing issue of our time is how online platforms enable civil rights violations, from discriminatory housing advertisements to discriminatory hiring systems. Law professors have been advocating for such reform. Privacy and communications scholar Olivier Sylvain has done crucial work to rethink § 230 in the face of algorithmic discrimination.[317] Election law and critical race scholar Spencer Overton has highlighted in scholarship and congressional testimony the importance of § 230 reform to prevent social media companies from enabling voter suppression.[318]

We have seen proposals in just that direction:

> [A] House bill, Civil Rights Modernization Act of 2021, would amend Section 230 of the Communications Act to make explicit that platforms that target housing, employment, financial services, and similar ads away from communities of color are not exempt from civil rights laws while the Senate bill, Safeguarding Against Fraud, Exploitation, Threats, Extremism and Consumer Harms (SAFE TECH) Act, would explicitly say that platforms aren't exempt from civil rights laws, cyber-stalking, targeted

[317] *See, e.g.*, Olivier Sylvain, *Intermediary Design Duties*, 50 CONN. L. REV. 203, 242, 270 (2018); Olivier Sylvain, *Platform Realism, Informational Inequality, and Section 230 Reform*, 131 YALE L.J.F. 475, 497 (2021); Olivier Sylvain, *Discriminatory Designs on User Data*, KNIGHT FIRST AMEND. INST. (Apr. 1, 2018), https://knightcolumbia.org/content/discriminatory-designs-user-data [https://perma.cc/7J5Z-TXQA]. Sylvain has brought his § 230, privacy, and communications law expertise to the FTC as Chairwoman Lina Khan's adviser. *See* Press Release, Fed. Trade Comm., FTC Chair Lina M. Khan Announces New Appointments in Agency Leadership Positions (Nov. 19, 2021), https://www.ftc.gov/news-events/news/press-releases/2021/11/ftc-chair-lina-m-khan-announces-new-appointments-agency-leadership-positions [https://perma.cc/2F99-XN75].

[318] *See* Spencer Overton, *State Power To Regulate Social Media Companies To Prevent Voter Suppression*, 53 U.C. DAVIS L. REV. 1793, 1812-13 (2020); *Hearing on "A Country in Crisis: How Disinformation Online Is Dividing a Nation" Before the Subcomm. on Commc'ns & Tech. and the Subcomm. on Consumer Prot. & Com. of the H. Comm. on Energy & Com.*, 116th Cong. 2 (2020), https://jointcenter.org/wp-content/uploads/2020/06/Overton-Final-Testimony-for-6-24-20-Disinformation-Hearing.pdf [https://perma.cc/F3DX-44LT] (testimony of Spencer Overton, Professor, George Washington University Law School). Overton is the President of the Joint Center for Political and Economic Studies, America's Black think tank. *Staff*, JOINT CTR. FOR POL. & ECON. STUDS., https://jointcenter.org/about/staff/ [https://perma.cc/534P-SMD7] (last visited Mar. 17, 2023).

> harassment, antitrust laws, international human rights laws, and wrongful death actions.[319]

As Spencer Overton remarked of the House bill, the proposed legislation is "thoughtful and tailored [and] addresses a real problem in advancing platform accountability."[320] He also said of the Senate bill:

> Platforms should not profit from targeting employment ads toward white users, or . . . voter suppression ads toward Black users. [The] bill makes it clear that Section 230 does not give platforms a free pass to violate civil rights laws, while also preserving the power of platforms to remove harmful disinformation.[321]

If Congress were game to adopt any of these approaches, I would support them.

CONCLUSION

If Congress were to pursue these reforms, they would find synergies elsewhere. Under the United Kingdom's proposed Online Safety Bill, online services hosting user-generated content would have a "duty of care" to improve online safety.[322] The bill would require online platforms to have appropriate systems and processes in place to protect users, including clear terms of service, mechanisms for accountability, and regularly issued transparency reports. Companies' duties of care would be tailored to their size and activities. The country's communications regulator, the Office of Communications, could issue fines of up to eighteen million British pounds or 10% of a company's annual global revenue; courts could order offending companies to cease operations in the United Kingdom.[323] If passed, the bill would apply to U.S. tech companies providing services and apps to U.K. citizens.[324] The European Commission's Digital Services Act has similar provisions.[325]

---

[319] *Both House and Senate Bills on Section 230 Reform Include Joint Center Proposal*, JOINT CTR. FOR POL. & ECON. STUD. (Feb. 10, 2021), https://jointcenter.org/both-house-and-senate-bills-on-section-230-reform-include-joint-center-proposals/ [https://perma.cc/HK3F-8ZLD]. Mary Anne Franks, Spencer Overton, and I worked with staff on these bills.

[320] *Id.*

[321] *Id.*

[322] *See* MINISTER OF STATE FOR DIGIT. & CULTURE, DRAFT ONLINE SAFETY BILL (2021), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/985033/Draft_Online_Safety_Bill_Bookmarked.pdf [https://perma.cc/65XS-YMKG].

[323] *See* Edina Harbinja, *U.K.'s Online Safety Bill: Not That Safe, After All?*, LAWFARE (July 8, 2021, 1:36 PM), https://www.lawfareblog.com/uks-online-safety-bill-not-safe-after-all [https://perma.cc/8B38-T4QC]; Thomas Reilly, Sam Jungyun Choi, Lisa Peets & Marty Hansen, *UK Government Plans for an Online Safety Bill*, COVINGTON: INSIDE PRIV. (Dec. 18, 2020), https://www.insideprivacy.com/international/united-kingdom/uk-government-plans-for-an-online-safety-bill/ [https://perma.cc/5TQP-7MBA].

[324] *See* Harbinja, *supra* note 323.

[325] *See The Digital Services Act: Ensuring a Safe and Accountable Online Environment*, EUR. COMM'N (Dec. 15, 2020), https://ec.europa.eu/info/strategy/priorities-2019-

Congress could move in this direction by giving a federal agency enforcement authority (in addition to rulemaking authority, as suggested in Part III). The FTC would be a wise choice—the agency's career staff are some of the smartest and most dedicated privacy professionals in the country. The FTC has been ably providing guidance on reasonable data security practices through reports, press releases, blog posts, and enforcement actions under § 5 of the Federal Trade Commission Act, which enables the agency to police unfair and deceptive trade practices.[326] For this to work, Congress would need to commit to providing sufficient resources because the FTC approximately 1,100 overtaxed employees with extensive privacy, fraud, and antitrust duties.[327]

As these moves outside the United States show, a duty of care has global resonance. Congress would do well to pay attention. And federal lawmakers are interested in tackling the problems created by the judiciary's overbroad interpretation of § 230.

It's gratifying to come to this moment. Section 230 reform isn't an off-the-wall idea. It is a real consideration, and career staff on Capitol Hill and their principals are moving forward on various proposals. We need to move carefully and strategically so that we tackle the real problems before us without doing harm to the very people we want to protect.

2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en [https://perma.cc/K334-YFWU].

[326] *See* Andrew Smith, *New and Improved FTC Data Security Orders: Better Guidance for Companies, Better Protection for Consumers*, FED. TRADE COMM'N: BUS. BLOG (Jan. 6, 2020), https://www.ftc.gov/news-events/blogs/business-blog/2020/01/new-improved-ftc-data-security-orders-better-guidance [https://perma.cc/94FE-3VBY].

[327] FED. TRADE COMM'N, FISCAL YEAR 2021: AGENCY FINANCIAL REPORT 8 (2021), https://www.ftc.gov/system/files/documents/reports/agency-financial-report-fy2021/ftc_fy2021_agency_financial_final.pdf [https://perma.cc/MFA6-RLNK].