THREE CONSTITUTIONAL THICKETS: WHY REGULATING ONLINE VIOLENT EXTREMISM IS HARD

This paper, part of the Legal Perspectives on Tech Series, was commissioned in conjunction with the Congressional Counterterrorism Caucus

DAPHNE KELLER
SEPTEMBER 2019
About the Program on Extremism

The Program on Extremism at George Washington University provides analysis on issues related to violent and non-violent extremism. The Program spearheads innovative and thoughtful academic inquiry, producing empirical work that strengthens extremism research as a distinct field of study. The Program aims to develop pragmatic policy solutions that resonate with policymakers, civic leaders, and the general public.

About the Author

Daphne Keller is the Director of Intermediary Liability at Stanford's Center for Internet and Society. Her work focuses on platform regulation and Internet users' rights. She has published both academically and in popular press; testified and participated in legislative processes; and taught and lectured extensively. Her recent work focuses on legal protections for users’ free expression rights when state and private power intersect, particularly through platforms’ enforcement of Terms of Service or use of algorithmic ranking and recommendations. Until 2015 Daphne was Associate General Counsel for Google, where she had primary responsibility for the company’s search products. She worked on groundbreaking Intermediary Liability litigation and legislation around the world and counseled both overall product development and individual content takedown decisions.

The views expressed in this paper are solely those of the author, and not necessarily those of the Program on Extremism or the George Washington University.
Introduction

In May of 2019, two months after an attacker horrified the world by livestreaming his massacre of worshippers in two New Zealand mosques, leaders of Internet platforms and governments around the world convened in Paris to formulate their response. In the resulting agreement, known as the Christchurch Call, they committed “to eliminate terrorist and violent extremist content online,” while simultaneously protecting freedom of expression.¹ The exact parameters of the commitment, and the means to balance its two goals, were left vague – unsurprising in a document embraced by signatories from such divergent legal cultures as Canada, Indonesia, and Senegal. The U.S. did not sign, though it endorsed similar language through G7 as recently as 2018, and will be asked to do so again in 2019.²

What might a law designed to meet these goals look like? International models abound – most of them establishing rules that, in the U.S., would not pass muster under the First Amendment. Australia’s post-Christchurch law was enacted with just 24 hours of public review, and imposes criminal penalties on executives of social media companies if they do not swiftly remove “abhorrent violent material.”³ A U.K. plan lists “extremist content and activity” as one of many areas to be regulated under to-be-determined rules by a to-be-determined government agency.⁴ Germany’s NetzDG gives platforms 24 hours to remove user posts that “manifestly” constitute “public incitement to crime” or encourage “the commission of a serious violent offence endangering the state[.]”⁵ The EU’s draft Terrorist Content Regulation allows police order platforms to take content down in just one hour.⁶ Officials have said that non-violent videos and religious poetry are among the things that must be removed.⁷ Other drafts of the EU’s planned law go much further, requiring hosting platforms of all sizes to adopt content filters – despite widespread concern about problems with the filters platforms use already.
In this paper, I review U.S. constitutional considerations for lawmakers seeking to balance terrorist threats against free expression online. The point is not to advocate for any particular rule. In particular, I do not seek to answer moral or norms-based questions about what content Internet platforms should take down. I do, however, note the serious tensions between calls for platforms to remove horrific but First-Amendment-protected extremist content—a category that probably includes the Christchurch shooter’s video—and calls for them to function as “public squares” by leaving up any speech the First Amendment permits. To lay out the issue, I draw on analysis developed at greater length in previous publications. This analysis concerns large user-facing platforms like Facebook and Google, and the word “platform” as used here refers to those large companies, not their smaller counterparts.

The paper’s first section covers territory relatively familiar to U.S. lawyers concerning the speech Congress can limit under anti-terrorism laws. This law is well-summarized elsewhere, so my discussion is quite brief. The second section explores a less widely-understood issue: Congress’s power to hold Internet platforms liable for their users’ speech. The third section ventures farthest afield, reviewing constitutional implications when platforms themselves set the speech rules, prohibiting legal speech under their Terms of Service (TOS). I will conclude that paths forward for U.S. lawmakers who want to both restrict violent extremist content and protect free expression are rocky, and that non-U.S. laws are likely to be primary drivers of platform behavior in this area in the coming years.

A. Prohibiting Speech

The U.S. First Amendment protects speech that is illegal in many other countries. Americans can share material that is widely considered offensive, harmful, or morally abhorrent. Congress does have latitude to prohibit some speech that supports terrorism, though. The Supreme Court upheld one of the key existing anti-terrorism laws, the “material support” statute at 18 USC § 2339B, against a First Amendment
challenge in 2010. American groups argued that the law’s ban on providing training and advice to foreign terrorist organizations violated their speech rights. The Court agreed that their rights were burdened, but upheld the law, noting that the Government’s “interest in combating terrorism is an urgent objective of the highest order[.]” Importantly, the Court observed, the material support law did not prevent U.S. speakers from independently advocating for foreign terrorist organizations or their goals – they just could not work with those organizations to do so.

Laws that did penalize speech purely because it depicted, advocated, or promoted violent extremism would face major constitutional hurdles. The Supreme Court held, in Brandenburg v. Ohio, that the government may not ban “the mere abstract teaching ... of the moral propriety or even moral necessity for a resort to force and violence.” To punish violence-promoting speech, the government must show that it is “directed to inciting or producing imminent lawless action and is likely to incite or produce such action.” This means that a blanket prohibition on material advocating terrorism would likely be struck down. And while some Internet users could be prosecuted for sharing particular material like the Christchurch video as an intentional means of inciting violence, the Brandenburg test would not support a ban on the video itself.

Some scholars have argued that this framework does not adequately respond to risks posed by present-day terrorists’ use of the Internet. Cass Sunstein, for example, has proposed relaxing the Brandenburg “imminence” requirement, given social media’s potential to “dramatically amplify the capacity of speech in one place to cause violence elsewhere at some uncertain time[.]” That approach tracks many non-U.S. lawmakers’ responses to Christchurch and other acts of violent extremism. However, it would require the Supreme Court to revisit not only the Brandenburg test but also its seminal Internet First Amendment case, Reno v. ACLU, which rejected the idea that Congress could restrict more speech online than offline.
B. Making Platforms Take Down Speech

The second important constitutional issue is specific to laws regulating intermediaries like Internet platforms. These companies face the same federal criminal liability as any other defendant as a statutory matter – immunities under laws like Communications Decency Act Section 230 (“CDA 230”) do not affect those claims.\(^\text{19}\) As a constitutional matter, though, special considerations apply when laws hold intermediaries responsible for their users’ speech.

The concern is that over-reaching intermediary liability laws will cause risk-averse platforms to take down users’ lawful speech – a phenomenon scholars have referred to as “collateral censorship.”\(^\text{20}\) A pair of important 20th century cases, *Smith v. California* and *Bantam Books v. Sullivan*,\(^\text{21}\) both cited this concern in striking down laws that held book stores liable for obscenity in books. As the Court noted, a bookseller facing strict liability is incentivized “to restrict the books he sells to those he has inspected,” leading to state-compelled censorship that is “hardly less virulent for being privately administered.”\(^\text{22}\)

In *Reno*, the Court identified a similar threat to Internet users who depend on private companies to convey their speech. The Internet speech regulation it struck down, the Court noted, empowered dishonest or mistaken accusers to exercise a “heckler’s veto,” causing intermediaries to silence users.\(^\text{23}\) Numerous quantitative studies have since documented platforms’ removal of lawful speech under “notice and takedown” systems like the Digital Millennium Copyright Act (“DMCA”), bearing out the Court’s concern.\(^\text{24}\) Accusers who have successfully gamed platforms’ fears of liability range from governments using bogus copyright claims to suppress evidence of misconduct\(^\text{25}\) to small businesses seeking to reduce competitors’ online presence.\(^\text{26}\)
This does not mean platforms can never be liable for online content, of course. Few question the constitutionality of existing U.S. laws like the DMCA or child pornography statutes, for example. But it does mean that a clumsily drafted law can fail on constitutional grounds. Lawmakers seeking to reduce this risk could, for example, create procedural protections for speakers, like giving them a chance to respond to accusations or appeal platforms’ takedown decisions. Such protections might be particularly important in the extremism context, given the similarities and connections between “terrorist content” and protected political speech.

A law requiring platforms to take down speech associated with terrorism would also face basic questions about its efficacy in advancing the government’s security goals, and about its effects on lawful speech. Both questions are hotly debated by experts.

On the question of security, researchers disagree about the causal connections between online content and real-world violence, and about whether removing extremist material from platforms makes us safer. Aggressive content removal might disrupt extremists’ public messages, but it might also drive them into echo chambers, undermine law enforcement investigations, and silence important moderate voices. If platforms over-remove and take down innocent speakers’ content, the errors – and resulting sense of frustration, disenfranchisement, or marginalization -- are likely to fall on the very groups whose perceived risk of radicalization drove the law’s adoption in the first place. Overall, we are far from empirical clarity about how platform content removals advance or impede security goals.

The question about damage to protected speech from platform takedowns should in principle be easier to answer. But, outside of anecdotal examples, facts are hard to come by. We know particularly little about automated tools like content filters – despite some companies’ sweeping claims about better content moderation through artificial intelligence. The best-known tool for detecting violent extremist content is the database
maintained by the Global Internet Forum to Counter Terrorism, which is used by at least thirteen platforms to automatically weed out user posts containing some 80,000 videos and images. As the ACLU and other civil liberties organizations pointed out, filters are notoriously bad at distinguishing material advocating violence from the same material used in other contexts like news reporting. “Almost nothing is publicly known about the specific content that platforms block using the Database,” they noted, “or about companies’ internal processes or error rates[.]” The groups cited concerning errors, like YouTube’s deletion of over 100,000 videos gathered by the public interest Syrian Archive for prosecution of human rights abuses.

Platforms supplement automated tools with ever-increasing armies of content moderators around the globe. But they, too, seem to make a lot of mistakes. Researchers cannot say how frequently platforms take down the wrong speech, though, without seeing what speech has been removed. For the most part, platforms do not share that information. Instead, their transparency reports share aggregate data reflecting only the platforms’ own characterization of the content they took down. This dearth of public information makes it very hard to say how well takedowns are working, how well they could work, and what a law tailored to protect First Amendment rights would look like.

C. Letting Platforms Take Down Speech

Regulating online speech connected to terrorism is constitutionally complex. But even legislative inaction can raise First Amendment questions. Internet users increasingly question platforms’ practices of removing legal speech based on their Terms of Service (TOS). Some argue that platforms are violating users’ speech rights, and that courts or Congress should intervene.
Platforms face strong incentives to take down speech that is protected by the First Amendment. Market forces push this way, since users and advertisers typically do not want to see lawful but highly violent or offensive material. Non-U.S. law pushes this way, too. In some cases, platforms may effectively export speech-restrictive laws voluntarily, by incorporating them in globally applied TOS – as Facebook, Twitter, YouTube, and Microsoft did in agreeing to the EU’s Hate Speech Code of Conduct. American politicians contribute to this dynamic, too, by urging platforms to remove lawful but horrific material or to use blunt tools like filters. The net result is an increasing array of TOS-based speech prohibitions, many reflecting widely held U.S. social norms but not reflecting U.S. law. As these rules expand, ever more decisions about speech move outside the purview of democratic legislation or public judicial review, and into platforms’ discretion.

Is platforms’ growing power over public discourse a constitutional problem? In the U.S. this is currently framed as a conservative issue, but concerns about platforms as information gatekeepers have long been raised by speakers across the political spectrum and around the world. In the violent extremism context, though, platforms’ freedom to adopt discretionary speech rules lets them do what many critics, including politicians, want: take down content that is considered offensive or harmful, but that is not illegal. If we argue that platforms should step outside the law and carry out values-based removals for extremist content, it is hard to argue that they should simultaneously have to respect users’ First Amendment rights.

Courts have so far been unreceptive to claims that platforms should be compelled to carry speech against their will. None of the thirty or more plaintiffs who have brought suits in the U.S. have won. It is hard to say what winning would even mean – would platforms really lose all discretion, and have to show users all the ugliest speech the First Amendment permits? Would that obligation affect their ranking algorithms as well as their takedown policies? Still, new suits keep being filed, and the issue is increasingly
prominent inside the Beltway. Implicit in the discussion, perhaps, is the idea that laws might force platforms to carry some legal speech, but allow or even require them to take down other legal speech.

Legally, that idea has some precedent. But it is not pretty. In laws governing owners of older information channels like broadcasters, Congress restricted some otherwise-legal speech -- and the Supreme Court accepted its power to do so based on medium-specific considerations like children’s easy access to content. It also required broadcasters and cable companies to carry content they did not want to. The fairness doctrine required broadcasters to air views on both sides of political issues, for example. Other rules required cable companies to tolerate a relative free-for-all of First Amendment-protected speech on public access channels, but not on the companies’ other channels. The rules Congress and the FCC crafted in an effort to balance rights of “platforms” (cable and broadcast companies) and “users” (creators and consumers of content they transmitted) are being litigated to this day. A case pending before the Supreme Court in 2019, Manhattan Community Access Corp. v. Halleck, has been billed as a dress rehearsal for future cases about users’ speech rights on Internet platforms.

If there are analogous rules for the Internet, though, it is not clear what they would look like or whether they could pass constitutional muster. In particular, laws restricting currently legal speech – or disfavoring it by giving platforms special incentives or leeway to take it down – would affect ordinary people's speech rights much more than cable and broadcast laws ever did. And the Supreme Court already said, in Reno, that the special First Amendment rules it approved for cable do not apply to the Internet. It seems unlikely that Congress would seriously try to devise a new regime along these lines anyway, given the sheer complexity of older Communications law and its reliance on a powerful regulatory agency, the FCC. The effort would be further complicated by conservatives’ long-standing position that rules like the fairness doctrine were unconstitutional.
In short, we are unlikely to see a constitutional law that simultaneously (1) requires platforms to carry legal speech, (2) meets many lawmakers’ anti-terrorist-content goals, and (3) survives constitutional review.

**Conclusion**

Regulating online violent extremist speech is hard. Most obviously, it is hard because the First Amendment constrains Congress’s choices, as discussed in this paper’s first two sections. But it is also hard if policymakers want platforms to take down more speech than the law requires in some cases, but hew to the First Amendment in others. Laws pursuing both goals at once are not a logical impossibility -- laws governing broadcasters and cable companies provide a precedent. But it is highly debatable whether similar laws applied to Internet platforms would be constitutional. And as a political matter, Congress seems highly unlikely to enact them.

The likely future of online terrorist content regulation, then, may look much like the present. In that scenario, major platforms would continue to prohibit more content than U.S. law does. By relying on unreliable instruments like filters, they would likely take down more material than intended, including news reporting or political speech. New requirements would flow from countries with more restrictive speech traditions than the U.S., and platforms would shape their global operations in response to those countries’ demands. U.S. law would not counterbalance this trend with laws requiring platforms to protect lawful speech, given the constitutional, political, and normative barriers to enacting them.
References

10. *Matal v. Tam*, 137 S. Ct. 1744, 1764 (2017) (plurality opinion) (rejecting government’s asserted interest in protecting listeners from race-based demeaning messages in trademarked terms); *Snyder v. Phelps*, 562 U.S. 443, 454 (2011) (First Amendment barred claim for emotional distress caused to soldier’s family by protestors at his funeral holding signs including “Thank God for IEDs” and “You're Going to Hell”).
15. Id. at 447-448. Other First Amendment exceptions for “true threats” or for speech used to plan or execute a crime would be relevant for laws prohibiting a much narrower class of terrorism-related speech. See Killion at 23-28.
19. 47 U.S.C. 230(e)(1). Multiple civil courts have rejected claims that platforms’ ordinary operations violate material support laws, mostly without reaching the question of CDA 230 immunity. See Eric


23 *Reno* at 880.


27 17 USC 512; 18 USC § 2258A.


29 Such a law would presumably be held to strict scrutiny, meaning it must be “narrowly tailored” to achieve a “compelling” state interest. *U.S. v. Stevens*, 559 U.S. 460, 473 (2010).

30 Keller 2018 at 20-26 (reviewing security research and concerns).


32 *Id.*

33 *Id.*

34 In fairness, it’s unclear that applying nuanced speech rules at Internet scale is actually possible. No government has ever attempted it, and courts spend months or years on questions platforms seek to resolve in hours or seconds.

35 A notable exception is the information held, mostly about DMCA takedowns, by Harvard’s Lumen Database. Lumen, About Us, [https://www.lumendatabase.org/pages/about](https://www.lumendatabase.org/pages/about).

36 *See* sources discussed in Keller 2019 at 11-22.


39 Courts in at least two countries – Brazil and Germany – have upheld users’ claims that platforms wrongfully removed their speech, and ordered platforms to reinstate it. These cases, however, build on legal traditions that are comparatively tolerant of state intervention in private business operations. David Meyer, “Court Tells Facebook: Stop Deleting ‘Offensive’ Comment,” ZDNet, April 13, 2018, [www.zdnet.com/article/court-tells-facebook-stop-deleting](www.zdnet.com/article/court-tells-facebook-stop-deleting).
-offensive-comment (describing German case); Civil Appeal No. 0000412-86.2016.8.24.0175 (Brazilian case).
40 See Keller 2019 at 11-12.


45 See Keller 2019 at 23-27 (discussing possible regulatory regimes and their drawbacks).