Policy Analysis

CATO INSTITUTE

November 28, 2017 | Number 828

What to Do about the Emerging Threat of Censorship Creep on the Internet

By Danielle Keats Citron

EXECUTIVE SUMMARY

Popular tech companies—Google, Facebook, Twitter, and others—have strongly protected free speech online, a policy widely associated with the legal norms of the United States. American tech companies, however, operate globally, and their platforms are subject to regulation by the European Union, whose member states offer less protection to expression than does the United States. European regulators are pressuring tech companies to control and suppress extreme speech. The regulators' clear warning is that, if the companies do not comply "voluntarily," they will face harsher laws and potential liability. This regulatory effort runs the risk of censorship creep, whereby a wide array of protected speech, including political criticism and newsworthy content, may end up being removed from online platforms on a global scale.

European regulators cannot be expected to pull back and adopt U.S. norms for speech. The tech company leaders may, however, reduce the risks to free speech by insisting on clear definitions of "hate speech," holding regulators accountable before the public, fostering detailed transparency of government actions, and appointing ombudsmen.

Danielle Keats Citron is the Morton and Sophia Macht Professor of Law at the University of Maryland Francis King Carey School of Law, an affiliate fellow at the Yale Information Society Project, and an affiliate scholar at the Stanford Center on Internet and Society.

"European lawmakers have pressed companies to ban 'fake news' to help combat extremist expression; they will press for the removal of far more, including political dissent and cultural commentary."

INTRODUCTION

For much of its history, Silicon Valley has been a full-throated champion of First Amendment values. When online platforms banned certain types of speech in terms-of-service (TOS) agreements, they proceeded cautiously, with a preference for an American-style approach to free expression. More recently, tech companies have tailored their speech policies to European norms rather than American ones. Ordinary market forces were not behind this shift. Instead, threatened legislation prompted the change.

In the wake of terrorist attacks in late 2015, European Union (EU) regulators warned tech companies that they would face prohibitively expensive fines and potential criminal penalties unless extremist and hateful content was swiftly removed. In response, the dominant social media platforms have altered their speech policies to ban extremist content in ways that risk censorship creep. Already, European lawmakers have pressed companies to ban "fake news" to help combat extremist expression.[1] No doubt, they will press for the removal of far more, including political dissent and cultural commentary. The impact will be far reaching. Because TOS agreements apply everywhere that platforms are accessed, the changes will affect free expression on a global scale.

This study offers potential safeguards to prevent censorship creep. Companies can and should adopt prophylactic protections designed to contain government overreach and censorship creep for the good of free expression. Censorship creep can be contained with definitional clarity, robust accountability, detailed transparency, and ombudsman oversight. The proposals that follow may be attractive to tech executives and the informed public interested in curtailing government overreach and conveying their commitment to users' free expression. As Apple's struggle with the U.S. government over encryption illustrated, tech companies enjoy public support when they defend fundamental freedoms.

FROM FREE SPEECH CHAMPIONS TO COERCED CENSORS

A decade ago, Sen. Joseph Lieberman (D-CT) publicly chastised YouTube for refusing to remove terrorist training videos. The senator's pressure failed to produce results because the company prioritized the protection of users' free speech over its popularity on Capitol Hill.[2] Crucially, the company knew that there was little that Congress could actually do given the First Amendment's robust protections against most viewpoint-based regulation. Long after the showdown with Senator Lieberman, American-style free speech values continued to guide tech companies' policies about what expression was permissible on their platforms. TOS agreements typically protected users' ability to express unpopular views while prohibiting targeted abuse that silenced individuals.[3]

Of late, however, Silicon Valley's commitment to free speech has eroded. The catalyst was a spate of terrorist attacks in Paris and Brussels in late 2015. European regulators blamed Silicon Valley for giving extremist groups access to potential recruits. They warned that unless online platforms guaranteed the swift removal of extremist or hateful speech, they would face prohibitively expensive fines and criminal penalties.[4] The regulators' threats were not idle: in the EU, unlike in the United States, there isn't a heavy presumption against speech restrictions.

To stave off threatened European regulation, tech companies have retreated from a strong free speech stance. In May 2016, Facebook, Microsoft, Twitter, and YouTube (referred to in the rest of the paper as "the Companies") signed an agreement with the European Commission to "prohibit the promotion of incitement to violence and hateful conduct."[5] The agreement defined "hateful conduct" as speech inciting violence or hatred against protected groups. Under the agreement, the Companies pledged to remove reported hate speech that violated TOS within 24 hours. The European Commission was given the right to review the companies' compliance with the agreement.

On December 5, 2016, the day before the European Commission issued a report sharply criticizing compliance with the hate-speech agreement, the Companies announced plans for an industry database of "hashes"—unique digital signatures—of extremist material banned on their platforms. The hash technology would enable the immediate flagging and removal of prohibited content on participating companies' platforms. According to the announcement, other companies will be given access to the database as soon as it is operational.[6] The European Commission hailed the industry database as the "next logical step" in a "public-private partnership" to combat extremism.[7]

This industry database indicates how much the debate has moved toward government oversight of digital speech. Just months before, executives in the tech industry dismissed calls for such a database on the grounds that "violent extremist material" was a malleable term and governments would surely pressure companies to include hashes that would silence far more than terrorist propaganda. To address such free speech concerns, the Companies have explained that content will be hashed only if it involves "the most extreme and egregious terrorist images and videos . . . content most likely to violate all of our respective companies' content policies."[8] Hashed material will not be deleted from participants' sites immediately. Instead, each company will review content included in the database under its own policies.[9]

Following the announcement of the hate speech agreement and the industry database, the demands of European leaders have only escalated. After a series of terrorist attacks in London in 2017, British Prime Minister Teresa May and French President Emmanuel Macron threatened steep fines for failure to remove extremist propaganda from online platforms.[10] Shortly thereafter, Google announced a four-part plan to address terrorist propaganda that included the increased use of technology to identify terrorist-related videos, the hiring of additional content moderators, the removal of advertising on objectionable videos, and the directing of potential terrorist recruits to counter-radicalization videos.[11] Facebook responded with a pledge to increase its use of artificial intelligence to stop the spread of terrorist propaganda and to hire 3,000 more people to review speech reported as TOS violations.[12]

The Companies have not chosen this path for efficiency's sake or to satisfy the concerns of advertisers and advocates. Instead, European regulators have extracted private speech commitments by threatening to pass new laws making platforms liable for extremist speech. Unlike in the United States, there isn't a heavy presumption against speech restrictions in the EU, although laws penalizing speech must satisfy a proportionality analysis.[13] No matter how often EU lawmakers describe the recent changes to private speech practices as "voluntary," the fact is that they were the product of government coercion. And such coercion may be expanding. Apart from such incentives for increased regulation, the Companies probably still prefer freedom of speech. How can they act on that commitment even as the EU seeks more coercion?

Censorship Creep at a Global Scale

To be sure, companies' changed policies may have some important benefits. With less terrorist propaganda and hate speech online, there might be fewer people joining ISIS (the Islamic State of Iraq and Syria) fighters in Syria or planting bombs in markets and restaurants. But the policy changes pose a risk of censorship creep as well.

Definitional ambiguity is part of the problem. "Hateful conduct" and "violent extremist material" are vague terms that can be stretched to include political dissent and cultural commentary. They could be extended to a government official's tweets, posts critiquing a politician, or a civil rights activist's profile.[14] Violent extremist material could be interpreted to cover violent content of all kinds, including news reports, and not just gruesome beheading videos.

Censorship creep isn't merely a theoretical possibility—it is already happening. European regulators' calls to remove "illegal hate speech" have quickly ballooned to cover expression

"'Hateful conduct' and 'violent extremist material' are vague terms that can be stretched to include political dissent and cultural commentary."

“Extremist and hateful speech adds valuable information to public discourse: the fact that such views exist can highlight the need to counter them.”

that does not violate existing EU law, including bogus news stories. Commenting on the hate-speech agreement, European Justice Commissioner Věra Jurová criticized the Companies for failing to remove “online radicalization, terrorist propaganda, and fake news.” [15] Legitimate debate could easily fall within Jurová’s characterization of hate speech.

As more expression is deemed to violate TOS agreements, more expression will be deleted. When content is reported as hate speech, the likely response will be removal.[16] Removal of reported content would forestall criticism and would be cheaper than the cost of complying with new laws. The pledge to review hate-speech reports within 24 hours will reinforce this tendency. Speed inevitably sacrifices thoughtful deliberation. Similarly, there surely will be pressure to remove content that other companies have designated as violent extremist expression. If that were the case, the industry database would become a “delete-it-all” program.[17]

These developments will have a far-reaching impact because TOS agreements typically apply globally.[18] Unlike a court order that applies only within the issuing country’s borders, a company’s decision about a TOS violation applies everywhere its services are accessed.[19] This is true for hate speech and violent extremist material included in the database. Removal for a TOS violation means worldwide removal. This sort of censorship is hard to circumvent.

The stakes for free expression are high. Content may be removed even though it is essential for public debate and the reporting of news.[20] A key insight of free speech theory is that individuals need to speak and listen to make decisions about the kind of society they want.[21] As the editorial board of the *Washington Post* wrote in response to social media companies’ removal of terrorist propaganda, “Citizens of every country deserve to know what is going on in the world and what people at both ends of the spectrum think about it—however hard that is to stomach.”[22]

Extremist and hateful speech adds valuable information to public discourse: the fact that such views exist can highlight the need to counter them.[23] As human rights activist Aryeh Neier has argued, “Freedom of speech itself serves as the best antidote to the poisonous doctrines of those trying to promote hate.”[24] The expression of hate or extremist views enables society to assert strong social norms rejecting them.[25]

Removal of extremist expression would undermine efforts designed to change people’s minds.[26] For example, Jigsaw, a Google-owned think tank, has developed a program that uses a combination of Google’s advertising algorithms and YouTube’s video platform to identify aspiring ISIS recruits and to offer alternatives to hateful ideologies. The program places advertising alongside results for keywords and phrases commonly searched for by people attracted to ISIS. The ads link to YouTube channels containing videos that have potential to undo ISIS’s brainwashing, such as testimonials from former extremists and imams denouncing ISIS’s corruption of Islam.[27]

Even if the majority of people embracing hateful ideas may not be open to counter speech, some may be.[28] In 2009, Megan Phelps-Roper, a member of the Westboro Baptist Church, developed a considerable following tweeting hateful views about lesbian, gay, bisexual, and transgender individuals. She connected online with people who explained the cruelty of her positions. Phelps-Roper’s interactions on Twitter ultimately led her to reject bigotry.[29] In a Brookings Institution study titled “The ISIS Twitter Census,” J. M. Berger and Jonathon Morgan found that “when we segregate members of ISIS social networks, we are, to some extent, also closing off potential exit ramps.”[30]

Moreover, the removal of expression denies disaffected individuals opportunities to let off steam that might stop them from turning to violence.[31] As noted by the United Nations General Assembly in its Plan of Action to Prevent Violent Extremism, blocking online activity fuels narratives of victimization and risks further isolating disaffected individuals. Aggrieved speakers may feel even more aggrieved and

more inclined to act on pent-up anger. Removing an ISIS Twitter account could "increase the speed and intensity of radicalization for those who do manage to enter the network."[32]

PROTECTIONS AGAINST CENSORSHIP CREEP

European regulators have effectively exerted power over the expression of people who do not live in their jurisdictions and cannot hold them accountable. The result is worldwide conformity with European speech values without meaningful accountability or oversight.[33] Given the success of these efforts, European regulators will continue to demand more "voluntary" changes to coerce conformity with desired speech norms. Such "public-private partnerships" are fruitful courses of action for state censors. They secure the adoption of governmental preferences without the burdens of formal process. EU regulators will hardly rein in their pressure on their own.

Silicon Valley may be our best protection against censorship creep. Tech companies can pursue several strategies to push back against government overreach: definitional clarity, robust accountability, detailed transparency, and ombudsmen oversight.

Definitional Clarity

Government requests to remove hate speech or to hash extremist material should be reviewed under a well-developed set of definitions. Clarity in the definition, meaning, and application of both terms would help constrain censorship creep. To that end, policies should provide specific examples of content deserving designation as hate speech or violent extremist material. This would help prevent the gradual broadening of the standards governing the removal of expression.

Some have suggested that companies look to international human rights law for guidance in defining both terms.[34] But human rights law is unlikely to provide clarity because it contains exceptionally flexible standards.[35] The Council of Europe's secretary general is drafting "common European standards for hate speech and terrorist material to better protect freedom of expression online."[36] That project will be helpful if it provides clear definitions and illustrations that curtail the malleability of both terms.[37] As tech companies work on their definitions of hate speech and extremist material, they should consider including human rights groups and academics in their efforts.[38] Civil liberties groups have argued for a role in helping companies understand "various meanings given to 'violent extremism' and related concepts, and the potential impact of ambiguity in this area on the promotion and protection of human rights."[39]

Those definitions, while designed for content moderators, should be shared publicly so that government actors can understand the limits of efforts to remove speech under TOS agreements and community guidelines. With those limits in mind, governmental authorities may be less inclined to try to silence unpopular but protected expression.

Robust Accountability

Rigorous accountability is essential to check government efforts to censor disfavored viewpoints and dissent. Removal requests by state authorities (or nongovernmental organizations [NGOs] acting on the state's behalf) should be subject to rigorous review. For instance, the European Commission worked with 12 NGOs to report hate speech and assess companies' compliance with the hate speech agreement. To start, government officials or NGOs acting on their behalf should be required to identify themselves when reporting content for TOS violations. Online platforms must know that they are dealing with governmental authorities or their surrogates. Companies should have a separate reporting channel for government authorities and any organizations working on the state's behalf. Twitter, for instance, has "intake channels dedicated for law enforcement and other authorized reporters" to file "legal requests."[40] All removal requests made by state actors or their surrogates, including TOS reports, should proceed through that channel.

"Silicon Valley may be our best protection against censorship creep."

"When state actors seek to suppress speech under terms of service agreements, reviewers should view their requests with a presumption against removal, or at least a healthy dose of skepticism."

Government requests should be viewed through a special lens. Governments raise particularly troubling concerns about silencing political dissent. To be sure, ordinary people can be hecklers, but the concern for governments is systematic efforts to silence dissent or unfavorable news. When state actors seek to suppress speech under TOS agreements, reviewers should view their requests with a presumption against removal, or at least a healthy dose of skepticism.[41] Content moderators should receive training about censorship creep, including past and present governmental efforts to silence critics. Training should focus on how to distinguish banned material from newsworthy content. This is not an easy task, but it is crucial nonetheless.

Decisions related to government requests should be accompanied by an explanation—decisionmakers who have to articulate their reasons are likely to think more carefully about their decisions.[42] When a moderator decides to grant a government request for removal on the basis of a TOS violation, that decision should automatically pass through a second layer of review. Individuals whose speech is removed should be notified about the removal and given a chance to appeal.

Even-stronger protections are essential to prevent governments from co-opting the industry database, which runs the risk of becoming a total blacklist as more companies participate. The Tech Companies could adopt a blanket rule that governments cannot contribute hashes to the database.[43] An alternative is to subject government requests to several layers of review and to condition the submission on the approval of senior staff.[44]

Detailed Transparency

Another check on censorship creep is for companies to provide detailed reports on governmental efforts to censor hate speech and extremist material through informal measures. Transparency reports enable public conversation about censorship. European users can contact lawmakers with concerns about authorities' attempts to use tech companies as censorship proxies. The more users understand about companies' efforts to protect their fundamental freedoms, the more users will trust the platforms they use. Human rights advocates can call attention to concerns about censorship creep. Ultimately, transparency reports can generate "productive discussion about the appropriate use and limits of [state] authority."[45]

Some social media companies have provided transparency about government requests to suppress speech. Twitter has been hailed for its transparency efforts, and rightfully so. The company's 2016 Transparency Report details the number of legal requests for content removal broken down by country.[46] Crucially, and uniquely, it discloses the number of government requests seeking removal of terrorism content for TOS violations.[47] Twitter is working to expand its reporting of all "known, non-legal government TOS requests we receive through our standard customer service intake channels, such as requests to remove impersonating accounts and other content that violates our Rules against abuse."[48]

Much as Twitter has done for terrorist content and expects to do far more of in the future, corporate transparency reports should detail the number, subject matter, and results of *all* government requests to remove content for TOS violations.[49] If governments are allowed to request the addition of hashes to the industry database, transparency reports should include details about those requests. Although transparency cannot solve the problem of censorship creep, it can help contain it, especially if strong standards and robust accountability procedures are adopted.

Ombudsmen Oversight

An acute concern of censorship creep is its potential to suppress newsworthy content. Governments may seek to remove terrorist or hateful content whose publication is in the legitimate public interest. To address this concern, companies should consider hiring or consulting ombudsmen whose life's work is the newsgathering process.[50] Ombudsmen, who are also known as public editors, work to

protect press freedom and to promote high-quality journalism. Their role is to "act in the best interest of the news consumer."[51]

Ombudsmen should have a special role in assessing government removal requests made through informal channels such as TOS or industry databases. They can help identify requests that would remove material that is important for public debate and knowledge. Then, too, because the industry database raises special concerns about the suppression of expression, ombudsmen could review all contributions to the database with the public interest in mind.

CONCLUSION

By pressuring Silicon Valley to alter private speech policies and practices, EU regulators have effectively set the rules for free expression across the globe. The question is whether tech companies will fight on behalf of their users to contain government overreach. My proposals for definitional clarity, robust accountability, detailed transparency, and ombudsmen oversight will help combat censorship creep.

NOTES

This paper is based on Danielle Keats Citron, "Extremist Speech, Compelled Conformity, and Censorship Creep," forthcoming, 2018, in the *Notre Dame Law Review*. Special thanks to Susan McCarty for her expert assistance and to the editors of *Notre Dame Law Review* for supporting this effort.

1. Cara McGoogan, "EU Accuses Facebook and Twitter of Failing to Remove Hate Speech," *Telegraph (London)*, December 5, 2016, http://www.telegraph.co.uk/technology/2016/12/05/eu-accuses-facebook-twitter-failing-remove-hate-speech/.

2. Timothy B. Lee, "YouTube Rebuffs Senator's Demand to Remove Islamist Videos," *Ars Technica*, May 20, 2008, https://arstechnica.com/tech-policy/2008/05/youtube-rebuffs-senatorss-demands-for-removal-of-islamist-videos/.

3. Danielle Keats Citron, *Hate Crimes in Cyberspace* (Cambridge, MA: Harvard University Press, 2014), p. 232; Kate Klonick, "The New Governors: The People, Rules, and Processes Governing Online Speech," *Harvard Law Review* (forthcoming, 2018).

4. Lizzie Plaugic, "France Wants to Make Google and Facebook Accountable for Hate Speech," *Verge*, January 27, 2015, https://www.theverge.com/2015/1/27/7921463/google-facebook-accountable-for-hate-speech-france.

5. European Commission, "European Commission and IT Companies Announce Code of Conduct on Illegal Online Hate Speech," news release, May 31, 2016, http://europa.eu/rapid/press-release_IP-16-1937_en.htm.

6. Liat Clark, "Facebook, Twitter, Microsoft, YouTube Launch Shared Terrorist Media Database," *Wired UK*, December 6, 2016, http://www.wired.co.uk/article/facebook-twitter-microsoft-youtube-launch-shared-terrorism-database.

7. European Commission, "EU Internet Forum: A Major Step Forward in Curbing Terrorist Content on the Internet," news release, December 8, 2016, http://europa.eu/rapid/press-release_IP-16-4328_en.htm.

8. Clark, "Facebook, Twitter, Microsoft, YouTube."

9. Facebook, "Partnering to Help Curb Spread of Online Terrorist Content," news release, December 5, 2016, http://newsroom.fb.com/news/2016/12/partnering-to-help-curb-spread-of-online-terrorist-content/.

10. Amanda Paulson and Eva Botkin-Kowaki, "In Terror Fight, Tech Companies Caught between US and European Ideals," *Christian Science Monitor*, June 23, 2017, https://www.csmonitor.com/Technology/2017/0623/In-terror-fight-tech-companies-caught-between-US-and-European-ideals.

11. Kent Walker, "Four Steps We're Taking Today

to Fight Terrorism Online," *Google in Europe* (blog), June 18, 2017, https://blog.google/topics/google-europe/four-steps-were-taking-today-fight-online-terror/.

12. Monika Bickert and Brian Fishman, "Hard Questions: How We Counter Terrorism," *Hard Questions* (blog), Facebook, June 15, 2017, https://newsroom.fb.com/news/2017/06/how-we-counter-terrorism/.

13. Article 19 of the International Covenant on Civil and Political Rights allows states to limit freedom of expression under circumstances that satisfy proportionality review. http://www.ohchr.org/EN/ProfessionalInterest/Pages/CCPR.aspx.

14. Sam Levin, "Facebook Temporarily Blocks Black Lives Matter Activist after He Posts Racist Email," *Guardian*, September 12, 2016, https://www.theguardian.com/technology/2016/sep/12/facebook-blocks-shaun-king-black-lives-matter; Tracy Jan and Elizabeth Dwoskin, "A White Man Called Her Kids the N-Word. Facebook Stopped Her from Sharing It," *Washington Post*, July 31, 2017, https://www.washingtonpost.com/business/economy/for-facebook-erasing-hate-speech-proves-a-daunting-challenge/2017/07/31/922d9bc6-6e3b-11e7-9c15-177740635e83_story.html?utm_term=.97d6e7103703.

15. Cara McGoogan, "EU Accuses Facebook and Twitter."

16. Jillian C. York, "European Commission's Hate Speech Deal with Companies Will Chill Speech" (blog of the Electronic Frontier Foundation), June 3, 2016, https://www.eff.org/deeplinks/2016/06/european-commissions-hate-speech-deal-companies-will-chill-speech.

17. Andy Greenberg, "Inside Google's Internet Justice League and Its AI-Powered War on Trolls," *Wired*, September 19, 2016, https://www.wired.com/2016/09/inside-googles-internet-justice-league-ai-powered-war-trolls/.

18. YouTube's description of its TOS is the same for inside the United States as outside it. "Terms of Service," YouTube, https://www.youtube.com/t/terms. The same is true for Twitter. "The Twitter Rules," Twitter, https://support.twitter.com/articles/18311.

19. Emma Llansó (director of free expression, Center on Democracy and Technology), in discussion with the author, January 15, 2017.

20. Courtney C. Radsch, "Privatizing Censorship in Fight against Extremism Is Risk to Press Freedom" (blog of the Committee to Protect Journalists), October 16, 2015, https://cpj.org/blog/2015/10/privatizing-censorship-in-fight-against-extremism-.php.

21. Citron, *Hate Crimes in Cyberspace*.

22. Editorial Board, "The Government Wants Social Media Sites to Take Down Terrorist Propaganda. Maybe They Shouldn't," *Washington Post*, September 16, 2016, https://www.washingtonpost.com/opinions/the-government-wants-social-media-sites-to-take-down-terrorist-propaganda-maybe-they-shouldnt/2016/09/16/148d75cc-7b77-11e6-ac8e-cf8e0dd91dc7_story.html?utm_term=.4a6a4fb8e07c.

23. Steven H. Shiffrin, "Racist Speech Outsider Jurisprudence and the Meaning of America," *Cornell Law Review* 80 (November 1994): 43.

24. Flemming Rose, *The Tyranny of Silence* (Washington, Cato Institute, 2014), p. 85.

25. C. Edwin Baker, "Autonomy and Hate Speech" in *Extreme Speech and Democracy*, ed. Ivan Hare and James Weinstein (New York: Oxford University Press, 2011), p. 151.

26. *Whitney v. California*, 274 U.S. 357, 377 (1927) (concurring, J. Brandeis) (remedy for bad speech is "more speech, not enforced silence").

27. Andy Greenberg, "Google's Clever Plan to Stop Aspiring ISIS Recruits," *Wired*, September 7, 2016, https://www.wired.com/2016/09/googles-clever-plan-stop-aspiring-isis-recruits/.

28. Adrian Chen, "Unfollow," *New Yorker*, November 23, 2015, http://www.newyorker.com/magazine/2015/11/23/conversion-via-twitter-westboro-baptist-church-megan-phelps-roper.

29. Megan Phelps-Roper, "I Grew Up in the Westboro Baptist Church; Here's Why I Left," *Ted Talk*, February 2017, https://www.ted.com/talks/megan_phelps_roper_i_grew_up_in_the_westboro_baptist_church_here_s_why_i_left?utm_campaign=social&utm_medium=referral&utm_source=facebook.com&utm_content=talk&utm_term=global-social%20issues#t-627390.

30. J. M. Berger and Jonathon Morgan, "The ISIS Twitter Census," Brookings Analysis Paper no. 20, March 2015, p. 58, https://www.brookings.edu/wp-content/uploads/2016/06/isis_twitter_census_berger_morgan.pdf.

31. *Whitney*, 274 U.S. at 375 (concurring, J. Brandeis); Vincent Blasi, "The Checking Value in First Amendment Theory," *American Bar Foundation Research Journal* 2, no. 3 (1977): 521.

32. Berger and Morgan, "The ISIS Twitter Census," p. 3.

33. Article 19 of the International Covenant on Civil and Political Rights allows states to limit freedom of expression under circumstances that satisfy proportionality review.

34. Scott Craig and Emma Llansó, "Pressuring Platforms to Censor Content Is Wrong Approach to Combatting Terrorism" (blog of the Center for Democracy & Technology), November 5, 2015, https://cdt.org/blog/pressuring-platforms-to-censor-content-is-wrong-approach-to-combatting-terrorism/ (arguing that when government seeks to police speech, notably extremism, through TOS, those requests should be grounded in legal frameworks rooted in international human rights rather than TOS).

35. Rose, *The Tyranny of Silence*, pp. 150–51. As Floyd Abrams explains in *The Soul of the First Amendment* (New Haven, CT: Yale University Press 2017), pp. 44–45, the European Court of Human Rights has upheld hate-speech convictions involving criticism of politicians and bigoted views expressed by politicians.

36. Council of Europe, "Council of Europe Secretary General Concerned about Internet Censorship: Rules for Blocking and Removal of Illegal Content Must Be Transparent and Proportionate," news release, June 1, 2016, http://www.coe.int/en/web/tbilisi/-/council-of-europe-secretary-general-concerned-about-internet-censorship-rules-for-blocking-and-removal-of-illegal-content-must-be-transparent-and-prop?desktop=false.

37. Ibid.

38. European Digital Rights, "Input on Human Rights and Preventing and Countering Violent Terrorism," March 18, 2016, https://edri.org/files/2016-UN-consultation.pdf.

39. Ibid.

40. "Removal Requests," Twitter, https://transparency.twitter.com/en/removal-requests.html.

41. Article 19, *Freedom of Expression and the Private Sector in the Digital Age: Article 19's Written Comments*, Office of the United Nations High Commissioner for Human Rights, http://www.ohchr.org/Documents/Issues/Expression/PrivateSector/Article19.pdf.

42. Danielle Keats Citron, "Technological Due Process," *Washington University Law Review* 85, no. 6 (2007): 1249.

43. Emma Llansó, (director of free expression, Center on Democracy and Technology), in interview with the author, January 15, 2017.

44. Emma Llansó, "Takedown Collaboration by Private Companies Creates Troubling Precedent"

(blog of the Center for Democracy & Technology), December 6, 2016, https://cdt.org/blog/takedown-collaboration-by-private-companies-creates-troubling-precedent/.

45. Liane Lovitt, "Why Transparency Reports Matter Now More Than Ever," *Medium*, May 13, 2016, https://medium.com/inflection-points/why-transparency-reports-matter-now-more-than-ever-9fb6ebe733fa.

46. "Removal Requests," Twitter.

47. "Government TOS Reports," Twitter, https://transparency.twitter.com/en/gov-tos-reports.html (in the six-month period from July 2016 to December 2016, Twitter received 716 reports from governments worldwide related to 5,929 accounts; 85 percent were removed by Twitter for terms-of-service violations related to violent extremism, https://transparency.twitter.com/en/removal-requests.html shows breakdown by country).

48. "Government TOS Reports," Twitter; "Content Removal Requests Report," Microsoft, https://www.microsoft.com/en-us/about/corporate-responsibility/crrr.

49. Freedom Online Coalition Working Group 3, *Submission to UN Special Rapporteur David Kaye: Study on Freedom of Expression and the Private Sector in the Digital Age*, Office of the United Nations High Commissioner for Human Rights, http://www.ohchr.org/Documents/Issues/Expression/PrivateSector/FreedomOnlineCoalition.pdf.

50. "About ONO," Organization of News Ombudsmen, http://newsombudsmen.org/about-ono.

51. Ibid.

RELATED PUBLICATIONS FROM THE CATO INSTITUTE

Staring at the Sun: An Inquiry into Compulsory Campaign Finance Donor Disclosure Laws by Eric Wang, Cato Institute Policy Analysis (forthcoming 2017)

The State of Free Speech and Tolerance in America: Attitudes about Free Speech, Campus Speech, Religious Liberty, and Tolerance of Political Expression by Emily Ekins, Cato Institute Survey Report (October 31, 2017)

Commercial Speech and the Values of Free Expression by Martin H. Redish, Cato Institute Policy Analysis no. 813 (June 19, 2017)

Freedom of Speech under Assault on Campus by Daniel Jacobson, Cato Institute Policy Analysis no. 796 (August 30, 2016)

Hate Speech Laws: Ratifying the Assassin's Veto by Robert Corn-Revere, Cato Institute Policy Analysis no. 791 (May 24, 2016)

Move to Defend: The Case against the Constitutional Amendments Seeking to Overturn *Citizens United* by John Samples, Cato Institute Policy Analysis no. 724 (April 23, 2013)

The DISCLOSE Act, Deliberation, and the First Amendment by John Samples, Cato Institute Policy Analysis no. 664 (June 25, 2010)

The Fallacy of Campaign Finance Reform by John Samples, University of Chicago Press (2006)

RECENT STUDIES IN THE CATO INSTITUTE POLICY ANALYSIS SERIES

827. **Corruption and the Rule of Law: How Brazil Strengthened Its Legal System** by Geanluca Lorenzon (November 20, 2017)

826. **Liberating Telemedicine: Options to Eliminate the State-Licensing Roadblock** by Alex Nowrasteh (November 15, 2017)

825. **Border Patrol Termination Rates for Discipline and Performance** by Alex Nowrasteh (November 2, 2017)

824. **The Coming Transit Apocalypse** by Randal O'Toole (October 24, 2017)

823. **Zoning, Land-Use Planning, and Housing Affordability** by Vanessa Brown Calder (October 18, 2017)

822. **Unforced Error: The Risks of Confrontation with Iran** by Emma Ashford and John Glaser (October 9, 2017)

821. **Why the United States Should Welcome China's Economic Leadership** by Colin Grabow (October 3, 2017)

820. **A Balanced Threat Assessment of China's South China Sea Policy** by Benjamin Herscovitch (August 28, 2017)

819. **Doomed to Repeat It: The Long History of America's Protectionist Failures** by Scott Lincicome (August 22, 2017)

818. **Preserving the Iran Nuclear Deal: Perils and Prospects** by Ariane Tabatabai (August 15, 2017)

817. **Reforming the National Flood Insurance Program: Toward Private Flood Insurance** by Ike Brannon and Ari Blask (July 19, 2017)

816. **Withdrawing from Overseas Bases: Why a Forward-Deployed Military Posture Is Unnecessary, Outdated, and Dangerous** by John Glaser (July 18, 2017)

815. **Cybersecurity or Protectionism? Defusing the Most Volatile Issue in the U.S.–China Relationship** by Daniel Ikenson (July 13, 2017)

814. **Step Back: Lessons for U.S. Foreign Policy from the Failed War on Terror** by A. Trevor Thrall and Erik Goepner (June 26, 2017)

813. **Commercial Speech and the Values of Free Expression** by Martin H. Redish (June 19, 2017)

812. **Would More Government Infrastructure Spending Boost the U.S. Economy?** by Ryan Bourne (June 6, 2017)

811. **Four Decades and Counting: The Continued Failure of the War on Drugs** by Christopher J. Coyne and Abigail R. Hall (April 12, 2017)

810. **Not Just Treading Water: In Higher Education, Tuition Often Does More than Replace Lost Appropriations** by Neal McCluskey (February 15, 2017)

809. **Stingray: A New Frontier in Police Surveillance** by Adam Bates (January 25, 2017)

CATO INSTITUTE

Published by the Cato Institute, Policy Analysis is a regular series evaluating government policies and offering proposals for reform. Nothing in Policy Analysis should be construed as necessarily reflecting the views of the Cato Institute or as an attempt to aid or hinder the passage of any bill before Congress. Contact the Cato Institute for reprint permission. All policy studies can be viewed online at www.cato.org. Additional printed copies of Cato Institute Policy Analysis are $6.00 each ($3.00 each for five or more). To order, please email catostore@cato.org.