

ARTICLE

REMEDIATING SOCIAL MEDIA: A LAYER-CONSCIOUS APPROACH

ANNEMARIE BRIDY¹

*“We didn’t focus on how you could wreck this system intentionally.”
– Vint Cerf, Co-inventor of the Internet Protocol²*

CONTENTS

INTRODUCTION	194
I. A HISTORY OF LAYER-CONSCIOUS INTERNET REGULATION	199
A. <i>The End-to-End Principle in Network Design</i>	199
B. <i>Regulation at the Network Layer: Common Carriage and Content-Agnosticism</i>	201
C. <i>Regulation at the Application Layer: Safe Harbors and Content-Awareness</i>	205
II. EVOLVING SOCIAL NORMS FOR SPEECH ON SOCIAL MEDIA	213
A. <i>Shifting Political Winds</i>	214
B. <i>Failures in Social Media’s “Marketplace of Ideas”</i>	216
III. BALANCING FREE SPEECH WITH INFORMATION QUALITY	219
A. <i>Clarity</i>	220
B. <i>Consistency</i>	222
C. <i>Appealability</i>	225
CONCLUSION	228

¹ Professor, University of Idaho College of Law; Affiliate Scholar, Stanford Center for Internet and Society; Affiliated Fellow, Yale Information Society Project. This article was prepared for the symposium “Governing the Internet: Public Access, Private Regulation,” held on February 2, 2018 at Boston University School of Law. The author would like to thank the editors of the Journal of Science & Technology Law and the faculty of BU Law for the invitation to participate. I owe additional thanks to the symposium’s other participants and to John Blevins, Casey Inge, Daphne Keller, Blake Reid, and Hal Singer.

² Lee Rainie & Janna Anderson, *The Fate of Online Trust in the Next Decade*, PEW RES. CTR., at 2 (Aug. 10, 2017), http://assets.pewresearch.org/wp-content/uploads/sites/14/2017/08/09163223/PI_2017.08.10_onlineTrustNextDecade_FINAL.pdf [<https://perma.cc/YK8P-ABYU>].

INTRODUCTION

2017 was a bad year for the Internet. For executives at Facebook, Google, and Twitter, the year culminated in Congressional testimony explaining their firms' roles in a covert Russian operation to skew the outcome of the 2016 presidential election.³ The Russians' objective was to use American social media platforms to spread disinformation and stoke racial resentment in ways the Russians believed would benefit the campaign of Donald J. Trump.⁴ Social media content created by Russian agents working around the clock for the secretive Internet Research Agency reached more than 126 million users on Facebook alone.⁵ The full extent of the operation will likely never be known.

Journalists trying to get to the bottom of the Russian election meddling story discovered pathologies of the Internet's attention economy that legal and media studies scholars have been writing about for the last several years.⁶ From filter bubbles and clickbait to revenge porn and "fake news," the antisocial effects of social media are now front and center in a serious public debate about the future of the Internet and the firms that have come to dominate it.⁷ As the public learns more about the ease with which the Internet's most popular platforms can be exploited to harass, deceive, and manipulate their users, there is a growing consensus that the Internet is broken and that tech titans dominating the Internet's edge are largely to blame.⁸

³ See Cecilia Kang et al., *Tech Executives Are Contrite about Election Meddling, but Make Few Promises on Capitol Hill*, N.Y. TIMES (Oct. 31, 2017), <https://www.nytimes.com/2017/10/31/us/politics/facebook-twitter-google-hearings-congress.html>.

⁴ Mike Isaac & Daisuke Wakabayashi, *Russian Influence Reached 126 Million Through Facebook Alone*, N.Y. TIMES (Oct. 30, 2017), <https://www.nytimes.com/2017/10/30/technology/facebook-google-russia.html>.

⁵ *Id.*

⁶ See, e.g., MARK BARTHOLOMEW, *ADCREEP: THE CASE AGAINST MODERN MARKETING* (2017) (exploring the inextricable connection between online advertising and pervasive surveillance technologies); TIM WU, *THE ATTENTION MERCHANTS: THE EPIC SCRAMBLE TO GET INSIDE OUR HEADS* (2016) (tracing the evolution of the Internet's dominant business model: the commercialization of mass attention); Zeynep Tufekci, *Facebook's Ad Scandal Isn't a 'Fail,' It's a Feature*, N.Y. TIMES (Sept. 23, 2017), <https://www.nytimes.com/2017/09/23/opinion/sunday/facebook-ad-scandal.html> (arguing that Facebook's advertising engine is perfectly designed for targeted behavioral manipulation in politics as well as commerce).

⁷ See, e.g., Jean M. Twenge, *Have Smartphones Destroyed a Generation?*, THE ATLANTIC, Sept. 2017, <https://www.theatlantic.com/magazine/archive/2017/09/has-the-smartphone-destroyed-a-generation/534198/> [<https://perma.cc/T93D-KFBL>] (connecting post-Millennial teens' pervasive use of smartphones and social media to rising anxiety and declining mental health).

⁸ The popularity of this thesis is reflected in a number of recent monographs written for a mass-market audience. See, e.g., FRANKLIN FOER, *WORLD WITHOUT MIND: THE EXISTENTIAL THREAT OF BIG TECH* (2017); SCOTT GALLOWAY, *THE FOUR: THE HIDDEN DNA OF AMAZON, APPLE, FACEBOOK, AND GOOGLE* (2017); JONATHAN TAPLIN, *MOVE FAST AND BREAK THINGS:*

The narrative that Big Tech is an existential threat to democracy is approaching the territory of moral panic, but underlying concern about the power of dominant platforms to shape and steer public discourse is justified. Facebook CEO Mark Zuckerberg famously said that Facebook has grown to be more like a government than a traditional firm.⁹ To make that statement more concrete, consider that over one-fourth of the world's population is now governed by Facebook's Terms of Service.¹⁰ Twitter's user base is nowhere near the size of Facebook's, but tweets are now the primary medium through which the President of the United States fitfully communicates the country's domestic and foreign policies to the world.¹¹ Google, for its part, owns around 75% of the global market for online searches,¹² and YouTube streams a billion hours of video daily to over a billion users worldwide.¹³ Network effects and platform economics have transformed these firms from information services into "functional sovereigns" whose power and reach surpass those of Westphalian nation states.¹⁴

HOW FACEBOOK, GOOGLE, AND AMAZON CORNERED CULTURE AND UNDERMINED DEMOCRACY (2017); SIVA VAIDHYANATHAN, *ANTI-SOCIAL MEDIA: HOW FACEBOOK HAS DISCONNECTED CITIZENS AND UNDERMINED DEMOCRACY* (2018).

⁹ DAVID KIRKPATRICK, *THE FACEBOOK EFFECT* 254 (2010).

¹⁰ See Ben Popper, *A Quarter of the World's Population Now Uses Facebook Every Month*, *THE VERGE* (May 3, 2017, 4:34 PM), <https://www.theverge.com/2017/5/3/15535216/facebook-q1-first-quarter-2017-earnings> [<https://perma.cc/RS79-WQKU>].

¹¹ See Elizabeth Landers, *White House: Trump's Tweets are 'Official Statements.'*, *CNN*, <https://www.cnn.com/2017/06/06/politics/trump-tweets-official-statements/index.html> [<https://perma.cc/C66N-KK5M>] (last updated June 6, 2017, 4:37 PM); see also Jonah Shepp, *How U.S. Foreign Policy is Being Shaped by Trump's Tweets*, *N.Y. MAG* (Jan. 19, 2018, 11:41 AM), <http://nymag.com/daily/intelligencer/2018/01/how-u-s-foreign-policy-is-being-shaped-by-trumps-tweets.html> [<https://perma.cc/3VKD-QBEY>].

¹² George Slefo, *Google's Search Domination Is Eroding Because of Amazon, Apps*, *ADAGE* (Jan. 10, 2018), <http://adage.com/article/digital/amazon-apps-chew-google-s-search-market-share/311910/> [<https://perma.cc/B9NW-6JKW>].

¹³ Jack Nicas, *YouTube Tops 1 Billion Hours of Video a Day, on Pace to Eclipse TV*, *WALL STREET J.* (Feb. 27, 2017, 1:40 PM), <https://www.wsj.com/articles/youtube-tops-1-billion-hours-of-video-a-day-on-pace-to-eclipse-tv-1488220851>.

¹⁴ See Frank Pasquale, *From Territorial to Functional Sovereignty: The Case of Amazon*, *L. & POL. ECON. BLOG* (Dec. 6, 2017), <https://lpeblog.org/2017/12/06/from-territorial-to-functional-sovereignty-the-case-of-amazon/> [<https://perma.cc/RNF8-LBDJ>] (arguing that major digital firms are "market makers, able to exert regulatory control over the terms on which others can sell goods and services" and that "they aspire to displace more government roles over time, replacing the logic of territorial sovereignty with functional sovereignty . . .").

The drumbeat for a regulatory response is getting louder. And it's coming from points across the political spectrum.¹⁵ Some are calling for interventions in the area of antitrust law.¹⁶ Others have proposed imposing at the Internet's application layer content neutrality rules that have historically applied only at the network layer.¹⁷ To describe such rules, conservative activist Phil Kerpen coined the term "layer-neutral net neutrality."¹⁸ Supporters of this approach assert that rules requiring social media platforms to behave like network infrastructure providers in their handling of users' content will enhance freedom of expression and limit the role of dominant platforms as gatekeepers of the privatized public sphere.¹⁹ Kerpen, who staunchly opposed the FCC's 2015 Open Internet Order imposing net neutrality rules at the network layer,²⁰ argues that social media

¹⁵ See, e.g., Pamela A. Maclean, *Google Resists Becoming Digital 'Town Square' in Censorship Spat*, BLOOMBERG (Mar. 15, 2018, 5:00 AM), <https://www.bloomberg.com/news/articles/2018-03-15/google-resists-becoming-digital-town-square-in-censorship-spat> ("Silicon Valley's social media giants are under attack from both the left and the right for not doing enough to police hate speech, terrorist propaganda and Russian election meddling.").

¹⁶ See, e.g., Steven Strauss, Opinion, *Is it Time to Break Up the Big Tech Companies?*, LA TIMES (June 30, 2016, 6:00 AM), <http://www.latimes.com/opinion/op-ed/la-oe-strauss-digital-robber-barons-break-up-monopolies-20160630-snap-story.html> [<https://perma.cc/3AXC-ZYVH>] ("Government antitrust interventions have promoted innovation in the past, and should again be employed to make conditions more favorable for competitors and consumers."). But see Greg Ip, *The Antitrust Case Against Facebook, Google and Amazon*, WALL STREET J. (Jan. 16, 2018, 11:52 AM), <https://www.wsj.com/articles/the-antitrust-case-against-facebook-google-amazon-and-apple-1516121561> (discussing the low likelihood of antitrust action against dominant digital platforms because of the limited reach of antitrust law's consumer welfare standard for assessing competitive harm).

¹⁷ See David McCabe, *One Idea for Regulating Google and Facebook's Control Over Content*, AXIOS (Aug. 18, 2017), <https://www.axios.com/one-idea-for-regulating-google-and-facebooks-control-over-content-1513304938-26b2f2ae-90b7-4f6a-b12f-012aad621e3b.html> [<https://perma.cc/XZY6-XEAQ>] (publishing a "Confidential Policy Memorandum" by Phil Kerpen that introduces and explains the concept of "layer-neutral net neutrality").

¹⁸ *Id.*

¹⁹ Kerpen and other conservatives, including Republican FCC Chairman Ajit Pai, allege that the content moderation policies of Google, Facebook, and Twitter discriminate against political and social conservatives, censoring speech on the right even as the platforms claim to embrace viewpoint neutrality. See *id.* (quoting the memo in which Kerpen accuses Google, Facebook, and Twitter of "represent[ing] themselves as politically neutral while systematically promoting liberal views and limiting or even banning conservatives"); Hamza Shaban, *FCC Chairman Says Social Media Platforms Lack Transparency in How They Restrict Conservative Content*, WASH. POST (Dec. 12, 2017), https://www.washingtonpost.com/news/the-switch/wp/2017/12/12/fcc-chairman-says-social-media-platforms-lack-transparency-in-how-they-restrict-conservative-content/?utm_term=.03cfc09fdefd (reporting on Pai's call for "more scrutiny" over decisions by Web platforms "to restrict political content with pro-Trump and conservative messaging").

²⁰ The 2015 Open Internet Order, which applied to providers of residential Broadband Internet Access Service, imposed rules at the network layer prohibiting providers from blocking

companies and network infrastructure operators should be subject to the same no-blocking rules because they are similarly situated in terms of market power and user lock-in effects.²¹ Former Democratic Senator Al Franken, a stalwart supporter of the 2015 Open Internet Order, offered the same rationale in an op-ed in *The Guardian*.²² Franken wrote that “no one company should have the power to pick and choose which content reaches consumers and which doesn’t. And Facebook, Google, and Amazon—like [Internet service providers]—should be ‘neutral’ in their treatment of lawful information and commerce on their platforms.”²³

The concept of layer-neutral net neutrality also has proponents in the legal academy. In an article exploring regulatory tools for promoting expressive freedom in the era of Big Tech, Frank Pasquale calls for “platform neutrality” in the form of a must-carry obligation for search and social media platforms.²⁴ In the article, Pasquale criticizes tech companies for the Janus-faced posture they have taken over the years with respect to their relationship to online speech.²⁵ These firms have resisted regulatory burdens by casting themselves as First Amendment speakers (for example, with respect to the ranking of search results).²⁶ At the same time, they have claimed regulatory immunities by casting themselves

lawful content, throttling lawful content, or discriminating against lawful content on the basis of its source or destination. Protecting and Promoting the Open Internet, 30 FCC Rcd. 5601, ¶ 4 (2015) [hereinafter 2015 Open Internet Order]. It also imposed transparency rules on Broadband Internet Access Service providers concerning the speed of their service and their network management practices. *Id.* ¶ 24.

²¹ McCabe, *supra* note 17.

²² Al Franken, *We Must Not Let Big Tech Threaten Our Security, Freedoms, and Democracy*, GUARDIAN (Nov. 8, 2017, 2:20 PM), <https://www.theguardian.com/commentis-free/2017/nov/08/big-tech-security-freedoms-democracy-al-franken> [<https://perma.cc/PM3T-3FX4>].

²³ *Id.*

²⁴ See Frank Pasquale, *Platform Neutrality: Enhancing Freedom of Expression in Spheres of Private Power*, 17 THEORETICAL INQUIRIES IN L. 487, 497-503 (2016) (proposing the adoption of must-carry rules for search and social media platforms) [hereinafter Pasquale, *Platform Neutrality*]. In the background of Pasquale’s article is a 2014 report from the French Digital Council calling for “platform neutrality” in the form of transparency, non-monopolization, and non-discrimination rules for edge providers, most notably Google. See *Platform Neutrality: Building an Open and Sustainable Digital Environment*, CNUM, at 11, 19 (2014), https://ec.europa.eu/futurium/en/system/files/ged/platformneutrality_va.pdf [<https://perma.cc/555G-6JY6>]. Among the report’s recommended neutrality rules is one requiring “non-discrimination between forms of expression and shared content.” *Id.* at 19.

²⁵ Pasquale, *Platform Neutrality*, *supra* note 24 at 496. (“Large technology platforms’ strategic, opportunistic, and contradictory self-characterizations take advantage of the siloed nature of legal disputes.”). See also Tarleton Gillespie, *The Politics of ‘Platforms’*, 12 NEW MEDIA & SOC’Y 347, 356 (2010) (discussing Google’s “effort to inhabit the middle [and be] rewarded for facilitating expression but not liable for [expression’s] excesses”).

²⁶ Pasquale, *Platform Neutrality*, *supra* note 24, at 494-96.

as mere conduits for third-party speech (for example, with respect to defamation and copyright infringement).²⁷ For Pasquale, the game is up, and the time has come to hold private platforms responsible for the central role they have assumed in public life and public discourse. One way of accomplishing that goal, he concludes, would be for regulators to subordinate corporate free speech rights to those of human speakers and regulate platforms as digital utilities with common carriage obligations.²⁸

This Article is a high-level effort to explain, in terms of both regulatory history and shifting public attitudes about online speech, why adopting a must-carry obligation for social media platforms is not what the Internet needs now.²⁹ Such a requirement would more likely exacerbate than remediate the problems with information quality and integrity described in the opening paragraphs above. Part I discusses the historical layer-consciousness of Internet regulation and explains the public policies underlying differential treatment of “core” and “edge” services. Part II considers evolving speech norms at the Internet’s edge and the increasing pressure on social media platforms to more actively address some demonstrable failures in social media’s “marketplace of ideas.” Part III argues that a must-carry rule for social media platforms is precisely the wrong regulatory approach for addressing those failures. The better prescription, I argue, is to breathe new life into the underused “Good Samaritan” provision in § 230 of the Communications Decency Act, which was intended to protect and promote good faith content moderation at the Internet’s edge. What the Internet needs

²⁷ *Id.* at 494.

²⁸ *See id.* at 503 (arguing that “massive internet platforms must take the bitter with the sweet: if they want to continue avoiding liability for intellectual property infringement and defamation, they should welcome categorization as a conduit for speech, rather than speaker status itself”).

²⁹ Whereas Pasquale uses the term “platform” very broadly to include search, social media, and ecommerce services, *id.*, I focus here exclusively on social media platforms. As a general matter, current discussions of how to regulate Big Tech would benefit from a more nuanced appreciation of both the heterogeneity of the platform landscape and the multi-platform nature of the Internet’s largest firms. Google, for example, is not one platform; it is a network of platforms, including Search, YouTube, AdSense, and Android. *Our Products*, GOOGLE, <https://www.google.com/about/products/> [<https://perma.cc/HD7M-BSHZ>] (last visited May 26, 2018). Non-discrimination rules may be appropriate for some types of platforms but not others, depending, for example, on the availability of alternative providers in the relevant market and whether the platform operator is in a position to give preferential treatment to its own vertically integrated properties. *Cf.* Annabelle Gawer & Michael A. Cusumano, *Industry Platforms and Ecosystem Innovation*, 31 J. PROD. INNOV. MANAG. 417, 421 (2013) (pointing out that “the rise of industry platforms [raises] complex social welfare questions regarding trade-offs between the social benefits of platform-compatible innovation versus the potentially negative effects of preventing competition in overall systems”).

now is an awakening to what James Grimmelmann has called “the virtues of moderation.”³⁰

I. A HISTORY OF LAYER-CONSCIOUS INTERNET REGULATION

The concept of neutrality has positive connotations of fairness and equity, and it has long been at the heart of our understanding of the Internet as an open information system. As Andrea Renda argues, however, there is danger in regarding neutrality as a totem—as a principle that should be translated into regulatory prescriptions for every kind of intermediary across the Internet ecosystem.³¹ Internet regulation has never been layer-neutral. This Part explains how and why providers of the Internet’s physical infrastructure (i.e., network operators) and providers of applications that run atop that infrastructure (i.e., edge providers) have historically been regulated differently. Recent proposals for “platform neutrality” rules should be evaluated with reference to this regulatory history and in the context of the policy goals underlying historical regulatory choices. In the context of previous debates over search engine neutrality, John Blevins argued—and I agree—that layers should matter for regulatory purposes because different layers of modern digital networks have very different economic and technological attributes.³²

A. *The End-to-End Principle in Network Design*

At the heart of both the Internet’s initial design and the concept of net neutrality is a basic principle of networking known as the end-to-end principle.³³ Stated in non-technical terms, the end-to-end principle provides that functional specialization and diversity should be located at the endpoints (or edges) of the network, while the core or center of the network should be functionally limited

³⁰ James Grimmelmann, *The Virtues of Moderation*, 17 YALE J. L. & TECH. 42 (2015). Grimmelmann defines “moderation” as “the governance mechanisms that structure participation in a community to facilitate cooperation and prevent abuse.” *Id.* at 47.

³¹ Andrea Renda, *Antitrust, Regulation, and the Neutrality Trap: A Plea for a Smart, Evidence-Based Internet Policy*, CEPS, at 2 (Apr. 2015), <https://www.ceps.eu/publications/antitrust-regulation-and-neutrality-trap> [<https://perma.cc/Y8V3-SVAG>].

³² John Blevins, *The New Scarcity: A First Amendment Framework for Regulating Access to Digital Media Platforms*, 79 TENN. L. REV. 353, 353 (2012). Blevins frames his argument for layer-conscious regulation in terms of infrastructural scarcity: “[A]pplication-layer platforms are far more competitive and contestable than network-layer platforms, even when a given platform becomes dominant. Access regulations at the application layer are therefore less necessary and will likely stifle both speech and innovation.” *Id.* at 358.

³³ See BARBARA VAN SCHEWICK, INTERNET ARCHITECTURE AND INNOVATION 96 (2010) (explaining the “broad version” of the end-to-end principle); Mark A. Lemley & Lawrence Lessig, *The End of End-to-End: Preserving the Architecture of the Internet in the Broadband Era*, 48 UCLA L. REV. 925, 930-31 (2001) (explaining the end-to-end arguments first articulated by Saltzer, Reed, and Clark in 1981).

to carrying and routing data.³⁴ The physical infrastructure of the network in an end-to-end model is underspecified, meaning that the network's hardware is not specialized to carry any particular type of data or to serve any particular type of application.³⁵ Instead, the core of the network is agnostic about the type of data it carries, and it treats all the data it carries in the same way.³⁶ Its sole purpose is to route traffic between endpoints according to instructions provided by the endpoints themselves.³⁷ David Isenberg described the early Internet's design as a "bits-in, bits-out" model: the bits go in one end and come out the other, no matter whether they're "voice, music, bank balances, e-mail or TV."³⁸ Isenberg contrasted this type of network, which he dubbed a "Stupid Network," with the public-switched telephone network—an "Intelligent Network" that was tightly specified for carrying voice signals and therefore suitable for not much else.³⁹

On the Internet, the key to getting data from one end of the network to the other is the Internet Protocol (IP).⁴⁰ IP is the open-standard networking protocol that allows heterogeneously configured local area networks from all over the world to interconnect with one another.⁴¹ IP is the lingua franca that makes the global Internet—with a capital "I"—possible by allowing packets of data to be routed accurately between any source and destination on the Internet.⁴² Each Internet endpoint has a unique IP address at which it can receive, and from which

³⁴ See VAN SCHEWICK, *supra* note 33, at 96 ("[A]pplication-specific functionality usually cannot—and preferably should not—be implemented in the lower layers of the network, the network's core. . . . [L]ower layers, or the core of the network, should provide only general services of broad utility across applications, whereas application-specific functionality should be implemented in the higher layers at the end hosts.").

³⁵ See *id.* at 107 ("The broad version of the end-to-end arguments advises that the network should not contain application-specific functionality and should provide only general services useful for a large variety of applications.").

³⁶ See *id.* at 72–73 (explaining that in an end-to-end network, the network is not application-aware and does not positively or negatively affect the execution of applications running over it).

³⁷ See *id.* at 73 ("The application-blindness of an end-to-end network prevents the network owner from discriminating against applications running on its network.").

³⁸ David S. Isenberg, *The Dawn of the "Stupid Network"*, ACM NETWORKER, Feb./Mar. 1998, at 28.

³⁹ *Id.* at 26–28.

⁴⁰ See *id.* at 28–29 (discussing the function and importance of IP); see also Nadeem Unuth, *What IP Means and How It Works*, LIFEWIRE, <https://www.lifewire.com/internet-protocol-explained-3426713> [<https://perma.cc/FW89-A5DV>] (last updated May 13, 2018) (explaining that IP "describes how data packets move through a network").

⁴¹ See Isenberg, *supra* note 38, at 28 ("The foremost design goal of IP is to cross multiple, physically different networks. To IP, it doesn't matter if the underlying transport is circuit, SONET, Ethernet, Bitnet, FDDI, or smoke signals.").

⁴² See *id.* at 28–29 (explaining that IP's routing function enables internetworking and that "the Internet . . . is a virtual network"—a "network of networks"—that is independent of wires and transport protocols").

it can send, data.⁴³ To quote Isenberg again, “IP makes the details of the network irrelevant,” because it allows transfers to occur between endpoints regardless of the underlying wires and transport protocols.⁴⁴ In this sense, the Internet is a virtual network that allows developers at the network’s edge to design and deploy new services and applications without having to rely on network operators to build any new functionality into the physical core of the network.⁴⁵ We say that application software (e.g., a web browser or email client) runs “on top” of the Internet because its design and operation are independent of the network’s underlying hardware and software layers.⁴⁶

B. Regulation at the Network Layer: Common Carriage and Content-Agnosticism

Putting a network’s users, as opposed to the network’s operators, in control of what gets sent and received over the network finds its legal analog in the regulatory principle of common carriage.⁴⁷ At early common law, a variety of businesses, including innkeepers, railroads, warehouses, package carriers, and ferry operators, were classified as common carriers.⁴⁸ For the most part, these businesses, as the term denotes, carried people or goods from place to place.⁴⁹ Historically, the essential attributes of common carriage have been nondiscriminatory public access and indifference to the nature of the goods carried.⁵⁰ Common carriers of goods are required to accept all lawful packages for transport to

⁴³ See Unuth, *supra* note 40 (“The IP protocol standardizes the way machines over the Internet or any IP network forward or route their packets based on their IP addresses.”).

⁴⁴ Isenberg, *supra* note 38, at 29.

⁴⁵ *Id.*

⁴⁶ See VAN SCHEWICK, *supra* note 33, at 71 (“Application autonomy implies a hierarchical relationship between applications and the network: the applications are in control, and the network has a serving role. Lower layers are responsible for providing very general building blocks, which can then be used by the application designer to realize application-specific needs.”).

⁴⁷ Sascha D. Meinrath & Victor W. Pickard, *Transcending Net Neutrality: Ten Steps Toward an Open Internet*, 12 J. INTERNET L. 1, 12 (2008) (asserting that network neutrality is “inextricably linked” to common carriage principles).

⁴⁸ Thomas B. Nachbar, *The Public Network*, 17 COMMLAW CONCEPTUS 67, 76 (2008); Kevin Werbach, *Only Connect*, 22 BERKELEY TECH. L.J. 1233, 1246 (2007).

⁴⁹ Nachbar, *supra* note 48, at 76.

⁵⁰ *Id.* at 107 (“Nondiscrimination has been implemented almost exclusively with regard to delivery of undifferentiated services, such as carriage. . . . The identity of the transported good is largely irrelevant.”); Werbach, *supra* note 48, at 1246 (explaining that “[a] common carrier cannot . . . differentiate in the treatment of similarly situated customers, evaluate the content of what it receives from its customers, or refuse to serve interested customers, even when that means building out its facilities to reach them”).

destinations specified by their senders, and they cannot modify the contents of those packages en route.⁵¹

In telecommunications law, common carriage rules apply to services that are classified as “telecommunications services” under Title II of the Telecommunications Act of 1996.⁵² In 2015, the FCC adopted the Open Internet Order, which classified last-mile broadband services as telecommunications services—and therefore common carriers—for regulatory purposes.⁵³ For the preceding decade, the FCC had classified broadband Internet access services as “information services” not bound by Title II common carriage rules.⁵⁴ Even before Title II reclassification, however, the FCC claimed and asserted regulatory authority over broadband providers “to ensure that providers of telecommunications for Internet access or Internet Protocol-enabled (IP-enabled) services are operated in a neutral manner.”⁵⁵ Title II reclassification thus changed the law on the books but not the law on the ground for network operators.⁵⁶ For all intents and purposes, the end-to-end principle (i.e., net neutrality) has always been at the heart of the Internet’s technical and policy architectures.

⁵¹ See, e.g., *Fed. Trade Comm’n v. Verity Int’l, Ltd.*, 443 F.3d 48, 58 (2d Cir. 2006) (explaining that “the definition of a common carrier coalesced into two requirements: (1) the entity holds itself out as undertaking to carry for all people indifferently; and (2) the entity carries its cargo without modification”).

⁵² For a compact history of telecommunications common carriage regulation, see Susan P. Crawford, *Transporting Communications*, 89 B.U.L. REV. 871 (2009).

⁵³ 2015 Open Internet Order, *supra* note 20, ¶ 403. Broadband providers challenged the legality of the 2015 Open Internet Order as soon as it was adopted. See *United States Telecom Ass’n v. Fed. Comm’n Comm’n*, 825 F.3d 674 (D.C. Cir. 2016) (challenging the 2015 Open Internet Order on statutory and constitutional grounds). The court upheld the 2015 Open Internet Order. See *id.* at 744 (denying the petitions for review).

⁵⁴ See *id.* at 692 (explaining that after 2005, when the Supreme Court affirmed the FCC’s classification of cable broadband as an information service, the FCC classified other types of broadband service, such as DSL and mobile broadband, as information services without a standalone offering of telecommunications). Prior to 2005, broadband DSL, which relied on phone lines to deliver Internet access, had been classified as a telecommunications service to which common carrier rules applied. See *id.* at 691.

⁵⁵ *Appropriate Framework for Broadband Access to the Internet Over Wireline Facilities*, 20 FCC Rcd. 14986, ¶ 4 (2005). Subject to “reasonable network management,” the neutrality principles the FCC laid out in 2005 were intended to ensure consumers had the right to (1) “access the lawful Internet content of their choice;” (2) “run applications and use services of their choice;” (3) “connect their choice of legal devices that do not harm the network;” and (4) enjoy “competition among network providers, application and service providers, and content providers.” *Id.*

⁵⁶ The history of the FCC’s regulatory treatment of broadband service is complex and littered with litigation. The 2015 Open Internet Order contains a detailed narrative of that history, culminating in the assertion that “[t]he Commission has steadily and consistently worked to protect the open Internet for the last decade, starting with the adoption of the *Internet Policy Statement* up through its recent *2014 Open Internet NPRM* following the D.C. Circuit’s *Verizon* decision.” 2015 Open Internet Order, *supra* note 20, ¶ 328.

In the 2015 Open Internet Order, the FCC promulgated three bright-line net neutrality rules to govern lawful transfers of data between ends on the Internet: (1) no blocking; (2) no throttling (i.e., no slowing or degrading); and (3) no paid prioritization (i.e., no “fast lanes” for edge services willing to pay for special treatment).⁵⁷ ISPs were permitted under the 2015 Open Internet Order to engage in reasonable network management to prevent congestion and to ensure quality of service for types of traffic, like streaming video, that are sensitive to network latency.⁵⁸ Unlike traditional telecommunications services, ISPs under the 2015 Open Internet Order were not subject to rate regulation, a forbearance the FCC described as consistent with its historical “light touch” approach to regulating Internet infrastructure.⁵⁹

The 2015 Open Internet Order’s net neutrality rules codified the end-to-end principle. They were premised on the belief that keeping the core of the Internet open and underspecified promotes innovation at the edge by making it cheap and relatively frictionless for users and application developers to do their own thing.⁶⁰ The “ability to ‘just do it,’” Isenberg wrote, “liberates huge amounts of innovative energy.”⁶¹ Openness at the Internet’s physical and network layers keeps barriers to entry low for new services at the application layer.⁶² And history has shown that such openness leads to a tremendous diversity of products and services at the Internet’s edge: telephony, video conferencing, instant messaging, blogs, e-commerce, streaming music and video, search, social media, and the rapidly growing Internet of Things.⁶³

End-to-end may soon be at an end, however, because the FCC repealed the 2015 Open Internet Order in late 2017.⁶⁴ It remains to be seen whether the repeal will take effect. More than a dozen parties—including state attorneys general, public interest groups, and smaller edge providers like Mozilla and Etsy—have

⁵⁷ 2015 Open Internet Order, *supra* note 20, ¶¶ 14–18.

⁵⁸ *See id.* ¶¶ 32, 69.

⁵⁹ *Id.* ¶ 5.

⁶⁰ *See id.* ¶ 76 (stating the FCC’s commitment to the principle “that the Internet’s openness promotes innovation, investment, competition, free expression, and other national broadband goals”).

⁶¹ Isenberg, *supra* note 38, at 29.

⁶² *See* VAN SCHEWICK, *supra* note 33, at 140–41 (explaining that in an end-to-end network developing new applications does not require any change to the core of the network, which results in low costs of innovation and no costs of system adaptation).

⁶³ *See* 2015 Open Internet Order, *supra* note 20, ¶ 77 (“The record . . . supports the proposition that the Internet’s openness continues to enable a virtuous cycle of innovation in which new uses of the network—including new content, applications, services, and devices—lead to increased end-user demand for broadband, which drives network improvements, which in turn lead to further innovative network uses.”) (internal quotation omitted).

⁶⁴ *See* Restoring Internet Freedom, 33 FCC Rcd. 311, ¶ 2 (2018).

sued to block the repeal.⁶⁵ In the meantime, Congress could moot the issue by amending the Telecommunications Act to codify the FCC's existing net neutrality rules in some form.⁶⁶ Network operators led by AT&T are backing a legislative amendment that would incorporate watered-down rules permitting, for example, paid prioritization.⁶⁷ With federal policy in a state of disarray, some state legislatures have intervened to require net neutrality from providers operating within their own borders.⁶⁸ A probable fight over federal preemption overshadows those efforts.⁶⁹

We don't know what the future of Internet regulation holds, but we do know what the past has enabled. We owe the rich and diverse edge economy we have today to both the end-to-end principle and the FCC's history of defending net neutrality in the face of relentless resistance from the telecommunications industry. Advances in network hardware and software enable today's ISPs to manage the data they carry in very granular ways.⁷⁰ To the extent that the core of the Internet has remained relatively underspecified, that constraint has for many

⁶⁵ Jon Brodtkin, *FCC Must Defend Net Neutrality Repeal in Court Against Dozens of Litigants*, ARS TECHNICA (Mar. 12, 2018, 12:15 PM), <https://arstechnica.com/tech-policy/2018/03/fcc-must-defend-net-neutrality-repeal-in-court-against-dozens-of-litigants/> [<https://perma.cc/B33J-GU8Q>].

⁶⁶ See Harper Neidig, *With Rules Repealed, What's Next for Net Neutrality?*, THE HILL (Feb. 23, 2018, 6:00 AM), <http://thehill.com/policy/technology/375185-with-rules-repealed-whats-next-for-net-neutrality> [<https://perma.cc/YG9D-P9SQ>] (reporting that some Republican lawmakers are backing legislation that would codify some net neutrality principles).

⁶⁷ See Jon Brodtkin, *AT&T CEO's Net Neutrality Plan Calls for Regulation of Websites*, ARS TECHNICA (Jan. 24, 2018, 5:35 PM), <https://arstechnica.com/tech-policy/2018/01/attceos-net-neutrality-plan-calls-for-regulation-of-websites/> [<https://perma.cc/L62B-6Z79>]; see also T.C. Sottek, *Comcast, Verizon, and AT&T Want Congress to Make a Net Neutrality Law Because They Will Write It*, THE VERGE (July 12, 2017, 1:17 PM), <https://www.theverge.com/2017/7/12/15959932/comcast-verizon-att-net-neutrality-day-of-action> [<https://perma.cc/DTL3-RNWC>].

⁶⁸ See, e.g., Jon Brodtkin, *Why the First State with a Net Neutrality Law Isn't Scared of Lawsuits*, ARS TECHNICA (Mar. 16, 2018, 10:45 AM), <https://arstechnica.com/tech-policy/2018/03/net-neutrality-supporting-lawmaker-tells-isps-were-ready-for-lawsuits/> [<https://perma.cc/M933-TL9J>] (reporting on net neutrality legislation adopted in Washington state); see also Neidig, *supra* note 66 (reporting that twenty-six state legislatures are considering state-level net neutrality legislation).

⁶⁹ See Brodtkin, *supra* note 68 (discussing the preemption issue).

⁷⁰ See generally Ralf Bendrath & Milton Mueller, *The End of the Net as We Know It? Deep Packet Inspection and Internet Governance*, 13 NEW MEDIA & SOC. 1142 (2011) (exploring the possible consequences of network operators' deployment of "intelligent" routers capable of deep packet inspection, which facilitates comprehensive data surveillance and discrimination as data packets transit the network); see also 2015 Open Internet Order, *supra* note 20, ¶ 85 ("Techniques used by broadband providers to identify and select traffic may include approaches based on packet payloads (using deep packet inspection), network or transport layer headers (e.g., port numbers or priority markings), or heuristics (e.g., the size, sequencing, and/or timing of packets).").

years been regulatory rather than technical. With the repeal of the 2015 Open Internet Order, diversity at the edge stands to suffer as broadband providers seek to maximize profits by imposing discriminatory pricing on edge providers in exchange for faster speeds or other preferential treatment.⁷¹

Economists are divided about the likely effects of paid prioritization on edge innovation under current market conditions.⁷² If the FCC's repeal of the Open Internet Order survives judicial scrutiny, the debate will no longer be purely academic. Whatever the future brings, regulation of the Internet's network layer has historically required content-agnosticism on the part of network operators in service of the end-to-end principle. The dominance of a few platforms in search, ecommerce, and social media should not obscure the fact that the Internet's edge remains extraordinarily rich and diverse. If existing net neutrality rules are weakened or eliminated, the titans at the edge will survive by dint of their size and wealth. The question is what will happen to the rest, including potential future disrupters of today's dominant players.

C. Regulation at the Application Layer: Safe Harbors and Content-Awareness

Unlike regulation at the network layer, regulation at the application layer has encouraged content-awareness. This regulatory bifurcation is consistent with the end-to-end principle, which assigns control over the contents of communications to users and application providers at the edge of the network. Disparate regulatory treatment of infrastructure and application providers also makes sense from the perspective of preventing, proving, and redressing speech-related harms. This is so because the application layer is the layer at which bits carried across the network surface as content—as intelligible words and images that cause legally actionable injuries, including defamation, harassment, infliction of emotional distress, invasion of privacy, and intellectual property infringement.⁷³ The

⁷¹ See *id.* ¶ 82 (discussing incentives for network operators to engage in paid prioritization and the likely negative effects of paid prioritization on net neutrality's "virtuous cycle" of innovation); Benjamin Warlick, *Net Neutrality Repeal May Put Tech Startups In The Slow Lane*, LAW360 (Feb. 22, 2018, 5:04 PM), <https://www.law360.com/articles/1015325/net-neutrality-repeal-may-put-tech-startups-in-the-slow-lane> ("[T]here is real concern that ISPs will set aside fast lanes for tech giants, leaving behind a second tier of content providers including tech startups, universities and governments.").

⁷² Compare, e.g., Richard T. B. Ma, et al., *Paid Prioritization and Its Impact on Net Neutrality*, 35 IEEE J. SELECTED AREAS IN COMM'NS 367, 367 (2017) ("From a welfare perspective, our results suggest that paid prioritization could be superior to the imposition of net neutrality regulations.") with Hong Guo & Robert F. Easley, *Network Neutrality Versus Paid Prioritization: Analyzing the Impact on Content Innovation*, 25 PRODUCTION & OPERATIONS MGMT. 1261 (2016) (arguing that net neutrality creates—and paid prioritization eliminates—a "pro bono innovation zone" that allows new content providers to enter the market without contributing to network provider profits).

⁷³ John Blevins argues that it is a category mistake—"layer confusion"—to treat network-level data transmissions as speech for First Amendment purposes. See Blevins, *supra* note 32,

application layer is the Internet's human-experiential layer. Users interact with each other, for good and ill, at the network's edge.

The two most important laws governing how edge service providers handle user-generated content are the Communications Decency Act (CDA) of 1996⁷⁴ and the Digital Millennium Copyright Act (DMCA) of 1998.⁷⁵ Both statutes were enacted in the early days of the commercial Internet. Together, they provide broad legal protection for edge service providers whose business models entail hosting and displaying large amounts of user-generated content.⁷⁶ Without such protection, Congress believed, innovative online services could not launch and scale, because their founders and investors could not afford to assume the risk of unlimited liability for their users' illegal speech.⁷⁷ Section 230 of the CDA shields covered service providers from being treated as speakers, publishers, or distributors of any illegal user-generated content they host, with the exception of content that infringes intellectual property rights.⁷⁸ Section 512 of the DMCA fills the gap that section 230 left with respect to claims involving intellectual property infringement by shielding covered service providers from liability for their users' online copyright infringements.⁷⁹

at 386 (“Network-layer transmission not only does not convey a message, it was self-consciously designed to ignore any such messages.”).

⁷⁴ 47 U.S.C. § 230 (2012).

⁷⁵ 17 U.S.C. § 512 (2012).

⁷⁶ See *BWP Media USA, Inc. v. Clarity Dig. Grp., LLC*, 820 F.3d 1175, 1177–78 (10th Cir. 2016) (“Section 512 of the DMCA contains a safe harbor provision protecting online and internet service providers (‘ISPs’) from monetary liability, only allowing for limited injunctive relief, when copyright infringement occurs through use of the service.”); *Zeran v. Am. Online, Inc.*, 129 F.3d 327, 330 (4th Cir. 1997) (“By its plain language, § 230 creates a federal immunity to any cause of action that would make service providers liable for information originating with a third-party user of the service.”).

⁷⁷ See *BWP Media*, 820 F.3d at 1178 (“The safe harbor provision is designed to “[provide] greater certainty to service providers concerning their legal exposure for infringements that may occur in the course of their activities.”) (quoting H.R. Rep. No. 105–551, pt. 2, at 49–50 (1998)); *Zeran*, 129 F.3d at 331 (“Congress’ purpose in providing the § 230 immunity was thus evident. Interactive computer services have millions of users. The amount of information communicated via interactive computer services is therefore staggering. The specter of tort liability in an area of such prolific speech would have an obvious chilling effect. . . . Congress considered the weight of the speech interests implicated and chose to immunize service providers to avoid any such restrictive effect.”).

⁷⁸ See § 230(c)(1) (“No 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.”); § 230(e)(2) (“Nothing in this section shall be construed to limit or expand any law pertaining to intellectual property.”).

⁷⁹ See § 512(a)–(d) (creating safe harbors from monetary damages for copyright infringement for providers that route, cache, or store users’ material and for providers that link to third-party material online). § 512 covers network operators performing routing functions only when they act in a content-agnostic way. See § 512(k) (“As used in subsection (a), the term ‘service provider’ means an entity offering the transmission, routing, or providing of

In addition to limiting providers' liability for their users' illegal content, both the CDA and the DMCA contain provisions intended to encourage providers to remove illegal and offensive content from their services.⁸⁰ This is a well-known fact about the DMCA, because the statute conditions safe harbors for service providers on notice-based removal of allegedly infringing content.⁸¹ Providers under the DMCA are not required to affirmatively monitor their services for infringing content, but Congress in drafting the statute anticipated that right holders and service providers would cooperatively develop "standard technical measures" for protecting content online.⁸² No such statutory measures were ever developed, but large platforms like Facebook, YouTube, and Vimeo now voluntarily screen user uploads against databases of reference files provided by copyright holders.⁸³ YouTube's Content ID system is the most widely known of these filtering systems.⁸⁴ Where uploads match reference files, Content ID automatically applies a business rule selected by the copyright holder, and the matched file is accordingly blocked, monitored, or monetized.⁸⁵ To help address concerns about over-blocking, both the DMCA and Content ID incorporate an appeal process for users who believe their content was wrongfully claimed to be infringing.⁸⁶

It is a less well-known fact that the CDA was intended to encourage covered providers to be content-aware in the operation of their services.⁸⁷ The CDA as a

connections for digital online communications, between or among points specified by a user, of material of the user's choosing, without modification to the content of the material as sent or received.").

⁸⁰ § 230 provides immunity for service providers "on account of any action voluntarily taken in good faith to restrict access to or availability of material that the provider . . . considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected." § 230(c)(2). § 512 of the DMCA requires service providers seeking safe harbor to "expeditiously . . . remove or disable access" to claimed copyright infringing material either upon notice from a right holder or when the provider otherwise becomes aware of it. § 512(c).

⁸¹ *See id.* (setting forth a "notice and takedown" protocol for right holders and service providers to follow).

⁸² *See* § 512(m) (providing that safe harbor cannot be conditioned on a service's affirmatively monitoring its system for infringing activity); § 512(i) (providing that services shall not interfere with standard technical measures employed by right holders to protect their content).

⁸³ *See generally* Annemarie Bridy, *Copyright's Digital Deputies: DMCA-plus Enforcement by Internet Intermediaries*, in *RESEARCH HANDBOOK ON ELECTRONIC COMMERCE LAW* (John Rothchild ed., 2017), at 195–198 (discussing the use of filtering technology for copyright enforcement on platforms).

⁸⁴ *See id.* (explaining the mechanics of Content ID).

⁸⁵ *Id.* at 196.

⁸⁶ *Id.* at 197.

⁸⁷ For example, a recent article in *WIRED* magazine, which is ordinarily accurate in its reporting on technology law and policy, incorrectly asserted that Facebook could lose its §

whole was addressed to emerging problems related to online information quality and parents' ability to limit their children's exposure to pornography and other adult content.⁸⁸ The original intent of the statute was to limit—by creating criminal liability for—the online distribution and display of sexually explicit content deemed “harmful to minors.”⁸⁹ Before the statute's liability provisions went into effect, however, they were struck down in court on First Amendment grounds.⁹⁰ When that happened, section 230, the CDA's exemption for service providers, became unmoored to some extent from its original context.⁹¹ Almost ab initio, section 230 was an exception without a rule. Over time, it came to be viewed predominantly as a means of avoiding liability for providers that *decline* to remove any content uploaded by users.⁹² In keeping with the Internet's historically strong free speech ethos, services hosting user-generated content have mostly

230 immunity if it were to edit content on its platform. See Nicholas Thompson & Fred Vogelstein, *Inside the Two Years That Shook Facebook—and the World*, WIRED (Feb. 12, 2018), <https://www.wired.com/story/inside-facebook-mark-zuckerberg-2-years-of-hell/> [<https://perma.cc/W6YP-BFVQ>]. In fact, editing does *not* abrogate immunity, however, unless the act of editing itself causes content to become illegal. See Fair Hous. Council of San Fernando Valley v. Roommates.Com, LLC, 521 F.3d 1157, 1169 (9th Cir. 2008) (“A website operator who edits user-created content—such as by correcting spelling, removing obscenity or trimming for length—retains his immunity for any illegality in the user-created content, provided that the edits are unrelated to the illegality.”).

⁸⁸ See Mark A. Lemley, *Rationalizing Internet Safe Harbors*, 6 J. TELECOMM. & HIGH TECH. L. 101, 101-02 (2007). Mark Lemley points out that § 230 arose almost by accident given that the real purpose of the CDA was to keep the Internet porn-free and, thereby, safe for kids. See *id.* at 103. § 230 was added as an afterthought by members of Congress who worried that providers could incur publisher liability for defamation if they stopped acting as neutral conduits for third-party speech and made good faith efforts to remove objectionable content. *Id.*

⁸⁹ See *Reno v. Am. Civil Liberties Union*, 521 U.S. 844 (1997) (deciding the constitutionality of the CDA's provisions enacted to protect minors from “indecent” and “patently offensive” communications on the Internet).

⁹⁰ See *id.* at 849 (affirming the district court's holding that the challenged provisions of the CDA violated the First Amendment).

⁹¹ See Danielle K. Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity*, 86 FORDHAM L. REV. 401, 403 (2017) (“The CDA was part of a broad campaign—rather ironically in retrospect—to restrict access to sexually explicit material online. Lawmakers thought they were devising a limited safe harbor from liability for online providers engaged in self-regulation.”).

⁹² *Id.* at 406 (describing section 230 as “a mighty fortress protecting platforms from accountability for unlawful activity on their systems. . .”). For a thorough empirical analysis of the section 230 case law, see David S. Ardia, *Free Speech Savior or Shield for Scoundrels: An Empirical Study of Intermediary Immunity under Section 230 of the Communications Decency Act*, 43 LOY. L.A. L. REV. 373 (2010).

relied on section 230 to sustain very speech-protective policies.⁹³ As those services scaled up, their libertarian approach to user speech had the significant business advantage of being inexpensive and easy to administer.

The part of section 230 that immunizes providers for regulating their users' online speech appears in section 230(c)(2) under the heading "Protection for 'Good Samaritan' blocking and screening of offensive material."⁹⁴ The statute identifies several categories of content to which a provider might in good faith restrict access, including speech that is "obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected."⁹⁵ This provision explicitly relieves service providers of any obligation to adopt content moderation policies that are coextensive with the protective reach of the First Amendment.⁹⁶ While the CDA's Good Samaritan provision is not quite the antithesis of a must-carry rule, it is a rather broad license to engage in the kind of content-based discrimination that is prohibited of common carriers. To borrow a turn of phrase from Rebecca Tushnet, section 230 "allows Internet intermediaries to have their free speech and everyone else's too."⁹⁷

While it is true that providers invoking the Good Samaritan provision have wide latitude to choose what content to block or remove, their discretion to decide for themselves what counts as "otherwise objectionable" speech within the meaning of section 230 is not absolute.⁹⁸ For example, in *National Numismatic Certification v. eBay* the court employed the interpretive canon of *ejusdem generis* to deny eBay immunity from claims arising from its decision to delist an auction because it believed the listed coins were counterfeit.⁹⁹ Similarly, in *Song*

⁹³ Cf. Matthew Ingram, *For Twitter Free Speech Matters—Not Real Names*, GIGAOM (Oct. 18, 2011) <https://gigaom.com/2011/10/18/for-twitter-free-speech-is-what-matters-not-real-names/> [<https://perma.cc/V49G-9XNT>] (quoting a Twitter executive who called Twitter "the free speech wing of the free speech party"); Thompson & Vogelstein, *supra* note 87 ("This notion that Facebook is an open, neutral platform is almost like a religious tenet inside the company.").

⁹⁴ 47 U.S.C. § 230(c) (2012).

⁹⁵ 47 U.S.C. § 230(c)(2).

⁹⁶ *Id.*

⁹⁷ Rebecca Tushnet, *Power Without Responsibility: Intermediaries and the First Amendment*, 76 GEO. WASH. L. REV. 986, 1002 (2008).

⁹⁸ See, e.g., *Song fi Inc. v. Google, Inc.*, 108 F. Supp. 3d 876, 884 (N.D. Cal. 2015) (rejecting a "completely subjective (and entirely unbounded) reading" of section 230(c)(2)); *Sherman v. Yahoo! Inc.*, 997 F. Supp. 2d 1129, 1138 (S.D. Cal. 2014) (declining "to broadly interpret 'otherwise objectionable' material to include any or all information or content.").

⁹⁹ *Nat'l Numismatic Certification, LLC. v. eBay, Inc.*, No. 6:08-CV-42-ORL-19GJK, 2008 WL 2704404, at *25 (M.D. Fla. July 8, 2008) ("It is difficult to accept, as eBay argues, that Congress intended the general term 'objectionable' to encompass an auction of potentially-counterfeit coins when the word is preceded by seven other words that describe pornography, graphic violence, obscenity, and harassment.").

fi v. Google the court held that Congress did not intend the Good Samaritan provision to give providers complete subjective discretion over the types of content they block or remove.¹⁰⁰ Citing legislative history, the dictionary meaning of “objectionable” (i.e., “harmful, undesirable”), and the canon of *ejusdem generis*, the court rejected Google’s argument that it should be entitled to immunity from claims arising from its decision to block public access to a YouTube video that Google believed had an artificially inflated view count.¹⁰¹ To paraphrase the holding of another court, section 230’s catch-all for “otherwise objectionable” content does not give providers a blank check to block whatever they want to.¹⁰²

The requirement of good faith in section 230(c)(2) serves as an additional limit on the scope of providers’ Good Samaritan discretion. Courts have found in that requirement a reason to consider both the legitimacy of a provider’s motivation for blocking and its willingness to provide an explanation to affected users. In *Zango v. Kaspersky Lab*, Judge Fisher of the Ninth Circuit said in a concurrence that blocking third-party content for an anticompetitive purpose or on “malicious whim” could constitute bad faith.¹⁰³ In *Smith v. Trusted Universal Standards in Electronic Transactions*, the court held that failing to offer an explanation for a content-blocking decision in response to a user’s request could constitute bad faith.¹⁰⁴

The CDA’s Good Samaritan provision was created for the purpose of legislatively overruling *Stratton Oakmont, Inc. v. Prodigy Servs. Co.*, one of the first cases to consider whether online services could be liable for the illegal speech of their users.¹⁰⁵ Prodigy, the provider of an actively moderated electronic bulletin board service (BBS), automatically screened user content for keywords it

¹⁰⁰ *Song fi*, 108 F. Supp. 3d at 884.

¹⁰¹ *Id.* (“[T]he ordinary meaning of ‘otherwise objectionable,’ as well as the context, history, and purpose of the Communications Decency Act all counsel against reading ‘otherwise objectionable’ to mean anything to which a content provider objects regardless of why it is objectionable.”).

¹⁰² *Sherman*, 997 F. Supp. 2d at 1138. It is important to keep in mind, however, that a provider’s contractual terms of service may give it more latitude to block user content than Section 230, as interpreted by courts, might accommodate. A provider that falls outside the scope of Section 230’s immunity for Good Samaritan blocking of “otherwise objectionable” content may still be able to fall back on its terms of service to avoid liability for a contested content removal.

¹⁰³ *See Zango, Inc. v. Kaspersky Lab, Inc.*, 568 F.3d 1169, 1178 (9th Cir. 2009) (Fisher, J., concurring) (expressing concern that “a blocking software provider might abuse that immunity to block content for anticompetitive purposes or merely at its malicious whim”).

¹⁰⁴ *Smith v. Trusted Universal Standards in Elec. Transactions, Inc.*, No. CIV09-4567RBKMW, 2010 WL 1799456, at *7 (D.N.J. May 4, 2010) (“One would expect that if an interactive computer service had acted in good faith, it could and would come forward with the legitimate basis for its actions when questioned. . . .”).

¹⁰⁵ *See S. Rep. No. 104–230*, at 194 (1996) (“One of the specific purposes of [§ 230] is to overrule *Stratton Oakmont v. Prodigy* and any other similar decisions. . . .”); *Zeran v. Am.*

wanted to block, and its board moderators deleted content for a range of reasons, including offensiveness, bad taste, bad advice, irrelevancy, and solicitation.¹⁰⁶ After deleting a user's post, the moderator would send a standard notice to the affected user stating the reason for the deletion.¹⁰⁷

The court in *Stratton Oakmont* held that Prodigy's active moderation of its bulletin boards involved the exercise of editorial discretion, which made it strictly liable as a publisher for any illegal content it failed to remove.¹⁰⁸ That liability arose by way of an analogy between Prodigy's BBS service and traditional print publications like magazines and newspapers, whose publishers are treated for legal purposes as the speakers of whatever words they elect to print.¹⁰⁹ The court pointed out that Prodigy marketed itself as a service that was discriminating about the content it chose to display: it "held itself out to the public and its members as controlling the content of its computer bulletin boards," and it "implemented this control through its automatic software screening program, and the Guidelines which Board Leaders are required to enforce."¹¹⁰ Applying intermediary liability principles from the world of brick-and-mortar publishing, the court said that Prodigy could have escaped publisher liability if it had acted solely as a neutral conduit for the distribution of its users' speech.¹¹¹

Stratton Oakmont's holding sent a clear message to online service providers like Prodigy that were trying to enforce community content guidelines: Don't edit or screen any user content at all unless you're prepared to risk liability for all of it.¹¹² Congress recognized the decision's probable consequences and included the Good Samaritan provision in section 230 in order to "remove disincentives" for edge providers like Prodigy to develop and deploy tools for moderating content on their services.¹¹³ Given the economic and expressive costs of

Online, Inc., 129 F.3d 327, 331 (4th Cir. 1997) ("Congress enacted § 230 to remove the disincentives to self-regulation created by the *Stratton Oakmont* decision.").

¹⁰⁶ *Stratton Oakmont, Inc. v. Prodigy Servs. Co.*, No. 31063/94, 1995 WL 323710, at *2–3 (N.Y. Sup. Ct. May 24, 1995).

¹⁰⁷ *Id.* at *3.

¹⁰⁸ *Id.* at *4.

¹⁰⁹ *Id.* at *3.

¹¹⁰ *Id.* at *4.

¹¹¹ *Id.* at *4–5.

¹¹² *See Zeran v. Am. Online, Inc.*, 129 F.3d 327, 333 (4th Cir. 1997) ("Any efforts by a service provider to investigate and screen material posted on its service would only lead to notice of potentially defamatory material more frequently and thereby create a stronger basis for liability.").

¹¹³ *See* 47 U.S.C. § 230(b)(4) (2012) ("It is the policy of the United States . . . to remove disincentives for the development and utilization of blocking and filtering technologies that empower parents to restrict their children's access to objectionable or inappropriate online material."); *Zeran*, 129 F.3d at 331 ("Another important purpose of § 230 was to encourage service providers to self-regulate the dissemination of offensive material over their services.").

active moderation, however, few sites hosting massive quantities of user-generated content had any appetite for the kind of monitoring that Prodigy was doing. Section 230 permitted but didn't require it, and the market didn't seem to demand it, so edge providers mostly didn't do it.¹¹⁴ Unlike the DMCA's safe harbor, which is conditioned on notice-based removal of content identified as illegal, the CDA's immunity is not conditioned on the removal of any content.¹¹⁵ It gives service providers a relatively free hand to manage (or not manage) their users' speech.

Over the years, courts have interpreted section 230 to provide very strong protection for edge providers that decline to remove allegedly illegal or offensive content.¹¹⁶ In close cases, courts err on the side of immunity.¹¹⁷ Some critics of this doctrinal development argue that courts deciding section 230 cases lost sight of the broader policy aims of the CDA and, in doing so, acquiesced in the development of a toxic online culture that Congress tried to prevent by including the Good Samaritan provision in section 230.¹¹⁸ They point to cases in which section 230 was held to shield dodgy service providers that not only tolerate but actively solicit users' abusive or defamatory speech.¹¹⁹ Some judges, too, have

¹¹⁴ The notable exceptions to this trend have always been obscenity and child pornography. The CDA does not give service providers immunity from prosecution for federal crimes involving obscenity or child abuse. *See* § 230(e)(1) ("Nothing in this section shall be construed to impair the enforcement of section 223 or 231 of this title, chapter 71 (relating to obscenity) or 110 (relating to sexual exploitation of children) of title 18, or any other Federal criminal statute.").

¹¹⁵ *Compare* *Viacom Int'l, Inc. v. YouTube, Inc.*, 676 F.3d 19, 27 (2d Cir. 2012) ("Section 512(c) also sets forth a detailed notification scheme that requires service providers to designate[] an agent to receive notifications of claimed infringement, and specifies the components of a proper notification, commonly known as a 'takedown notice,' to that agent.") (internal quotations omitted), *with* *Zeran*, 129 F.3d at 333 ("Because the probable effects of distributor liability on the vigor of Internet speech and on service provider self-regulation are directly contrary to § 230's statutory purposes, we will not assume that Congress intended to leave liability upon notice intact.").

¹¹⁶ *See generally* *Ardia*, *supra* note 92 (surveying and analyzing over a decade's worth of § 230 case law).

¹¹⁷ *See* *Fair Hous. Council of San Fernando Valley v. Roommates.Com, LLC*, 521 F.3d 1157, 1174 (9th Cir. 2008) ("Such close cases, we believe, must be resolved in favor of immunity, lest we cut the heart out of section 230 by forcing websites to face death by ten thousand duck-bites, fighting off claims that they promoted or encouraged—or at least tacitly assented to—the illegality of third parties.").

¹¹⁸ *See* *Citron & Wittes*, *supra* note 91, at 403.

¹¹⁹ *See, e.g.*, *Jones v. Dirty World Entm't Recordings LLC*, 755 F.3d 398, 402 (6th Cir. 2014) (stating that users of the site, who colloquially refer to themselves as "The Dirty Army," may submit "dirt"—i.e., content that may include text, photographs, or video about any subject"); *Blumenthal v. Drudge*, 992 F. Supp. 44, 51 (D.D.C. 1998) ("AOL issued a press release making clear the kind of material Drudge would provide to AOL subscribers—gossip and rumor—and urged potential subscribers to sign onto AOL in order to get the benefit of the Drudge Report.").

expressed discomfort with the broad application of section 230 in cases involving defendants that clearly promote the spread of rumors and scandal.¹²⁰

Beyond the narrow domain of DMCA notice-and-takedown, which is highly choreographed by statute, social media platforms have had little incentive to develop correspondingly structured takedown and appeal procedures for other kinds of disputed content. Although the Good Samaritan provision of section 230 was intended to encourage responsible, systematic content moderation, edge providers' takedown practices have tended to be reactive and haphazard, often to the detriment of marginalized users and communities of users.¹²¹ Looking at the current state of online discourse, one would be hard pressed to argue that the Good Samaritan provision has done the work that Congress thought it would do. Responsible, systematic moderation on social media platforms is arguably needed now more than ever to limit the spread of disinformation and to curb such forms of abuse as revenge porn, threats of violence, targeted harassment, and doxing. Such moderation would largely be foreclosed by a must-carry rule at the application layer, because a wide swath of abusive and harmful speech is lawful by First Amendment standards. Accomplishing responsible moderation at scale without running roughshod over expressive freedoms is the defining challenge of this stage in the Internet's evolution.

II. EVOLVING SOCIAL NORMS FOR SPEECH ON SOCIAL MEDIA

Adopting a must-carry rule for social media would also be out of step with public opinion. In growing numbers, members of the public believe that social media platforms are not doing enough to address online harassment.¹²² Eight in ten respondents to a Pew Research Center survey said that online platforms should be responsible for policing and preventing abuse online.¹²³ When asked about balancing the ability of individuals to speak freely online and the creation

¹²⁰ *Id.* (“If it were writing on a clean slate, this Court would agree with plaintiffs. AOL . . . has affirmatively promoted Drudge as a new source of unverified instant gossip on AOL. Yet it takes no responsibility for any damage he may cause.”).

¹²¹ See, e.g., Tushnet, *supra* note 97, at 996–1002 (describing a “purge” in 2007 of supposed pedophiles on LiveJournal that resulted in account suspensions for readers of Nabokov’s *Lo-lita*, incest survivors, and writers of Harry Potter fan fiction); Jessica Anderson, et al., *Un-friending Censorship*, ONLINECENSORSHIP.ORG (Mar. 31, 2016), https://s3-us-west-1.amazonaws.com/onlinecensorship/posts/pdfs/000/000/044/original/Onlinecensorship.org_Report_-_31_March_2016.pdf?1459436925 [<https://perma.cc/68AA-X5MH>] (reporting the results of a 2015 survey of 161 users of six social media platforms—Facebook, Flickr, Google+, Instagram, Twitter, and YouTube—about their experiences with content-related account suspensions and content removals).

¹²² Maeve Duggan, *Online Harassment 2017*, PEW RES. CTR., at 4 (July 11, 2017), http://www.pewinternet.org/2017/07/11/the-broader-context-of-online-harassment/?utm_content=buffer4173e&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer [<https://perma.cc/Q4C9-8EWF>].

¹²³ *Id.* at 5.

of a welcoming environment for others, over half of respondents prioritized a welcoming environment.¹²⁴ In addition to being concerned about harassment, respondents in large numbers worried about the damaging effects of online exposure to false or inaccurate information.¹²⁵ This research suggests that we are experiencing a measurable shift in social norms away from the libertarian speech values that have historically permeated the Internet's edge and an emerging public preference for more active content moderation on social media platforms. In 2018, a "net neutrality" rule for social media feels like retrograde motion.

A. *Shifting Political Winds*

Political opponents of Big Tech are seeking to capitalize on the public's growing anxiety about the state of the online information economy. For the first time since the CDA's enactment, those who advocate narrowing the scope of section 230 immunity have found success on Capitol Hill. The expedient vehicle for their broader ambition was a controversial piece of legislation called the Fight Online Sex Trafficking Act (FOSTA), which became law in 2018.¹²⁶ FOSTA's enactment followed a years-long effort by state attorneys general and civil claimants to hold online classified services like Backpage.com and Craigslist liable for criminal activity associated with advertising for "erotic services."¹²⁷ Over the years, section 230 has consistently shielded these providers and others like them in litigation.¹²⁸ FOSTA removes from the protection of section 230

¹²⁴ *Id.* at 6.

¹²⁵ *Id.* at 56-58.

¹²⁶ Colin Lecher, *Senate Passes Controversial Anti-Sex Trafficking Bill*, THE VERGE (Mar. 21, 2018, 4:23 PM), <https://www.theverge.com/2018/3/21/17147688/senate-sesta-fosta-vote-anti-sex-trafficking> [<https://perma.cc/TTG5-7T7H>].

¹²⁷ *See, e.g.*, Jonathan J. Cooper, *California Pursues New Pimping Charges Against Backpage.com*, BUSINESS INSIDER, (Dec. 23, 2016, 5:59 PM), <http://www.businessinsider.com/ap-california-pursues-new-pimping-charges-against-backpagecom-2016-12> [<https://perma.cc/6KML-EKCY>] (reporting on efforts by California Attorney General Kamala Harris to prosecute Backpage.com executives notwithstanding § 230); Brad Stone, *Craigslist Sues South Carolina Attorney General*, N.Y. TIMES (May 20, 2009, 10:21 AM), <https://bits.blogs.nytimes.com/2009/05/20/pushed-against-wall-craigslist-sues-south-carolina-attorney-general/> [<https://perma.cc/4D2V-SXQE>] (reporting on Craigslist's filing for declaratory judgment against South Carolina Attorney General Henry McMaster, who publicly threatened to investigate and criminally prosecute the site's operator).

¹²⁸ *See, e.g.*, *Jane Doe No. 1 v. Backpage.com, LLC*, 817 F.3d 12, 24 (1st Cir. 2016), *cert. denied*, 137 S. Ct. 622 (2017) (affirming dismissal of the appellants' sex trafficking claims); *Craigslist, Inc. v. McMaster*, No. 2:09-cv-01308-CWH (D.S.C., Aug. 5, 2010) (entering declaratory judgment for Craigslist against South Carolina's attorney general); *Dart v. Craigslist, Inc.*, 665 F. Supp. 2d 961, 970 (N.D. Ill. 2009) (entering judgment on the pleadings for Craigslist).

websites that “promote or facilitate” prostitution.¹²⁹ Putting aside the merits of the legislation, which critics legitimately believe will produce a range of unintended consequences, including *less* content moderation by edge providers,¹³⁰ FOSTA’s passage is an important sign of shifting political winds for service providers that have long relied on section 230’s broad immunity without doing all that much in the way of Good Samaritan blocking.¹³¹ Social media platforms like YouTube, Facebook, and Twitter recognize that they are operating in a dramatically changed political and cultural environment from the one they operated in five years ago. Their industry trade group, the Internet Association, initially opposed FOSTA in any form but later changed its position, possibly fearing alternative legislation that would have further narrowed the scope of section 230.¹³²

The pressure platforms now face to do more in the way of content moderation is even more intense abroad.¹³³ In Europe, where free speech protections are markedly weaker than they are in the United States, there has never been an equivalent to section 230’s broad immunity. Service provider safe harbors in the E-Commerce Directive have long been conditioned on the removal of all types of illegal content—not just copyright-infringing content.¹³⁴ In 2016, at the urging of the European Commission, Facebook, YouTube, Twitter, and Microsoft

¹²⁹ Allow States and Victims to Fight Online Sex Trafficking Act of 2017, H.R. 1865, 115th Cong. (2017); Stop Enabling Sex Traffickers Act of 2017 (SESTA), S. 1693, 115th Cong. (2018) (FOSTA’s companion bill in the Senate).

¹³⁰ For a discussion of problems with FOSTA/SESTA and likely unintended consequences, see Daphne Keller, *SESTA and the Teachings of Intermediary Liability*, THE CTR. FOR INTERNET & SOC’Y (Nov. 2, 2017), <http://cyberlaw.stanford.edu/files/publication/files/SESTA-and-IL-Keller-11-2.pdf> [<https://perma.cc/2RP9-AXD9>].

¹³¹ See Cecilia Kang, *House Passes Online Sex Trafficking Bill After Big Tech Companies Back Off*, N.Y. TIMES (Feb. 27, 2018), <https://www.nytimes.com/2018/02/27/business/online-sex-trafficking-bill.html> (“The bill’s passage in the House signaled an important turning point for an industry that has largely operated free of regulations. Pressure has been mounting for social media companies and other internet giants to be better stewards of their powerful platforms.”).

¹³² See Tom Jackman, *Internet Companies Drop Opposition to Bill Targeting Online Sex Trafficking*, WASH. POST (Nov. 7, 2017), https://www.washingtonpost.com/news/true-crime/wp/2017/11/07/internet-companies-drop-opposition-to-bill-targeting-online-sex-trafficking/?utm_term=.65afa579bca6.

¹³³ See, e.g., Natasha Lomas, *Europe Keeps up the Pressure on Social Media over Illegal Content Takedowns*, TECHCRUNCH (Jan. 9, 2018), <https://techcrunch.com/2018/01/09/europe-keeps-up-the-pressure-on-social-media-over-illegal-content-takedowns/> [<https://perma.cc/6NRL-SZAB>] (“Last fall the Commission said it would monitor tech giants’ progress vis-a-vis content takedowns over the next six months to decide whether to take additional measures—such a drafting legislation.”).

¹³⁴ The EU’s intermediary safe harbors can be found in Articles 12–15 of the E-Commerce Directive. See Directive 2000/31/EC of the European Parliament and of the Council of 8 June

agreed to a voluntary code of conduct requiring them to review and act on complaints about hate speech on their platforms within twenty-four hours.¹³⁵ The twenty-four hour target tightened the E-Commerce Directive's more flexible standard of "expeditious" action.¹³⁶ Most recently, EU regulators have demanded that platforms adopt a similar code of conduct, with the same twenty-four hour turnaround, for "terrorist" speech, threatening regulation if voluntary compliance is not forthcoming.¹³⁷

B. Failures in Social Media's "Marketplace of Ideas"

Another important dimension of the current reckoning for online platforms is a growing recognition that free speech and censorship work differently in the age of social media than they did in the era when public discourse was dominated by traditional media outlets controlling scarce broadcast and print resources.¹³⁸ Social scientist Zeynep Tufekci argues that contemporary censorship manifests not as the restriction of speech but as the manipulation of attention in an environment flooded with speech:

The most effective forms of censorship today involve meddling with trust and attention, not muzzling speech itself. As a result, they don't look much like the old forms of censorship at all. They look like viral or coordinated harassment campaigns, which harness the dynamics of viral outrage to impose an unbearable and disproportionate cost on the act of speaking out. They look like epidemics of disinformation, meant to undercut the credibility of valid information sources. They look like bot-fueled campaigns of

2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce'), art. 12–15, 2000 O.J. (L 178). For services that cache and host user-generated content, safe harbor is conditioned on the expeditious removal of any illegal content of which the provider gains knowledge. *See id.* As in the DMCA, the E-Commerce Directive does not permit member states to condition safe harbor on an affirmative monitoring requirement. *Compare id.*, art. 15 with 17 U.S.C. § 512(m) (2012).

¹³⁵ Alex Hern, *Facebook, YouTube, Twitter and Microsoft Sign EU Hate Speech Code*, THE GUARDIAN (May 31, 2016, 8:16 AM), <https://www.theguardian.com/technology/2016/may/31/facebook-youtube-twitter-microsoft-eu-hate-speech-code> [<https://perma.cc/NXP2-BSE4>].

¹³⁶ *See* Directive on electronic commerce, *supra* note 134, art. 12–14 (setting forth conditions for safe harbor).

¹³⁷ *See* Stephanie Bodoni, *EU Warns Tech Giants to Remove Terror Content in 1 Hour—or Else*, BLOOMBERG (Mar. 1, 2018), <https://www.bloomberg.com/news/articles/2018-03-01/remove-terror-content-in-1-hour-or-else-eu-warns-tech-giants> [<https://perma.cc/UE82-Z5TR>] ("The European Union issued internet giants an ultimatum to remove illegal online terrorist content within an hour, or risk facing new EU-wide laws.")

¹³⁸ *See generally* Jack Balkin, *Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation*, 51 U.C. DAVIS L. REV. 1149 (2018).

trolling and distraction, or piecemeal leaks of hacked materials, meant to swamp the attention of traditional media.¹³⁹

The technical architecture of social media platforms, which is optimized for engagement, creates a speech environment far different from the “marketplace of ideas” that John Stuart Mill envisioned and Oliver Wendell Holmes enshrined in our First Amendment jurisprudence.¹⁴⁰ In that idealized marketplace, everyone has an opportunity to be speak and be heard. Participants encounter competing rational arguments for and against controversial propositions, and the most-well-reasoned arguments ultimately win the day. Truth triumphs over falsehood, and reason defeats emotion. That process of truth-finding through truth-testing bears little resemblance to the algorithmic sorting that creates winners and losers in social media’s attention sweepstakes.¹⁴¹

To promote engagement, social media platforms avoid exposing users to content they might find irrelevant or uninteresting.¹⁴² Each user’s timeline (Twitter) or newsfeed (Facebook) is populated with content based on data about that user’s personal preferences that the platform extracts and aggregates over time—primarily for the benefit of advertisers seeking to reach ever more particularized cohorts of consumers.¹⁴³ Users’ timelines and newsfeeds are engineered to keep them logged in and interacting with content on the platform for as long as possible.¹⁴⁴ This algorithmic personalization contributes to the filter bubble effect that social scientists have linked to increasing social polarization and identity

¹³⁹ See Zeynep Tufekci, *It’s the (Democracy-Poisoning) Golden Age of Free Speech*, WIRED (Jan. 16, 2018, 6:00 AM), <https://www.wired.com/story/free-speech-issue-tech-turmoil-new-censorship/> [<https://perma.cc/HU75-3AVE>]

¹⁴⁰ See *Abrams v. United States*, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting) (“But when men have realized that time has upset many fighting faiths, they may come to believe . . . that the ultimate good desired is better reached by free trade in ideas—that the best test of truth is the power of the thought to get itself accepted in the competition of the market. . . .”).

¹⁴¹ See Will Oremus, *Who Controls Your Facebook Feed*, SLATE (Jan. 3, 2016), http://www.slate.com/articles/technology/cover_story/2016/01/how_facebook_s_news_feed_algorithm_works.html [<https://perma.cc/CRT3-XKYZ>] (explaining how Facebook’s newsfeed algorithm selects and ranks posts for individual users).

¹⁴² See *id.* (explaining that the Facebook algorithm assigns a “relevancy” score to every post it could possibly show a particular user at a particular point in time).

¹⁴³ See *id.* (“Once every possible post in your feed has received its relevancy score, the sorting algorithm can put them in the order that you’ll see them on the screen. The post you see at the top of your feed, then, has been chosen over thousands of others as the one most likely to make you laugh, cry, smile, click, like, share, or comment.”).

¹⁴⁴ See *id.* (“[C]licks, likes, shares, and comments are what make posts go viral, turn individual users into communities, and drive traffic to the advertisers that Facebook relies on for revenue”).

politics.¹⁴⁵ Platforms show users only what they feel confident will hold users' attention.

One of the more disturbing findings to come out of recent studies of social media use is that users find false and inflammatory content more engaging and shareable than true and uncontroversial content.¹⁴⁶ As social media platforms currently operate, they are finely tuned to propagate and amplify extreme and outrageous speech.¹⁴⁷ On Twitter, fake followers and accounts controlled by bots exacerbate this problem, giving extremist viewpoints the appearance of greater public support than they actually have.¹⁴⁸ In addition, social media ad platforms make it cheap and easy to promote extreme viewpoints and junk information.¹⁴⁹ Anyone willing to pay to promote a message can extend his or her reach, regardless of the merits of the message.¹⁵⁰ Ad tech also makes it easy to customize audiences for promoted content based on legally suspect classifications like age, race, and ethnicity.¹⁵¹ During the 2016 presidential campaign,

¹⁴⁵ Cass Sunstein argues that social media's algorithmic personalization undermines our capacity for engaging in a well-functioning system of deliberative democracy. See CASS R. SUNSTEIN, #REPUBLIC: DIVIDED DEMOCRACY IN THE AGE OF SOCIAL MEDIA 4–5 (2017) (advocating an “architecture of serendipity” to mitigate the effect of filter bubbles).

¹⁴⁶ See Soroush Vosoughi et al., *The Spread of True and False News Online*, 359 SCIENCE 1146, 1146–1149 (2018) (finding that “[f]alsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information, and the effects were more pronounced for false political news than for false news about terrorism, natural disasters, science, urban legends, or financial information”).

¹⁴⁷ See M.J. Crockett, *Moral Outrage in the Digital Age*, NATURE HUMAN BEHAVIOR (Sept. 18, 2017) (finding that “outrage-inducing content appears to be more prevalent and potent online than offline”).

¹⁴⁸ See Nicholas Confessore et al., *The Follower Factory*, N.Y. TIMES (Jan. 27, 2018), <https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html> (reporting on the widespread use of fake accounts and bots to amplify and simulate support for messages on Twitter and Facebook).

¹⁴⁹ See, e.g., Zeke Faux, *How Facebook Helps Shady Advertisers Pollute the Internet*, BLOOMBERG BUSINESSWEEK (Mar. 27, 2018, 5:00 AM), <https://www.bloomberg.com/news/features/2018-03-27/ad-scammers-need-suckers-and-facebook-helps-find-them> [<https://perma.cc/6G7H-2Z4G>] (reporting on a conference in Berlin that sells training for aspiring Facebook ad affiliates in the art of selling “miracle diet pills, instant muscle builders, brain boosters, [and] male enhancers.”).

¹⁵⁰ See *id.*

¹⁵¹ See Julia Angwin et al., *Facebook (Still) Letting Housing Advertisers Exclude Users by Race*, PROPUBLICA (Nov. 21, 2017, 1:23 PM), <https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin> [<https://perma.cc/F4RQ-QNJB>] (reporting that Facebook allowed placement of ads that excluded from their targeted audiences categories of users protected by the Fair Housing Act, including African Americans, Jews, people with children, and the disabled).

groups supporting the Trump campaign paid to serve racially divisive, anti-Clinton ads to African American Facebook users.¹⁵² Their goal was not to win support for Trump by convincing voters that he had good ideas; it was to suppress Democratic voter turnout by creating distaste for Clinton.¹⁵³

Under pressure from lawmakers, major advertisers, and the public, Facebook and YouTube have announced that they will commit substantially more resources to the development and implementation of internal mechanisms for improving the quality and civility of the information environments they have created.¹⁵⁴ The challenges associated with doing this responsibly and consistently are daunting, and missteps have been common. Both companies are relying on a combination of human review and algorithmic flagging or filtering of prohibited content.¹⁵⁵ However much these companies may prefer the *laissez-faire* speech norms they grew up on, increased content moderation is their new normal. The trend toward more actively moderated services is actually consistent with the policy goals underlying the CDA. Congress in 1996 *embraced* an editorial role for the kinds of public-facing edge services that have evolved into today's social media mega-platforms. The days of dial-up Internet access are long behind us, but Prodigy's legacy lives on in the CDA's under-operationalized Good Samaritan provision.

III. BALANCING FREE SPEECH WITH INFORMATION QUALITY

Recent experience teaches that social media platforms people once felt confident would strengthen community and democracy can easily be manipulated by individual and state actors to achieve corrosive effects. Social media companies are suffering from a loss of public trust, even as people continue to rely on them for news and social connection.¹⁵⁶ At this moment in the Internet's evolution,

¹⁵² Isaac & Wakabayashi, *supra* note 4.

¹⁵³ See Antonio G. Martínez, *How Trump Conquered Facebook—Without Russian Ads*, WIRED (Feb. 23, 2018, 10:06 AM), <https://www.wired.com/story/how-trump-conquered-facebook-without-russian-ads/> [<https://perma.cc/9ZSV-XGT7>] (explaining how the Trump campaign used Facebook's ad auctions and its Custom Audiences and Lookalike Audiences tools to target likely Clinton voters with ads that were "engaging but dispiriting").

¹⁵⁴ April Glaser, *Want a Terrible Job? Facebook and Google May Be Hiring.*, SLATE (Jan. 18, 2018, 11:44 AM), <https://slate.com/technology/2018/01/facebook-and-google-are-building-an-army-of-content-moderators-for-2018.html> [<https://perma.cc/CQV9-P8SX>] (reporting on plans at Google to hire 10,000 additional moderators and plans at Facebook to hire 4,000).

¹⁵⁵ See Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1636-38 (2018) (discussing the difference between *ex ante* algorithmic moderation and *ex post* human moderation).

¹⁵⁶ In twenty-one of twenty-eight international markets evaluated in the 2018 Edelman Trust Barometer, public trust in social media platforms dropped, and almost half of respondents said they don't trust platforms. See EDELMAN INTELLIGENCE, 2018 EDELMAN TRUST

when social media platforms are being exploited by foreign agents and domestic troll armies to heighten social conflict and spread disinformation, a must-carry rule for social media platforms is precisely the wrong prescription.¹⁵⁷

Such a rule would not only run counter to the Internet's history of layer-conscious regulation, it would hamper constructive efforts by platforms to curb the use of their services to spread harmful, antisocial speech. A more productive approach would be to consider how platforms can (or can be made to) responsibly operationalize the CDA's Good Samaritan provision to more effectively and consistently address problems of information quality associated with disinformation, hate speech, threats of violence, harassment, doxing, and other recognized forms of online abuse. For social media platforms that have adopted contractual community guidelines, as all of the major ones have, we need some ground rules. Practices better than those that platforms currently demonstrate in this area include increased definitional clarity with respect to categories of prohibited speech, greater consistency with respect to content removals, and implementation of efficient processes that allow users both to flag potential violations and to contest removals they believe are unjustified.¹⁵⁸

A. Clarity

As Danielle Citron has pointed out, clarity in definitions for terms like “hate speech” and “terrorist material” is critical to prevent censorship creep—the expansion of speech policies beyond their original goals.¹⁵⁹ Definitional clarity has other benefits, too. These include notice to users about the kind of speech culture a platform is trying to foster and facilitation of consistent enforcement by the platform. Citron suggests definitions of hate speech drawn from domestic tort law, domestic civil rights law, or international human rights law.¹⁶⁰ Such definitions have the benefit of existing consensus and are supported by bodies of decisional law that clarify their boundaries in specific cases. Vague policies concerning the prevention of “abuse” and the removal of “abusive” or “inappropriate” content create uncertainty for users and provide a poor basis for platform moderators to make principled decisions about removals.

BAROMETER EXECUTIVE SUMMARY, at 8-9 (2018), http://cms.edelman.com/sites/default/files/2018-02/2018_Edelman_TrustBarometer_Executive_Summary_Jan.pdf [<https://perma.cc/3FR3-FAKL>] (reporting a five-point increase in trust in journalism and a two-point decrease in trust in social media platforms from 2017 to 2018).

¹⁵⁷ Different considerations are relevant for non-social-media platforms like search and e-commerce. In conversations about “platform” policy, where online platforms are defined broadly in terms of two-sided markets (*i.e.*, users and advertisers), it is as important to recognize differences between platforms as it is to recognize differences between network-layer and application-layer service providers.

¹⁵⁸ *Cf.* Balkin, *supra* note 138 (arguing that social media platforms should behave as “information fiduciaries,” with attendant obligations).

¹⁵⁹ Danielle K. Citron, *Extremist Speech, Compelled Conformity, and Censorship Creep*, 93 NOTRE DAME L. REV. 1035, 1050, 1062 (2018).

¹⁶⁰ *Id.* at 1062–63.

Takedown mishaps arising from vaguely defined prohibitions on nudity illustrate the need for precision in the definition of impermissible content. Workable definitional lines between health information, art, and pornography are sometimes difficult to draw, which means that platforms' policies concerning nudity and sexual content must delimit the bounds of those categories in as clear-but-nuanced a way as possible. On Facebook, examples of controversial takedowns for nudity crop up regularly: Images of women breastfeeding have been treated as violations.¹⁶¹ Images of women's bare backs have been removed, but images of men's bare chests have not.¹⁶² An image of the prehistoric statue the Venus of Willendorf was taken down.¹⁶³ So, too, was an image of Gustave Courbet's nineteenth-century painting "L'Origine du Monde."¹⁶⁴

Some of these examples involve sexualized nudity; others do not. The Venus of Willendorf is a highly stylized stone fertility icon with little anatomical detail beyond a pair of very large breasts.¹⁶⁵ Courbet's painting, by contrast, is an anatomically realistic depiction of a mostly naked woman lying on her back with her legs open and her genitalia fully visible.¹⁶⁶ Both are recognized works of art. But whereas the Venus statue falls quite clearly on the "safe" side of the art-pornography divide, the Courbet painting is much harder to consign to one side or the other. Neither work is obscene by First Amendment standards, but those are not the standards that Facebook has adopted.¹⁶⁷ And, as discussed above, section 230 permits that choice.

¹⁶¹ Caitlin Dewey, *Facebook Is Embroiled in Yet Another Breastfeeding Photo Controversy*, WASH. POST (Feb. 26, 2015), https://www.washingtonpost.com/news/the-intersect/wp/2015/02/26/facebook-is-embroiled-in-yet-another-breastfeeding-photo-controversy/?utm_term=.a4a970b74edb.

¹⁶² Sapna Maheshwari & Sheera Frenkel, *Facebook Lets Ads Bare a Man's Chest. A Woman's Back is Another Matter.*, N.Y. TIMES (Mar. 1, 2018), <https://www.nytimes.com/2018/03/01/business/media/facebook-ads-gender.html>.

¹⁶³ Jesus Diaz, *Facebook Censors 30,000-Year-Old Masterpiece of Prehistoric Art*, CO.DESIGN (Mar. 1, 2018), <https://www.fastcodesign.com/90162556/facebook-censors-30000-year-old-masterpiece-of-prehistoric-art> [<https://perma.cc/HS5G-QBAQ>].

¹⁶⁴ Damian Sharkov, *Facebook to Face Court in France over Painting of a Woman's Crotch*, NEWSWEEK (Feb. 2, 2018, 4:35 AM), <http://www.newsweek.com/facebook-face-court-france-over-sublime-painting-womans-crotch-797955> [<https://perma.cc/UJB6-SB7S>].

¹⁶⁵ See Diaz, *supra* note 163 (displaying a copy of the figure).

¹⁶⁶ See Sharkov, *supra* note 164 (displaying a copy of the painting).

¹⁶⁷ See *Community Standards*, FACEBOOK, https://www.facebook.com/communitystandards/objectionable_content/adult_nudity_sexual_activity [<https://perma.cc/4TVC-T2YD>] (last visited May 9, 2018) (explaining Facebook's policy for content involving "Adult Nudity and Sexual Activity") [hereinafter Facebook Community Standards].

The French Facebook user whose account was suspended after he posted the Courbet image sued Facebook in French court and prevailed on a breach of contract claim.¹⁶⁸ Such a claim would not have survived a motion to dismiss in a U.S. court because of the cover Facebook gets from section 230's Good Samaritan provision.¹⁶⁹ The room for good faith error that section 230 allows is necessary given the scale at which social media platforms operate and the difficulty of line-drawing for many kinds of unwanted content. But that room for error should not be exploited as a license to err. In their content moderation operations, platforms should take seriously section 230's good faith requirement and should interpret that requirement as an obligation to learn from past mistakes and iterate toward more refined and implementable definitions of prohibited content.¹⁷⁰ Courts, for their part, can support the norm of definitional clarity by treating it as an element of good faith when evaluating a provider's assertion of Good Samaritan immunity.

B. Consistency

Another important element of revitalized Good Samaritanism is consistency. It's one thing for social media platforms to adopt well-defined content removal policies; it's another for them to enforce those policies consistently given the almost inconceivable scale at which they now operate. Twitter, for example, processes 500 million tweets per day.¹⁷¹ YouTube ingests over 300 hours of user-uploaded video per minute.¹⁷² As public scrutiny of platforms' moderation practices has increased, so too have concerns about their fairness and consistency. In the current era of political polarization, platforms need to be able to show that content removals and account suspensions are justified with reference to specific policies. The production of a clear justification for every removal or suspension can counter claims of arbitrariness and political bias.

¹⁶⁸ Philippe Sotto, *French Court Issues Mixed Ruling in Facebook Nudity Case*, U.S. NEWS (Mar. 15, 2018, 7:07 PM), <https://www.usnews.com/news/business/articles/2018-03-15/french-court-issues-mixed-ruling-in-facebook-nudity-case>.

¹⁶⁹ See *Facebook Statement of Rights and Responsibilities*, FACEBOOK, <https://www.facebook.com/terms.php> [<https://perma.cc/MV6Z-XLCQ>] (last visited Mar. 26, 2018) (“[Facebook] can remove any content or information [users] post on Facebook if we believe that it violates this Statement or our policies.”).

¹⁷⁰ Facebook claims to have embraced this approach. See *id.* (“Our nudity policies have become more nuanced over time. We understand that nudity can be shared for a variety of reasons, including as a form of protest, to raise awareness about a cause, or for educational or medical reasons. Where such intent is clear, we make allowances for the content.”).

¹⁷¹ Salman Aslam, *Twitter by the Numbers: Stats, Demographics & Fun Facts*, OMNICORE (Jan. 1, 2018), <https://www.omnicoreagency.com/twitter-statistics/> [<https://perma.cc/QRE2-REKS>].

¹⁷² Salman Aslam, *YouTube by the Numbers: Stats, Demographics & Fun Facts*, OMNICORE (Feb. 5, 2018), <https://www.omnicoreagency.com/youtube-statistics/> [<https://perma.cc/7563-DZAV>].

Claims of political bias are becoming more frequent and are often cited by proponents of a must-carry rule. Some conservatives allege that platforms are engaging in politically motivated takedowns and account suspensions.¹⁷³ FCC Chairman Ajit Pai recently responded to critics of his net neutrality repeal by accusing edge providers of being a greater threat to the open Internet than network operators are.¹⁷⁴ Pai's whataboutism is an obvious attempt to shift the policy conversation about network openness from the Internet's core to its edge—a strategic instance of what John Blevins calls “layer confusion.”¹⁷⁵ For the technical and policy reasons discussed in Part I above, the law has never demanded neutrality from social media platforms, and it shouldn't start demanding neutrality now. What it should demand, however, is reasonable consistency in the enforcement of platforms' published community speech guidelines.

Critics who accuse social media platforms of systematically targeting conservative speakers point to sporadic examples of “alt-right” activists whose accounts were suspended for speech-related violations.¹⁷⁶ In 2016, Twitter suspended the accounts of a handful of far-right provocateurs, including Richard Spencer.¹⁷⁷ The suspensions, Twitter said, were the result of increased activity violating Twitter's policies prohibiting abuse and targeted harassment.¹⁷⁸ Spencer and his supporters claimed that the actions were politically motivated, casting doubt on Twitter's good faith in enforcing its policies.¹⁷⁹ The controversy raises legitimate questions about selective enforcement that Twitter and all social media platforms must be prepared to answer. However, disparate impact does not necessarily equate with intentional discrimination. It is possible that consistent, even-handed enforcement of community guidelines concerning hate speech, targeted harassment, and the spread of misinformation will have uneven

¹⁷³ See, e.g., Alina Selyukh, *Feeling Sidelined by Mainstream Social Media, Far-Right Users Jump to Gab*, NPR (May 21, 2017, 6:46 AM), <https://www.npr.org/sections/alltechconsidered/2017/05/21/529005840/feeling-sidelined-by-mainstream-social-media-far-right-users-jump-to-gab> [<https://perma.cc/A5RG-4QYB>]; see also Cathy Young, *How Facebook, Twitter Silence Conservative Voices Online*, THE HILL (Oct. 28, 2016, 12:55 PM), <http://thehill.com/blogs/pundits-blog/media/303295-how-facebook-twitter-are-systematically-silencing-conservative> [<https://perma.cc/A5RG-4QYB>].

¹⁷⁴ Seth Fiegerman, *Trump's FCC Chairman Accuses Twitter of Silencing Conservatives*, CNN (Nov. 28, 2017), <http://money.cnn.com/2017/11/28/technology/ajit-pai-tech/index.html> [<https://perma.cc/UQ95-N254>].

¹⁷⁵ Blevins, *supra* note 32, at 378, 386.

¹⁷⁶ See, e.g., Travis M. Andrews, ‘A Great Purge?’: *Twitter Suspends Richard Spencer, Other Prominent Alt-Right Accounts*, WASH. POST (Nov. 16, 2016), https://www.washingtonpost.com/news/morning-mix/wp/2016/11/16/a-great-purge-twitter-suspends-richard-spencer-other-prominent-alt-right-accounts/?utm_term=.85fb4e0289bb.

¹⁷⁷ *Id.*

¹⁷⁸ *Id.*

¹⁷⁹ *Id.*

effects across the political spectrum of social media users. Naked allegations of political bias should not chill platforms from enforcing their policies.

Two empirical studies designed to map the spread of misinformation and hate speech on social media help explain why unbiased enforcement of speech guidelines may disparately impact speakers on the far right end of the political spectrum. In 2016, researchers at Oxford University conducted a three-month study of junk news and political polarization among groups of U.S. Twitter and Facebook users.¹⁸⁰ They found that content from “junk news sources”—defined as those that “deliberately publish misleading, deceptive or incorrect information purporting to be real news about politics, economics or culture”—is more often shared by conservatives than by left-leaning account holders.¹⁸¹ Trump supporters accounted for 55% of junk news traffic in the Twitter study sample and 58% of junk news traffic in the Facebook study sample.¹⁸²

In a similar study conducted in 2013 to map hate speech on Twitter, researchers at Humboldt State University geotagged 150,000 tweets containing racist, homophobic, or anti-disability words that were used in a hateful way.¹⁸³ They found that tweets containing such speech were more likely to emanate from small towns and rural areas, which voted overwhelmingly for Donald Trump in the 2016 election.¹⁸⁴ The results of these studies suggest that conservatives may be disparately, yet not unfairly, impacted by platforms’ good faith attempts to check the spread of hate speech and misinformation. In the current climate of intense political polarization, and in light of public skepticism about platforms’ motives for content removal, platforms should be able to demonstrate that their moderators are applying community guidelines without pretext or intent to discriminate based on factors outside the scope of the guidelines.

A significant step toward greater consistency in moderation is a shift over time from standards to rules, which Kate Klonick documents in her qualitative

¹⁸⁰ See Vidya Narayanan et al., *Polarization, Partisanship and Junk News Consumption over Social Media in the US* (Feb. 6, 2018), <http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/02/Polarization-Partisanship-JunkNews.pdf> [<https://perma.cc/WF5V-229U>].

¹⁸¹ See *id.* at 2, 4-5.

¹⁸² *Id.* at 5.

¹⁸³ See Alexis Kleinman, *Twitter Hate Speech Map Pinpoints Racist, Homophobic Hotspots across U.S.*, HUFFPOST (May 13, 2013, 11:43 AM), https://www.huffingtonpost.com/2013/05/13/twitter-hate-speech_n_3265916.html [<https://perma.cc/WR9G-UZCC>] (describing the study conducted by geography students at Humboldt State University under the supervision of Dr. Monica Stephens).

¹⁸⁴ Dan Balz, *Rural America Lifted Trump to the Presidency. Support is Strong But Not Monolithic.*, WASH. POST (June 17, 2017), https://www.washingtonpost.com/politics/rural-america-lifted-trump-to-the-presidency-support-is-strong-but-not-monolithic/2017/06/16/df4f9156-4ac9-11e7-9669-250d0b15f83b_story.html?utm_term=.bea100e9da6d (reporting that Donald Trump won 60% of the vote to Hillary Clinton’s 34% in the 2,332 counties that make up small-town and rural America); see also Kleinman, *supra* note 183.

survey of the current landscape in social media moderation.¹⁸⁵ In Facebook's earlier days, moderation guidelines were sparse, and moderators made decisions about content removals based on their gut reactions.¹⁸⁶ One Facebook employee Klonick interviewed described the then-prevailing standard as "Feel bad? Take it down."¹⁸⁷ It is easy to see how the subjectivity and flexibility inherent in a loose, standards-based approach could lead to uneven, unpredictable decision-making.¹⁸⁸ Platforms' transition to a rules-based approach may make moderation more time-intensive, particularly if the rules in question have many exceptions, but it should also increase consistency.¹⁸⁹

Ensuring consistency is especially important given the extent to which social media platforms outsource the human component of their content moderation operations to third-party contractors, many of which pay low wages and have high turnover due to the stressful nature of the work.¹⁹⁰ In today's environment of increasing moderation—and backlash from quarters accustomed to the more free-wheeling speech environment of the early Internet—it is important for platforms to be able to demonstrate to affected users that content removals are justified and undertaken with fairness and consistency.

C. Appealability

As platforms turbocharge their moderation operations in response to political pressure and shifting public norms, instances of mistaken content blocks and removals will inevitably increase. In addition to regulating categories of speech within the contemplation of section 230's drafters, Facebook is now targeting fake identities, fake audiences, false facts, and false narratives in the interest of

¹⁸⁵ See Klonick, *supra* note 155.

¹⁸⁶ *Id.* at 1631.

¹⁸⁷ *Id.*

¹⁸⁸ See Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L.J. 557, 607–08 (1992) (asserting that rules are less adaptable than standards to the idiosyncrasies of particular circumstances but will tend to provide clearer notice than standards to individuals at the time they decide how to act).

¹⁸⁹ See Duncan Kennedy, *Form and Substance in Private Law Adjudication*, 89 HARV. L. REV. 1685, 1688 (1976) ("[T]he two great social virtues of formally realizable rules, as opposed to standards or principles, are the restraint of official arbitrariness and certainty.").

¹⁹⁰ See Adrian Chen, *The Laborers Who Keep Dick Pics and Beheadings Out of Your Facebook Feed*, WIRED (Oct. 23, 2014, 6:30 AM), <https://www.wired.com/2014/10/content-moderation/> [<https://perma.cc/JW5A-TN99>] (reporting on working conditions for employees of content moderation contractors in the Philippines); Annalee Newitz, *Will Facebook Actually Hire 3,000 Content Moderators, Or Will They Outsource?*, ARS TECHNICA (May 4, 2017, 8:56 AM), <https://arstechnica.com/tech-policy/2017/05/facebook-promises-to-hire-3000-people-to-moderate-content/> [<https://perma.cc/2TG2-CUN7>] (reporting on Facebook's outsourcing of content moderation).

protecting election security and integrity.¹⁹¹ With this growth in the range of content subject to blocking and removal comes an increased risk of widespread private censorship. The movement in the EU toward tightened timelines—like the twenty-four hour turnaround in the hate speech code of conduct described in Part II.A above—will only exacerbate the over-removal problem. Assuring that social media users have a way to dispute content blocks or removals that they believe are unjustified must therefore be a core component of reinvigorated Good Samaritanism.

As Daphne Keller has pointed out, section 512 of the DMCA provides an existing model for how appeals of Good Samaritan removals could work.¹⁹² Under section 512, providers that receive statutory takedown notices from copyright holders notify the affected user when the content in question is removed.¹⁹³ If the user believes the allegation is mistaken or abusive, she can submit a counter-notice to the service provider.¹⁹⁴ Upon receipt of the counter-notice, the provider notifies the notice sender, who then has ten days to file a lawsuit against the counter-notice sender.¹⁹⁵ If the notice sender doesn't file suit within ten days, then the provider must restore the disputed content no later than fourteen days after it was removed.¹⁹⁶ Through the notice-and-counter-notice process, the dispute between the notice sender and the accused user either moves to the courts, or it is resolved extra-judicially—with the content in question either staying down (if no counter-notice is sent) or going back up (if a counter-notice goes unanswered).

The DMCA process is initiated by third parties, but its protections for users could also apply to removals undertaken on a provider's own initiative. As an example, YouTube's Content ID filtering system incorporates a counterclaim process even though Content ID blocks are not notice-driven.¹⁹⁷ In cases involving disputes over algorithmically-driven Good Samaritan removals, human review of the material in question should be the norm, because only human reviewers are capable of understanding contextual nuances of language and images.¹⁹⁸

¹⁹¹ *Hard Questions: What is Facebook Doing to Protect Election Security?*, FACEBOOK.COM (Mar. 29, 2018), <https://newsroom.fb.com/news/2018/03/hard-questions-election-security/> [<https://perma.cc/P7FL-TYC7>].

¹⁹² *Cf.* Keller, *supra* note 130, at 6-7.

¹⁹³ 17 U.S.C. § 512(g) (2012).

¹⁹⁴ *Id.*

¹⁹⁵ *Id.*

¹⁹⁶ *Id.*

¹⁹⁷ *See* Bridy, *supra* note 83, at 197.

¹⁹⁸ *See* Alexis C. Madrigal, *Inside Facebook's Fast-Growing Content-Moderation Effort*, ATLANTIC (Feb. 7, 2018), <https://www.theatlantic.com/technology/archive/2018/02/what-facebook-told-insiders-about-how-it-moderates-posts/552632/> [<https://perma.cc/7SGK-SF7F>] (“The current stable of machine-learning technologies is not good at looking at the *context* of a given post or user or community group.”).

In addition to establishing counter-notice procedures, section 512 creates a civil cause of action for targets of takedown notices that are sent in bad faith.¹⁹⁹ At the time of the DMCA's enactment, Congress recognized that section 512's notice-and-takedown procedure could be abused by people seeking to suppress speech for purposes unrelated to copyright infringement.²⁰⁰ In a cause of action for "knowing material misrepresentation" that content is infringing, a prevailing user can recover attorney's fees and costs from an abusive notice sender.²⁰¹

The DMCA's speech-protective provisions are far from perfect in their operation.²⁰² Notably, few users actually file counter-notices, because doing so results in disclosure of the user's identity to the notice sender, who may be a bad actor.²⁰³ The DMCA also requires the user to affirmatively consent to personal jurisdiction in federal court and receipt of service of process, which is an alarming prospect for users with little or no knowledge of copyright law.²⁰⁴ The "knowing material misrepresentation" provision has also seen little use.²⁰⁵ The knowledge standard is subjective, and therefore difficult to prove.²⁰⁶ And statutory damages, which relieve right holders of the burden to show actual damages in infringement suits, are unavailable to users claiming that a takedown notice was sent in bad faith.²⁰⁷ Actual damages in such cases are difficult to prove and likely to be only nominal.²⁰⁸

¹⁹⁹ § 512(f).

²⁰⁰ See S. REP. NO. 105-190 (1998), at 21 ("The Committee was acutely concerned that it provide all end-users—whether contracting with private or public sector online service providers—with appropriate procedural protections to ensure that material is not disabled without proper justification.").

²⁰¹ § 512(f).

²⁰² See Urban et al., *Notice and Takedown in Everyday Practice* 117-118 (Berkeley Law Sch. Pub. Law Research Paper, Paper No. 2755628), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2755628 (describing due process failures in the DMCA's counter-notice procedures).

²⁰³ See § 512(g) (providing that the service provider must provide the notice sender with a copy of the user's counter-notice, which must contain the user's name, address, and telephone number).

²⁰⁴ See *id.* (providing that the counter-notice must contain "a statement that the subscriber consents to the jurisdiction of Federal District Court for the judicial district in which the address is located . . . and that the subscriber will accept service of process from the person who provided notification. . .").

²⁰⁵ See Urban et al., *supra* note 202, at 43 n.121 (collecting cases).

²⁰⁶ *Rossi v. Motion Picture Ass'n of Am., Inc.*, 391 F.3d 1000, 1003-04 (9th Cir. 2004).

²⁰⁷ Urban et al., *supra* note 202, at 129 (citing the unavailability of statutory damages as an impediment to misrepresentation claims under section 512(f)).

²⁰⁸ *Cf. Lenz v. Universal Music Corp.*, 815 F.3d 1145, 1156 (9th Cir.), *cert. denied*, 137 S. Ct. 416 (2016), *and cert. denied*, 137 S. Ct. 2263 (2017) (rejecting the defendant's claim that a plaintiff in a section 512(f) case must prove actual damages and allowing a claim for nominal damages "due to an unquantifiable harm suffered").

Despite the infrequency with which the DMCA's user-protective provisions are invoked in practice, the framework was designed with the clear goal of protecting the expressive rights of users whose content is targeted for removal. Procedures for protecting users' expressive rights must figure prominently in the design of any fair system for enforcing community speech guidelines on social media platforms. Adapting the DMCA's user protections for Good Samaritan removals should entail not only a process for appealing removal decisions to human reviewers but also a mechanism for penalizing third parties who make clearly abusive claims, assuming the platform receiving an abusive third-party complaint has "jurisdiction" over the complainant (i.e., the abusive complainant has an account with the provider in question).

CONCLUSION

Supporters of "platform neutrality" appeal to notions of regulatory equity and symmetry. They say what's good for the goose at the Internet's network layer is good for the gander at the application layer. What they fail to recognize, however, is that personalization—i.e., content discrimination—is central, not merely incidental, to the design of social media platforms. Personalization provides value to both users and advertisers, albeit not in equal measure. That asymmetry underlies some of the Internet's most pressing information quality problems.²⁰⁹

The recent trend toward more systematic content moderation responds to the public's justified perception that the speech culture of social media has become antisocial in demonstrably damaging ways. Increased moderation is a positive development that could help rebuild public trust in social media. But it requires strong safeguards—e.g., clarity, consistency, and appealability—to protect users' expressive freedoms.

Active content moderation on social media platforms is fully supported by the Internet's longstanding policy architecture, which is built on the end-to-end principle. Section 230 of the CDA and section 512 of the DMCA reflect a policy choice of content-awareness at the Internet's application layer, where machine-readable data surfaces as intelligible—and potentially harmful—speech. Net neutrality rules for network operators reflect a policy choice of content-agnosticism at the Internet's network layer, where common carriage rules for data packets ensure low barriers to entry and promote innovation at the edge. As we contemplate how better to regulate Big Tech, we should not lose sight of the

²⁰⁹ Giving users more control over what they see in their timelines and newsfeeds would help address criticism that platforms' content-selection algorithms are opaque and optimized to display the wrong kinds of content. In drafting section 230, Congress imagined that providers would develop filtering and selection tools *for users*. See 47 U.S.C. § 230(b)(3) (2012) ("It is the policy of the United States...to encourage the development of technologies which maximize user control over what information is received by individuals, families, and schools who use the Internet and other interactive computer services."). Like the Good Samaritan provision, the user control aspect of section 230 has been under-operationalized.

2018]

REMEDIATING SOCIAL MEDIA

229

enduring relevance of the end-to-end principle. Neutrality on the Internet has its place. That place is not social media platforms. Not now.