

**EMERGING
THREATS**

DISCRIMINATORY DESIGNS ON USER DATA

BY OLIVIER SYLVAIN

**KNIGHT
FIRST AMENDMENT
INSTITUTE**

at Columbia University

EMERGING THREATS

ABOUT EMERGING THREATS

The Knight First Amendment Institute's *Emerging Threats* series invites leading thinkers to identify and grapple with newly arising or intensifying structural threats to the system of free expression. These threats may be caused by changes in the forms and applications of technology, in the means and economics of communication, in the norms and practices of politics, or in legal doctrine. The papers in the series explore ways to address these threats and preserve the foundations of democracy essential to healthy open societies, including the United States.

The *Emerging Threats* series is edited by David Pozen, professor at Columbia Law School and inaugural visiting scholar at the Knight Institute.

ABOUT THE KNIGHT INSTITUTE

The Knight First Amendment Institute is a non-partisan, not-for-profit organization established by Columbia University and the John S. and James L. Knight Foundation to defend the freedoms of speech and press in the digital age through strategic litigation, research, and public education.

For more information, please visit www.knightcolumbia.org.

ABOUT THE AUTHOR

Olivier Sylvain is an Associate Professor of Law at Fordham University School of Law. He teaches legislation & regulation, administrative law, information law, and law related to communications technologies. He is also the Director of the McGannon Center for Communications Research. Additionally, Sylvain is part of a team of research engineers and social entrepreneurs to whom the National Science Foundation in fall 2017 awarded a three-year one-million-dollar grant to prototype a secure, affordable, and energy-efficient computing network that is to be owned and operated as a "commons resource" for Harlem residents. Sylvain was a Karpatkin Fellow in the National Legal Office of the American Civil Liberties Union in New York City and a litigation associate at Jenner & Block, LLC, in Washington, D.C. He is on the board of directors for the New York affiliate of the American Civil Liberties Union.

TABLE OF CONTENTS

Discriminatory Designs on User Data

I. Section 230 Immunity: A Brief Overview	5
II. The Lived Human Costs of “Unfettered” Online Speech: The Example of Nonconsensual Pornography	8
III. More than a Conduit: Online Intermediaries’ Designs on User Data	11
<i>A. Intermediary Designs and User Experiences</i>	<i>11</i>
<i>B. Discriminatory Designs on User Content and Data: The Example of Online Housing Marketplaces</i>	<i>13</i>
<i>C. Doctrinal Responses — and Resources</i>	<i>17</i>
IV. Toward a More Nuanced Immunity Doctrine	19

DISCRIMINATORY DESIGNS ON USER DATA

Olivier Sylvain

The stated aim of online intermediaries like Facebook, Twitter, and Airbnb is to provide the platforms through which users freely meet people, purchase products, and discover information.¹ As “conduits” for speech and commerce,² intermediaries such as these are helping to create a more vibrant and democratic marketplace for goods and ideas than any the world has seen before.³

That, at least, is the theory on which Congress enacted Section 230 of the Communications Decency Act (CDA) in 1996.⁴ One of the central objectives of Section 230’s drafters was to ensure that intermediaries are “unfettered” by the obligation to police third-party user content.⁵ They believed that conventional tort principles and regulatory rules were simply not workable in an environment in which so much user content flows,⁶ and they doubted that intermediaries would be able to create new value for users if they constantly had to monitor, block, or remove illicit content. In the words of free speech doctrine, members of Congress worried the intermediaries would be “chilled” by the fear that they could be held legally responsible for content posted by users.⁷

Section 230 of the CDA therefore protects intermediaries from liability for distributing third-party user content. Courts have read Section 230 broadly, creating an immunity for intermediaries who do all but “materially contribute” to the user content they distribute.⁸ That is, courts have read the statute’s protections to cover services that “augment[]” user content, but not services that demonstrably “help” to develop the alleged illegal expressive

⁺ Associate Professor, Fordham Law School.

¹ See, e.g., Mark Zuckerberg, *Bringing the World Closer Together*, Facebook (June 22, 2017), <http://www.facebook.com/zuck/posts/10154944663901634>; *About Us*, Airbnb, <http://www.airbnb.com/about/about-us> (last visited Feb. 23, 2018); Ricardo Castro, *A Better Way to Connect with People*, Twitter Blog (May 3, 2016), http://blog.twitter.com/official/en_us/a/2016/a-better-way-to-connect-with-people.html.

² *Zeran v. America Online, Inc.*, 129 F.3d 327, 332 (4th Cir. 1997).

³ See Orly Lobel, *The Law of the Platform*, 101 Minn. L. Rev. 87, 89 (2016) (discussing “the digital platform revolution”).

⁴ 47 U.S.C. § 230 (2012).

⁵ *Id.* § 230(b)(2).

⁶ See, e.g., 104 Cong. Rec. H8469 (statements of Rep. Cox and Rep. Wyden); H.R. Rep. No. 104-58, at 194 (1996); see also Eugene Volokh, *Freedom of Speech in Cyberspace from the Listener’s Perspective: Private Speech Restrictions, Libel, State Action, Harassment, and Sex*, 1996 U. Chi. Legal F. 377, 405–06 (1996); Alan H. Bomser, *A Lawyer’s Ramble down the Information Superhighway*, 64 Fordham L. Rev. 697, 799–800 (1996).

⁷ See Anthony Ciolli, *Chilling Effects: The Communications Decency Act and the Online Marketplace of Ideas*, 63 U. Miami L. Rev. 137, 148 (2008); Seth F. Kreimer, *Censorship by Proxy: The First Amendment, Internet Intermediaries, and the Problem of the Weakest Link*, 155 U. Pa. L. Rev. 11, 28–29 (2006); Rebecca Tushnet, *Power Without Responsibility: Intermediaries and the First Amendment*, 76 Geo. Wash. L. Rev. 986, 991, 998–99, 1006–09, 1015–16 (2008); Felix T. Wu, *Collateral Censorship and the Limits of Intermediary Immunity*, 87 Notre Dame L. Rev. 293, 300, 315–18 (2011).

⁸ See, e.g., *Jones v. Dirty World Entm’t Recordings*, 755 F.3d 398, 413 (6th Cir. 2014); *Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1167–71 (9th Cir. 2008) (en banc).

⁹ *Roommates.com*, 521 F.3d at 1167–68.

conduct.⁹ Many believe that the internet would not be as dynamic and beguiling today were it not for the protection that Section 230 has been construed to provide for online intermediaries.¹⁰

This may be true. But Section 230 doctrine has also had a perverse effect. By providing intermediaries with such broad legal protection, the courts' construction of Section 230 effectively underwrites content that foreseeably targets the most vulnerable among us. In their ambition to encourage an "unfettered" market for online speech, the developers of Section 230 immunity have set up a regime that makes online engagement more difficult for children, women, racial minorities, and other predictable targets of harassment and discriminatory expressive conduct. Examples abound: the gossip site that enabled users to anonymously post salacious images of unsuspecting young women;¹¹ the social media site through which an adult male lured a young teenage girl into a sexual assault;¹² the classifieds site that has allegedly facilitated the sex trafficking of minors;¹³ the online advertising platform that allows companies to exclude Latinos from apartment rentals and older people from job postings;¹⁴ the unrelenting social media abuse of feminist media critics¹⁵ and a prominent black female comedian;¹⁶ the live video stream of a gang rape of a teenage girl.¹⁷

The standard answer to the charge that current immunity doctrine enables these acts is that the originators of the illicit content are to blame, not the "neutral" services that facilitate online interactions.¹⁸ Intermediaries, in this position, merely pass along user speech; they do not encourage its production or dissemination, and, in any case, Section 230 immunity exists to protect against a different problem: the "collateral censorship" of lawful content.¹⁹

¹⁰ See, e.g., *id.* at 1180 (McKeown, J., concurring in part and dissenting in part) ("We have underscored that this broad grant of webhost immunity gives effect to Congress's stated goals 'to promote the continued development of the Internet and other interactive computer services' and 'to preserve the vibrant and competitive free market that presently exists for the Internet and other interactive computer services.'" (quoting *Carafano v. Metroplash.com*, 339 F.3d 1119, 1123 (9th Cir. 2003)).

¹¹ *Jones v. Dirty World Entm't Recordings, LLC*, 755 F.3d 398 (6th Cir. 2014).

¹² *Doe v. MySpace*, F. Supp. 2d 843 (W.D. Tex. 2007).

¹³ *Doe v. Backpage.com, LLC*, 817 F.3d 12 (1st Cir. 2016), cert. denied, 137 S. Ct. 622 (2017).

¹⁴ Julia Angwin & Terry Parris Jr., *Facebook Lets Advertisers Exclude Users by Race*, ProPublica (Oct. 28, 2016), <http://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race>.

¹⁵ Nick Wingfield, *Feminist Critics of Video Games Facing Threats in 'GamerGate' Campaign*, N.Y. Times (Oct. 15, 2014), <http://www.nytimes.com/2014/10/16/technology/gamergate-women-video-game-threats-anita-sarkeesian.html>.

¹⁶ Anna Silman, *A Timeline of Leslie Jones's Horrific Online Abuse*, Cut (Aug. 24, 2016), <http://www.thecut.com/2016/08/a-timeline-of-leslie-joness-horrific-online-abuse.html>.

¹⁷ Emanuella Grinberg, *Police: At Least 40 People Watched Teen's Sexual Assault on Facebook Live*, CNN (Mar. 22, 2017), <http://www.cnn.com/2017/03/21/us/facebook-live-gang-rape-chicago/index.html>.

¹⁸ See Rob Goldman, *This Time, ProPublica, We Disagree, Facebook Newsroom* (Dec. 20, 2017), <http://newsroom.fb.com/news/h/addressing-targeting-in-recruitment-ads>.

¹⁹ See Wu, *supra* note 7, at 315–18.

This answer, however, is either glib or too wedded to an obsolete conception of how online intermediaries operate. Intermediaries today do much more than passively distribute user content or facilitate user interactions. Many of them elicit and then algorithmically sort and repurpose the user content and data they collect. The most powerful services also leverage their market position to trade this information in ancillary or secondary markets.²⁰

Intermediaries, moreover, design their platforms in ways that shape the form and substance of their users' content. Intermediaries and their defenders characterize these designs as substantively neutral technical necessities, but as I explain below, recent developments involving two of the most prominent beneficiaries of Section 230 immunity, Airbnb and Facebook, suggest otherwise. Airbnb and Facebook have enabled a range of harmful expressive acts, including violations of housing and employment laws, through the ways in which they structure their users' interactions.

At a minimum, companies should not get a free pass for enabling unlawful discriminatory conduct, regardless of the social value their services may otherwise provide. But more than this, I argue here,²¹ Section 230 doctrine requires a substantial reworking if the internet is to be the great engine of democratic engagement and creativity that it should be. Section 230 is no longer serving all the purposes it was meant to serve. The statute was intended at least in part to ensure the vitality and diversity, as well as the volume, of speech on new communications platforms. By allowing intermediaries to design their platforms without internalizing the costs of the illegal speech and conduct they facilitate, however, the statute is having the opposite effect.

This paper has four parts. The first discusses the basic contours of the prevailing doctrine, including the legislative purposes behind Section 230 and the logic courts have relied on to support broad immunity for intermediaries. The second part identifies ways in which the doctrine, in assuming that intermediaries are passive disseminators of information, may accelerate the mass distribution of content that harms vulnerable people and members of historically subordinated groups. I focus in particular on the distribution of nonconsensual pornography as a species of content that not only exacts a discrete reputational or privacy toll on victims but also fuels the circulation of misogynist views that harm young women in particular.

The third part of the paper turns to the designs that intermediaries employ to structure and enhance their users' experience, and how these designs themselves can further discrimination. While the implications of this analysis reach beyond injuries to historically marginalized groups, my goal is to explain how the designs employed by two of the most prominent intermediaries today, Airbnb and Facebook, have enabled unlawful discrimination.

²⁰ See generally Kenneth Bamberger & Orly Lobel, *Platform Market Power*, 32 Berk. Tech. L.J. 1, 37–39 (2018) (discussing ways in which intermediaries leverage their market position to exploit user data in different markets); Lina M. Khan, Note, *Amazon's Antitrust Paradox*, 126 Yale L.J. 710 (2017) (discussing ways in which intermediaries may raise antitrust concerns to the extent they cultivate their position as "essential infrastructure" for commerce across industries).

²¹ This argument builds on my recent writing. See Olivier Sylvain, *Intermediary Design Duties*, 50 Conn. L. Rev. 202 (2018) [hereinafter Sylvain, *Design Duties*]; Olivier Sylvain, *AOL v. Zeran: The Cyberlibertarian Hack of Section 230 Has Run Its Course*, Law.com (Nov. 10, 2017), <http://www.law.com/therecorder/sites/therecorder/2017/11/10/aol-v-zeran-the-cyberlibertarian-hack-of-%C2%A7230-has-run-its-course>.

The fourth and final part of the paper proposes a reform to the doctrine: I argue that courts should account for the specific ways in which intermediaries' designs do or do not enable or cause harm to the predictable targets of discrimination and harassment. As recent developments underscore, Section 230 immunity doctrine must be brought closer in line with longstanding equality and universality norms in communications law.²²

I. Section 230 Immunity: A Brief Overview

The immunity that intermediaries enjoy under Section 230 of the CDA²³ has helped to bring about the teeming abundance of content in today's online environment. The prevailing interpretation of Section 230 bars courts from imposing liability on intermediaries that are the "mere conduits" through which user-generated content passes.²⁴ This doctrine protects services that host all kinds of content—everything from customer product reviews to fake news to dating profiles.

Congress invoked a very old concept when it drafted this law. The central provision of Section 230, titled "Protection for 'Good Samaritan' blocking and screening of offensive material,"²⁵ resembles laws in all the states that in one way or another shield defendants from liability arising from their good-faith efforts to help those in distress.²⁶ Good Samaritan laws are inspired by the Biblical parable that praises the do-gooder who risks ridicule and censure to help a stranger left for dead.²⁷

Section 230's drafters applied this concept to online activity. They created an exception under tort law, which traditionally holds publishers liable for distributing material they know to be unlawful, but does not hold them liable if they lack notice about the illegality of the communicative act at issue.²⁸ Proponents of Section 230 worried that,

²² On these norms, see generally Olivier Sylvain, *Network Equality*, 67 *Hastings L.J.* 443 (2016).

²³ The pertinent language provides as follows:

(1) Treatment of publisher or speaker

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.

(2) Civil liability. No provider or user of an interactive computer service shall be held liable on account of—

(A) any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected; or

(B) any action taken to enable or make available to information content providers or others the technical means to restrict access to material described in paragraph (1).

47 U.S.C. § 230(c) (2012). The statute excludes from immunity intermediaries that are "responsible, in whole or in part, for the creation or development" of illicit user content. *Id.* § 230(f)(3). Applying this language, courts have subjected defendant intermediaries to liability when they "materially contribute" to the offending content. See, e.g., *Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1167–71 (9th Cir. 2008) (en banc). In principle, the "material contribution" standard limits the scope of the protection to services that are only conduits of content. In practice, however, it is a very high bar for plaintiffs to clear.

²⁴ *Zeran v. America Online, Inc.*, 129 F.3d 327, 332 (4th Cir. 1997).

²⁵ 47 U.S.C. § 230(c).

²⁶ See Benjamin C. Zipursky, *Online Defamation, Legal Concepts, and the Good Samaritan*, 51 *Val. U. L. Rev.* 1, 31 (2016); Benjamin C. Zipursky, *Thinking in the Box in Legal Scholarship: The Good Samaritan and Internet Libel*, 50 *J. Legal Educ.* 55, 60 (2016).

²⁷ *Luke* 10:23-37 ("[A] Samaritan, as he traveled, came where the man was; and when he saw him, he took pity on him. He went to him and bandaged his wounds, pouring on oil and wine. Then he put the man on his own donkey, brought him to an inn and took care of him.").

²⁸ *Zeran*, 129 F.3d at 330–32.

without this legislation, claims for secondary liability would either stifle expressive conduct in the then-nascent medium or discourage intermediaries from policing content altogether.²⁹ They further insisted that government regulators such as the Federal Communications Commission should play no role in deciding what sorts of content prevailed online; viewers (and their parents) should make those decisions for themselves.³⁰

While an interest in both free speech and the Good Samaritan concept drove Congress to enact Section 230, courts interpreting the statute have been far more influenced by the free speech concerns. In contrast to the nuanced requirements of the Digital Millennium Copyright Act's notice-and-takedown regime,³¹ online intermediaries have not been required under Section 230 to block or screen offensive material in any particular way. Today, Section 230 doctrine provides a near-blanket immunity to intermediaries for hosting tortious third-party content. Long-established internet companies like America Online and Craigslist that host massive amounts of user content have been clear beneficiaries. Relying on Section 230, courts have immunized them from liability for everything from defamatory posts on electronic bulletin boards to racially discriminatory solicitations in online housing advertisements.³² Leading opinions have reasoned that the scale at which third-party content passes through online services makes that content infeasible to moderate; requiring services to try would not only chill online speech but also stunt the internet's development as a transformative medium of communication.³³ This immunity now applies to a wide range of online services that host and distribute user content, including Twitter's microblogging service, Facebook's flagship social media platform, and Amazon's online marketplace. Thanks to Section 230, these companies have no legal obligation to block or remove mendacious tweets, fraudulent advertisements, or anticompetitive customer reviews by rivals.³⁴

As a result, most targets of illicit online user content in the United States have little to no effective recourse under law to have that content blocked or removed. They can sue the original posters of the content. But such litigation often presents serious challenges, including the cost of bringing a lawsuit, the difficulty of discovering the identities of anonymous posters, and, even if the suit is successful on the merits, the difficulty of obtaining remedies that are commensurate with the harm.³⁵ Targets can also enlist services like search engine optimizers that make it harder to find the offending material. They can complain to the intermediaries about offending posts. And they can press intermediaries to improve their policies generally. If none of these strategies succeeds, users can boycott the service, as many people did recently—for one day—to protest the failure of Twitter to

²⁹ 104 Cong. Rec. H8469 (statement of Rep. Wyden).

³⁰ *Id.* (statement of Rep. Cox).

³¹ See 17 U.S.C. § 512(c) (2012); see also *Viacom v. YouTube*, 676 F.3d 19 (2d. Cir. 2012).

³² See, e.g., *Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157 (9th Cir. 2008) (en banc); *Zeran v. America Online, Inc.*, 129 F.3d 327 (4th Cir. 1997).

³³ See, e.g., *Zeran*, 129 F.3d at 333.

³⁴ See, e.g., *Klayman v. Zuckerberg*, 753 F.3d 1354 (D.C. Cir. 2014); *Joseph v. Amazon.com*, 46 F. Supp. 3d 1095 (W.D. Wash. 2014); *Goddard v. Google*, 640 F. Supp. 2d 1193 (N.D. Cal. 2009).

³⁵ See Danielle Keats Citron, *Hate Crimes in Cyberspace* 122 (2014).

protect women from “verbal harassment, death threats, and doxing.”³⁶ Even if effective, however, this last option sometimes feels far from optimal, given that the promise of the internet is understood to lie in its unrivaled opportunities for commercial engagement and social integration. Exit would only exacerbate extant disparities.³⁷

The threat of losing consumers, it must be said, is potent enough to have moved many intermediaries to develop content-governance protocols and automated systems for content detection. Even though Section 230 doctrine has removed any legal duty to moderate third-party content, certain companies routinely block or remove content when its publication detracts from the character of the service they mean to provide. And so, for instance, Google demotes or delists search engine optimizers and sites that host “fake news” and offensive content.³⁸ Facebook removes clickbait articles and has now partnered with fact-checking organizations like Snopes and PolitiFact to implement a notification process for removing “fake news.”³⁹

The reform that the news aggregation and discussion site Reddit undertook in 2015 is especially striking in this regard. Reddit, which had been evangelical about its laissez-faire approach to user-generated content, implemented rules that ban “illegal” content, “involuntary pornography,” material that “[e]ncourages or incites violence,” and content that “[t]hreatens, harasses, or bullies or encourages others to do so.”⁴⁰ Many “redditors” rebelled, voting up user comments that addressed Reddit’s Asian-American female CEO in racist and misogynist ways.⁴¹ These posts were popular enough among redditors to make it to the site’s front page, the prime position on the site that touts itself as “the first page of the Internet.” Reddit subsequently buttressed its restrictions on violent and harassing content.⁴² Moreover, it recently banned a “subreddit” of self-identified misogynists.⁴³ Reddit’s reforms have been met with fierce resistance from self-styled free speech enthusiasts.⁴⁴ But the company does not appear to be backpedaling at this time.

³⁶ See Debbie Chachra, *Twitter’s Harassment Problem Is Baked into Its Design*, Atlantic (Oct. 16, 2017), <http://www.theatlantic.com/technology/archive/2017/10/twitters-harassment-problem-is-baked-into-its-design/542952>.

³⁷ See Sylvain, *supra* note 22, at 462–64.

³⁸ See *Search King v. Google*, 2003 WL 21464568 (W.D. Okla. 2003). See generally Deepa Seetharaman, *Google Retools Search Engine to Demote Hoaxes, Fake News*, Wall St. J. (Apr. 25, 2017), <http://www.wsj.com/articles/google-retools-search-engine-to-down-play-hoaxes-fake-news-1493144451>.

³⁹ Erin Griffith, *Facebook Can Absolutely Control Its Algorithm*, Wired (Sept. 26, 2017), <http://www.wired.com/story/facebook-can-absolutely-control-its-algorithm>; Amber Jamieson & Olivia Solon, *Facebook to Begin Flagging Fake News in Response to Mounting Criticism*, Guardian (Dec. 15, 2016), <http://www.theguardian.com/technology/2016/dec/15/facebook-flag-fake-news-fact-check>.

⁴⁰ *Reddit Content Policy*, Reddit, <http://www.reddit.com/help/contentpolicy> (last visited Feb. 23, 2018); see also *Removing Harassing Subreddits*, Reddit (June 10, 2015), http://np.reddit.com/r/announcements/comments/39bpam/removing_harassing_subreddits.

⁴¹ Charlie Warzel, *Reddit Is a Shrine to the Internet We Wanted and That’s a Problem*, BuzzFeed (June 19, 2015), <http://www.buzzfeed.com/charliwarzel/reddit-is-a-shrine-to-the-internet-we-wanted-and-thats-a-pro>.

⁴² ModNews, *Update on Site-Wide Rules Regarding Violent Content*, Reddit (Oct. 25, 2017), http://www.reddit.com/r/modnews/comments/78p7bz/update_on_sitewide_rules_regarding_violent_content.

⁴³ See Aja Romano, *Reddit Just Banned One of Its Most Toxic Forums. But It Won’t Touch the Donald*, Vox (Nov. 13, 2017), <http://www.vox.com/culture/2017/11/13/16624688/reddit-bans-incels-the-donald-controversy>.

⁴⁴ *Id.*

As this example indicates, and as new scholarship illuminates,⁴⁵ attention to consumer demand and a sense of corporate responsibility have motivated certain intermediaries to moderate certain user content. It may be tempting to conclude that reforms to Section 230 law are therefore unnecessary. Unregulated intermediaries might be the best gauges of authentic user sentiment about what is or is not objectionable. Section 230 doctrine, on this view, allows users to express and learn from each other in a dynamic fashion, without the distortions that may be caused by tort liability or government mandates. This is part of why free speech enthusiasts ascribe so much significance to the statute: Section 230 doctrine for them is premised on a noble faith in the moral and democratic power of unregulated information markets.⁴⁶

II. The Lived Human Costs of “Unfettered” Online Speech: The Example of Nonconsensual Pornography

These arguments for near-blanket immunity only go so far, though. As much as some intermediaries may try, the fact is that many others do not make any effort to block or remove harmful expressive conduct. According to their critics, sites like Backpage (a classified site through which users are known to engage in the sex trafficking of minors) or TheDirty (a gossip site known for soliciting derogatory content about unsuspecting young women) are unabashed solicitors and distributors of a species of content that attacks members of historically subordinated groups. Under current doctrine, they are immune for acting in this way. They are just as immune under Section 230 as are ostensibly content-conscience intermediaries like Facebook and Twitter that purport to remove or block various categories of illicit user content but nevertheless sometimes distribute it.⁴⁷ The prevailing justification for this approach is to protect against the “collateral censorship” of lawful content.⁴⁸ This view holds that slippage in the direction of occasionally hosting hurtful material is the price of ensuring free speech online.

It may be correct that tolerating harmful content every now and again is the cost of promoting the statutory objective of an “unfettered” online speech environment. But just as a wide range of offline expressive acts like fraud, sexual harassment, and racially discriminatory advertisements for housing are not entitled to legal protection, we might wonder whether online services should be entirely immune for similar behaviors by their users.⁴⁹ To be

⁴⁵ See Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 113 Harv. L. Rev. (forthcoming 2018), <http://ssrn.com/abstract=2937985> (manuscript at 32–37); Karen Levy & Solon Barocas, *Designing Against Discrimination in Online Markets*, 32 Berkeley Tech. L.J. (forthcoming 2018), <http://ssrn.com/abstract=3084502>.

⁴⁶ See Derek Khanna, *The Law That Gave Us the Modern Internet—and the Campaign to Kill It*, Atlantic (Sept. 12, 2017), <http://www.theatlantic.com/business/archive/2013/09/the-law-that-gave-us-the-modern-internet-and-the-campaign-to-kill-it/279588>.

⁴⁷ See, e.g., Ariana Tobin et al., *Facebook’s Uneven Enforcement of Hate Speech Rules Allows Vile Posts to Stay Up*, ProPublica (Dec. 27, 2017), <http://www.propublica.org/article/facebook-enforcement-hate-speech-rules-mistakes>; Julia Angwin et al., *Facebook (Still) Letting Housing Advertisers Exclude Users by Race*, ProPublica (Nov. 21, 2017), <http://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin>.

⁴⁸ See sources cited *supra* note 7.

⁴⁹ There also is a toll on the human moderators responsible for censoring illicit content. Recent reporting suggests that these workers are traumatized by the material they censor. See Lauren Weber & Deepa Seetharaman, *The Worst Job in Technology: Staring at Human Depravity to Keep It off Facebook*, Wall St. J. (Dec. 27, 2017), <http://www.wsj.com/articles/the-worst-job-in-technology-staring-at-human-depravity-to-keep-it-off-facebook-1514398398>.

sure, there is a significant qualitative and quantitative difference between the reach of offline and online expressive acts: The latter travel further and faster than the former by a long shot. But this fact hardly removes the need to regulate harmful online behaviors. Quite the contrary. The human costs of “unfettered” online speech may be aggravated by the internet’s reach, and the costs themselves are disproportionately shouldered by those who are most likely to be the targets of attacks and abuse both online and off. That is to say, the victims of online abuse tend to be the same sorts of people who have always been subject to attack and harassment offline in the United States and elsewhere—in particular, young women, racial minorities, and sexual “deviants.”⁵⁰

The harm that these users experience is made worse by the way in which illicit or inflammatory content, once distributed, can spread across the internet at a speed and scale that is hard, if not impossible, to control. This unforgiving ecology raises the stakes of occasional slippage for the predictable targets and systemic victims of harmful content. The internet thus reinforces some of the classic arguments for the regulation of assaultive speech acts that target members of historically subordinated groups.⁵¹ The vitriolic content that flows through online intermediaries affects members of these groups distinctively, discouraging them from participating fully in public life online and making their social and commercial integration even more difficult than it might otherwise be.⁵²

Consider nonconsensual pornography, the distribution of nude images of a person who never authorized their distribution. On the internet, such images are generally shared in order to humiliate or harass the depicted person. In some instances, third parties then exploit the images to extort the victim, as in the case of sites that require a fee to take the images down.⁵³ Other parties discover and distribute such images for free, without necessarily knowing anything about the depicted individual.

The injuries caused by nonconsensual pornography are clear and are felt most immediately and painfully by its victims. Section 230 jurisprudence is riddled with cases that illustrate these harms. In one of the more cited ones, *Barnes v. Yahoo!, Inc.*,⁵⁴ a young woman sued Yahoo! for failing to remove a false dating site profile of her created by her ex-boyfriend. The profile contained her work phone number and address, as well as nude and suggestive photographs accompanied by promises of sex. Would-be suitors and predators soon came looking for her at work. The harm caused by this cruel hoax was plain.

⁵⁰ See Citron, *supra* note 35, at 13–16.

⁵¹ See, e.g., Mari J. Matsuda et al., *Words that Wound: Critical Race Theory, Assaultive Speech, and the First Amendment* (1993); Charles R. Lawrence, III, *Crossburning and the Sound of Silence: Antisubordination Theory and the First Amendment*, 37 *Vill. L. Rev.* 787 (1992).

⁵² Cf. Richard Delgado & Jean Stefancic, *Understanding Words that Wound* 217–18 (2004) (advocating a “new approach” that “points out how speech and equality stand in reciprocal relation; neither can thrive without the other. Speech without equality is a lecture, a sermon, a rant. Speech, in other words, presumes equality, or something like it, among participants in a dialogue”).

⁵³ See Margaret Talbot, *The Attorney Fighting Revenge Porn*, *New Yorker* (Dec. 5, 2016), <http://www.newyorker.com/magazine/2016/12/05/the-attorney-fighting-revenge-porn>.

⁵⁴ 570 F.3d 1096 (9th Cir. 2009).

Victims of nonconsensual pornography may experience many other indignities. Once posted, the offending image takes on a life of its own, exacting something that resembles an endlessly repeating privacy invasion. Danielle Citron and Mary Anne Franks, who have been thinking and writing compellingly about the issue for almost a decade now, explain the phenomenon:

Today, intimate photos are increasingly being distributed online, potentially reaching thousands, even millions of people, with a click of a mouse. A person's nude photo can be uploaded to a website where thousands of people can view and repost it. In short order, the image can appear prominently in a search of the victim's name. It can be e-mailed or otherwise exhibited to the victim's family, employers, coworkers, and friends. The Internet provides a staggering means of amplification, extending the reach of content in unimaginable ways.⁵⁵

The scale of distribution magnifies the harm to depicted individuals far beyond what is possible through other communications technologies. In this environment, taking down nonconsensual pornography, once it has been posted on an online intermediary, often becomes a futile and agonizing game of whack-a-mole.

In addition to the direct harms to those whose images are being exploited, the distribution of nonconsensual pornography also exacts a more general harm that mirrors and reinforces the routine subjugation of young women.⁵⁶ It is different in this regard from defamatory user posts, the prototypical subject of Section 230 jurisprudence, in which the injury caused by the defamatory posts are reputational in nature.⁵⁷ Nonconsensual pornography sweeps its victims into a network of blogs, pornography sites, social media groups, Tumblrs, and Reddit discussion threads that enthusiastically traffic in the collective humiliation of young women.⁵⁸

And yet, Section 230 doctrine relieves online intermediaries of any legal obligation to block or remove nonconsensual pornography. When sued for distributing such images and videos, the intermediaries cite Section 230 to justify their passive role. Courts have generally sided with them, explaining that the immunity is not contingent on sites' policing of illicit user content.⁵⁹ The result is not only grief for the predictable victims of online abuse and harassment but also a regulatory regime that helps to reinforce systemic subordination.

⁵⁵ Danielle Keats Citron & Mary Anne Franks, *Criminalizing Revenge Porn*, 49 Wake Forest L. Rev. 345, 350 (2014).

⁵⁶ See Danielle Keats Citron, *Law's Expressive Value in Combating Cyber Gender Harassment*, 108 Mich. L. Rev. 373 (2009); see also Clare McGlynn et al., *Beyond 'Revenge Porn': The Continuum of Image-Based Sexual Abuse*, 25 Feminist Legal Stud. 25 (2017); Catherine Buni & Soraya Chemaly, *The Unsafety Net: How Social Media Turned Against Women*, Atlantic (Oct. 9, 2014), <http://www.theatlantic.com/technology/archive/2014/10/the-unsafety-net-how-social-media-turned-against-women/381261>.

⁵⁷ See, e.g., *Barrett v. Rosenthal*, 146 P.3d 510 (Cal. 2006); *Batzel v. Smith*, 333 F.3d 1018 (9th Cir. 2003); *Zeran v. America Online, Inc.*, 129 F.3d 327 (4th Cir. 1997). See generally Joel R. Reidenberg et al., Ctr. on Law & Info. Pol'y at Fordham Law Sch., *Section 230 of the Communications Decency Act: A Survey of the Legal Literature and Reform Proposals* (Apr. 25, 2012), <http://ssrn.com/abstract=2046230> (surveying sixteen years of Section 230 cases).

⁵⁸ See Citron, *supra* note 35, at 127 ("Cyber harassment reinforces gender stereotypes by casting women as sex objects that are unfit for life's important opportunities.")

⁵⁹ See, e.g., *Barnes v. Yahoo!, Inc.*, 570 F.3d 1096 (9th Cir. 2009); *Jones v. Dirty World Entm't Recordings, LLC*, 755 F.3d 398 (6th Cir. 2014).

III. More than a Conduit: Online Intermediaries' Designs on User Data

As pernicious as it is, cyberharassment does not reflect the full scope of the threat that such broad legal protection for online intermediaries poses to vulnerable persons. This is because, today, most if not all intermediaries affirmatively shape the form and substance of user content. Adding to the arguments that scholars like Citron and Franks have ably made, I want to call attention here to this crucial way in which Section 230 immunity entrenches extant barriers to social and commercial integration for historically subordinated groups. I want to suggest, furthermore, that over two decades into the development of the networked information economy, online intermediaries should not be able to claim blissful indifference when their designs predictably elicit or even encourage expressive conduct that perpetuates discrimination and subjugation.

I make these arguments in this part in several sections. In section A, I illustrate the ways in which intermediaries pervasively influence users' online experiences. In section B, I explain how such designs can enable and exacerbate certain categories of harmful expressive acts. Section C looks at the courts' responses.

A. Intermediary Designs and User Experiences

Popular services like Facebook, Twitter, and Airbnb offer good examples of how intermediary designs interact with user experiences. Twitter immediately distributes its users' posts (tweets) after the users type them. But its user interface affects the nature and content of those tweets. Twitter's 280-character limitation, for example, has generated its own abbreviated syntax and conventions (for example, hashtags and subtweets).⁶⁰ The company also allows pseudonyms, effectively allowing users to be anonymous. This liberal approach to attribution invites creativity and useful provocation but also the harassment and targeted attacks mentioned above. Twitter knows this, and in many cases it will take down such attacks after the fact and remove users who routinely violate the company's no-harassment policy.

These superficial interface design features are distinct from the designs on content that occur behind (so to speak) the user interface. Some companies are intentionally deceptive about how they acquire or employ content. Take, for example, the online marketing company that placed deceptive information about its clients' products on affiliated "fake news" sites.⁶¹ Or consider the online sleuthing company that, in response to solicited user requests for information about people, routinely contracted with third-party researchers to retrieve information in ways it allegedly knew violated privacy law.⁶²

⁶⁰ Twitter recognizes the significance of its character limitation; it increased the limitation from 140 in November 2017 to improve the user experience. See Aatif Sulleyman, *Twitter Introduces 280 Characters to All Users*, Independent (Nov. 7, 2017), <http://www.independent.co.uk/life-style/gadgets-and-tech/news/twitter-280-characters-tweets-start-when-get-latest-a8042716.html>.

⁶¹ See *FTC v. LeadClick Media*, 838 F.3d 158 (2d Cir. 2016).

⁶² See *FTC v. Accusearch*, 570 F.3d 1187 (10th Cir. 2009).

Without necessarily resorting to outright deception, many more intermediaries administer their platforms in obscure or undisclosed ways that are meant to influence how users behave on the site.⁶³ Many intermediaries, for example, employ user interfaces designed to hold user attention by inducing something like addictive reliance.⁶⁴ Facebook employs techniques to ensure that each user sees stories and updates in her “News Feeds” that she may not have seen on her last visit to site.⁶⁵ And its engineers constantly tweak the algorithms that manage the user experience.⁶⁶ In addition, many intermediaries analyze, sort, and repurpose the user content they elicit. Facebook and Twitter, for example, employ software to make meaning out of their users’ “reactions,” search terms, and browsing activity in order to curate the content of each user’s individual feed, personalized advertisements, and recommendations about “who to follow.” (A *Wired* magazine headline of three years ago comes to mind: “How Facebook Knows You Better than Your Friends Do.”⁶⁷) Intermediaries ostensibly do all of these things to improve user experiences, but their practices are often problematic and opaque to the outside world.⁶⁸ As very recent revelations involving Cambridge Analytica underscore, Facebook for years shared its unrivaled trove of user data with third-party researchers, application developers, and data brokers in the interest of deepening user engagement.⁶⁹ Facebook reportedly took 30 percent of developer profits in the process.⁷⁰

This is all to say that intermediaries now have near-total control of users’ online experience. They design and predict nearly everything that happens on their site, from the moment a user signs in to the moment she logs out. The lure of “big” consumer data pushes them to be ever more aggressive in their efforts to attract new users, retain existing users, and generate information about users that they can mine and market to others. It is neither surprising nor troubling that companies make handsome profits in this way. But these developments undermine any notion that online intermediaries deserve immunity because they are mere conduits for, or passive publishers of, their users’ expression. Online intermediaries pervasively shape, study, and exploit communicative acts on their services.

⁶³ See Frank Pasquale, *The Black Box Society* 3, 28–31 (2015).

⁶⁴ See Adam Alter, *Irresistible: The Rise of Addictive Technology and the Business of Keeping Us Hooked* (2017); Paul Lewis, *Our Minds Can Be Hijacked: The Tech Insiders Who Fear a Smartphone Dystopia*, *Guardian* (Oct. 6, 2017), <http://www.theguardian.com/technology/2017/oct/05/smartphone-addiction-silicon-valley-dystopia>. See generally Tim Wu, *The Attention Merchants: The Epic Scramble to Get Inside Our Heads* (2016); Nir Eyal, *Hooked: How to Build Habit-Forming Products* (2014).

⁶⁵ Noam Cohen, *Silicon Valley Is Not Your Friend*, *N.Y. Times* (Oct. 13, 2017), <http://www.nytimes.com/interactive/2017/10/13/opinion/sunday/Silicon-Valley-Is-Not-Your-Friend.html>.

⁶⁶ Julia Carrie Wong, *Facebook Overhauls News Feed in Favor of ‘Meaningful Social Interactions,’* *Guardian* (Jan. 11, 2018), <http://www.theguardian.com/technology/2018/jan/11/facebook-news-feed-algorithm-overhaul-mark-zuckerberg>.

⁶⁷ Issie Lapowsky, *How Facebook Knows You Better Than Your Friends Do*, *Wired* (Jan. 13, 2015), <http://www.wired.com/2015/01/facebook-personality-test>.

⁶⁸ See Christina Passariello, *Facebook: Media Company or Technology Platform?*, *Wall St. J.* (Oct. 30, 2016), <http://www.wsj.com/articles/facebook-media-company-or-technology-platform-1477880520>.

⁶⁹ See Matthew Rosenberg et al., *How Trump Consultants Exploited the Facebook Data of Millions*, *N.Y. Times* (Mar. 17, 2018), <http://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html>; see also Paul Lewis, *‘Utterly Horrifying’: Ex-Facebook insider Says Covert Data Harvesting Was Routine*, *Guardian* (Mar. 20, 2018), <http://www.theguardian.com/news/2018/mar/20/facebook-data-cambridge-analytica-sandy-parakilas>.

⁷⁰ Lewis, *supra* note 69.

All of this, moreover, belies the old faith that such services operate at too massive a scale to be asked to police user content. Online intermediaries are already carefully curating and commoditizing this content through automated “black box” processes that would seem unworkable were they not working so well. The standard justifications for broad immunity under Section 230—grounded in fears of imposing excessive burdens on intermediaries and chilling their distribution of lawful material—have become increasingly divorced from technological and economic realities. As intermediaries have figured out how to manage and distribute user data with ever greater precision, the traditional case for Section 230 immunity has become ever less compelling, if not altogether inapt.

B. Discriminatory Designs on User Content and Data: The Example of Online Housing Marketplaces

These developments in intermediary design have been underway for over a decade now and have become far-reaching and consequential enough in themselves to warrant rethinking of Section 230 doctrine. The problems with the doctrine, however, are made worse when intermediaries’ designs facilitate expressive conduct that harms vulnerable people and members of historically subordinated groups.⁷¹ We often hear about the dangerous content that intermediaries automatically distribute by algorithm, as in the notorious ways in which Facebook and Twitter facilitated the targeted dissemination of “fake news” in the months leading up to the 2016 presidential election,⁷² or the advertisement that Instagram made of a user’s personal photo of a violently misogynist threat she had received through her account.⁷³ My point here, however, is that the stakes of automated intermediary designs are especially high for certain predictable communities. Unpoliced, putatively neutral online application and service designs can entrench longstanding racial and gender disparities.

Consider Airbnb’s popular home-sharing service. Quite unlike Twitter’s liberal approach to personal attribution, Airbnb’s main service requires each guest to create an online profile with certain information, including a genuine name and phone number. It also encourages inclusion of a real photograph.⁷⁴ For Airbnb, the authenticity of this profile information is vital to the operation of the service, as it engenders a sense of trust and connection between hosts and guests. Guests’ physical characteristics may contain social cues that instill either familiarity and comfort, on the one hand, or suspicion and distrust, on the other. The sense of authentic connection that Airbnb is adamant about cultivating, however, has dangerous consequences in a market long plagued by discrimination against racial and ethnic minorities. In its more insidious manifestations, access to a guest’s name and profile picture affords hosts the ability to assess the trustworthiness of a guest based on illicit biases—against, say, Latinos or blacks—that do not accurately predict a prospective guest’s reliability as a tenant. In this way, Airbnb’s service directly reinforces discrimination when it requires users to share information that suggests their own race.

⁷¹ Cf. Solon Barocas & Andrew D. Selbst, *Big Data’s Disparate Impact*, 104 Calif. L. Rev. 671 (2016) (discussing ways in which algorithmic analysis and machine learning may produce discriminatory impacts); Levy & Barocas, *supra* note 45 (discussing ways in which intermediary designs may have discriminatory impacts).

⁷² See, e.g., Nancy Scola & Josh Meyer, *Twitter Takes Its Turns in the Russian Probe Spotlight*, Politico (Sept. 28, 2017), <http://www.politico.com/story/2017/09/28/twitter-russia-probe-spotlight-243239>.

⁷³ Sam Levin, *Instagram Uses ‘I Will Rape You’ Post as Facebook Ad in Latest Algorithm Mishap*, Guardian (Sept. 21, 2017), <http://www.theguardian.com/technology/2017/sep/21/instagram-death-threat-facebook-olivia-solon>.

⁷⁴ Airbnb also gives users the option of importing information from users’ Facebook accounts.

That race would matter so much to Airbnb hosts should not be a surprise. Race, after all, has long played an enormous—and pernicious—role in U.S. housing markets, online as well as offline. SketchFactor, the crowdsourced neighborhood safety rating application, for example, became little more than a platform for users to share racist stereotypes about “shady” parts of town.⁷⁵ Match.com, the ostensibly race-neutral online dating application, facilitates users’ discrimination against blacks.⁷⁶ Similarly, Airbnb hosts use the home-sharing service to discriminate against racial minorities whose identities as such are suggested in their profiles. Guests have complained publicly about this phenomenon, giving rise to the hashtag #AirbnbWhileBlack.⁷⁷ One guest reported that a host abruptly cancelled her reservation after sending an unambiguously bigoted explanation: “I wouldn’t rent to u if u were the last person on earth. One word says it all. Asian.”⁷⁸ Researchers at the Harvard Business School have substantiated individual claims like these, finding that Airbnb guests “with distinctively African-American names are 16 percent less likely to be accepted relative to identical guests with distinctively White names.”⁷⁹ Airbnb felt compelled to commission a well-regarded civil rights attorney to conduct a study on the topic. Her review, too, found a distinct pattern of host discrimination against users whose profiles suggest they are a member of a racial minority group.⁸⁰

The difference between these racially discriminatory patterns as they appear on Airbnb versus dating or neighborhood rating apps is that the former are illegal because they violate fair housing laws. The 1968 Fair Housing Act (FHA), for example, specifically forbids home sellers or renters, as well as brokers, property managers, and agents, from distributing advertisements “that indicate[] any preference, limitation, or discrimination based on race, color, religion, sex, handicap, familial status, or national origin.”⁸¹ States have similar laws. In light of the mounting evidence that hosts use its service to discriminate unlawfully, Airbnb has augmented its efforts to police discriminatory behavior by hosts. In addition to requiring users to forswear that practice, the company now also requires new users to agree “to treat everyone in the Airbnb community—regardless of their race, religion, national origin, ethnicity, skin color, disability, sex, gender identity, sexual orientation or age—with

⁷⁵ Andrew Marantz, *When an App Is Called Racist*, *New Yorker* (July 29, 2015), <http://www.newyorker.com/business/currency/what-to-do-when-your-app-is-racist>. See generally Anthony G. Greenwald & Linda Hamilton Krieger, *Implicit Bias: Scientific Foundations*, 94 *Calif. L. Rev.* 945 (2006); Jerry Kang & Kristin Lane, *Seeing through Colorblindness: Implicit Bias and the Law*, 58 *UCLA L. Rev.* 465 (2010).

⁷⁶ See Emanuella Grinberg, *When It Comes to Dating Sites, Race Matters*, *CNN* (Jan. 13, 2016), <http://www.cnn.com/2016/01/13/living/where-white-people-meet-feat/index.html>. This is to say nothing of sites like WhereWhitePeopleMeet that openly exploit this phenomenon. *Id.*

⁷⁷ See, e.g., Kristen Clarke, *Does Airbnb Enable Racism?*, *N.Y. Times* (Aug. 23, 2016), <http://www.nytimes.com/2016/08/23/opinion/how-airbnb-can-fight-racial-discrimination.html>; Carla Javier, *A Trump-Loving Airbnb Host Canceled This Woman’s Reservation Because She’s Asian*, *Splinter News* (Apr. 6, 2017), <http://splinternews.com/a-trump-loving-airbnb-host-canceled-this-womans-reserva-1794086239>; Carla Herreria, *Amsterdam Airbnb Host Accused of Pushing South African Down Stairs Is Arrested*, *Huffington Post* (July 13, 2017), http://www.huffingtonpost.com/entry/amsterdam-airbnb-host-pushes-guest-stairs-racist_us_59680a7de4b03389bb164286.

⁷⁸ Javier, *supra* note 78.

⁷⁹ Benjamin Edelman et al., *Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment*, 9 *Am. Econ. J.: Applied Econ.* 1, 2 (2017).

⁸⁰ Laura Murphy, Laura Murphy & Assocs., *Airbnb’s Work to Fight Discrimination and Build Inclusion: A Report Submitted to Airbnb* 16–17 (2016), http://blog.atairbnb.com/wp-content/uploads/2016/09/REPORT_Airbnbs-Work-to-Fight-Discrimination-and-Build-Inclusion.pdf.

⁸¹ 42 U.S.C. § 3604(c) (2012).

respect, and without judgment or bias.”⁸² Airbnb has also promoted its “instant bookings” service as an alternative to its main service.⁸³ “Instant bookings” does not require elaborate profiles (including racially suggestive names or pictures) to complete transactions.

However, Airbnb still facilitates discrimination through its main service to the extent that it continues to rely on names and pictures. The “instant bookings” feature, paired with the main service, creates a “two-tiered reservations system”: In one system (instant bookings), guests lose a sense of conviviality with hosts but obtain some peace of mind in knowing that they will not be discriminated against on the basis of race, while in the other system (the main service), discrimination is inevitable but also exploited to promote “authentic” connections.⁸⁴

Section 230 doctrine arguably insulates Airbnb’s design choices from antidiscrimination law’s scrutiny. The company and its defenders have routinely cited Section 230 as a protection against liability for a wide range of illicit host activities, including discrimination that violates fair housing laws.⁸⁵ In their view, the statutory immunity is robust enough to protect Airbnb from liability for these expressive acts by third-party hosts because the company only facilitates transactions between users.⁸⁶ It does not contribute anything material to the transactions themselves.⁸⁷

Airbnb is far from alone in deploying designs that routinely generate serious forms of discrimination. Late in 2016, ProPublica published the first in a series of illuminating reports on Facebook Ads, the social media company’s powerful microtargeted advertising platform. This service enables advertisers to customize campaigns to social media users based on the information that Facebook gathers about those users. Facebook Ads is a bargain (at a clip of \$30 for each advertisement) compared to the going rate of top social media marketing and advertising firms. It can be a great help to entrepreneurs of all sizes because it identifies salient market segments in real time.

Facebook Ads is also distinctive because the company employs software, first, to analyze the unrivaled troves of user data that it collects and, second, to create dozens of categories from which advertisers may choose. These include targeted classifications within geographic locations, demographics, friendship networks, and online user

⁸² Airbnb, *General Questions About the Airbnb Community Commitment*, <http://www.airbnb.com/help/article/1523/general-questions-about-the-airbnb-community-commitment> (last visited Feb. 23, 2018).

⁸³ Airbnb, *Business Is Better with Instant Book*, <http://www.airbnb.com/host/instant> (last visited Feb. 23, 2018).

⁸⁴ Katie Benner, *Airbnb Adopts Rules to Fight Discrimination by Its Hosts*, N.Y. Times (Sept. 8, 2016), <http://www.nytimes.com/2016/09/09/technology/airbnb-anti-discrimination-rules.html>. As Nancy Leong and Aaron Belzer have recently shown, moreover, the guest-rating systems on online platforms like Airbnb and Uber further entrench discrimination by aggregating illicit biases over time. See Nancy Leong & Aaron Belzer, *The New Public Accommodations: Race Discrimination in the Platform Economy*, 105 Geo. L.J. 1271, 1293–95 (2017).

⁸⁵ Tracey Lien, *Airbnb’s Legal Argument: Don’t Hold Us Accountable for the Actions of Our Hosts*, L.A. Times (June 29, 2016), <http://www.latimes.com/business/technology/la-fi-tn-airbnb-free-speech-20160629-snap-story.html>.

⁸⁶ *Id.*; see also Julia Carrie Wong, *How a Failed Attempt to Get Porn off the Internet Protects Airbnb from the Law*, Guardian (June 29, 2016), <http://www.theguardian.com/technology/2016/jun/29/airbnb-lawsuit-san-francisco-regulation-internet-porn>.

⁸⁷ On the other hand, Airbnb’s decision to settle in some of these cases may suggest that the company worries about its role in perpetuating discrimination, irrespective of whether Section 230 supplies immunity. Cf. Sam Levin, *Airbnb Gives in to Regulator’s Demand to Test for Racial Discrimination by Hosts*, Guardian (Apr. 27, 2017), <http://www.theguardian.com/technology/2017/apr/27/airbnb-government-housing-test-black-discrimination>.

behaviors.⁸⁸ Among the more notorious categories in the recent past were ones that “enabled advertisers to direct their pitches to the news feeds of almost 2,300 people who expressed interest in the topics of ‘Jew hater,’ ‘How to burn jews,’ or ‘History of “why jews ruin the world.””⁸⁹ No human at Facebook created these specific anti-Semitic classifications. Facebook’s algorithms determined that they were salient based on user interest at the time.⁹⁰

Facebook’s algorithms likewise seem to have created various controversial demographic classifications for “ethnic” or “multicultural” affinities, a category that does not connote race as such so much as users’ cultural associations and inclinations.⁹¹ These classifications are predictive proxies, however, for race and ethnicity. Recent news reports have shown that, through these classifications, Facebook Ads has enabled building managers and employers to exclude racial minorities from advertisements about apartment rentals and to exclude older people from advertisements about jobs.⁹² When faced with stories of discrimination on the advertising platform in late 2016, Facebook immediately announced a plan to stamp out the practice.⁹³ Among other things, Facebook now requires advertisers to certify that they do not discriminate in contravention of civil rights laws.⁹⁴ But, as with Airbnb, reports of illicit use of the site continue to surface.⁹⁵

Critics and victims of these practices would greatly prefer to seek relief and reform from the intermediary itself—from Facebook—rather than from thousands of individual users. Aggrieved parties have thus filed federal class action lawsuits against Facebook alleging fair housing and employment discrimination violations.⁹⁶ Predictably, Facebook has cited Section 230 to defend its advertising platform. It argues that the company does not control the reach or content of targeted ads; third-party advertisers do. According to Facebook, its platform is nothing more than a “neutral tool” to help these advertisers “target their ads to groups of users most likely to be interested in the goods or services being offered.”⁹⁷ This activity, it asserts, falls squarely in the category of “publishing” for which companies like Facebook are granted immunity under the CDA.

⁸⁸ See Facebook Business, *Facebook Ads*, <http://www.facebook.com/business/products/ads> (last visited Feb. 23, 2018).

⁸⁹ Julia Angwin et al., *Facebook Enabled Advertisers to Reach ‘Jew Haters,’* ProPublica (Sept. 14, 2017), <http://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters>.

⁹⁰ Jaclyn Peiser, *Anti-Semitism’s Rise Gives the Forward New Resolve*, N.Y. Times (Oct. 8, 2017), <http://www.nytimes.com/2017/10/08/business/media/the-forward-antisemitism.html>.

⁹¹ ProPublica, which first broke the story about this practice, see Angwin & Parris, *supra* note 14, has not reported on whether Facebook generates these categories manually or by algorithm. I do not take up the question here, but the roles of automation and machine learning raise difficult questions about proof of intention under current nondiscrimination law.

⁹² See Julia Angwin et al., *Facebook Job Ads Raise Concerns About Age Discrimination*,

N.Y. Times (Dec. 20, 2017), <http://www.nytimes.com/2017/12/20/business/facebook-job-ads.html>; Angwin & Parris, *supra* note 14.

⁹³ Sapna Maheshwari & Mike Isaac, *Facebook Will Stop Some Ads from Targeting Users by Race*, N.Y. Times (Nov. 11, 2016), <http://www.nytimes.com/2016/11/12/business/media/facebook-will-stop-some-ads-from-targeting-users-by-race.html>.

⁹⁴ Rachel Goodman, *Facebook’s Ad Targeting Problems Prove How Easy It Is to Discriminate Online*, NBC News (Nov. 30, 2017), <http://www.nbcnews.com/think/opinion/facebook-s-ad-targeting-problems-prove-how-easy-it-discriminate-ncna825196>.

⁹⁵ *Id.*

⁹⁶ See, e.g., Complaint, *Mobley v. Facebook, Inc.*, No. 5:16-cv-06440-EJD (N.D. Cal. Nov. 3, 2016), 2016 WL 6599689. cause substantial emotional distress.” 18 U.S.C § 2261A(2)(B).

⁹⁷ Defendant’s Notice of Motion and Motion to Dismiss First Amendment Complaint; Memorandum of Points and Authorities in Support Thereof at 10, *Mobley v. Facebook, Inc.* (N.D. Cal.) (June 1, 2017) (No. 5:16-cv-06440-EJD), available at <http://assets.documentcloud.org/documents/4333515/Outten-FB-FB-Motion-to-Dismiss-4-3-17.pdf>. Facebook also asserts that the plaintiffs lack standing and that, in any event, it is not discriminating within the meaning of the pertinent civil rights laws. *Id.* at 14–25.

C. Doctrinal Responses—and Resources

Section 230 doctrine could very well lead courts to side with Facebook on this matter. But it is hardly obvious that it should, given that the alleged discrimination would not be possible but for the way in which Facebook leverages its unrivaled access to social media user data to generate the illicit categories. In Facebook’s favor, courts have read Section 230 to immunize intermediaries that host racially discriminatory advertisements or solicitations. In 2008, the U.S. Court of Appeals for the Seventh Circuit explained that the popular classifieds site Craigslist could not be held liable for hosting third-party housing advertisements that overtly expressed preferences for people on the basis of race, sex, religion, sexual orientation, and family status.⁹⁸ The panel explained that Congress enacted the statute to protect services exactly like Craigslist. The company neither had a hand in the authorship of the discriminatory advertisements nor caused or induced advertisers to post such content.⁹⁹ Craigslist, the panel reasoned, acts as nothing more than a publisher of (sometimes racist) user content and, as such, could not be liable under federal fair housing law.¹⁰⁰ Had Congress meant to include an exception under Section 230 for such laws, it would have said so.¹⁰¹

But the Section 230 case law also contains some resources and opportunities for plaintiffs like those in the current Facebook Ads case. In the same year that the Seventh Circuit ruled in favor of Craigslist, the Ninth Circuit sitting *en banc* held that an important design element of Roommates.com, a website that also brokers connections between people in the housing market, was not immune under Section 230.¹⁰² As a condition of participation on the site, Roommates.com required subscribers to express preferences that are strictly forbidden under fair housing law.¹⁰³ Among other things, the site’s developers designed a dropdown menu that listed gender, sexual orientation, and family status as potential options. (Notably, the menu did not include race among the listed items.) A participant had to share such a preference to find a match. The Ninth Circuit held that this design feature “materially contributed” to a fair housing law violation every time a user expressed a preference for one of those prohibited classifications.¹⁰⁴ This conclusion flowed from language in Section 230 that does not extend protection to intermediaries that help to “create or develop” illicit third-party content.¹⁰⁵

As important as the *Roommates.com* opinion has become in limiting the scope of immunity under Section 230, it is worth noting that the Ninth Circuit was very careful in how it discussed its holding. The court made a point of limiting its no-immunity conclusion to the dropdown menu. The plaintiffs had argued that a separate,

⁹⁸ *Chicago Lawyers’ Comm. for Civil Rights v. Craigslist, Inc.*, 519 F.3d 666 (7th Cir. 2008).

⁹⁹ *Id.* at 671.

¹⁰⁰ *Id.*

¹⁰¹ *Id.*

¹⁰² *Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157 (9th Cir. 2008) (*en banc*).

¹⁰³ *Id.* at 1165–72.

¹⁰⁴ *Id.*

¹⁰⁵ 47 U.S.C. § 230(f)(3) (2012). After remand, a three-judge panel did nothing to alter this conclusion in its ruling four years later. *Fair Hous. Council of San Fernando Valley v. Roommate.com, LLC*, 666 F.3d 1216 (9th Cir. 2012). In this later opinion, the panel held that, while the immunity under Section 230 did not bar the suit against Roommates.com for its drop-down menu, Roommates.com’s specific conduct at issue did not violate the FHA because “the FHA doesn’t apply to the sharing of living units” as opposed to “the sale or rental of a dwelling.” *Id.* at 1222 (discussing the scope of 42 U.S.C. § 3604(c)).

blank dialogue box that Roommates.com makes available to subscribers also permits them to express bigoted preferences and share information in violation of fair housing law.¹⁰⁶ For example, subscribers had posted comments that they “prefer white Male roommates,” that “the person applying for the room MUST be a BLACK GAY MALE,” or that they are “NOT looking for black muslims.”¹⁰⁷ The court held that Section 230 immunizes Roommates.com from liability for statements like these. It is not enough, the court reasoned, that the site encourages subscribers to share preferences and information, as this is “precisely the kind of situation for which section 230 was designed to provide immunity.”¹⁰⁸ Roommates.com only “passively displayed” the statements and had “no way to distinguish unlawful discriminatory preferences from perfectly legitimate statements.”¹⁰⁹ This conclusion jibes with the Seventh Circuit’s approach to Craigslist.¹¹⁰ Indeed, these two opinions neatly mapped out the basic contours of Section 230 doctrine when they were decided in 2008. The *Roommates.com* opinion, in particular, is now routinely cited as authority for the “material contribution” standard.¹¹¹

The Ninth Circuit’s other notable conclusion in that case, decided a couple of years after a post-remand trial court finding for Roommates.com, was that the plaintiff civil rights organization, the Fair Housing Council of the San Fernando Valley (FHC), had standing to seek relief even if it was not itself the victim of a discrete discriminatory act.¹¹² FHC had alleged that Roommates.com was strictly liable for designing its site in a way that discriminated against prospective renters. It claimed standing to sue, however, because its research into the company’s discriminatory designs was a drain on its resources and frustrated its mission.¹¹³ The Ninth Circuit agreed, holding that FHC had suffered an actual injury sufficient to have standing.

In essence, the court determined that the organization could stand in for a hypothetical Roommates.com subscriber who would be harmed by users’ discriminatory preferences and postings.¹¹⁴

This holding makes good sense, as discriminatory targeted advertisements and solicitations subjugate racial minorities even when their victims do not witness or otherwise experience the discriminatory act directly. Civil rights laws often reach beyond discrete acts of exclusion in order to redress systemic patterns of subordination and exclusion. Roommates.com’s design choices, FHC had argued, facilitated communicative acts of discrimination in a market long plagued by that very problem. And if not for FHC’s intervention, the court reasoned, these patterns of bias would continue.

¹⁰⁶ *Roommates.com*, 521 F.3d at 1173.

¹⁰⁷ *Id.*

¹⁰⁸ *Id.* at 1174.

¹⁰⁹ *Id.*

¹¹⁰ *See id.* at 1173 n.33 (explaining that the court’s holding is consistent with the Seventh Circuit’s *Craigslist* opinion).

¹¹¹ *See, e.g., Jones v. Dirty World Entm’t Recordings*, 755 F.3d 398, 410–12 (6th Cir. 2014); *FTC v. Accusearch*, 570 F.3d 1187, 1200 (10th Cir. 2009).

¹¹² *Fair Hous. Council of San Fernando Valley v. Roommate.com, LLC*, 666 F.3d 1216, 1219 (9th Cir. 2012).

¹¹³ *Id.*

¹¹⁴ *Id.*

IV. Toward a More Nuanced Immunity Doctrine

The *Roommates.com* opinion, issued a decade ago, helps to show the way forward. The Ninth Circuit’s careful treatment of the two contested features of the website design of Roommates.com demonstrated an appreciation for the diversity of ways in which the company elicits content from users, and its standing ruling demonstrated an appreciation for the realities of civil rights harms.

However, the Ninth Circuit’s opinion did not go far enough; it did not address the increasingly subtle and tentacular kinds of control that online intermediaries exert over users’ experiences today. The system through which Facebook, for example, algorithmically sorts and repurposes user data to support microtargeted advertising is a far cry from the clumsy dropdown menu in the *Roommates.com* case. Two decades after the CDA’s enactment, it has become increasingly implausible to equate this powerful manipulation of users’ data and content with traditional publishing under Section 230.

Section 230 doctrine must be adapted to the political economy of contemporary online information flows. Judges and litigants already have a rich set of tools from antidiscrimination and consumer protection law for determining liability and providing remedies for harmful expressive conduct. But the current Section 230 doctrine cuts cyberspace off from these other bodies of law, foreclosing liability analysis for companies whose service designs routinely facilitate or even encourage illicit content.

It is important to emphasize, moreover, that holding intermediaries to account for such designs does not require anything like strict liability for the harms caused by nonconsensual pornography or any other user-generated content. Consistent with the neglected Good Samaritan goal of the statute, Section 230 can quite comfortably be interpreted to provide a safe harbor for intermediaries that try in good faith to block or take such content down.¹¹⁵ That is, after all, precisely what the text of Section 230(c)(2)(A) says, at least with regard to “objectionable” speech.¹¹⁶ At the same time, courts could allow plaintiffs to seek redress from intermediaries that knowingly or negligently facilitate the distribution of harmful content. As the Ninth Circuit’s ruling against Roommates.com shows, we do not need new statutory language to assess intermediary liability when the user interface at issue enables illegal online conduct.

But the experience of two decades of Section 230 litigation does suggest that new statutory language could help, particularly since the prevailing view prevents the plain meaning of the Good Samaritan title and Section 230(c)(2)(A) from doing any meaningful work. The statute itself, moreover, fails to give clear direction on the kinds of torts it covers. Nor, for that matter, does the statute address the extent to which a defendant must “create[] or develop[]”

¹¹⁵ See Facebook Business, *Take the Work out of Hiring*, <http://www.facebook.com/business/news/take-the-work-out-of-hiring> (last visited Feb. 23, 2018).

¹¹⁶ See 47 U.S.C. § 230(c)(2)(A) (2012) (“No provider or user of an interactive computer service shall be held liable on account of . . . any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected.”).

the offending material.¹¹⁷ This has been left to the courts to sort out. Distressed by the wide scope of the doctrine and some of these textual gaps, legislators and activists have been promoting amendments to Section 230 that would create exceptions for prostitution, nonconsensual pornography, and the sex trafficking of minors.¹¹⁸ There is no reason why Congress couldn't also write in an explicit exception to Section 230 immunity for violations of civil rights laws.

Such proposals will face substantial pushback from intermediaries and others.¹¹⁹ A company like Facebook, for example, has a lot to lose from any change that would require it to be more careful about how it distributes user content or generates personal or targeted advertisements.¹²⁰ Even a shift to what some are now calling “contextual advertising,” where an advertiser buys the context in which social media users engage with each other rather than individual users’ profiles, could cost a company like Facebook billions of dollars.¹²¹ And to be sure, apart from the commercial interests at stake, there are important free speech arguments for keeping Section 230 broad: The content and data flowing through the online speech environment may not be as abundant in a world in which intermediaries are held to account for their users’ content and their own designs on user data. But then again, it is difficult to weigh this “chilling” concern against the chilling of members of historically subordinated groups that is already happening under existing law.¹²²

Whether legal reform in this area takes place in the legislature or the judiciary or both, reform is necessary. Judges, lawyers, and legislators should stop shielding intermediaries from liability on the basis of implausible assumptions about their neutrality or passivity—and should instead start looking carefully at how intermediaries’ designs on user content do or do not result in actionable injuries. This attention to design will further sensitize intermediaries to the ways in which their services perpetuate systemic harms. Equipped with a more nuanced approach to intermediary immunity, we might come to expect an online environment that is hospitable to all comers.

¹¹⁷ *Id.* § 230(f)(3). See generally Sylvain, *Design Duties*, *supra* note 21, at 239–42.

¹¹⁸ See Stop Enabling Sex Traffickers Act of 2017, S. 1693, 115th Cong. (2017); Allow States and Victims to Fight Online Sex Trafficking Act of 2017, H.R. 1865, 115th Cong (2017).

¹¹⁹ See, e.g., Internet Ass’n, *Intermediary Liability*, <http://internetassociation.org/positions/intermediary-liability> (last visited Feb. 23, 2018); see also Electronic Frontier Found’n, *Stop SESTA: Congress Doesn’t Understand How Section 230 Works* (Sept. 7, 2017), <http://www.eff.org/deeplinks/2017/09/stop-sesta-congress-doesnt-understand-how-section-230-works>.

¹²⁰ John Battelle, *Facebook Can’t Be Fixed*, NewsCo (Jan. 5, 2018), <http://shift.newco.co/its-the-advertising-model-stupid-b843cd7edbe9>.

¹²¹ *Id.*

¹²² It is also difficult to disentangle this free speech argument from the intermediaries’ commercial interests. European regulators, for instance, fined Google almost two and a half billion Euros last summer for abusing its market dominance in search to give “an illegal advantage to another Google product.” European Commission, Press Release, *Antitrust: Commission Fines Google €2.42 Billion for Abusing Dominance as Search Engine by Giving Illegal Advantage to Own Comparison Shopping Service* (June 27, 2017), http://europa.eu/rapid/press-release_IP-17-1784_en.htm.