

# REGULATING THE SOCIAL PUPPETEERS: § 230 & MARGINALIZED SPEECH

KYLIE THOMPSON\*

INTRODUCTION.....	462
I. WHAT IS § 230?.....	465
A. <i>Legislative History</i> .....	465
B. <i>Expansive Scope</i> .....	466
C. <i>Threading the § 230 needle</i> .....	468
II. THE POWER OF SOCIAL PLATFORMS .....	471
A. <i>Harassment &amp; Disinformation</i> .....	471
B. <i>Content Moderation Practices</i> .....	474
C. <i>Whose Content is it Anyway?</i> .....	477
III. REFORMING § 230 .....	478
CONCLUSION .....	481

*This paper addresses social platforms’ immunity under § 230 of the Communications Decency Act (CDA) in light of recent scholarship on content moderation and curation practices. I first became interested in this topic when I heard Danielle Citron speak about it at the Silicon Flatirons’ Internet Platforms conference in the spring of 2019. Since then, § 230 has been increasingly scrutinized by both democrats and republicans. What was intended to encourage platforms to moderate content has since been employed to grant platforms broad immunity, allowing platforms to refrain from any moderation despite the legislature’s intent. Many see § 230 as the lifeblood of the Internet, and fear that collateral censorship will generate more harm without it. But many see it as an unnecessary shield for harms perpetrated online under the cloak of anonymity, disproportionately impacting marginalized speech. Views of § 230 on all sides are entangled with the complexities of First Amendment doctrine. This note seeks to highlight how the modern internet enshrines hierarchies of power and control, and argues that*

---

\* J.D. University of Colorado, Class of 2021; B.A. University of Dayton, Class of 2018. Special thanks to the CTLJ members for their work on this note, and to my professors for their support and feedback along the way.

*platforms should not be afforded broad, unfettered, and unprecedented immunity while they obscure machinations that perpetuate inequality and reap the benefits of chaos on the platform. It assesses paths to reforming § 230 in light of platforms' content moderation and curation practices.*

## INTRODUCTION

The Internet is different than any technology we've encountered before: different in its reach and speed of dissemination; different in its pervasive and ubiquitous presence; different in that it is not the Internet of 1996 anymore. Since 1996, intermediaries have evolved from the bulletin boards of service providers into innovative platforms that allow users to connect and share with each other for “free” and on a global scale.<sup>1</sup> The market therefore changed from one in which revenues were generated by traditional models of profit with the end user as the customer, to an attention market in which the advertiser is the customer and the end user becomes the product.<sup>2</sup>

While the Internet has changed global interactions, the use of discrimination on the basis of gender, sexual identity, and race remains embedded in cultural values as a tool of division and control.<sup>3</sup> Women, BIPOC, and gender non-conforming individuals are disproportionately subject to the worst online abuses like impersonation, doxing, stalking,<sup>4</sup> revenge porn,<sup>5</sup> sexual assault threats, and blackmail, to name a few.<sup>6</sup> While it's no shock that gender-based, sexuality-based, and race-based discrimination are

---

1. See generally TIM WU, *THE ATTENTION MERCHANTS: THE EPIC SCRAMBLE TO GET INSIDE OUR HEADS* 5–6 (2016) (describing the evolution of rapid commercialization, advertisement, and social media and its effects on daily lives and the economy).

2. *Id.* at 335–36.

3. See, e.g., DANIELLE KEATS CITRON, *HATE CRIMES IN CYBERSPACE* 73–80 (Harvard U. Press, 2014).

4. See Soraya Chemaly, *There's No Comparing Male and Female Harassment Online*, TIME (Sept. 9, 2014, 10:55 AM), <https://time.com/3305466/male-female-harassment-online/> [<https://perma.cc/WWA7-P2KG>] (“70 percent of those stalked online are women. More than 80 percent of cyber-stalking defendants are male.”).

5. See *id.* (“[A] study of 1,606 revenge porn cases showed that 90% of those whose photos were shared were women, targeted by men.”).

6. See *id.*; see also RUHA BENJAMIN, *RACE AFTER TECHNOLOGY* 23 (Polity Press, 2019) (providing an example of racial targeting online, “[White nationalists] are especially fond of Twitter and use it to spread their message, grow their network, disguise themselves online, and generate harassment campaigns that target people of color, especially Black women.”).

expressed online,<sup>7</sup> it is shocking to realize how these harms are transformed by the ecosystem of cyberspace. The immediate, widespread, and permanent nature of the Internet exacerbates identity-based harms, making them more difficult to remedy.<sup>8</sup> On top of that, platforms might be incentivized to ignore online abuse that otherwise generates traffic and engagement on the platform.<sup>9</sup>

This note raises concerns with social media platforms in particular because of their unique influence over end users and the platforms' economic incentives to maintain user attention. These social platforms, such as Facebook and Twitter, have employed algorithmic designs that tap into human psychology in order to better tailor content to particular users and generate more interest in the platform—essentially, to make users addicted to the site.<sup>10</sup> These websites' algorithmic designs exacerbate discrimination harms and can chill expression because social values that perpetuate inequities are embedded in the design.<sup>11</sup> If destructive and harmful content gets attention, then platforms have an interest in keeping that content.<sup>12</sup>

Social platforms have begun to create complex content moderation rule sets enforced by human moderators and artificial intelligence (AI) in response to external pressures.<sup>13</sup> Outside

---

7. See generally *Online violence: Just because it's virtual doesn't make it any less real*, GLOBAL FUND FOR WOMEN, <https://www.globalfundforwomen.org/online-violence-just-because-its-virtual-doesnt-make-it-any-less-real/> [<https://perma.cc/LR4U-Y232>].

8. See generally *WMC Speech Project: Online Abuse 101*, WOMEN'S MEDIA CTR., <https://www.womensmediacenter.com/speech-project/online-abuse-101/> [<https://perma.cc/C2SF-6K4G>] (explaining kinds of online abuse from a civil rights perspective).

9. See Danielle Keats Citron & Mary Anne Franks, *The Internet as a Speech Machine and Other Myths Confounding Section 230*, 2020 U. OF CHI. LEGAL F. 45, 53–54 (2020) (“Yet the online advertising business model continues to incentivize revenue-generating content that causes significant harm to the most vulnerable among us. Online abuse generates traffic, clicks, and shares because it is salacious and negative. Deep fake pornography sites as well as revenge porn and gossip sites thrive thanks to advertising revenue.”).

10. See generally *THE SOCIAL DILEMMA* (Argent Pictures 2020) (the documentary-drama hybrid explores the dangerous human impact of social networking); see also, Olivier Sylvain, *Discriminatory Designs on User Data*, KNIGHT FIRST AMENDMENT INST. (Apr. 1, 2018), <https://knightcolumbia.org/content/discriminatory-designs-user-data> [<https://perma.cc/BS2Z-8SGH>] (“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.”).

11. See Sylvain, *supra* note 10; see also SAFIYA UMOJA NOBLE, *ALGORITHMS OF OPPRESSION 1* (N.Y. U. Press 2018) (arguing that search engines reinforce racism through algorithms).

12. See Citron & Franks, *supra* note 9, at 46.

13. *Id.* at 53 (“What often motivates [banning, filtering, and blocking decisions] is pressure from the European Commission to remove hate speech and terrorist activity.

influence in moderation decisions further underscore the undesirable power imbalance that exists between individual user and powerful state and private actors. While content moderation decisions are to some extent grounded in First Amendment principles,<sup>14</sup> social platforms are not bound by the First Amendment and thus platforms have ultimate say over content on their sites.<sup>15</sup> This note accepts that platforms are in the best position to address harms given the complexity of the cyberspace ecosystem and because platforms control the means and method of communication on their service.

However, platforms cannot be relied upon to make the best decision for consumers because of economic incentives to maximize profits. While platforms should not be liable for any and all content as a newspaper would be, they also should not receive sweeping immunity for all content provided by third parties on the platform, especially when the platforms have the power to make a difference in the lives of the vulnerable individuals from whom they profit.

This note draws on recent literature to highlight how social platforms and the immunity § 230 provides them enshrine hierarchies of power that disproportionately impact marginalized speech. It assesses why platforms' obscure machinations that perpetuate harassment and inequality should lead to the conclusion that § 230 needs reform. Section I provides necessary background: it maps out CDA § 230, explaining the legislative history and the scope of the Act. Following is a discussion of why reform is necessary to preserve the intent of the legislation. Section II explains the problems of harassment and disinformation online and assesses cyber civil rights in light of recent scholarship on platforms' content moderation and content curation practices. Section III argues for a potential avenue to reform the language of § 230, drawing on and augmenting Citron and Wittes' proposal to condition immunity on reasonableness.<sup>16</sup>

---

The same companies have banned certain forms of online abuse...in response to pressure from users, advocacy groups, and advertisers. They have expended resources to stem abuse when it has threatened their bottom line.”)

14. Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1621 (2018) (“American lawyers trained and acculturated in American free speech norms and First Amendment law oversaw the development of company content-moderation policy. Though they might not have ‘directly imported First Amendment doctrine,’ the normative background in free speech had a direct impact on how they structured their policies.”).

15. See *State Action Requirement*, LEGAL INFO. INS., [https://www.law.cornell.edu/wex/state\\_action\\_requirement](https://www.law.cornell.edu/wex/state_action_requirement) [https://perma.cc/5JU5-CGEQ].

16. See generally Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity*, 86 FORDHAM L. REV. 401 (2017).

## I. WHAT IS § 230?

To make sense of the debate surrounding § 230, it is first important to understand what the law does and why it was enacted. Section A provides some background on the legislative history surrounding § 230; and Section B addresses the broad scope given the provision through court interpretation. Section C describes the important strands in the debate and threads the needle between repeal and preservation. It recognizes that § 230 reflects imbalances in the First Amendment and thus should be reformed.

### A. *Legislative History*

Section 230 is the only remaining provision of the CDA, which was passed as part of an effort to curtail pornography on the Internet.<sup>17</sup> The other provisions of the Act were struck down in violation of the First Amendment,<sup>18</sup> but § 230 remains codified in Title V of the Telecommunications Act of 1996.<sup>19</sup> This note concerns section (c) of the provision, which is the meat and potatoes of the legislation.<sup>20</sup> It grants safe harbor to providers and users of “interactive computer services” from publisher liability for information provided by a third party.<sup>21</sup> It also immunizes platforms from liability for taking actions to moderate in “good faith.”<sup>22</sup>

A few months after Senator Exon’s CDA proposal, Stratton Oakmont, a securities investment banking firm, successfully sued Prodigy, an online bulletin board, for libel.<sup>23</sup> The court treated Prodigy as a publisher rather than a distributor and held that because Prodigy had made efforts to censor some third-party content, it was strictly liable as a publisher for all content.<sup>24</sup> This

---

17. John Bergmayer, *What Section 230 Is and Does – Yet Another Explanation of One of the Internet’s Most Important Laws*, PUB. KNOWLEDGE (May 14, 2019), <https://www.publicknowledge.org/blog/what-section-230-is-and-does-yet-another-explanation-of-one-of-the-internets-most-important-laws/> [https://perma.cc/ZPC3-AL4X] (“After all, this is why it was enacted as part of the Communications Decency Act, most of the rest of which was struck down as unconstitutional, but which was broadly aimed at scrubbing the internet of porn.”).

18. *Id.*

19. 47 U.S.C. § 230 (2018); *see also* Blake Reid, *Section 230 of...what?*, BLAKE.E.REID (Sept. 4, 2020), <https://blakereid.org/section-230-of-what/> [https://perma.cc/XK5B-SQCD].

20. *See* 47 U.S.C. § 230(c) (2018).

21. *See id.* § 230(c)(1).

22. *See id.* § 230(c)(2).

23. *See* Stratton Oakmont, Inc. v. Prodigy Servs. Co., 1995 WL 323710 (N.Y. Sup. Ct. May 24, 1995).

24. *Id.* at \*5.

decision followed a prior case, *Cubby, Inc. v. CompuServe Inc.*, 776 F. Supp. 135 1991, where the court found that CompuServe, an internet service provider (ISP) that refrained from moderating its bulletin board, was like a public library, and therefore more like a distributor than a publisher.<sup>25</sup> The *Cubby* court's treatment of CompuServe as a distributor meant liability turned on evidence that the intermediary had knowledge of the content at issue.<sup>26</sup> Thus, under *Cubby*, a platform that takes a hands-off approach to moderation would not be held liable under a distributor standard so long as it didn't have knowledge of the specific content that created the harm;<sup>27</sup> however, under *Prodigy*, any content moderation would give rise to strict publisher liability.<sup>28</sup> As a result of the *Cubby* and *Stratton Oakmont* cases, online platforms were incentivized to refrain from moderating content on platforms rather than walk a tightrope between distributor and publisher standards, and to instead turn a blind eye in order to avoid potential liability.<sup>29</sup>

House Representatives Chris Cox and Ron Wyden saw the decisions in the *Cubby* and *Stratton Oakmont* rulings as nonsensical because the intermediaries essentially provided the same services, but Prodigy had attempted to moderate third-party content.<sup>30</sup> Representatives Cox and Wyden thought that the *Stratton Oakmont* ruling would disincentivize investment in the tech sector for fear of being held liable for the content that others put online.<sup>31</sup> So they teamed up to write an amendment to the CDA—§ 230—that was intended to foster growth in the technology sector and to encourage “interactive computer services” to engage in content moderation without fear of liability.<sup>32</sup>

### B. *Expansive Scope*

Several cases followed the adoption of § 230 that defined the scope of the law and etched broad immunity for platforms. Both *Zeran* and *Blumenthal* involved defamation suits against AOL in which AOL received summary judgment; however, each court dealt

---

25. See Bergmayer, *supra* note 17 (quoting the court in *Cubby*, “[a] lower standard of liability to an electronic news distributor such as CompuServe than that which is applied to a public library, bookstore, or newsstand would impose an undue burden on the free flow of information.”).

26. See *id.*; see also JEFF KOSSEFF, THE TWENTY-SIX WORDS THAT CREATED THE INTERNET 42 (2019).

27. See KOSSEFF, *supra* note 26, at 42–43.

28. *Id.* at 52.

29. See *Id.* at 52–56.

30. *Id.* at 59.

31. *Id.* at 60.

32. *Id.* at 61.

with the § 230 immunity claim in different ways.<sup>33</sup> The Fourth Circuit in *Zeran* framed its decision in terms of policy, claiming it would be an “impossible burden” for online service providers to look into every defamation claim and make an accurate determination as to the merits of the claim or risk being held liable as a distributor.<sup>34</sup> In contrast, the district court for D.C. in *Blumenthal* felt bound to the text of § 230 despite skepticism over the practical implications of the statute in context.<sup>35</sup> These cases illustrate how the text of § 230 has been interpreted and applied by courts to protect intermediaries in the face of defamation claims. This is despite factual circumstances in both cases showing the intermediaries’ potential role in the harm, i.e., delayed removal and promotion of harmful content, respectively.

In contrast, the *Roommates.com* decision signaled some limit on the breadth of § 230.<sup>36</sup> Roommates.com provided a platform service that connected potential roommates through individual profiles. The court held that Roommates.com was not immune from liability under § 230 because it had a hand in crating the discriminatory content at issue—it developed a discriminatory questionnaire as a condition of service that was subsequently used to conduct a filtering process based on subscribers’ answers.<sup>37</sup> The court remarked that § 230 “was not meant to create a lawless no-man’s-land on the Internet.”<sup>38</sup>

More recently, the Supreme Court likened the Internet to the modern public square for purposes of First Amendment doctrine in *Packingham*.<sup>39</sup> Although that case concerned government regulation of speech, it raised alarm that the Court could potentially apply that analogy to cases concerning platform liability in the future, which would mean that platforms could be likened to public utilities and thus subject to First Amendment constraints that are traditionally reserved to government actors.<sup>40</sup> This

---

33. See *Zeran v. America Online, Inc.*, 129 F.3d 327 (4th Cir. 1997); but see *Blumenthal v. Drudge*, 992 F.Supp. 44 (D.D.C. 1998).

34. *Zeran*, 129 F.3d at 332–33.

35. *Blumenthal*, 992 F.Supp. at 51.

36. See *Fair Hous. Council of San Fernando Valley v. Roommates.com, L.L.C.*, 521 F.3d 1157 (9th Cir. 2008).

37. *Id.* at 1164 (“By requiring subscribers to provide the information as a condition of accessing its service, and by providing a limited set of pre-populated answers, Roommate becomes much more than a passive transmitter of information provided by others; it becomes the developer, at least in part, of that information.”).

38. *Id.*

39. *Packingham v. North Carolina*, 137 S. Ct. 1730, 1737 (2017).

40. *Id.*; see, e.g., *Packingham v. North Carolina*, 131 HARV. L. REV. 233, 233 (Nov. 10, 2017), <https://harvardlawreview.org/2017/11/packingham-v-north-carolina/> [https://perma.cc/QD7R-X4C3].

decision underscores society's changing conception of the pivotal role that platforms assume in modern communication and engagement.

### C. *Threading the § 230 needle*

Many supporters of § 230 assert that the Internet would not have been able to develop into what it is, or continue to exist as we know it today, without § 230.<sup>41</sup> This narrative thrives on the notion that we should be afraid of what the Internet will look like or become if § 230 is revised, or worse, repealed in its entirety. Some who contend that § 230 should be repealed argue that tort law will provide enough protection for platforms in the absence of immunity.<sup>42</sup> A traditional publication tort requires that fault amounting to negligence be shown, meaning the plaintiff must prove causation to recover.<sup>43</sup> This means platforms will not automatically be held liable for third-party tort harms without § 230.<sup>44</sup> But § 230 singles out the Internet as special and provides it more protections than other traditional media sources, and more protections than any other industry in our history.<sup>45</sup>

The argument for returned reliance on tort has intuitive appeal; however, a total repeal of § 230 could lead to negative consequences. Without § 230, the risk of returning to the messy distributor/publisher distinction at work in both the *Cubby* and *Stratton Oakmont* cases will likely incentivize platforms to either refrain from content moderation altogether or to over-moderate and engage in collateral censorship.<sup>46</sup> Rather than risk being held strictly liable as a publisher, platforms might step back and wash their hands of moderating any content without the protections of §

---

41. KOSSEFF, *supra* note 26, at 9; *but see* Carrie Goldberg (@cagoldberglaw), TWITTER (Dec. 30, 2020, 1:43 PM), <https://twitter.com/cagoldberglaw/status/1344383688507879426?s=20> [<https://perma.cc/J7F2-DUMD>] (“Section 230 did not create the internet as we know it. The shift from subscription based profit models to ‘free’ user data-mining and advertising profit models is what created the internet. It’s when the user stopped being the customer and started being the commodity.”).

42. *Recode Decode: CDA 230*, DECODER WITH NILAY PATEL (Aug. 23, 2019), <https://www.podchaser.com/podcasts/decoder-with-nilay-patel-100800/episodes/recode-decode-cda-230-43792630> [<https://perma.cc/UC2H-NM9A>].

43. *Id.*

44. *Id.*

45. *Id.*

46. *See* Bergmayer, *supra* note 17 (arguing that case law was underdeveloped before Section 230, as evidenced by *Cubby* and *Stratton Oakmont* and that “[s]imple repeal could lead to unmoderated cesspools on the one hand, and responsible platforms beset by lawsuits and crippled by damages on the other”).

230.<sup>47</sup> This is not desirable because, as victims' rights lawyers, cyber civil rights activists, and others have made clear, platforms are in the best position to address and moderate harmful content.<sup>48</sup> Alternatively, platforms might begin to censor more content than desirable or stop allowing users to contribute content altogether because of the risk that a platform will be held liable under a distributor standard for the inevitable mistakes made in choosing which content to take down.<sup>49</sup> Admittedly, this publisher/distributor distinction is only relevant in cases that pertain directly to speech and involve publication tort suits.<sup>50</sup> The problem underlying calls for repeal and reform is that § 230 jurisprudence has expanded protection beyond this narrow realm.<sup>51</sup> Still, it is better that the legislature speak to how courts should proceed than to leave it up to courts to decide the publication status of the most powerful corporations of our age.

Supporters often contend that § 230 should be preserved in its entirety because it appropriately balances the desire for platforms to moderate and the reality that platforms make filtering mistakes, against the concern that collateral censorship will occur without immunity.<sup>52</sup> As the argument goes, Congress made the following judgment in enacting § 230: “[t]he mistakes caused by liability are worse than the mistakes caused by immunity”<sup>53</sup> and thus platforms should be allowed the freedom to adopt moderation schemes at will without fear of liability. This argument succumbs to “the gravitational pull of the First Amendment” and asserts that all

---

47. *See id.*

48. *See Recode Decode: CDA 230, supra* note 42.

49. *But see* Carrie Goldberg (@cagoldberglaw), TWITTER (May 12, 2018, 3:26 PM), <https://twitter.com/cagoldberglaw/status/995415010678624257> [<https://perma.cc/E2L3-SE9E>] (arguing that platforms could buy insurance to avoid the projected financial burdens from increased tort liability).

50. *See generally* Bergmayer, *supra* note 17 (describing the previously discussed leading cases against online platforms before Section 230, one treating the platform as more akin to a publisher or speech and the other treating them as a distributor of speech).

51. Citron & Franks, *supra* note 9, at 59 (“When ‘courts routinely interpret Section 230 to immunize all claims based on third-party content,’ –including civil rights violations; ‘negligence; deceptive trade practices, unfair competition, and false advertising; the common law privacy torts; tortious interference with contract or business relations; intentional infliction of emotional distress; and dozens of other legal doctrines’ –they go far beyond existing First Amendment doctrine, and grant online intermediaries an unearned advantage over offline intermediaries.”).

52. *See, e.g.,* James Grimmelman, *To Err Is Platform*, KNIGHT FIRST AMENDMENT INST. (Apr. 6, 2018), <https://knightcolumbia.org/content/err-platform> [<https://perma.cc/Q9QU-SPPM>].

53. *Id.*

ensorship is necessarily bad and will result in less speech.<sup>54</sup> However, the First Amendment is not absolute and regulation might even serve to produce more speech—a greater diversity of speech—not less speech.<sup>55</sup> Franks argues that the First Amendment “has created a free speech dystopia in which only the powerful are truly at liberty to speak and the pursuit of truth has been rendered virtually impossible.”<sup>56</sup> Thus, the concern that liability will result in collateral censorship is contorted and made somewhat perverse with the understanding that only some speech is currently valued in our society: the speech of the powerful, white, male majority.<sup>57</sup> The next section explores in greater depth how harassment online leads to the suppression of marginalized speech.

As Citron and Wittes’ aptly named article proposes, I too believe that “the Internet will not break” from reforming § 230 to protect the victims of online abuse and harassment.<sup>58</sup> The benefits of § 230’s immunity “could have been secured at a slightly lesser price.”<sup>59</sup> The question of how to reform § 230 to address abuse against individual users and larger societal harms is admittedly difficult because of misinformation about the law and opaqueness concerning what role platforms have in developing or creating harmful content. However, that should not hinder efforts to recraft immunity in such a way that balances the burdens of plaintiffs against the burdens of platforms.<sup>60</sup> This middle ground would be

---

54. Mary Anne Franks, *The Free Speech Black Hole: Can the Internet Escape the Gravitational Pull of the First Amendment?* KNIGHT FIRST AMENDMENT INST. (Aug. 21, 2019), <https://knightcolumbia.org/content/the-free-speech-black-hole-can-the-internet-escape-the-gravitational-pull-of-the-first-amendment> [<https://perma.cc/EJC2-ZP36>] (“The assertion that regulating speech inevitably chills speech is false: given that some forms of speech themselves inflict chilling effects, regulating those forms of speech may actually serve free speech interests.”).

55. Citron & Franks, *supra* note 9, at 68.

56. Franks, *supra* note 54.

57. *Id.*; but see Daphne Keller, *Toward a Clearer Conversation About Platform Liability*, KNIGHT FIRST AMENDMENT INST. (Apr. 6, 2018), <https://knightcolumbia.org/content/toward-clearer-conversation-about-platform-liability> [<https://perma.cc/9RL3-GVGY>] (“So while [it] is right to say that vulnerable groups suffer disproportionately when platforms take down too little content, they also suffer disproportionately when platforms take down too much.”)

(“[W]hile...vulnerable groups suffer disproportionately when platforms take down too little content, they also suffer disproportionately when platforms take down too much.”); Citron & Franks, *supra* note 9, at 67 (“Section 230 already has a mechanism to address the unwarranted silencing of viewpoints. Under Section 230(c)(2), users or providers of interactive computer services enjoy immunity from liability for over-filtering or over-blocking speech only if they acted in ‘good faith.’”).

58. See generally Citron & Wittes, *supra* note 16.

59. *Id.* at 410.

60. See generally Evelyn Douek, *Governing Online Speech: From “Posts-As-Trumps” to Proportionality & Probability*, 121 COLUM. L. REV. 759 (2021) (advocating for content limitations proportionate to societal interests).

more desirable than letting current harms persist because of the Internet's central role in facilitating communication and civic engagement and the need to protect marginalized speech.

The question then becomes how to reform § 230 to address inequities while mitigating the risk that platforms might begin to over-moderate marginalized speech in the face of new regulation. It is a hard question because there is no perfect way to quantify both the harms to individuals and the benefits of § 230, and then to balance them against each other.<sup>61</sup> While it may be true that we cannot quantify the harms and benefits, and that the balancing judgment is based on personal values, I think we can, and should, do better. § 230 should not be read so broadly, and because it has already been read broadly by many courts, it should be amended by Congress to provide greater protection to vulnerable consumers.

## II. THE POWER OF SOCIAL PLATFORMS

Recent scholarship produced useful insights into what platforms are currently doing to address harmful content on the platform.<sup>62</sup> These scholars urge that any suggestions for reforming § 230 take into account the current practices that platforms follow.<sup>63</sup> But before looking to the practices of platforms, it is important to first have a better sense of the harms that occur online against largely marginalized groups. Thus, Section A describes online harassment and disinformation as a steadfast reprisal of old harms that are amplified in the internet ecosystem. Sections B and C draw on recent scholarship to understand platforms' current content moderation and curation practices and attempt to situate that understanding within the context of cyber civil rights activists' calls for § 230 reform.

### A. *Harassment & Disinformation*

The debate over § 230 produced some useful comparisons to historical events, such as the women's rights movement of the 60s

---

61. *Id.* at 42–43.

62. *See, e.g., id.* at 44–45; Klonick, *supra* note 14.

63. *See* Klonick, *supra* note 14, at 1603 (“If this fails and regulation is needed, it should be designed to strike a balance between preserving the democratizing forces of the internet and protecting the generative power of our New Governors, with a full and accurate understanding of how and why these platforms operate, as presented here.”); *see also* Douek, *supra* note 60, at 7 (“But changing the regulatory environment without a proper understanding of content moderation in practice will make the laws ineffective or, worse, create unintended consequences. Regulators need to understand the inherent characteristics of the systems they seek to reform.”); Sylvain *supra* note 10 (arguing that “[j]udges, lawyers, and legislators should...start looking carefully at how intermediaries' designs on user content do or do not result in actionable injuries.”).

and 70s. In the mid 1980s, radical feminist and lawyer Catharine MacKinnon argued that pornography should not be constitutionally protected as speech because it legitimizes abusive acts and suppresses the speech rights of women.<sup>64</sup> In particular, MacKinnon argued that pornography suppressed women's expression and their ability to speak out against abuse because it degrades and subordinates women as a class, effectively silencing them.<sup>65</sup> Even though the Supreme Court left only obscenity and child pornography outside the protections of the First Amendment, Mackinnon's argument is reprised in the debate surrounding harms perpetrated against women online.<sup>66</sup>

Harassers and abusers drive their victims offline by instilling legitimate fears of continued harassment, which leave victims effectively silenced.<sup>67</sup> And women and racial minorities are disproportionately the targets of some of the most egregious cyberattacks.<sup>68</sup> Thus, already marginalized speech is getting quashed by bad actors and impacting civic engagement on and off-line. § 230 enables platforms to turn a blind eye to this sort of censorship in that it immunizes even platforms that refuse to moderate illegal acts facilitated on the platform. Yet, at the same time, § 230 is somehow argued to protect against censorship—namely, the collateral censorship of predominantly privileged voices to begin with. What's more, big tech is allowed to reap the benefits of these harms and is a presumed innocent bystander without second thought of any sort of accomplice liability.

In the 1960s and 1970s, women collectively protested domestic violence and sexual harassment: practices entrenched in social norms of the day.<sup>69</sup> These norms followed narratives of victim-blaming and maintaining the status quo: what happens in the home stays in the home and 'boys will be boys.'<sup>70</sup> Women began to debunk these social beliefs through systematic and organized movements, calling attention to the inequities and harms produced.<sup>71</sup> Citron posits that the next frontier for attaining women's equality is online.<sup>72</sup> The entrenchment and normalization of revenge

---

64. KOSSEFF, *supra* note 26, at 210–11.

65. *Id.*

66. *Id.*

67. *See, e.g.*, Citron & Wittes *supra* note 58, at 420; *see also* Citron & Franks, *supra* note 9, at 55.

68. *See generally* CARRIE GOLDBERG, *NOBODY'S VICTIM: FIGHTING PSYCHOS, STALKERS, PERVS, AND TROLLS* (Plume 2019); *see also* KOSSEFF, *supra* note 26, at 209.

69. CITRON, *supra* note 3, at 95.

70. *Id.*

71. *Id.* at 96–99.

72. *Id.* at 100.

pornography online serves as compelling evidence. In order to secure equity online, we must continue to change social attitudes and dispel victim-blaming.<sup>73</sup> Just as “stay home” was not an acceptable response to workplace harassment, “stay offline” should not be an acceptable response to violence against women and minorities online.<sup>74</sup>

Importantly, the nature of harassment online shifted from random bad actors’ attacks on individual users to systematic attacks on communities or groups of people as a tool of disinformation campaigns.<sup>75</sup> These systemic, cross-platform attacks are often carried out by state actors who use tactics like trolling to spread disinformation and enlist “useful idiots” or average citizens to spread the message.<sup>76</sup> The targeting of specific communities aims to exacerbate social divisions and further polarize people on political issues. Thus, a dynamic persists where disinformation is rampant and harassment tactics are used to further divide people along socioeconomic lines, illustrating a larger societal problem beyond isolated harms to individuals.

Additionally, the ability for users to remain anonymous amplifies harassing and abusive behavior.<sup>77</sup> Identifying anonymous posters of harmful content is often difficult and unsuccessful.<sup>78</sup> In order for a plaintiff to unmask their anonymous attacker, the plaintiff must file suit against the anonymous defendant, subpoena the website to turn over data about the user—such as an IP address—and if the IP address is obtained, the plaintiff must request the name of the subscriber from the internet service provider that hosts that IP address.<sup>79</sup> This process poses several difficulties: not all websites require real names or email addresses, or keep track of IP addresses; the subpoena can be challenged, in which case complex First Amendment balancing tests are used to determine enforceability; and, even if the subpoena is enforced, the information often leads to a dead end.<sup>80</sup> Because

---

73. *Id.*

74. *Id.*

75. Brittan Heller, *Enlisting Useful Idiots: The Ties Between Online Harassment and Disinformation*, 19.1 COLO. TECH. L.J. 19, 20 (2021).

76. *Id.* at 26.

77. Citron & Franks, *supra* note 9, at 68 (“The Internet lowers the costs of engaging in abuse by providing abusers with anonymity and social validation, while providing new ways to increase the range and impact of that abuse. The online abuse of women in particular amplifies sexist stereotyping and discrimination, compromising gender equality online and off.”).

78. KOSSEFF, *supra* note 26, at 221.

79. *Id.*

80. *Id.* at 221–22.

platforms have far more power and ability to control what happens on the platform—who uses the platform, the terms and conditions of use, and what information they keep on users—platforms should be incentivized to take greater responsibility in protecting vulnerable consumers.

### B. *Content Moderation Practices*

Platforms enjoy unprecedented immunity from publisher liability under § 230, even though they maintain and regulate what *Packingham* likened to the modern-day public square for purposes of First Amendment law.<sup>81</sup> Cyber civil rights activists have called to reign in this immunity for over a decade.<sup>82</sup> In a recent article Klonick uncovered how and why social media platforms moderate content, and proposed that we understand and treat platforms as new types of governance, separate and apart from traditional First Amendment categorical analogies:

[P]latforms should be thought of as operating as the New Governors of online speech. These New Governors are part of a new triadic model of speech that sits between the state and speakers- publishers. They are private, self-regulating entities that are economically and normatively motivated to reflect the democratic culture and free speech expectations of their users.<sup>83</sup>

Klonick frames the “governance” that platforms engage in as an iterative process reflecting the “interplay between user and platform.”<sup>84</sup> However, “governance” can be boiled down to something much simpler, something we ought not lose sight of. “Governance,” in its simplest form, implies control and authority over a group backed by the threat of punishment. Facebook asserts control of its users by setting the terms and conditions of engagement on the app—more specifically, through the Abuse Standards.<sup>85</sup> The standards are non-negotiable for entrance and participation, and the punishment for violating those standards can rise to dismissal from the platform indefinitely—a serious punishment resulting in the loss of access to a scarce medium for

---

81. See *Packingham v. North Carolina*, 137 S. Ct. 1730, 1732 (2017).

82. See generally, e.g., CITRON, *supra* note 3.

83. Klonick, *supra* note 14, at 1603.

84. *Id.* at 1617.

85. See *id.* at 1644.

speech.<sup>86</sup> While platforms have not been and should not be treated as government speakers, platforms should also not be allowed to control a scarce venue for modern communication and engagement completely free from traditional tort liability.

The moderation or “governance” process evolves constantly behind the scenes as platforms attempt the impossible—to keep pace with rapidly changing expectations about speech.<sup>87</sup> Platforms could not possibly take into account every user’s changing expectation so what or who has the greatest influence on policy iterations? Klonick “discusses four major ways platforms’ content-moderation policies are subject to outside influence: (1) government request, (2) media coverage, (3) third-party civil society groups, and (4) individual users’ use of the moderation process.”<sup>88</sup> All four of these categories reflect the embeddedness of power and privilege, victims lack of access to justice. Governments influence content decisions of platforms by threatening to regulate platforms or cut off access to the platform entirely.<sup>89</sup> The media exerts influence over content decisions by evoking public outcry and collective action.<sup>90</sup> Third party groups exert influence over content decisions by advocating for the interests of those they represent and meeting collectively with industry players to discuss content guidelines.<sup>91</sup> While this category initially appears to be a win for individual users, third party groups operate at a level removed from users and cannot be said to adequately represent every user that has been harmed. The moderation process itself is also problematic because not all users will have access to this process;<sup>92</sup> not everyone is afforded technological due process. Therefore, the voices of victims are still unheard, and they have little to no recourse by which to impact the policy decisions of platforms.

Moreover, individual users that exhaust all formal avenues to complain about moderation decisions often turn to informal tactics. Rory Van Loo discusses the limits of user’s informal tactics to shape moderation decisions, like taking to social media to complain: “An assault victim should not have to take to social media and reveal a very private and painful event to the world to get a response. Moreover, users with few followers have less social media influence.

---

86. *See generally id.* at 1661 (“In the years since Reno, the hold of certain platforms has arguably created scarcity—if not of speech generally, undoubtedly of certain mediums of speech that these platforms provide.”).

87. *See generally id.* at 1629.

88. *Id.* at 1649.

89. *See id.* at 1650–52.

90. *See id.* at 1652–53.

91. *See id.* at 1655–56.

92. *See id.* at 1657.

Appealing to the CEO may go nowhere.”<sup>93</sup> Without power and influence, or collective public outcry, individual users may have no recourse to the injustices suffered on social platforms when the platform refuses to help. And even with public outcry and collective action, those with power and privilege in society may succeed in keeping up content that would otherwise be removed or vice versa. Thus, fundamental problems of control and power continue to shape the moderation process outside the public view and without legal teeth mandating the transparency of moderation policy decisions.

It’s important to keep in mind that platforms are first and foremost companies, not public utilities, which means that fundamental to their existence—and thus any scheme of “governance”—is the drive to maximize profits. It makes sense then to examine the financial incentives of social platforms in moderating content particularly when § 230 does not currently require that platforms do any moderating in order to receive immunity from tort liability. Klonick argues that platforms have financial incentive to moderate content according to the expectations of users: “[p]latforms have created a voluntary system of self-regulation because they are economically motivated to create a hospitable environment for their users in order to incentivize engagement.”<sup>94</sup> Certainly, if engagement suffers because users are made uncomfortable by particular content on the platform, then so too will advertising revenues.<sup>95</sup> But these economic incentives are complicated because users are the commodity and advertisers are the customers in the revenue models of social platforms.<sup>96</sup> Platforms like Facebook generate revenue from advertisers, not users, and thus are incentivized to protect advertisers before users, and potentially at the expense of users.<sup>97</sup> Additionally, the fear of users leaving because of unhospitable conditions is countered by strong network effects and concentration of power in the market, as well as the argument that abusive material actually generates traffic and attention.<sup>98</sup> So again it becomes evident that the powerful are protected at the expense of the vulnerable who do not

---

93. Rory Van Loo, *Federal Rules of Platform Procedure*, U. OF CHI. L. REV. (forthcoming) (manuscript at 31), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3576562](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3576562) [<https://perma.cc/U7QC-6TTY>].

94. Klonick, *supra* note 14, at 1615.

95. *See id.* at 1627.

96. *See* Citron & Franks, *supra* note 9, at 52.

97. *See* Van Loo, *supra* note 93, at 30–31 (giving an example of TripAdvisor taking down bad reviews in order to protect advertisers, which resulted in harms to individuals who relied on good reviews).

98. *See* Citron & Franks, *supra* note 9, at 53.

have adequate access or power to influence content moderation policy.

### C. *Whose Content is it Anyway?*

Beyond content removal decisions, social platforms make curation decisions about how and where content gets placed on the platform. Douek discusses this shift in content moderation policy in the context of Facebook:

Facebook is increasingly relying not on the blunt content moderation tools of removing posts or pages, but on the subtle tools of limiting their reach and exposure. For ‘borderline’ content in each of its harmful categories, Facebook works to ‘distribute that content less’ to reduce the incentive to post such content.<sup>99</sup>

If transparency is a problem in the more concrete decisions to take down or leave up content, it is even more of a problem in the context of curation algorithms that determine placement. These decisions, while meant to increase revenue by increasing engagement, also end up “[shaping] the form and substance of their users’ content” in several notable and problematic ways.<sup>100</sup> Platforms’ designs on user data under the new attention markets have been described as manipulative because the algorithms used deploy principles of human psychology to alter human behavior by getting users to visit the platform more often.<sup>101</sup> Sophisticated users who recognize and understand this process may purposefully change their behavior while participating on the platform in order to influence the algorithms curating their content in one way or another.<sup>102</sup> Additionally, the curation of targeted content gives rise to filter bubbles or echo chambers that reinforce particular viewpoints and keep users isolated from content outside their comfort zone.<sup>103</sup> Taking all this into account, it becomes increasingly difficult to separate the tortious or otherwise illegal content of third parties from the platform that promotes, filters, and profits from user engagement with such content.

---

99. Evelyn Douek, *Facebook’s ‘Oversight Board.’ Move Fast with Stable Infrastructure and Humility*, 21 N.C.L.J. & TECH. 1, 42–43 (2019); see also Klonick, *supra* note 14, at 1660 (“For the content that stays up—like a newspaper determining what space to allot certain issues—platforms also have intricate algorithms to determine what material a user wants to see and what material should be minimized within a newsfeed, homepage, or stream.”).

100. Sylvain, *supra* note 10, at 2.

101. THE SOCIAL DILEMMA (Argent Pictures 2020).

102. See generally Jillian Warren, *This is How the Instagram Algorithm Works in 2021*, LATER (Jan. 4, 2021), <https://later.com/blog/how-instagram-algorithm-works/> [<https://perma.cc/M5ZK-DJCX>].

103. See, e.g., Klonick *supra* note, 14 at 1667.

While the development of AI technologies holds the potential to solve many complex problems pertaining to online participation and interaction, such as moderating discriminatory or illegal content, its use is not without intense controversy. The nascent technologies are still developing and make many mistakes without human oversight.<sup>104</sup> Even so, the algorithmic designs on user data deployed by big tech illustrates that platforms are capable of sifting through large volumes of content, archiving user data, and filtering and curating content for particular users. But the impetus behind these capabilities is the capitalistic incentive to increase profits; thus, the artificial intelligence technologies developed by platforms are being deployed and developed to achieve the goal of profit maximization. If we allow AI technology to continue developing for the single-minded goal of maximizing the capitalist's profit, then we might miss opportunities to apply this technology to a different set of problems.<sup>105</sup>

If the attention revenue model and resulting data practices shape the substance of content on the platform that creates subsequent harms to users, then broad immunity under § 230 seems unreasonable because the platform did contribute to the harm. At the very least, a plaintiff should be able to bring a civil claim against a platform and receive a response from the platform before the lawsuit is dismissed on § 230 grounds. Although, the influence of algorithms is probably not enough evidence to overcome a § 230 defense in court, like in *Rommates.com* or in *Backpage*, because the judiciary is not well situated to understand the intricacies of algorithms and their psychological effects.

### III. REFORMING § 230

Society won't fall apart without blanket immunity for platforms: the Internet will fight to adapt.<sup>106</sup> Platforms are more than just mere conduits for user content, they are active players in shaping communication online.<sup>107</sup> Whether or not they have a hand

---

104. Elizabeth Dwoskin & Nitasha Tiku, *Facebook sent home thousands of human moderators due to the coronavirus. Now the algorithms are in charge*, WASH. POST (Mar. 24, 2020, 3:55 PM), <https://www.washingtonpost.com/technology/2020/03/23/facebook-moderators-coronavirus/> [<https://perma.cc/83HT-YE4C>].

105. Note that Ruha Benjamin expresses the genuine concern that the deployment of new technologies in the tech world perpetuates racial inequalities through what she terms “the New Jim Code,” or “the employment of new technologies that reflect and reproduce existing inequities but that are promoted and perceived as more objective or progressive than the discriminatory systems of a previous era.” Benjamin, *supra* note 6, at 5–6.

106. *See generally* Citron & Wittes, *supra* note 58.

107. Sylvain, *supra* note 10.

in a particular harm to a particular user is a more difficult question. But we ought to be weary of giving platforms a free pass on liability, before ever getting to the merits. Citron and Wittes have proposed altering the language of § 230(c)(1) in the following way:

No provider or user of an interactive computer service that *takes reasonable steps to prevent or address unlawful uses of its services* shall be treated as the publisher or speaker of any information provided by another information content provider *in any action arising out of the publication of content provided by that information content provider*.<sup>108</sup>

These changes would effectively make platform immunity contingent upon platforms having a reasonable policy or process to prevent or address unlawful uses of the platform. Citron and Franks later expanded on how this might play out in the context of a platform's motion to dismiss on § 230 grounds: "The question would not be whether a platform acted reasonably with regard to a specific use. . . [but rather,] . . . whether the [platform] engaged in reasonable content moderation practices writ large with regard to unlawful uses that clearly create serious harm to others."<sup>109</sup> Further, Citron and Wittes suggest that an analysis of what constitutes "reasonable" would take into account factors such as volume of content, whether unlawful actions were encouraged, and whether requests to remove content were addressed in order to account for the differences between ISPs, social media platforms, and other interactive computer services.<sup>110</sup>

This would be a careful and well-balanced first step to addressing the challenges created by § 230 because it would (1) require that all social platforms, and other types of platforms, adopt reasonable moderation capabilities and policies, eliminating the problems elucidated by *Herrick v Grindr*;<sup>111</sup> and, (2) it would not necessarily prompt over moderation because the question focuses on the reasonable efforts to moderate writ large rather than in the particular instance. However, given the reality of the tech sector in that a few firms have vast market power and that many of these

---

108. Citron & Wittes, *supra* note 58, at 419.

109. Citron & Franks, *supra* note 9, at 22.

110. See Citron & Wittes, *supra* note 58, at 419.

111. *Herrick v. Grindr L.L.C.*, 765 Fed. Appx. 586 (2019); Carrie Goldberg, *Herrick v. Grindr: Why Section 230 of the Communications Decency Act Must Be Fixed*, LAWFARE (Aug. 14, 2019), <https://www.lawfareblog.com/herrick-v-grindr-why-section-230-communications-decency-act-must-be-fixed> [<https://perma.cc/QJT7-M3WN>] (230 immunity granted to Grindr in a products liability tort suit alleging that Grindr harmed Herrick by not taking down a profile impersonating him, and not even having the capability to do so built into the architecture of the platform).

firms have already developed content moderation policies, it is likely that platforms will still be able to succeed in dismissing cases relatively easily under this § 230 framework. It also does not get at the problem of platform transparency and accountability because individual moderation and curation decisions could continue to play out behind closed doors at the behest of powerful actors.

To incentivize greater transparency and accountability, I also propose changing § 230(c)(2)(A) to read that “no provider or user of an interactive computer service shall be held liable on account of any action *reasonably taken, and that is made in accordance with a reasonably transparent process*, to restrict access to or availability of material...” (italics are proposed changes). What constitutes “reasonable” action here might include a consideration of the degree to which the content was considered and whether it conformed to a written public policy of the platform. I suggest writing into the statute a requirement for platforms to make aspects of the moderation process more transparent to quell concerns that platforms may be censoring certain voices disproportionately.<sup>112</sup> The goal is to create a more legitimate system of accountability without recreating harms.

A reasonableness standard combined with transparency would force platforms that wish to continue benefitting from § 230 to develop clear public facing policies and to explain decisions about content moderation, which might help to expose any effort to censor on the basis of race, gender, class, or political affiliation. However, public outcry would still be necessary to shame platforms into curtailing censorship on discriminatory grounds because private parties are free to infringe the First Amendment rights of others and engage in hate speech. Transparency would at least pave the way to greater accountability on the part of platforms.

Of course, these reforms will likely prompt First Amendment challenges. Citron and Wittes contend that conditioning § 230(c)(1) immunity upon reasonable efforts to moderate does not burden free speech interests because it merely rolls back an immunity that is not required by the First Amendment.<sup>113</sup> On the other hand, one could make the creative argument that this change would be a type of compelled speech in that the government is dictating that platforms must engage in some level of moderation. I don’t think this counterargument is likely to win out because it seems well

---

112. See DAVID KAYNE, *SPEECH POLICE: THE GLOBAL STRUGGLE TO GOVERN THE INTERNET* 10–11 (2019) (Kayne argues, as an alternative to regulation, that platforms should be more transparent about how they arrive at policy choices, how they make decisions when moderating content, and how their algorithms make decisions).

113. Citron & Wittes, *supra* note 58, at 419–20.

settled that systems of liability necessarily encourage and discourage certain behaviors. Thus, it appears plausible that Congress can constitutionally encourage moderation by dangling the carrot of immunity under § 230. Conditional immunity would force platforms to think carefully about how to moderate content and build the architecture of the platform with safeguards in place if they want to benefit from immunity.

Adding a reasonable standard in both §§ 230(c)(1) and (c)(2) might make it less likely that a platform could get away with disproportionately moderating the content of vulnerable groups. However, a transparency requirement might also be challenged on First Amendment grounds. I think the transparency requirement would withstand scrutiny because, likewise to the recommended provision in (c)(1), it only modifies or conditions an immunity that is not guaranteed by the First Amendment. Like the reasonableness requirement, a transparency requirement seeks to encourage transparent behavior in order to receive the carrot of immunity. Intermediaries would not be forced to reveal proprietary information but would have to implement and produce some evidence of a reasonable process by which users could inquire after moderation decisions that impact them, and potentially reverse the decision.<sup>114</sup>

## CONCLUSION

If §230 remains as is, then the victims of cyberattacks—largely, women and minorities—will continue to be driven offline by harassment and abuse with little to no recourse for justice in the most likely event that they cannot identify the individual perpetrator. And if victims do bring a claim against a platform for some theory of harm, then the broad reading that courts have given § 230 may end the claim's life at the pleading stage. The threat doesn't dissipate, it persists, forcing victims offline which is an essential method of communication and civic involvement today and for the foreseeable future. Platforms have little incentive to provide help when they are immune from expansive tort liability, financially benefitting from the attention created by harassment and abuse, and powerful enough to withstand backlash from the minority that are affected. Therefore, to incentivize platforms to

---

114. For a more in-depth process discussion, see generally Rory Van Loo, *Federal Rules of Platform Procedure*, U. OF CHI. L. REV. (forthcoming) (manuscript at 31), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3576562](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3576562) [<https://perma.cc/VW4B-M9H3>] (arguing that today's platforms need mandated procedures and legal standards for dispute resolution to foster transparency and accountability similar to financial institutions before.)

exclude bad actors and protect vulnerable populations' ability to engage in democracy, § 230 should be revised. First, immunity under § 230 should be conditioned on the platform taking reasonable steps to moderate online content. This means that (c)(1) should be revised to make exemption from speaker/publisher treatment contingent upon reasonable efforts to moderate, and (c)(2) should be revised to make immunity from liability for actions taken to moderate content contingent upon reasonableness and transparency.