
NOTE

SECTION 230 AS FIRST AMENDMENT RULE

Section 230 of the Communications Decency Act of 1996¹ has been lauded as “the most important law protecting internet speech” and called “perhaps the most influential law to protect the kind of innovation that has allowed the Internet to thrive.”² The law’s tremendous importance stems from the shield it provides to websites against suits based on torts committed by users. For instance, Wikipedia cannot be held liable for defamation posted by a user. This intermediary liability protection encourages websites to engage in content moderation without fear that their efforts to screen content will expose them to liability for defamatory material that slips through. Without this protection, websites would have an incentive to censor constitutionally protected speech in order to avoid potential lawsuits.³

But § 230 is under attack on multiple fronts.⁴ From the popular media⁵ to Capitol Hill,⁶ some view the law with disdain. Various scholars have also heavily criticized § 230, saying amending the law would help to reduce defamation online.⁷ And, in the courts, 2016 was perhaps a nadir for § 230, as judges repeatedly adopted narrow readings of the law.⁸

¹ 47 U.S.C. § 230 (2012).

² *CDA 230: The Most Important Law Protecting Internet Speech*, ELECTRONIC FRONTIER FOUND., <https://www.eff.org/issues/cda230> [<https://perma.cc/JN9Y-TVNT>]; accord Jack M. Balkin, *Old-School/New-School Speech Regulation*, 127 HARV. L. REV. 2296, 2313 (2014) (“Section 230 immunity . . . ha[s] been among the most important protections of free expression in the United States in the digital age.”); David Post, *A Bit of Internet History, or How Two Members of Congress Helped Create a Trillion or So Dollars of Value*, WASH. POST: VOLOKH CONSPIRACY (Aug. 27, 2015), <http://wapo.st/1K9AmTh> [<https://perma.cc/S4LN-WE9P>].

³ See *infra* Part III, pp. 2032–47.

⁴ Cindy Cohn & Jamie Williams, *20 Years of Protecting Intermediaries: Legacy of “Zeran” Remains a Critical Protection for Freedom of Expression Online*, LAW.COM: THE RECORDER (Nov. 10, 2017, 8:31 AM), <https://www.law.com/therecorder/sites/therecorder/2017/11/10/20-years-of-protecting-intermediaries-legacy-of-zeran-remains-a-critical-protection-for-freedom-of-expression-online/> [<https://perma.cc/U7ER-JPN3>].

⁵ See, e.g., Arthur Chu, *Mr. Obama, Tear Down This Liability Shield*, TECHCRUNCH (Sept. 29, 2015), <https://techcrunch.com/2015/09/29/mr-obama-tear-down-this-liability-shield/> [<https://perma.cc/C9QW-K965>].

⁶ See, e.g., Eric Goldman, *How SESTA Undermines Section 230’s Good Samaritan Provisions*, TECH. & MARKETING L. BLOG (Nov. 7, 2017), <http://blog.ericgoldman.org/archives/2017/11/how-sesta-undermines-section-230s-good-samaritan-provisions.htm> [<https://perma.cc/YJ75-343D>] (addressing congressional efforts to amend § 230).

⁷ See, e.g., Ann Bartow, *Section 230 Keeps Platforms for Defamation and Threats Highly Profitable*, LAW.COM: THE RECORDER (Nov. 13, 2017, 12:19 PM), <https://www.law.com/therecorder/sites/therecorder/2017/11/10/section-230-keeps-platforms-for-defamation-and-threats-highly-profitable/> [<https://perma.cc/ZMJ3-DEAN>].

⁸ Eric Goldman, *Ten Worst Section 230 Rulings of 2016 (Plus the Five Best)*, TECH. & MARKETING L. BLOG (Jan. 4, 2017), <http://blog.ericgoldman.org/archives/2017/01/ten-worst-section-230-rulings-of-2016-plus-the-five-best.htm> [<https://perma.cc/4N9G-3UTU>] (collecting cases).

Against this current, this Note provides the first thorough argument that the First Amendment requires § 230's bar on holding websites liable for the defamation of their users. While the First Amendment does not "require" the *federal statute*, of course, this Note argues that the First Amendment rule should be the same as § 230's rule. Under the Supreme Court's First Amendment case law on defamation, the private censorship produced by defamation liability for internet intermediaries cannot be justified by a government interest in defamation law. Recognizing § 230's more stable constitutional provenance explains why courts traditionally adopted a broad reading of the law, demonstrates the law's substantive importance, and helps predict what might occur should detractors succeed in achieving amendment by Congress.

Part I describes secondary liability for defamation and § 230. Part II explains the prevailing assumption among judges and scholars that the First Amendment does not require § 230. Part III then challenges this assumption, arguing that the Constitution protects internet intermediaries from liability for defamation committed by their users. The censorship that would result from internet intermediary liability for defamation cannot be saved by the government's interest in imposing liability.⁹ Part IV discusses this Note's implications and concludes.

I. DEFAMATION, INTERMEDIARY LIABILITY, AND § 230

Defamation is a common law tort that protects individuals against the publication of harmful false statements about them.¹⁰ "Publication" includes intentional and unreasonable failure to remove defamatory material under one's control.¹¹ Distributors, such as booksellers, may be held liable for defamation they transmit if they knew or had reason to know of its defamatory nature, but are not under a general duty to screen the items they retail.¹²

In the 1990s, courts began to apply these doctrines to internet services. In *Cubby, Inc. v. CompuServe Inc.*,¹³ a district court held that an internet service provider was not liable for allegedly defamatory content in one of its online forums because it had "no more editorial control" than would "a public library, book store, or newsstand,"¹⁴ and therefore was a mere distributor that did not know or have reason to know of the content.¹⁵ Later, in *Stratton*

⁹ This Note seeks to demonstrate the constitutional relevance of the policy-based arguments in favor of § 230, though it does not itself engage in a full-fledged policy analysis.

¹⁰ RESTATEMENT (SECOND) OF TORTS § 558 (AM. LAW INST. 1977).

¹¹ *Id.* § 577(2).

¹² *Id.* § 581 & cmts. d & e.

¹³ 776 F. Supp. 135 (S.D.N.Y. 1991).

¹⁴ *Id.* at 140.

¹⁵ *Id.* at 140-41. The court held there was no genuine issue of material fact as to knowledge. *Id.* at 141.

Oakmont, Inc. v. Prodigy Services Co.,¹⁶ a state court held that because an owner of online bulletin boards had exercised “editorial control” over offensive content, it could be held liable as a publisher of defamatory posts.¹⁷ This pair of cases posed a troubling choice for websites. If they took a hands-off approach to moderation, they received significant protection from liability. However, if they sought to proactively regulate content on their websites, they might face liability.¹⁸ This dilemma “created a minor sensation.”¹⁹

These concerns were heard on Capitol Hill when Congress enacted section 509 of the Communications Decency Act (codified at 47 U.S.C. § 230), which overruled *Stratton Oakmont*.²⁰ Section 230 provides that no website that relies on user-generated content “shall be treated as the publisher or speaker of any information provided by another information content provider.”²¹ Therefore, a website cannot be held liable for defamation posted by a user even if the website knows or has reason to know of the defamatory content.²² Of course, if an intermediary website *itself* created defamatory content, it could be held liable²³ — for example, if Facebook itself wrote a blog post on its website defaming the creators of Google Plus. In other words, websites are not immune from defamation claims. They are merely protected from being held secondarily liable for the defamatory statements of others.

¹⁶ 1995 WL 323710 (N.Y. Sup. Ct. May 24, 1995).

¹⁷ *Id.* at *4–5. The website also held itself out as engaging in moderation. *Id.*

¹⁸ *Zeran v. Am. Online, Inc.*, 129 F.3d 327, 331 (4th Cir. 1997) (noting that *Stratton Oakmont* created “disincentives to selfregulation [sic]”).

¹⁹ David R. Sheridan, *Zeran v. AOL and the Effect of Section 230 of the Communications Decency Act upon Liability for Defamation on the Internet*, 61 ALB. L. REV. 147, 159 (1997) (citing Robert Cannon, *The Legislative History of Senator Exon's Communications Decency Act: Regulating Barbarians on the Information Superhighway*, 49 FED. COMM. L.J. 51, 62 nn.51–52 (1996)).

²⁰ *Id.* at 150–51; Cannon, *supra* note 19, at 61–63, 62 nn.51–52.

²¹ 47 U.S.C. § 230(c)(1) (2012).

²² *Zeran*, 129 F.3d at 331–33.

²³ See 47 U.S.C. § 230(f)(3) (defining “information content provider”); *id.* § 230(c)(1) (exempting websites from liability only for information “provided by *another* information content provider” (emphasis added)).

In interpreting § 230, courts have largely followed through on congressional hopes of providing intermediary liability protection to websites for defamation claims.²⁴ For example, in the “seminal”²⁵ case *Zeran v. America Online, Inc.*,²⁶ then-Chief Judge Wilkinson held that § 230 protected America Online from a defamation claim based on messages posted on its bulletin boards.²⁷ Judge Wilkinson explained § 230 succinctly: it “creates a federal immunity to any cause of action that would make service providers liable for information originating with a third-party user of the service.”²⁸ The law bars suits that would hold websites liable for decisions about whether and how to moderate user-generated content.²⁹ As to congressional purpose, Judge Wilkinson identified first that the “specter of tort liability in an area of such prolific speech would have an obvious chilling effect” and second that § 230 encourages websites to moderate content without fear of liability.³⁰

II. THE ASSUMPTION THAT THE FIRST AMENDMENT DOES NOT REQUIRE § 230

Judges and academics are nearly in consensus in assuming that the First Amendment does not require § 230.³¹ Since the enactment of § 230, courts have had little reason to reach this constitutional question. In *Cubby*, decided before the enactment of § 230, while the court cited a First Amendment case to support its holding, it did not discuss the notion that the First Amendment might provide even more protection

²⁴ See David S. Ardia, *Free Speech Savior or Shield for Scoundrels: An Empirical Study of Intermediary Immunity Under Section 230 of the Communications Decency Act*, 43 LOY. L.A. L. REV. 373, 452 (2010) (“Defamation-type claims were far and away the most numerous claims in the section 230 case law, and the courts consistently held that these claims fell within section 230’s protections.” (footnotes omitted)); see also, e.g., *Batzel v. Smith*, 333 F.3d 1018 (9th Cir. 2003); *Green v. Am. Online*, 318 F.3d 465 (3d Cir. 2003). Courts have reached inconsistent results in nontraditional cases outside defamation law. See, e.g., *Doe v. Internet Brands, Inc.*, 824 F.3d 846 (9th Cir. 2016) (holding that § 230 did not protect the owner of the website Model Mayhem from a failure-to-warn claim); Recent Case, *Doe v. Internet Brands, Inc.*, 824 F.3d 846 (9th Cir. 2016), 130 HARV. L. REV. 777, 777 (2016) (criticizing the Ninth Circuit’s decision for “declin[ing] to adopt an alternative understanding of the statute more in line with the law’s stated policy objectives”).

²⁵ Cathy Gellis, *The First Hard Case: “Zeran v. AOL” and What It Can Teach Us About Today’s Hard Cases*, LAW.COM: THE RECORDER (Nov. 10, 2017, 1:02 AM), <https://www.law.com/therecorder/sites/therecorder/2017/11/10/the-first-hard-case-zeran-v-aol-and-what-it-can-teach-us-about-todays-hard-cases/> [<https://perma.cc/2U75-X7J8>]; see Patrick J. Carome & Cary A. Glynn, *Serendipity and Internet Law: How the “Zeran v. AOL” Landmark Almost Wasn’t*, LAW.COM: THE RECORDER (Nov. 10, 2017, 8:29 AM), <https://www.law.com/therecorder/sites/therecorder/2017/11/10/serendipity-and-internet-law-how-the-zeran-v-aol-landmark-almost-wasnt/> [<https://perma.cc/J95X-ZRXB>].

²⁶ 129 F.3d 327.

²⁷ *Id.* at 328.

²⁸ *Id.* at 330.

²⁹ *Id.* at 331.

³⁰ *Id.*

³¹ For courts, see, for example, *Batzel v. Smith*, 333 F.3d 1018, 1020 (9th Cir. 2003).

to websites.³² In *Stratton Oakmont*, the court acknowledged that the website's moderation system "may have a chilling effect on freedom of communication in Cyberspace," even though the court in effect required this type of website to employ similar moderation to avoid liability.³³ There too the court did not consider First Amendment concerns. As one district court put it, "Section 230 reflects a 'policy choice,' not a First Amendment imperative, to immunize ISPs from defamation . . . driven, in part, by free speech concerns."³⁴ More recently, in *Gonzalez v. Google, Inc.*,³⁵ the court stated in passing that "[i]n the absence of the protection afforded by section 230(c)(1), one who published or distributed speech online" may be liable for defamation even if the website had no knowledge of the content.³⁶ In 2016, a First Circuit panel acknowledged that "First Amendment values . . . drive" § 230, but wrote that this rule could be amended via mere legislation.³⁷

Academics share the assumption that the First Amendment does not require § 230.³⁸ As Professor Rebecca Tushnet writes, the "First Amendment does not currently require a particular solution" for internet intermediary defamation liability.³⁹ In defending § 230, Professor Jeff Koseff admits that its "immunity extends beyond intermediary protections provided by the First Amendment."⁴⁰ And Professor William H. Freivogel puts it bluntly: "It would not be accurate to argue that the First Amendment requires Section 230."⁴¹ In canvassing the First Amendment options for addressing how internet platforms moderate content, one scholar does not address the possibility of § 230 as a First Amendment rule.⁴² Other commentators seem to share this assumption

³² *Cubby, Inc. v. CompuServe, Inc.*, 776 F. Supp. 135, 139 (S.D.N.Y. 1991) (citing *Smith v. California*, 361 U.S. 147, 152–53 (1959)).

³³ *Stratton Oakmont, Inc. v. Prodigy Servs. Co.*, 1995 WL 323710, at *5 (N.Y. Sup. Ct. May 24, 1995).

³⁴ *Gucci Am., Inc. v. Hall & Assocs.*, 135 F. Supp. 2d 409, 421 (S.D.N.Y. 2001) (quoting *Zeran*, 129 F.3d at 330).

³⁵ No. 16-CV-03282, 2017 WL 4773366 (N.D. Cal. Oct. 23, 2017).

³⁶ *Id.* at *4 (citing *Batzel*, 333 F.3d at 1026–27) (ignoring First Amendment question).

³⁷ *Jane Doe No. 1 v. Backpage.com, LLC*, 817 F.3d 12, 29 (1st Cir. 2016), *cert. denied*, 137 S. Ct. 622 (2017). To be sure, perhaps the panel would differentiate between the constitutional law of defamation and potential criminal liability for sex trafficking.

³⁸ See, e.g., Jack M. Balkin, *The Future of Free Expression in a Digital Age*, 36 PEPP. L. REV. 427, 434 (2009) ("[Section 230] is not required by First Amendment doctrine."); Rebecca Tushnet, *Power Without Responsibility: Intermediaries and the First Amendment*, 76 GEO. WASH. L. REV. 986, 1008 n.95 (2008) ("Before the CDA, the assumption in the law reviews tended to be that the [*New York Times v. Sullivan*] standard was the best to be hoped for as a constitutional matter.")

³⁹ Tushnet, *supra* note 38, at 988.

⁴⁰ Jeff Koseff, *Defending Section 230: The Value of Intermediary Immunity*, 15 J. TECH. L. & POL'Y 123, 136 (2010).

⁴¹ William H. Freivogel, *Does the Communications Decency Act Foster Indecency?*, 16 COMM. L. & POL'Y 17, 48 (2011).

⁴² Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1613–15 (2018).

as well.⁴³ Moreover, the many scholars who have criticized § 230 do not seem to believe that a response is necessary against the charge that the rule is mandated by the Constitution. For instance, two critics simply write that § 230 is “not required by the First Amendment.”⁴⁴

III. WHY THE FIRST AMENDMENT REQUIRES § 230

This Part begins by explaining First Amendment scrutiny of defamation law and then argues that, under that case law, imposing defamation liability on internet intermediaries is unconstitutional.

A. Defamation and the First Amendment

Like § 230, the First Amendment operates as a constraint on the scope of defamation law. While some regulations of speech may be reviewed, for example, under the “generic” strict scrutiny test,⁴⁵ other types of speech are governed by specific tests devised by the Court “on a largely ad hoc basis.”⁴⁶ The specific rules that the Court devised to govern defamation law, for instance in the 1964 landmark case *New York Times Co. v. Sullivan*,⁴⁷ exemplify this ad hoc approach.

In *New York Times*, the Supreme Court held that under the First Amendment public officials alleging defamation must show the defendant acted with “actual malice” — knowledge of falsity or recklessness toward this potential.⁴⁸ The Court reasoned that not requiring actual malice could stifle vital discourse because of the fear of civil liability.⁴⁹ “A rule compelling the critic of official conduct to guarantee the truth of all his factual assertions,” the Court feared, leads to “self-censorship.”⁵⁰ Potential defendants might worry that they could not prove in court the legality of their statements or afford expensive litigation and therefore “make only statements which ‘steer far wider of the unlawful zone.’”⁵¹

⁴³ See, e.g., James Grimmelman, *No ESC*, LAW.COM: THE RECORDER (Nov. 10, 2017, 2:03 AM), <https://www.law.com/therecorder/sites/therecorder/2017/11/10/no-esc/> [<https://perma.cc/B2ZW-HQHJ>] (referring to § 230 as “subconstitutional free speech law”). But cf. Cecilia Ziniti, Note, *The Optimal Liability System for Online Service Providers: How Zeran v. America Online Got It Right and Web 2.0 Proves It*, 23 BERKELEY TECH. L.J. 583, 605–06 (2008) (arguing briefly that a notice-based system would be unconstitutional).

⁴⁴ Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity*, 86 FORDHAM L. REV. 401, 419 (2017); see also Heather Saint, Note, *Section 230 of the Communications Decency Act: The True Culprit of Internet Defamation*, 36 LOY. L.A. ENT. L. REV. 39, 69 (2015).

⁴⁵ Richard H. Fallon, Jr., *Strict Judicial Scrutiny*, 54 UCLA L. REV. 1267, 1292 (2007).

⁴⁶ *Id.* at 1291.

⁴⁷ 376 U.S. 254 (1964).

⁴⁸ *Id.* at 279–80.

⁴⁹ *Id.* at 277–80.

⁵⁰ *Id.* at 279.

⁵¹ *Id.* (quoting *Speiser v. Randall*, 357 U.S. 513, 526 (1958)).

Later, in *Gertz v. Robert Welch, Inc.*,⁵² the Court held that private individuals alleging defamation did not need to meet an actual malice requirement.⁵³ The Court noted that “punishment of error runs the risk of inducing a cautious and restrictive exercise of the constitutionally guaranteed freedoms of speech and press.”⁵⁴ The Court explained that the interest supporting defamation law is “the compensation of individuals for the harm inflicted on them by defamatory falsehood.”⁵⁵ This interest, the Court explained, emanated from the importance of protecting individuals’ reputations.⁵⁶ In resolving the “tension” between this interest and freedom of speech, the Court sought “breathing space” for the right to free speech by bestowing “strategic protection” under the *New York Times* standard.⁵⁷ The Court distinguished *New York Times* on two grounds. First, public officials and figures are better able to engage in counterspeech, whereas private individuals find it more difficult to refute published falsehoods.⁵⁸ Second, public officials and figures, unlike private individuals, voluntarily assume the risk of being subject to falsehoods.⁵⁹ Additionally, because the Court “require[d] that state remedies for defamatory falsehood reach no farther than is necessary to protect the legitimate interest involved”⁶⁰ in order to balance “compensating private individuals for wrongful injury to reputation”⁶¹ with “the constitutional command of the First Amendment,”⁶² it held unconstitutional punitive damages awarded with no actual malice.⁶³

In devising the rules governing defamation claims, and in other areas of First Amendment doctrine, the Supreme Court has engaged in a methodology of constitutional reasoning grounded in optimizing practical results. As Professor Richard Fallon explains, in developing various areas of constitutional doctrine, the Supreme Court must make determinations about empirical matters that inform the rules it crafts.⁶⁴ In *New York Times* and *Gertz*, Fallon recounts, the Court did not merely “bal-

⁵² 418 U.S. 323 (1974).

⁵³ *Id.* at 347.

⁵⁴ *Id.* at 340.

⁵⁵ *Id.* at 341.

⁵⁶ *Id.* (citing *Rosenblatt v. Baer*, 383 U.S. 75, 92 (1966) (Stewart, J., concurring)).

⁵⁷ *Id.* at 342 (quoting *NAACP v. Button*, 371 U.S. 415, 433 (1963)).

⁵⁸ *Id.* at 344.

⁵⁹ *Id.* at 345.

⁶⁰ *Id.* at 349.

⁶¹ *Id.* at 348.

⁶² *Id.* at 349.

⁶³ *Id.* at 349–50. The Court held that awarding punitive damages, compensation *beyond* actual injury, was both less valuable and more prone to abuse. *Id.* at 350.

⁶⁴ Richard H. Fallon, Jr., *The Supreme Court, 1996 Term — Foreword: Implementing the Constitution*, 111 HARV. L. REV. 54, 62 (1997).

ance, in an abstract way,” freedom of speech and the interest undergirding defamation law.⁶⁵ Instead, it also made “more concrete, empirical, and predictive assessments” regarding the “proclivity of the press to engage in self-censorship under alternative liability regimes,” “the proportion of truthful and untruthful assertions that would be chilled by such regimes,” “the harms that would be done by false speech,” and “the benefits of truthful speech that would be forgone under various imaginable rules.”⁶⁶ More dramatically, Professor Daniel Farber identifies *New York Times* as an example of the notion that “First Amendment doctrines reflect the fear that certain laws overdeter speech and thus lead to a suboptimal amount of total information disseminated in society,”⁶⁷ in order to demonstrate that First Amendment doctrines embody “public choice theory — that is, the application of economics methodology to political institutions.”⁶⁸ Finally, implementing this policy-based method of constitutional reasoning often involves what Professor David Faigman terms “constitutional fact-finding,” the Court’s use of empirical claims to create constitutional law.⁶⁹ As Fallon agrees, *New York Times* and *Gertz* are not “atypical in their reliance on empirical, predictive calculations,”⁷⁰ and Faigman demonstrates that the Supreme Court routinely makes assumptions about empirical propositions to support constitutional decisionmaking.⁷¹

In employing this practical optimization methodology in *New York Times*, the Court was comfortable calibrating a rule for public officials that intentionally “overenforce[s]” constitutional goals.⁷² Indeed, as Professor David Strauss observes, in constitutional law, prophylactic rules are both ubiquitous and necessary.⁷³ Strauss notes that from *Miranda* warnings to strict scrutiny, constitutional law is replete with rules aimed at protecting rights through overenforcement.⁷⁴ Expressly building on Strauss’s foundation, Professor Daryl Levinson identifies “[d]efamation law [as] another clear example of a First Amendment prophylactic

⁶⁵ *Id.* at 63.

⁶⁶ *Id.* This accommodation is evidenced in *Gertz*’s tailoring of its rule to require actual malice for punitive damages.

⁶⁷ Daniel A. Farber, *Free Speech Without Romance: Public Choice and the First Amendment*, 105 HARV. L. REV. 554, 568 (1991).

⁶⁸ *Id.* at 555.

⁶⁹ See David L. Faigman, “Normative Constitutional Fact-Finding”: Exploring the Empirical Component of Constitutional Interpretation, 139 U. PA. L. REV. 541, 550 (1991).

⁷⁰ Fallon, *supra* note 64, at 63.

⁷¹ See Faigman, *supra* note 69, at 550.

⁷² Fallon, *supra* note 64, at 63, 65.

⁷³ David A. Strauss, *The Ubiquity of Prophylactic Rules*, 55 U. CHI. L. REV. 190, 190 (1988).

⁷⁴ *Id.* at 205, 208, 209.

rule.”⁷⁵ Agreeing that prophylactic rules are ubiquitous, Levinson explains that constitutional rules necessarily “depend on such factors as the administrability and expense of a more precise rule and the error costs of false negatives and false positives.”⁷⁶

The practical optimization the Supreme Court employed in *New York Times* and *Gertz* to calibrate such a First Amendment prophylactic rule suggests that the constitutionality of internet intermediary defamation liability should be assessed along two dimensions that mirror the analysis in those cases: the degree to which this type of defamation liability, first, impinges on protected speech and, second, promotes a governmental interest. Those cases addressed the First Amendment constraints on setting mental states for defamation liability, whereas this Note employs their framework to promote First Amendment constraints on secondary liability for defamation. This Note contends that the censorship that would result from internet intermediary liability for defamation cannot be saved by the government’s interest in imposing liability. In contrast to scholars and jurists who have paid these First Amendment questions relatively little attention, this Note intends to demonstrate the constitutional relevance of the policy-based arguments in favor of § 230, though this Note does not itself engage in a full-fledged policy analysis.

B. Collateral Censorship

Without § 230 as the constitutional rule, internet intermediaries would limit a significant amount of constitutionally protected speech. The *New York Times* Court feared that without the requirement of actual malice, “would-be critics of official conduct” would hesitate to speak.⁷⁷ Internet intermediary liability implicates a specific variety of self-censorship — collateral censorship — which the *New York Times* Court explained by quoting *Smith v. California*⁷⁸ at length.⁷⁹ What Professor Jack Balkin has termed “collateral censorship” arises not when individuals limit their *own* speech based on a fear of liability, but rather “when *A* censors *B* out of fear that the government will hold *A* liable for the effects of *B*’s speech.”⁸⁰ In *Smith*, the Court held unconstitutional an ordinance that prohibited bookstores from possessing obscene books.⁸¹ In rejecting that strict liability rule, the Court explained that many “legal devices and doctrines, in most

⁷⁵ Daryl J. Levinson, *Rights Essentialism and Remedial Equilibration*, 99 COLUM. L. REV. 857, 902 n.186 (1999).

⁷⁶ *Id.* at 904.

⁷⁷ *New York Times Co. v. Sullivan*, 376 U.S. 254, 279 (1964).

⁷⁸ 361 U.S. 147 (1959).

⁷⁹ *New York Times*, 376 U.S. at 278–79 (quoting *Smith*, 361 U.S. at 153–54).

⁸⁰ J.M. Balkin, Essay, *Free Speech and Hostile Environments*, 99 COLUM. L. REV. 2295, 2296 (1999).

⁸¹ *Smith*, 361 U.S. at 148, 155.

applications consistent with the Constitution, . . . cannot be applied in settings where they have the collateral effect of inhibiting the freedom of expression, by making the individual the more reluctant to exercise it.”⁸² While obscenity is not protected by the First Amendment, the ordinance’s lack of a scienter requirement jeopardized citizens’ access to a variety of protected speech.⁸³ *New York Times* quoted from the following key passage⁸⁴:

For if the bookseller is criminally liable without knowledge of the contents, and the ordinance fulfills its purpose, he will tend to restrict the books he sells to those he has inspected; and thus the State will have imposed a restriction upon the distribution of constitutionally protected as well as obscene literature. . . . And the bookseller’s burden would become the public’s burden, for by restricting him the public’s access to reading matter would be restricted. . . . The bookseller’s limitation in the amount of reading material with which he could familiarize himself, and his timidity in the face of his absolute criminal liability, thus would tend to restrict the public’s access to forms of the printed word which the State could not constitutionally suppress directly. The bookseller’s self-censorship, compelled by the State, would be a censorship affecting the whole public, hardly less virulent for being privately administered.⁸⁵

As in *Smith*, exposing internet intermediaries to liability for defamation communicated by their users would lead to collateral censorship.

First, content moderation to cope with intermediary liability is difficult, and therefore costly.⁸⁶ When a website confronts potentially defamatory user-generated content, it must resolve questions of both law and fact. As to questions of law, there is no national law of defamation but instead a fifty-state patchwork.⁸⁷ Therefore, websites must resolve the choice of law inquiry regarding which state’s law applies and then determine what that state’s rule is.⁸⁸ Moreover, defamation law abounds with privileges and exceptions. Even if a website determined that certain content would support a prima facie case for defamation, it would still need to determine the applicability of various privileges and

⁸² *Id.* at 150–51.

⁸³ *Id.* at 153.

⁸⁴ *New York Times*, 376 U.S. at 278–79 (quoting *Smith*, 361 U.S. at 153–54).

⁸⁵ *Smith*, 361 U.S. at 153–54 (footnote omitted).

⁸⁶ Aaron Perzanowski, Comment, *Relative Access to Corrective Speech: A New Test for Requiring Actual Malice*, 94 CALIF. L. REV. 833, 858 n.172 (2006).

⁸⁷ James R. Pielemeier, *Constitutional Limitations on Choice of Law: The Special Case of Multistate Defamation*, 133 U. PA. L. REV. 381, 384–391 (1985).

⁸⁸ See *id.* at 391; Philip Adam Davis, Note, *The Defamation of Choice-of-Law in Cyberspace: Countering the View that the Restatement (Second) of Conflict of Laws Is Inadequate to Navigate the Borderless Reaches of the Intangible Frontier*, 54 FED. COMM. L.J. 339, 340–42 (2002); Corey Omer, Note, *Intermediary Liability for Harmful Speech: Lessons from Abroad*, 28 HARV. J.L. & TECH. 289, 316–18 (2014).

exceptions.⁸⁹ Questions of fact are also difficult for websites to resolve, involving “considerable costs of investigation.”⁹⁰ For example, a statement that a business often fails to meet its commercial obligations is not easily verifiable. To the extent that it is difficult for judges and juries to determine the truthfulness of potentially defamatory statements, it is even more difficult for intermediary websites to do so.⁹¹ Even upon receiving notice that a statement is allegedly defamatory, a website does not know whether a complainant is correct or merely hoping to illegitimately induce takedown.⁹² In the copyright context, a large number of takedown requests to websites are illegitimate.⁹³ Some websites have experimented with artificial intelligence algorithms to moderate content.⁹⁴ However, algorithms have struggled to correctly moderate content: for example, differentiating between impermissible nudity and fine art.⁹⁵ It would be even more difficult for artificial intelligence to properly identify defamation and quite costly to develop that software. And humans are not happy performing the task.⁹⁶ It is difficult to quickly determine whether certain speech is merely critical or actionable defamation. These difficulties are amplified by the volume of content websites face. As *Zeran* recognized about moderating “millions of postings,”⁹⁷ “[a]lthough this might be feasible for the traditional print publisher, the sheer number of postings on interactive computer services

⁸⁹ Cf. Meera Nair, *Adjudication by Algorithm*, FAIR DUTY (Jan. 3, 2018, 8:33 AM), <https://fair-duty.wordpress.com/2018/01/03/adjudication-by-algorithm/> [<https://perma.cc/BQ5U-WHF6>] (explaining that in the copyright context, the “entire list of exceptions is extensive and should be part of any algorithmic effort to” moderate and remove potentially copyrighted content).

⁹⁰ See Perzanowski, *supra* note 86, at 858 n.172.

⁹¹ See Felix T. Wu, *Collateral Censorship and the Limits of Intermediary Immunity*, 87 NOTRE DAME L. REV. 293, 301 (2011); Paul Sieminski & Holly Hogan, *Why (Allegedly) Defamatory Content on WordPress.com Doesn't Come Down Without a Court Order*, TECHDIRT (Feb. 7, 2018, 1:32 PM), <https://www.techdirt.com/articles/20180206/10271639166/why-allegedly-defamatory-content-wordpresscom-doesnt-come-down-without-court-order.shtml> [<https://perma.cc/46P7-QCKY>].

⁹² See Seth F. Kreimer, *Censorship by Proxy: The First Amendment, Internet Intermediaries, and the Problem of the Weakest Link*, 155 U. PA. L. REV. 11, 86 & n.238 (2006).

⁹³ See *Takedown Hall of Shame*, ELECTRONIC FRONTIER FOUND., <https://www EFF.org/takedowns> [<https://perma.cc/PKB9-PR3N>].

⁹⁴ See *Are Algorithms the Future of Content Moderation?*, WEBPURIFY (July 23, 2015) <https://www.webpurify.com/blog/algorithms-future-content-moderation/> [<https://perma.cc/Z3C7-J9LD>].

⁹⁵ *Id.*; see also Nair, *supra* note 89; NATASHA DUARTE ET AL., CTR. FOR DEMOCRACY & TECH., MIXED MESSAGES? THE LIMITS OF AUTOMATED SOCIAL MEDIA CONTENT ANALYSIS 5 (2017) <https://cdt.org/files/2017/11/Mixed-Messages-Paper.pdf> [<https://perma.cc/M7Z7-PP5K>]; Rhett Jones, *Man's YouTube Video of White Noise Hit with Five Copyright Claims*, GIZMODO (Jan. 5, 2018, 10:05 AM), <https://gizmodo.com/man-s-youtube-video-of-white-noise-hit-with-five-copyri-1821804093> [<https://perma.cc/882V-LXGS>].

⁹⁶ 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, 10:42 PM), <https://www.wsj.com/articles/the-worst-job-in-technology-staring-at-human-depravity-to-keep-it-off-facebook-1514398398> [<https://perma.cc/JG62-63XN>].

⁹⁷ *Zeran v. Am. Online, Inc.*, 129 F.3d 327, 331 (4th Cir. 1997).

would create an impossible burden in the Internet context.”⁹⁸ Efforts to surmount these difficulties, and thus increase the accuracy of moderation to avoid intermediary liability, would be costly because those efforts require investments in labor, time, or technology.

Second, as *Smith* recognized, the difficulties and costs created by intermediary liability would cause many websites to engage in various forms of collateral censorship — often the least costly method of avoiding liability.⁹⁹ In general, websites would err on the side of caution, defaulting to removing allegedly defamatory content instead of engaging in costly legal and factual investigation.¹⁰⁰ The cost to websites of collaterally censoring is very low, whereas the cost of not censoring content is much higher because that decision risks expensive litigation and adverse judgments.¹⁰¹ Websites “may be deterred from” permitting certain content, as *New York Times* explained, “even though it is believed to be true and even though it is in fact true, because of doubt whether it can be proved in court or fear of the expense of having to do so.”¹⁰² Individual website employees are unlikely to face repercussions for playing it safe but could face ramifications for allowing content that later leads to litigation expenses. Whether or not websites believe a potential lawsuit is meritorious, they will often default to removal because of the potential costs of litigation or an adverse result.¹⁰³ Even websites, like Facebook, that can “afford” high moderation and litigation costs would still prefer to avoid them, and this judgment will likely influence their moderation. Therefore, in the words of *New York Times*, websites would tend to permit “only statements which ‘steer far wider of the unlawful zone.’”¹⁰⁴

More generally, some websites might decide not to allow entire categories of content that will be more likely to expose them to liability. For example, politically controversial speech or business and product reviews may be more likely to lead to defamation actions than more mundane content.¹⁰⁵ Or bloggers might decline to include a comment section.¹⁰⁶

Worse still, some websites might never launch.¹⁰⁷ Because of their business models, perhaps to focus solely on particularly controversial content, the anticipated costs of moderation and litigation could prevent

⁹⁸ *Id.* at 333 (citing *Auvil v. CBS 60 Minutes*, 800 F. Supp. 928, 931 (E.D. Wash. 1992)).

⁹⁹ See *Smith v. California*, 361 U.S. 147, 153 (1959) (noting that imposing liability on the bookseller would decrease the books on offer).

¹⁰⁰ Perzanowski, *supra* note 86, at 858 n.172.

¹⁰¹ *Id.*

¹⁰² *New York Times Co. v. Sullivan*, 376 U.S. 254, 279 (1964).

¹⁰³ *Zeran*, 129 F.3d at 333.

¹⁰⁴ 376 U.S. at 279 (quoting *Speiser v. Randall*, 357 U.S. 513, 526 (1958)).

¹⁰⁵ Balkin, *supra* note 38, at 436.

¹⁰⁶ *Id.*

¹⁰⁷ *Id.*

them from even securing capital or launching.¹⁰⁸ This issue might be termed *complete* collateral censorship — where an intermediary fails to come into existence because of a fear of being held liable for the speech of others. Various websites credit § 230 with their very existence.¹⁰⁹

Additional collateral censorship will result from mistakes. Because the imposition of liability would lead to more moderation and removal, websites are more likely to make mistakes in removal decisions. Websites may make technical mistakes (perhaps from a user's accidental clicking of a "report" button). But given the difficulty of factual investigation, they are also likely to make fundamental mistakes about the factual basis of defamation claims — removing content based on incorrect understandings of the veracity of users' allegations. Moreover, websites will make mistakes of law. Fearing these mistakes, websites may default to adherence to the strictest state laws, thus censoring more speech and allowing the most speech-restrictive states to govern the entire internet. If websites employ algorithms to shoulder this legal burden, they expose themselves to the inaccuracies in those programs.

Due to the problems noted above, opportunistic lawyers or other individuals will attempt to exploit websites' vulnerabilities. Businesses and individuals that do not like posts about them on websites will request that the posts be taken down whether they are defamatory or not.¹¹⁰ Individuals and businesses hoping to have material taken down will learn how to manipulate intermediaries.¹¹¹ Websites would face difficulties dealing with even good faith reports of defamation, let alone handling individuals who allege defamation as a cynical tactic to remove the content they dislike.¹¹² If a business wants to hide a bad review or an individual hopes to conceal a piece of truthful but unflattering information, the business or individual can notify the website that the content is false and threaten to sue. Even if a website does not immediately capitulate, it will incur large costs investigating these claims and may reach the incorrect conclusion. During the investigation period, the

¹⁰⁸ See MATTHEW LE MERLE ET AL., BOOZ & CO., THE IMPACT OF U.S. INTERNET COPYRIGHT REGULATIONS ON EARLY-STAGE INVESTMENT 19 (2011), <https://www.strategyand.pwc.com/media/uploads/Strategyand-Impact-US-Internet-Copyright-Regulations-Early-Stage-Investment.pdf> [<https://perma.cc/FP3E-NW3C>]; Jerry Berman, *Policy Architecture and Internet Freedom*, LAW.COM: THE RECORDER (Nov. 10, 2017, 3:53 AM), <https://www.law.com/therecorder/sites/therecorder/2017/11/10/policy-architecture-and-internet-freedom/> [<https://perma.cc/T9C5-8PNB>] ("Without § 230 . . . speech would be limited and new applications might never have emerged if required to finance costly legal overhead to do business on the Internet.")

¹⁰⁹ See, e.g., Adi Kamdar, *CDA 230 Success Cases: Wikipedia*, ELECTRONIC FRONTIER FOUND. (July 26, 2013), <https://www.eff.org/deeplinks/2013/07/cda-230-success-cases-wikipedia> [<https://perma.cc/6CTN-A33M>].

¹¹⁰ See, e.g., Christina Mulligan, *Technological Intermediaries and Freedom of the Press*, 66 SMU L. REV. 157, 182 (2013); Wu, *supra* note 91, at 301; Sieminski & Hogan, *supra* note 91.

¹¹¹ Perzanowski, *supra* note 86, at 858 n.172.

¹¹² See *id.*

website may take down the content, which would also inhibit speech. For potentially defamatory posts, websites might decide to implement a delay so that they can prescreen content for defamation.

For these reasons, notice-based liability is problematic. As then-Chief Judge Wilkinson explained in *Zeran*, “liability upon notice has a chilling effect on the freedom of Internet speech” “[b]ecause service providers would be subject to liability only for the publication of information, and not for its removal, [so] they would have a natural incentive simply to remove messages upon notification, whether the contents were defamatory or not.”¹¹³

Third, the nondefamatory speech lost to collateral censorship is often valuable. In cases like *Reno v. ACLU*,¹¹⁴ the Supreme Court has demonstrated an appreciation for the vital role internet speech plays in modern society. The Court lauded the then-nascent internet’s “vast democratic forums.”¹¹⁵ It described the internet as a “dynamic, multifaceted category of communication includ[ing] not only traditional print and news services, but also audio, video, and still images, as well as interactive, real-time dialogue.”¹¹⁶ It noted that “any person with [internet access] can become a town crier with a voice that resonates farther than it could from any soapbox.”¹¹⁷ In addition, the Court observed that because of the tremendous scale of the internet, speech regulations that threatened liability for certain acts could limit many types of protected speech.¹¹⁸ More recently, in *Packingham v. North Carolina*,¹¹⁹ the Supreme Court held unconstitutional a statute that prohibited registered sex offenders from accessing social networking websites, like Facebook or Twitter, that allow children to have accounts.¹²⁰ The Court explained that “to foreclose access to social media altogether is to prevent the user from engaging in the legitimate exercise of First Amendment rights.”¹²¹ It deemed the internet “the *most important* place[] (in a spatial sense) for the exchange of views.”¹²² The Court continued that an understanding of the internet “informs the analysis”¹²³ of a law in question:

Social media offers “relatively unlimited, low-cost capacity for communication of all kinds.” On Facebook, for example, users can debate religion

¹¹³ *Zeran v. Am. Online, Inc.*, 129 F.3d 327, 333 (4th Cir. 1997); see Sieminski & Hogan, *supra* note 91.

¹¹⁴ 521 U.S. 844 (1997).

¹¹⁵ *Id.* at 868.

¹¹⁶ *Id.* at 870.

¹¹⁷ *Id.*

¹¹⁸ *Id.* at 875–78.

¹¹⁹ 137 S. Ct. 1730 (2017).

¹²⁰ *Id.* at 1733, 1738.

¹²¹ *Id.* at 1737.

¹²² *Id.* at 1735 (emphasis added).

¹²³ *Id.* at 1736.

and politics with their friends and neighbors or share vacation photos. On LinkedIn, users can look for work, advertise for employees, or review tips on entrepreneurship. And on Twitter, users can petition their elected representatives and otherwise engage with them in a direct manner. . . . In short, social media users employ these websites to engage in a wide array of protected First Amendment activity on topics “as diverse as human thought.”

. . . While we now may be coming to the realization that the Cyber Age is a revolution of historic proportions, we cannot appreciate yet its full dimensions and vast potential to alter how we think, express ourselves, and define who we want to be.¹²⁴

The Supreme Court’s veneration of internet speech suggests special caution before permitting laws that limit it.¹²⁵

More specifically, the nondefamatory speech lost to collateral censorship will often be vulnerable speech.¹²⁶ Individuals who want certain speech taken down sometimes file illegitimate content takedown requests.¹²⁷ This dynamic allows the majority to suppress minority views or could constitute a potential heckler’s veto.¹²⁸ The speech that is the first to be collaterally censored may be the most vulnerable and least likely to appear through alternative channels. At its core, the First Amendment seeks to protect unpopular views¹²⁹ — unobjectionable views are less frequently jeopardized. As noted above, because of the cost of additional content moderation, some websites may turn to algorithms for assistance. Yet recently, algorithms have fared no better in protecting marginalized speech: Google’s artificial intelligence moderation system that seeks to highlight toxic speech accidentally flags sentences such as “I am a gay woman.”¹³⁰

Other vulnerable speech includes speech of little immediate personal benefit but that, when part of a community, provides a large public benefit — such as business reviews or Wikipedia edits. Some of the most socially beneficial forms of speech that can pose defamation concerns

¹²⁴ *Id.* at 1735–36 (citations omitted).

¹²⁵ *Id.*

¹²⁶ If websites default to accepting a mere prima facie case for defamation as grounds to censor content, some of the collaterally censored speech will be communication that fits into the various well-justified exceptions and privileges to defamation claims.

¹²⁷ See Jeffrey Cobia, Note, *The Digital Millennium Copyright Act Takedown Notice Procedure: Misuses, Abuses, and Shortcomings of the Process*, 10 MINN. J.L. SCI. & TECH. 387, 391 (2009).

¹²⁸ Brett G. Johnson, *The Heckler’s Veto: Using First Amendment Theory and Jurisprudence to Understand Current Audience Reactions Against Controversial Speech*, 21 COMM. L. & POL’Y 175, 176–77 (2016).

¹²⁹ See JOHN HART ELY, DEMOCRACY AND DISTRUST 112 (1980); *Turner Broad. Sys., Inc. v. FCC*, 512 U.S. 622, 641 (1994).

¹³⁰ Elliot Harmon & Jeremy Gillula, *Stop SESTA: Whose Voices Will SESTA Silence?*, ELECTRONIC FRONTIER FOUND. (Sept. 13, 2017), <https://www.eff.org/deeplinks/2017/09/stop-sesta-whose-voices-will-sesta-silence> [<https://perma.cc/A5KG-NE8A>].

are consumer reviews, such as those on Yelp. These websites have flourished because of § 230.¹³¹ Facing liability, review websites would become more cautious and manipulable, and therefore less accurate, thus decreasing competition. Nonprofits like Wikipedia also depend on § 230 to freely provide accurate content.¹³²

Ultimately, the threat of defamation liability will often cause websites to seek to avoid liability by overcensoring valuable user speech.

C. Interest

The second area of First Amendment analysis concerns the government's interest underlying defamation law. In *Gertz*, the Court held that the "legitimate state interest underlying the law of libel is the compensation of individuals for the harm inflicted on them by defamatory falsehood."¹³³ However, the Court articulated a rationale for the compensation interest that spoke to a broader purpose: each individual has the "right to the protection of his own good name."¹³⁴ This reputational rationale is broader than the interest in compensation because it undergirds a larger swath of defamation law. For example, a reform that would increase only the deterrent effect of defamation law could not be supported by the compensation interest because that reform would not necessarily increase the likelihood of compensation; however, it would certainly promote the reputational rationale by decreasing the prevalence of defamation through deterrence.

In general, a reputational *interest* is a much more natural understanding of the justification for defamation law. The Court should adopt reputation protection, which involves deterrence, not mere compensation, as the interest justifying defamation laws. As the Court explained in *Rosenblatt v. Baer*,¹³⁵ "underl[ying] the law of defamation [is an] interest in *preventing* and redressing attacks upon reputation."¹³⁶ Would one prefer an ideal world in which every victim of defamation was compensated or one in which defamation law deterred all defamation before it took place, thus protecting all individuals' reputations? More realistically, the objective of defamation law should be reducing instances of defamation as much as possible while compensating individuals who are nonetheless defamed.¹³⁷ Analogously, the interest underlying "battery" is not merely securing a remedy for those who have

¹³¹ *CDA § 230 Success Case: Yelp*, ELECTRONIC FRONTIER FOUND., <https://www EFF.org/issues/cda230/successes/yelp> [<https://perma.cc/THP3-V64S>].

¹³² Kamdar, *supra* note 109.

¹³³ *Gertz v. Robert Welch, Inc.*, 418 U.S. 323, 341 (1974).

¹³⁴ *Id.* (citing *Rosenblatt v. Baer*, 383 U.S. 75, 92 (1966) (Stewart, J., concurring)).

¹³⁵ 383 U.S. 75.

¹³⁶ *Id.* at 86 (emphasis added).

¹³⁷ To be sure, this consideration might require a tradeoff between overall reductions in defamation and increases in the proportion of those who are defamed but do not receive compensation.

been battered but also reducing the occurrence of that tortious action.¹³⁸ This distinction matters because it expands the denominator: if one contemplates a broader interest than compensation alone, different laws may pass or fail constitutional muster. For instance, as argued below, § 230 does limit compensation, but the law mitigates this limitation because it encourages websites to remove defamation. The net effect on a general reputational interest is greater than the effect on compensation. When a legitimate interest is artificially narrowed, it can promote the constitutionality of laws that could fail as rights infringing under a more naturally broad interest.¹³⁹

Intermediary defamation liability does not serve this interest well because it would not significantly reduce defamation beyond the status quo. First, in the status quo, many websites moderate their content and remove defamatory content even without the threat of intermediary liability.¹⁴⁰ They make this decision because of “a sense of corporate social responsibility, but also, more importantly, because their economic viability depends on meeting users’ speech and community norms.”¹⁴¹ Websites have significant existing incentives to remove defamatory material. And, “[b]ecause they seek to please their customers, intermediaries are more likely than courts to develop content standards that conform to basic community values.”¹⁴² Second, some defamation may be persistent in the face of intermediary liability. Consider, for instance, the extreme amount of copyright infringement that persists on the internet even though federal law imposes liability on intermediaries for copy-

¹³⁸ See John C.P. Goldberg, *Twentieth-Century Tort Theory*, 91 GEO. L.J. 513, 525 (2003).

¹³⁹ In a sense, allowing for a broader view of the interest at stake requires policies that can accomplish the same “level” of interest fulfillment but with less rights abridgement.

¹⁴⁰ Klonick, *supra* note 42, at 1601, 1615. Some may wonder: If websites will collaterally censor when facing intermediary liability, why will websites not be effective in taking down defamation? And if websites already moderate, why is there not already collateral censorship such that imposing liability will not make a dramatic difference? This Note acknowledges both that intermediary liability may lead some websites to take down marginally more defamation and also that websites already limit the constitutionally protected speech of their users. However, it also posits that while websites already target defamation, increased liability will increase the proportion of protected speech that is removed in an effort to reduce defamation. Websites already make significant efforts to remove defamation. As liability and then moderation increase, there would be diminishing marginal returns to defamation reduction and correspondingly increasing marginal collateral censorship of constitutionally protected content. In order to be sure all defamatory content was removed, websites would remove much lawful content. Therefore, as the interest in defamation decelerates, collateral censorship accelerates. If intermediary liability effectively removes a significant amount of defamation, it would come at the cost of very dramatic collateral censorship. On the other hand, a website’s *independent* decision to take down content that is not illegal is not collateral censorship but merely an editorial decision (at least from a constitutional perspective).

¹⁴¹ *Id.* at 1625.

¹⁴² Kosseff, *supra* note 40, at 153.

right infringement committed by their users.¹⁴³ Persistent users will often be able to disseminate whatever information they want by using multiple accounts, anonymous accounts, or other websites. Certain bad-actor websites will also persist by remaining outside the jurisdiction of U.S. courts.¹⁴⁴ Third, intermediary liability could lead to less of a reduction in defamation because some websites will meet the “Moderator’s Dilemma”¹⁴⁵ posed by *Stratton Oakmont* by taking a more hands-off approach to content. In other words, instead of attempting to avoid liability by overcensoring their users, they will reduce the screening they engage in to avoid acquiring knowledge that might subject them to liability.¹⁴⁶ If they otherwise would have moderated content and removed some defamation, this choice renders defamation law *less* effective.

Those who have been defamed still retain various tools that may mitigate the harms of defamation. Section 230 does not prevent a defamed person from engaging in counterspeech.¹⁴⁷ Nor does it prevent plaintiffs from suing the party that originally defamed them.¹⁴⁸ In fact, an empirical study found that in a majority of § 230 cases, plaintiffs “were able to identify and sue the original source of the content that caused them harm.”¹⁴⁹ Additionally, the same study revealed that even if potential plaintiffs do not recover in court, they are often successful in getting the content in question removed.¹⁵⁰ While these options are sometimes of limited efficacy, they are at minimum marginally mitigating.

The considerable collateral censorship that intermediary liability would cause is not worth the meager benefit to the reputational interest such liability might provide. The fact that all plaintiffs could not achieve compensation is insufficient to reject this rule — *New York Times* has the same consequence. As the Court there explained, “erroneous statement is inevitable in free debate, and . . . it must be protected if the freedoms of expression are to have the ‘breathing space’ that they ‘need . . . to survive.’”¹⁵¹ The Court creates broad prophylactic rules,

¹⁴³ See Annemarie Bridy, *Is Online Copyright Enforcement Scalable?*, 13 VAND. J. ENT. & TECH. L. 695, 709 (2011).

¹⁴⁴ See Alan M. Trammell & Derek E. Bambauer, *Personal Jurisdiction and the “Interwebs,”* 100 CORNELL L. REV. 1129, 1188–89 (2015).

¹⁴⁵ Eric Goldman, *Congress Probably Will Ruin Section 230 This Week (SESTA/FOSTA Updates)*, TECH. & MARKETING L. BLOG (Feb. 26, 2018), <https://blog.ericgoldman.org/archives/2018/02/congress-probably-will-ruin-section-230-this-week-sestafosta-updates.htm> [https://perma.cc/NCH4-JRDG].

¹⁴⁶ *Zeran v. Am. Online, Inc.*, 129 F.3d 327, 333 (4th Cir. 1997).

¹⁴⁷ Perzanowski, *supra* note 86, at 860–61.

¹⁴⁸ See *Zeran*, 129 F.3d at 330 (“None of this means, of course, that the original culpable party who posts defamatory messages would escape accountability.”).

¹⁴⁹ Ardia, *supra* note 24, at 382; see *id.* at 486–88, 493.

¹⁵⁰ *Id.* at 489.

¹⁵¹ *New York Times Co. v. Sullivan*, 376 U.S. 254, 271–72 (1964) (second omission in original) (quoting *NAACP v. Button*, 371 U.S. 415, 433 (1963)).

“breathing space,” to protect the freedom of expression through intentional overenforcement of the constitutional right.¹⁵² *Gertz* consciously devised an “accommodation of the competing values at stake in defamation suits,”¹⁵³ and “attempt[ed] to reconcile state law with a competing interest grounded in the constitutional command of the First Amendment.”¹⁵⁴ To this analysis must be added the Court’s more recent statements on the importance of internet speech and the need for restraint in regulating it.¹⁵⁵ Given the new “relationship between the First Amendment and the modern Internet,” the Court has warned that it “must exercise extreme caution before suggesting that the First Amendment provides scant protection.”¹⁵⁶ For the First Amendment, intermediary liability imperils a significant amount of constitutionally protected speech through the collateral censorship explained above. Collateral censorship may be even more troublesome than the self-censorship feared in *New York Times* because the censored speakers do not themselves decide when to refrain from speaking.¹⁵⁷ For the interest in enforcing defamation law, imposing intermediary liability will be of limited utility because websites already moderate content, much defamation will persist in the face of intermediary liability, and intermediary liability might encourage some websites to decrease their moderation. The Court must require confidence in the benefits of the defamation law, especially when the speech at stake may be so valuable. Here, the gains for defamation law are doubtful whereas the harms to speech are significant. Therefore, under the Court’s defamation, collateral censorship, and internet speech case law, the First Amendment requires the prophylactic rule of § 230.

Applying the First Amendment in the untrodden ground of (1) internet (2) intermediary (3) defamation liability combines three areas of doctrine. By (1) recognizing the value *and* vulnerability of internet speech (*Reno* and *Packingham*), (2) identifying the First Amendment harm — collateral censorship — that intermediary liability imposes (*Smith*), and (3) employing the framework the Court uses to evaluate the constitutionality of defamation laws (*New York Times* and *Gertz*), the optimal constitutional rule comes into focus. To be sure, *Packingham* merely lauded internet speech, *Smith* rejected only strict liability, and *New York Times* calibrated a mental state (actual malice) and not secondary liability. However, § 230’s rule is the best extension of these precedents into the new context of internet intermediary defamation, for the reasons detailed above.

¹⁵² See Fallon, *supra* note 64, at 63; Levinson, *supra* note 75, at 902 & n.186.

¹⁵³ *Gertz v. Robert Welch, Inc.*, 418 U.S. 323, 348 (1974).

¹⁵⁴ *Id.* at 349.

¹⁵⁵ See *Packingham v. North Carolina*, 137 S. Ct. 1730, 1735–36 (2017).

¹⁵⁶ *Id.* at 1736.

¹⁵⁷ Wu, *supra* note 91, at 304.

D. Section 230's Critics

By way of framing potential critiques of § 230, as Cathy Gellis brilliantly explains, “§ 230 is potentially in jeopardy of becoming a victim of its own success,” because its benefits are less salient than are particular instances of defamation.¹⁵⁸ As she notes, “§ 230 has done so well creating a new normalcy that it’s much harder to see just how much it has allowed to go right,” such that “when things do go wrong . . . we are always at risk of letting our outrage at the specific injustice cause us to be tempted to kill the golden goose by upending something that on the whole has enabled so much good.”¹⁵⁹

Some might argue that § 230 unacceptably creates a different constitutional standard for online, versus offline, speech.¹⁶⁰ However, the proposed rule would be equally desirable in truly analogous offline contexts. More importantly, the Court has been willing to set different rules under the First Amendment for different forms of media based on their different factual contexts.¹⁶¹ The Court treats the regulation of adult content, for example, differently across different types of media such as newspapers, broadcast, and cable.¹⁶² More broadly, much of this line drawing is based on sound factual distinctions between various types of media. Here, for instance, internet intermediary liability would be less successful than offline intermediary liability in reducing defamation and is therefore less constitutionally desirable. And, as the Court has explained, given the relatively new “relationship between the First Amendment and the modern Internet,” it “must exercise extreme caution before suggesting that the First Amendment provides scant protection.”¹⁶³

Some critics of § 230 argue that the statute has unacceptable distributional consequences. Professor Mary Anne Franks, in particular, has written thoughtfully about the concern that § 230 may shield defamation that “disproportionately burden[s] vulnerable private citizens including women, racial and religious minorities, and the LGBT community.”¹⁶⁴ This Note accepts this claim. However, First Amendment doctrine is not necessarily concerned with disproportionately distributed harm¹⁶⁵ and

¹⁵⁸ Gellis, *supra* note 25.

¹⁵⁹ *Id.*

¹⁶⁰ See Jenna K. Stokes, *The Indecent Internet: Resisting Unwarranted Internet Exceptionalism in Combating Revenge Porn*, 29 BERKELEY TECH. L.J. 929, 930–31 (2014).

¹⁶¹ See Thomas W. Hazlett et al., *The Overly Active Corpse of Red Lion*, 9 NW. J. TECH. & INTELL. PROP. 51, 62 (2010).

¹⁶² See *id.* (comparing *Reno v. ACLU*, 521 U.S. 844 (1997), with *United States v. Playboy Entm’t Grp., Inc.*, 529 U.S. 803 (2000)).

¹⁶³ *Packingham v. North Carolina*, 137 S. Ct. 1730, 1736 (2017).

¹⁶⁴ Mary Anne Franks, *Moral Hazard on Stilts: Zeran’s Legacy*, LAW.COM: THE RECORDER (Nov. 10, 2017, 8:34 AM), <https://www.law.com/therecorder/sites/therecorder/2017/11/10/moral-hazard-on-stilts-zerans-legacy/> [<https://perma.cc/RE3X-SHHA>].

¹⁶⁵ See *R.A.V. v. City of St. Paul*, 505 U.S. 377, 391–93 (1992).

may be particularly skeptical of laws explicitly aimed at remedying it.¹⁶⁶ Yet the First Amendment should be particularly skeptical of laws that disproportionately hurt the *speech* of certain marginalized groups. Intermediary liability has this potential, as it would provide a heckler's veto to those who object to minority speech. Content moderation has "shut down conversations among women of color about the harassment they receive online," "censor[ed] women who share childbirth images in private groups," and "disappeared documentation of police brutality, the Syrian war, and the human rights abuses suffered by the Rohingya."¹⁶⁷ Intermediary liability would increase websites' incentive to cautiously accede to takedown requests targeting vulnerable private citizens. Liability may increase the use of moderation algorithms, and "[d]ecisions based on automated social media content analysis risk further marginalizing and disproportionately censoring groups that already face discrimination."¹⁶⁸ While marginalized communities may be particularly vulnerable to online defamation, they are also particularly vulnerable to the collateral censorship that would result from intermediary liability. In addition, even if a repeal of § 230 *would* generally benefit defamation plaintiffs, it is unclear whether *these* plaintiffs would benefit. Given the cost of litigation, our most marginalized citizens are the ones least likely to be able to take advantage of a new liability regime. Most importantly, as argued above, collateral censorship is a major threat to vulnerable voices online. Therefore, it is at best uncertain which regime has superior distributional consequences.

IV. THE IMPLICATIONS OF A CONSTITUTIONAL RULE

Several implications flow from the idea that the First Amendment requires internet intermediary liability protection. First, regardless of whether one is an internet exceptionalist,¹⁶⁹ this Note demonstrates how constitutional questions regarding the internet occasionally require unique answers at least due to dramatically changed factual circumstances. The volume of internet speech and its resistance to regulation produce a potentially surprising result for defamation law. Second, understanding § 230 as being equal to the constitutional requirement helps explain why courts have generally taken a broad view of the statute and consistently held

¹⁶⁶ See *Am. Booksellers Ass'n v. Hudnut*, 771 F.2d 323, 324–25 (7th Cir. 1985) (striking down an ordinance prohibiting certain explicit material considered as discriminating against women), *aff'd mem.*, 475 U.S. 1001 (1986).

¹⁶⁷ Corynne McSherry et al., *Private Censorship Is Not the Best Way to Fight Hate or Defend Democracy: Here Are Some Better Ideas*, ELECTRONIC FRONTIER FOUND. (Jan. 30, 2018), <https://www.eff.org/deeplinks/2018/01/private-censorship-not-best-way-fight-hate-or-defend-democracy-here-are-some> [https://perma.cc/ULV6-UX2Z].

¹⁶⁸ DUARTE ET AL., *supra* note 95, at 4.

¹⁶⁹ See generally Mark Tushnet, *Internet Exceptionalism: An Overview from General Constitutional Law*, 56 WM. & MARY L. REV. 1637 (2015).

against defamation claims. This realization also might explain why courts at first provided broad protection under the statute against defamation claims and then began to grow more reluctant in cases where *speech* seems less directly implicated, such as failure-to-warn claims. Third, recognizing the First Amendment as requiring § 230 shows how § 230 may be reminiscent of other federal statutes that would now likely constitute the rule required by the Constitution.¹⁷⁰ This type of statute demonstrates how Congress can enforce constitutional law prior to the courts and also how statutory experimentation can yield enduring norms. Fourth, in new cases on the edge of § 230's protections, this First Amendment underpinning provides a rationale, perhaps via constitutional avoidance, for interpreting immunity broadly. Fifth, § 230 covers more claims than defamation. If the First Amendment requires intermediary liability protection from defamation suits, other claims may also be implicated. Sixth, though this Note argues for shielding certain editorial decisions of websites, this legal argument should not preclude public debate regarding their practices. As discussed, many websites laudably expend resources seeking to remove defamation. But many websites should make more strides, seeking to provide a "fair opportunity to participate" and "direct accountability."¹⁷¹ Finally, if Congress amends or repeals § 230,¹⁷² courts should be willing to step in with the First Amendment if warranted.

This Note finds for § 230 enduring constitutional footing.¹⁷³ Given the risk of collateral censorship and meager gains in stopping defamation that an alternate rule would produce, the First Amendment cannot permit holding websites liable for the defamation of their users. When and if the time comes, courts should be willing to recognize the importance of this protection and hold it provided for by the Constitution.

¹⁷⁰ Consider the laws that govern administrative or civil procedure and procedural due process, the fair use doctrine and the First Amendment, and civil or voting rights laws and the Reconstruction Amendments.

¹⁷¹ Klonick, *supra* note 42, at 1603.

¹⁷² Eric Goldman, *Senate's "Stop Enabling Sex Traffickers Act of 2017" — And Section 230's Imminent Evisceration*, TECH. & MARKETING L. BLOG (July 31, 2017), <http://blog.ericgoldman.org/archives/2017/07/senates-stop-enabling-sex-traffickers-act-of-2017-and-section-230s-imminent-evisceration.htm> [<https://perma.cc/KF9B-TN7K>]; see also Elliot Harmon, *Amended Version of FOSTA Would Still Silence Legitimate Speech Online*, ELECTRONIC FRONTIER FOUND. (Dec. 11, 2017), <https://www.eff.org/deeplinks/2017/12/amended-version-fosta-would-still-silence-legitimate-speech-online> [<https://perma.cc/2YB4-ZYWL>].

¹⁷³ Some may argue that instead of basing this protection in First Amendment doctrine, the protection should be internal to defamation law. As a policy matter, these changes may be desirable as an *addition* to constitutional doctrine. However, crafting this constitutional rule is preferable to a solution based purely on the common law of torts. For one, a national constitutional rule gives more confidence to intermediaries and reduces litigation costs — thus decreasing the chance of collateral censorship.