Taming the Internet Pitchfork Mob: Online Public Shaming, the Viral Media Age, and the Communications Decency Act

ABSTRACT

Accompanying the explosive growth of the Internet, one lamentable trend is the rise of online public shaming. While online public shaming may positively incentivize individuals to modify their behavior in accordance with socially acceptable norms, there has also been the emergence of an online "pitchfork mob" that can have a real impact on individuals' livelihoods and overall wellbeing. Due to the lack of legal remedies available to victims of certain types of online shaming, this Note suggests that web hosts are empowered by the expansive protections of the Communications Decency Act to develop and implement policies to curb the prevalence of online public shaming.

TABLE OF CONTENTS

In 2013, Justine Sacco, former head of public relations for InterActiveCorp, forever shattered her career and personal life in roughly 140 characters.[1] With a mere 170 Twitter followers, Sacco carelessly posted a racist tweet regarding her upcoming trip to South Africa.[2] While undoubtedly in poor taste, the resulting firestorm sent shockwaves through the Internet.[3] Within hours of posting the fateful tweet, Sacco garnered rage that only the Internet could provide, and she became the number-one trending topic on Twitter with users fiercely calling for her resignation.[4] The fallout from Sacco's tweet is still infamous, even years later.[5] Sacco lost her job at InterActiveCorp, and despite her efforts to avoid social media, the same *Gawker* magazine journalist who re-tweeted Sacco's initial lethal tweet updated *Gawker* online magazine subscribers whenever Sacco tried to get a new job, effectively sabotaging Sacco's efforts to move on from her past gaffe.[6]

The Justine Sacco incident is noteworthy for many reasons. It demonstrates the necessity for individuals to "think before they speak," or, in this case, to think before they tweet, post, or comment online. It also demonstrates the fierce and swift nature of online

1. Jon Ronson, *How One Stupid Tweet Blew up Justine Sacco's Life*, N.Y. TIMES (Feb. 12, 2015), https://www.nytimes.com/2015/02/15/magazine/how-one-stupid-tweet-ruined-justine-saccos-life.html?_r=1 [https://perma.cc/G5JU-KFEJ].

2. *Id.*

3. *Id.*

4. *Id.*

5. Patrick Blanchfield, *Twitter's Outrage Machine Should Be Stopped.*, WASH. POST (Feb. 24, 2015), https://www.washingtonpost.com/posteverything/wp/2015/02/24/twitters-rage-mob-should-be-stopped-but-justine-sacco-is-the-wrong-poster-child/ [https://perma.cc/U9R4-62NJ].

6. Ronson, *supra* note 1; *see also* Sam Biddle, *Justine Sacco Is Back*, GAWKER (June 17, 2014, 9:45 AM), http://valleywag.gawker.com/justine-sacco-is-back-1591951969 [https://perma.cc/DX5W-84ZV].

vengeance against those who have committed behavior the Internet deems unacceptable.

Online public shaming since the Sacco incident has increased and changed in nature. In late September 2015, a new version of online public shaming emerged with the "Peeple app." Dubbed "Yelp for People," the app initially sought to grant users the ability to rate an individual on a one-to-five-star scale without the individual's consent, leading to anyone becoming a target of users' ratings.[7] Another infamous example of "shaming gone wrong" occurred after Minnesota dentist Walter Palmer killed beloved Cecil the Lion on an African hunting trip; outraged individuals subsequently trashed Palmer's Yelp page for his dental practice, filling it with vitriol related not to his abilities as a dentist, but for his actions in Africa.[8]

Nonetheless, online public shaming has created social positives in the consumer context. For example, frustrated consumers have increasingly been turning to Twitter to voice concerns about a company's detrimental policies or poor customer service.[9] Unfortunately, these social positives do not translate to the individual context as smoothly. While companies have substantial resources available to handle their public relations, including entire social media and marketing departments, most average citizens do not. Consequently, the ills of online public shaming fall especially hard on private citizens who become infamous on the internet overnight due to instances of online shaming gone viral.

One trend has become clear: the Internet "pitchfork mob" can have very real and sometimes wildly disproportionate effects on individuals as compared to their offensive behavior.[10] Sacco lost her job and continues to be professionally exiled; a Google search of her

---

7. Caitlin Dewey, *Everyone You Know Will Be Able to Rate You on the Terrifying 'Yelp for People'*, WASH. POST (Sept. 30, 2015), https://www.washingtonpost.com/news/the-intersect/wp/2015/09/30/everyone-you-know-will-be-able-to-rate-you-on-the-terrifying-yelp-for-people-whether-you-want-them-to-or-not/ [http://perma.cc/8GRN-T4YK].

8. Dale Lately, *A One-Star Human Being*, SLATE (Aug. 21, 2015), http://www.slate.com/articles/technology/future_tense/2015/08/lion_killing_dentist_walter_palmer_s_yelp_page_and_the_business_of_internet.html [https://perma.cc/TFQ6-R4NP]; *see also* Ed Payne, *Walter Palmer, the Man Who Killed Cecil the Lion, Returns to His Dental Practice*, CNN, http://www.cnn.com/2015/09/08/us/walter-palmer-dentist-cecil-lion-return/ [https://perma.cc/PXP6-8N2Y] (last updated Sept. 8, 2015).

9. Joe Silver, *Shamed on Twitter, Corporations Do an About-Face*, ARS TECHNICA (Apr. 21, 2014, 2:21 PM), http://arstechnica.com/business/2014/04/shamed-on-twitter-corporations-do-an-about-face/ [https://perma.cc/7MA7-ELBM].

10. For the purposes of this Note, a "pitchfork mob" refers to the general phenomenon of an angry mob going after an unpopular figure. For more information on the meaning and history of the pitchfork mob, see *Torches and Pitchforks*, TV TROPES, http://tvtropes.org/pmwiki/pmwiki.php/Main/TorchesAndPitchforks [https://perma.cc/GXT7-P8FC] (last visited Feb. 7, 2017).

name still reveals the fallout from the tweet.[11] Palmer faces a trashed reputation for his dentistry practice, which arguably has nothing to do with his hunting exploits.[12]

Unfortunately, the recourse for those shamed online is limited. Depending on the type of online shaming, victims may not be able to pursue any existing form of legal remedy.[13] For example, even though stating an unfavorable opinion online about someone's actions—disapproval of Justine Sacco's Tweet, for example—may be catalogued in large quantities on Google in perpetuity, wreaking untold havoc on both professional and personal reputations, such an action may not fall within the traditional definition of the tort of defamation.[14] Even if a victim can indeed state a claim for defamation against an online harasser, tracking down the online tortfeasor is notoriously difficult.[15] Adding further insult to injury, even in the instance that an individual is able to identify the harasser and state a cause of action, the harasser may be insolvent and unable to pay damages.

This Note explores the recourse available to individuals publicly shamed online. Part I analyzes the nature of online public shaming, the tangible harms it creates for individuals, and the unique nature of online public shaming compared to other harms. Part II explores the Communications Decency Act, which immunizes Internet Service Providers (ISPs) for the publication or distribution of certain tortious content created by third parties and also enables ISPs to create policies to filter online content without facing liability. Part III advocates for ISPs to utilize these expansive protections that the CDA provides to develop and implement new policies designed to deter the prevalence of shaming. Part III advances one model in which ISPs could achieve this goal by utilizing algorithms to identify when a user is engaging in shaming activities and prompt that user with a warning before completing the act. This Note therefore seeks to highlight the powerful, yet unrealized, role that ISPs have in lessening the prevalence of online mob shaming.

---

11. Google Search for Justine Sacco, GOOGLE, https://www.google.com/?gws_rd=ssl#q=justine+sacco [https://perma.cc/N5KC-ZGFL] (last visited Jan. 23, 2016) (search: "Justine Sacco").

12. *See* Lately, *supra* note 8.

13. *See infra* Section I.B.2 (comparing online shaming to other torts).

14. *See id.* (discussing how online shaming does not fit into the definition of the tort of defamation).

15. *See* Russell Brandom, *Finding Fuboy: One Man Spent Four Years and $35,000 to Unmask His Internet Troll*, VERGE (Nov. 23, 2015, 8:50 AM), http://www.theverge.com/2015/11/23/9772824/commenter-defamation-lawsuit-identity-revealed [https://perma.cc/EE28-DJD4].

I. GRAB YOUR PITCHFORKS: HOW ONLINE PUBLIC SHAMING HAS EXPLODED IN POPULARITY

A. Colonial Public Shaming in America

While the concept of online public shaming is a recent phenomenon fostered by the mainstream use of the Internet, the concept of public shaming itself has existed for centuries.[16] In colonial America, public shaming was particularly popular as a form of punishment.[17] In those days, the town square was used as a place to lock individuals in stockades and post signs stating the nature of the infraction committed by the individual.[18] Nathaniel Hawthorne's novel *The Scarlet Letter* exemplifies the nature of public shaming in colonial days, in which protagonist Hester Prynne is shamed by being forced to wear a scarlet "A" on her chest, signifying her status as an individual who had an affair.[19] Public shaming punishments were especially feared in small towns where individuals largely knew everyone who lived in the area.[20] Urbanization and the Industrial Revolution increased mobility of Americans, and public shaming punishments subsided as the centrality and importance of town centers in local communities decreased.[21] Additionally, the emergence of the prison system decreased the prevalence of public shaming punishments.[22]

B. Public Shaming 2.0: The Rise of Online Public Shaming

While the use of stockades in town squares to shame antisocial behavior is a relic of the past, the use of public shaming techniques has evolved and made a fierce comeback in the digital age. One important difference between state-sanctioned public shaming punishments and online public shaming is that online public shaming can occur for non-criminal acts.[23] Legal scholar Daniel Solove has explained this form of online public shaming as "norm enforcement."[24]

---

16. DANIEL J. SOLOVE, THE FUTURE OF REPUTATION 91 (2007).

17. Lauren M. Goldman, Note, *Trending Now: The Use of Social Media Websites in Public Shaming Punishments*, 52 AM. CRIM. L. REV. 415, 418 (2015) (explaining how colonial America favored utilizing public shaming punishments).

18. *Id.*

19. NATHANIEL HAWTHORNE, THE SCARLET LETTER 43–44 (Brian Harding & Cindy Weinstein eds., Oxford World's Classics 2007) (1850).

20. SOLOVE, *supra* note 16, at 91.

21. *Id.*; Goldman, *supra* note 17, at 421.

22. SOLOVE, *supra* note 16, at 92.

23. *Id.* at 84–85.

24. *Id.*

Norm enforcement occurs when individuals seek to correct behavior that does not comply with the perceived norm, or rule of conduct in society.[25] It can also occur silently, leaving the offender unaware of the alleged norm infraction.[26] This "quiet" norm enforcement is the genesis of several trends in online public shaming.[27] The kinds of norms enforced online can vary from etiquette norms to perceived norms about one's appearance and habits.[28] Various iterations of "norm-enforcing" tools have emerged online in past years, including websites cataloguing bad men to date,[29] delinquent taxpayers,[30] and the difficulty level of college professors.[31]

Further exemplifying this trend, the Peeple app debuted in 2015 as a new holistic norm enforcement tool. Initially advertised as a "positivity app," Peeple was marketed on the platform that individuals could rate anyone on a scale of one to five.[32] The app would enable users to create a profile *for* a third party whom they wanted to rate, regardless of whether the third party wished to be a part of Peeple.[33] Furthermore, accounts created without the consent of the third party did not have a "delete" option.[34] Dubbed "Yelp for People" by the *Washington Post*, Internet backlash regarding the app was swift and almost entirely negative.[35] After the backlash, the creators of the app considerably changed the app's core functions, instead requiring that individuals opt *in* before receiving reviews on the app and giving the app's users the ability to hide negative comments.[36]

---

25. *Id.*

26. *Id.*

27. *Id.* at 86.

28. *Id.*

29. *See* Molly McHugh, *Rating Men on LuLu Isn't Atoning for the Web's Chauvinism, It's Just Cruel*, DIGITAL TRENDS (June 19, 2013, 1:00 PM), http://www.digitaltrends.com /social-media/rating-men-on-lulu-isnt-atoning-for-the-webs-chauvinism-its-just-cruel/ [https://perma.cc/748H-XK5Y].

30. *Tax Delinquents List*, MASS.GOV, http://www.mass.gov/dor/individuals/tax -delinquents-list.html [https://perma.cc/HB73-G6Z2] (last visited Feb. 7, 2017).

31. RATE MY PROFESSORS, http://ratemyprofessors.com [https://perma.cc/KF23-QX3B] (last visited Feb. 4, 2017).

32. Dewey, *supra* note 7.

33. *Id.*

34. *Id.*

35. Dewey, *supra* note 7; Maddy Meyers, *The "Peeple" App, Referred to as Yelp for People, Sounds Like a Bad Dream*, MARY SUE (Sept. 30, 2015, 5:17 PM), http://www.themarysue.com/peeple-app-bad-dream/ [http://perma.cc/WXL3-2LTJ]; *see also* Elle Hunt, *Peeple Review People: The User-Review App You Didn't Dare Ask For*, GUARDIAN (Oct. 1, 2015, 4:48 AM), http://www.theguardian.com/technology/2015/oct/01/peeple-review-people-the -user-review-app-you-didnt-dare-ask-for [https://perma.cc/RMC7-KL7X].

36. Caitlin Dewey, *Peeple, the Terrifying "Yelp for People," Is (Sort of) Launching March 7*, WASH. POST (Mar. 4, 2016), https://www.washingtonpost.com/news/the-intersect/wp

While online norm enforcement can take the form of specialized apps—Peeple, for instance—individuals commonly enforce social norms through a diverse array of methods online, such as "liking" Facebook postings, sharing articles, or commenting on YouTube videos.[37] However, the meaning of those online actions can be indeterminate and can change over time.[38] For example, a 2016 incident involving the killing of a gorilla named Harambe in a Cincinnati zoo after a child entered the gorilla's enclosure sparked outrage online.[39] Initial online postings and petitions after the gorilla's death shamed the child's mother for a perceived lack of accountability over her child's actions.[40] Meanwhile, others shamed the role of zoos, claiming that it was captivity that had ultimately claimed the gorilla's life.[41] However, notably, some individuals "shamed the shamers," and outrage *about* the Internet's reaction to the gorilla's death ultimately eclipsed outrage about the actual incident.[42] By mid-summer 2016, the "Harambe meme" was born, with journalists noting that the meme persists "because it is, at its heart, a criticism of the online cultural environment that creates a phenomenon like the outrage at Harambe's death."[43]

One possible explanation for the upswing in online norm enforcement that has resulted in online "outrage culture" is the fact that it provides an opportunity for individuals to passively indicate that they do not endorse certain socially offensive behavior.[44] By

/2015/10/05/after-internet-backlash-peeple-co-founder-will-revise-her-app-to-make-it-positive/ [http://perma.cc/56C5-MSEH].

37. *See* Kate Klonick, *Re-Shaming the Debate: Social Norms, Shame, and Regulation in an Internet Age*, 75 MD. L. REV. 1029, 1053 (2016) (noting that acts of online norm enforcement can have indeterminate social meaning).

38. *Id.*

39. Mike McPhate, *Gorilla Killed After Child Enters Enclosure at Cincinnati Zoo*, N.Y. TIMES (May 29, 2016), http://www.nytimes.com/2016/05/30/us/gorilla-killed-after-child-enters-enclosure-at-cincinnati-zoo.html [https://perma.cc/UZM5-9HPH].

40. Sheila Hurt, *Justice for Harambe*, CHANGE.COM, https://www.change.org/p/cincinnati-zoo-justice-for-harambe [https://perma.cc/B6QY-8BW4] (last visited Jan. 23, 2017).

41. Marais Jacon-Duffy, *PETA Says Harambe's Death Is an Example of 'Captivity Taking an Animal's Life'*, WCPO CINCINNATI, http://www.wcpo.com/news/local-news/peta-says-harambes-death-is-an-example-of-captivity-taking-an-animals-life [http://perma.cc/V2EJ-M5RD] (last updated May 29, 2016, 6:45 PM).

42. Alex Abad-Santos, *Harambe the Gorilla: The Zoo Killing That's Set the Internet on Fire, Explained*, VOX (June 1, 2016, 2:42 PM), http://www.vox.com/2016/5/31/11813640/harambe-gorilla-cincinnati-zoo-killed [https://perma.cc/LQW6-Q5FN].

43. Abby Ohlheiser, *The Internet Won't Let Harambe Rest in Peace*, WASH. POST (July 27, 2016), https://www.washingtonpost.com/news/the-intersect/wp/2016/07/27/the-internet-wont-let-harambe-rest-in-peace/ [http://perma.cc/N5NR-CVNY].

44. Jillian Jordan, Paul Bloom, Moshe Hoffman & David Rand, *What's the Point of Moral Outrage?*, N.Y. TIMES (Feb. 26, 2016), http://www.nytimes.com/2016/02/28/opinion/sunday/whats-the-point-of-moral-outrage.html [http://perma.cc/BL25-F4J9].

reprimanding offensive behavior, individuals can effectively and easily signal to others that they are trustworthy because they punished a norm violator.[45] Posting a comment condemning the actions of Justine Sacco or Walter Palmer may not only signal to others that the commenter does not endorse their behavior, but also that the commenter is credible because he or she took actions to shame the norm violator. However, as the "Harambe meme" demonstrates, the meaning of acts of online norm enforcement is neither definite, nor static. Someone sharing a story about a gorilla's death may be doing so for a variety of reasons: out of genuine concern for the animal, ironically, or without much thought at all.[46] Regardless of the precise meaning of acts of online norm enforcement, this online commentary can be made cheaply, quickly, and indelibly, much to the detriment of shamed individuals.

1. The Critical Role of Anonymity and Shaming's Disproportionate Effects

Online anonymity and few cost barriers have contributed to the surge in online public shaming. The anonymity that the Internet provides enables some individuals to dodge accountability and any kind of tangible fallout from derisive statements they make about other individuals online.[47] Engaging in an act of online norm enforcement is also inexpensive; an individual simply needs Internet access, which 3.2 billion individuals are estimated to have.[48] This mentality has produced an increase in "drive-by relationships," a term describing a user's ability to anonymously and cheaply post *ad hominem* attacks and then quickly duck out of the situation.[49] This leaves the victim of the drive-by stuck with unsavory reputation-damaging content online, with little recourse to identify the poster. The concept of a "drive-by relationship" is simple; it is much easier to criticize someone's actions when you can do so anonymously.

---

45. *Id.*

46. *See* Klonick, *supra* note 37, at 1053 ("[I]ndeterminacy is true of the vast majority of online acts of social norm enforcement—re-Tweeting a message; sharing a link; commenting anonymously on a blog; liking something on Facebook—the acts themselves are so small, discrete, and instant that they do not necessarily have a clear social meaning.").

47. SOLOVE, *supra* note 16, at 140.

48. *Internet Users*, INTERNET LIVE STATS, http://www.internetlivestats.com/internet-users/#trend [https://perma.cc/VM4K-MW3M] (last visited Jan. 23, 2016).

49. SOLOVE, *supra* note 16, at 141; *see also* ROBERT D. PUTNAM, BOWLING ALONE: THE COLLAPSE AND REVIVAL OF THE AMERICAN COMMUNITY 177 (2000).

One way for an individual to track down an online poster is through the poster's Internet Protocol, or IP, address.[50] Every machine has a unique IP address that facilitates communications with other computers.[51] However, IP addresses are not unique to an individual *person*; an IP address, by itself, is nothing more than a string of numbers.[52] Currently, there is no universal online "login" system that attaches specific credentials to each individual online user; IP addresses are assigned to computers, not to people.[53] Consequently, the existence of an IP address alone, without additional identifying information, cannot pinpoint the absolute identity of an online poster.[54] Due to these constraints, victims of online attacks often have limited means to unmask their tormentors.

In contrast to the ease and low costs of shaming norm violators online, the consequences of shaming on its victims are disproportionately severe. The effects of shaming often balloon out of proportion when compared to the alleged norm violation.[55] Instances of online shaming do not provide due process rights to the victim; they are afforded no hearing to establish facts regarding the alleged norm infraction.[56] The shaming that occurred after the 2016 Cincinnati Zoo incident exemplifies this lack of due process; was the mother inattentive, or did the child simply manage to slip away? Many shamed the mother of the child *without* establishing these crucial facts.[57]

The lack of fact-establishing safeguards associated with shaming can skew news coverage of instances of shaming, thereby unintentionally adding to the pile-on effect by drawing more attention to the alleged norm violator.[58] The harms of shaming are further amplified by the indelible nature of shaming content that can be catalogued in perpetuity online. Exemplifying this harm, the Sacco Twitter incident is still the first item retrieved after performing a

---

50. *See* Joshua J. McIntyre, Comment, *Balancing Expectations of Privacy: Why Internet Protocol (IP) Addresses Should Be Protected as Personally Identifiable Information*, 60 DEPAUL L. REV. 895, 896 (2011).

51. *Id.*

52. *Id.* at 908.

53. *Id.* at 910.

54. *Id.*

55. Klonick, *supra* note 37, at 1045; SOLOVE, *supra* note 16, at 95–96.

56. *See* SOLOVE, *supra* note 16, at 96–98.

57. *See* Laura Coates, *Gorilla Shooting: When a Child's Death Draws Less Outrage than Harambe*, CNN (June 1, 2016, 8:31 AM), http://www.cnn.com/2016/06/01/opinions/harambe-gorilla-shooting-coates/ [https://perma.cc/EZ8T-GGHC].

58. Kelly McBride, *Journalism and Public Shaming: Some Guidelines*, POYNTER (Mar. 11, 2015), http://www.poynter.org/2015/journalism-and-public-shaming-some-guidelines/326097/ [https://perma.cc/X5Y8-29CP].

Google search of Sacco's name, despite Sacco's atonement and the passing of several years since the incident.[59] Consequently, online public shaming can serve as a "one-two" punch: the shaming can occur without any critical fact-establishing due process safeguards and can continually impact an individual's future due to the everlasting nature of online content, even if that content is distorted or factually incorrect.

2. The Important Distinction Between Online Shaming, Defamation, and Other Torts

While there is currently no tort of "undue shaming," the tort of defamation exists to curb undue harm to individuals' reputations. Defamation is a state-law claim, so the applicable elements vary slightly from state to state.[60] The Restatement of Torts states that a prima facie case of defamation requires a plaintiff to prove the existence of: (a) a false and defamatory communication concerning another, (b) an unprivileged publication[61] to a third party, (c) fault amounting to at least negligence on behalf of the publisher, and (d) either the actionability of the statement irrespective of special harm or the existence of special harm caused by the publication.[62]

The tort of defamation extends liability beyond the original defamer.[63] Individuals who repeat or otherwise republish defamatory communications are subject to the same degree of liability as the original speaker of such content.[64] For example, a newspaper that actively selects and edits content for publication would be considered a publisher and face liability for publishing defamatory content.[65] However, mere distributors of defamatory content, such as bookstores or libraries, generally only face liability for defamatory content when the distributor knows or has reason to know of the defamatory communication but transmits such a communication nonetheless.[66]

---

59. Google Search for Justine Sacco, GOOGLE, https://www.google.com/?gws_rd=ssl#q=justine+sacco [https://perma.cc/N5KC-ZGFL] (last visited Jan. 23, 2016) (search: "Justine Sacco").

60. 50 AM. JUR. 2D *Libel and Slander* § 15 (2016) ("An individual's interest in his or her reputation is a basic concern, but its reflection in the laws of defamation is solely a matter of state law.").

61. RESTATEMENT (SECOND) OF TORTS § 577 cmt. a (AM. LAW INST. 1977) ("Any act by which the defamatory matter is intentionally or negligently communicated to a third person is a publication.").

62. *Id.* § 558.

63. *Id.* § 578.

64. *Id.*

65. *See* RODNEY A. SMOLLA, 1 LAW OF DEFAMATION § 4:92 (2d ed. 2016).

66. *Id.*; *see also* RESTATEMENT (SECOND) OF TORTS § 581 (AM. LAW INST. 1977).

Since distributors serve merely as the physical means to the republication of defamatory content, a distributor who transmits a defamatory communication without knowledge of its defamatory nature is consequently shielded from liability.[67] It is important to note that imposing liability upon publishers and distributors of defamatory content has had large implications with the advent of the Internet, where web hosts regularly post and edit third-party content. The consequences of imposing publisher and distributor liability upon ISPs that host third-party tortious content is discussed in full detail in Part II.

Despite being designed to remedy reputational harms, the tort of defamation cannot necessarily help victims of online public shaming. The tort of defamation hinges upon the issuance of a *falsehood*—stating an unpopular opinion, or criticizing someone for his or her actions, does not equate to uttering a falsehood about someone.[68] For example, posting a tweet that states "Shame on Walter Palmer for killing Cecil the Lion" or "Check out how ignorant Justine Sacco's tweet is" is not inherently defamatory—the poster is not stating falsehoods about either individual. Even if a plaintiff is able to state a claim for defamation—for example, a scenario wherein a blogger falsely accuses an individual of having sexually transmitted diseases—there are further difficulties litigating such cases, as litigation is expensive and time consuming, making it an unlikely avenue for the average individual to pursue.

Online public shaming victims could also pursue the tort of intentional infliction of emotional distress (IIED).[69] The elements of a successful IIED claim are: (1) extreme and outrageous conduct with either the intention of, or reckless disregard for, causing emotional distress; (2) the suffering by the plaintiff of severe or extreme emotional distress; and (3) actual or proximate causation.[70] While certain extreme instances of shaming could conceivably satisfy the basic elements of an IIED claim, plaintiffs face the same issues as those in litigating other online torts: tracking down a defendant in such cases can be difficult, if not impossible; litigation is expensive and time consuming; and the volume of harmful online content

67. SMOLLA, *supra* note 65, at § 4:92; RESTATEMENT (SECOND) OF TORTS § 581 (AM. LAW INST. 1977).

68. "To create liability for defamation there must be . . . a *false* and defamatory statement concerning another." RESTATEMENT (SECOND) OF TORTS § 588(a) (AM. LAW INST. 1977) (emphasis added).

69. 136 AM. JUR. PROOF OF FACTS 3D *Proof of Intentional Infliction of Emotional Distress* § 1 (2013).

70. *Id.* § 4.

renders success against one individual defendant inadequate in the face of the tidal wave of harmful online postings stored elsewhere.

Online public shaming can also evolve into other types of online harms. Far from simply voicing disapproval of a norm violator's behavior, shaming can take a dark turn into "doxing,"[71] and even calling for sexual violence or other physical harms against an individual.[72] These actions can be categorized as cyber harassment, especially when online mobs repeatedly act to target a specific individual.[73] This raises questions as to when acts of online public shaming and norm enforcement break down into forms of online harassment. Legal scholar Kate Klonick notes that "[o]nline shaming often turns into cyber bullying and harassment the more attenuated the social actions become from the nexus of social norm enforcement."[74] Consequently, the development of policies to deter the prevalence of online public shaming should be included in any conversation about how to address online harassment more generally.

II. THE COMMUNICATIONS DECENCY ACT AND SUBSEQUENT DEVELOPMENTS IN ISP LIABILITY FOR THIRD-PARTY STATEMENTS

Several options exist for reducing the prevalence of online public shaming. First, individuals could engage in online shaming less often. However, due to the low costs of entry, online anonymity, and perhaps a lack of awareness or consideration of long-term repercussions of shaming, it is unlikely that individuals will unilaterally choose to reduce shaming behavior unless the behavior becomes unfashionable or otherwise detrimental.[75] Second, web hosts could remove shaming posts; however, policing shaming in this manner would likely be largely impracticable and lead to inevitable free speech concerns. Third, web hosts could espouse a moderate approach by warning users about the consequences of shaming by employing algorithms and filters that detect shaming activity. Web hosts that take such actions to filter content for its propensity to

71. "Doxing" is the act of intentionally publishing someone's personal data, such as a home address or social security number online. Sameer Hinduja, *Doxing and Cyberbulling*, CYBERBULLYING RES. CTR. (Sept. 16, 2015), http://cyberbullying.org/doxing-and-cyberbullying [https://perma.cc/T8QF-GC5H].

72. *See* Klonick, *supra* note 37, at 1034.

73. *See* DANIELLE KEATS CITRON, HATE CRIMES IN CYBERSPACE 3 (2014) ("[C]yber harassment is often understood to involve the intentional infliction of substantial emotional distress accomplished by online speech that is persistent enough to amount to a 'course of conduct' rather than an isolated instance.").

74. Klonick, *supra* note 37, at 1034.

75. The idea that online public shaming becomes less fashionable is plausible and arguably already occurring, given the birth of the "Harambe meme." *See supra* Section I.B.

shame would not become liable as publishers or distributors for third-party online content under Section 230 of the Communications Decency Act. Consequently, web hosts have an untapped potential to reduce the prevalence of online public shaming. The following section explores, in detail, the Communications Decency Act and its revolutionary impact upon liability for the publication or distribution of third-party statements online.

A. The Advent of the Communications Decency Act

The Internet presented uncharted territory to both lawmakers and courts alike. While the concept of publisher and distributor liability for the reproduction of defamatory statements is fairly straightforward in the context of traditional media, such as newspapers,[76] it was unclear if the same rules would be practicable in the age of the Internet.

In 1991, the first recorded "cybertort" case,[77] *Cubby, Inc. v. CompuServe*,[78] held ISP CompuServe not liable for defamatory comments made by one of its posters when CompuServe did not screen the content its users posted.[79] After *Cubby*, it appeared that ISPs would be held to the distributor liability standard: so long as an ISP did not know that a third party's content was defamatory, it would face no liability in distributing the content online.[80] However, it remained unclear as to whether traditional publisher liability would apply in the same way to online defamatory communications.[81]

Distinguishable from *Cubby,* another initial case involving Internet defamation, *Stratton Oakmont v. Prodigy Services,* imposed publisher liability on a defendant ISP that posted an individual user's defamatory statements in 1995.[82] In *Stratton Oakmont*, a subscriber to ISP Prodigy Services posted a comment on its online bulletin board accusing Stratton Oakmont, a securities investment banking firm, of making fraudulent securities offerings.[83] When Stratton Oakmont sued ISP Prodigy, the court ruled that Prodigy could be held liable

76. See discussion of publisher and distributor liability for reproducing defamation, *supra* Section I.B.2.

77. *See* Michael L. Rustad & Thomas H. Koenig, *Rebooting Cybertort Law*, 80 WASH. L. REV. 335, 364 (2005) ("The first cybertort case was decided in 1991, when CompuServe, Inc. was held not liable for a third party's publication of defamatory statements on its services.").

78. *See* Cubby, Inc. v. CompuServe, Inc., 776 F. Supp. 135, 137–40 (S.D.N.Y. 1991).

79. *Id.* at 141.

80. *See* Rustad & Koenig, *supra* note 77, at 366.

81. *Id.* at 365.

82. *See* Stratton Oakmont v. Prodigy Servs. Co., No. 31063/94, 1995 WL 323710, at *1 (N.Y. Sup. Ct. May 24, 1995); *Cubby*, 776 F. Supp. at 140.

83. *Stratton Oakmont*, 1995 WL 323710, at *1.

under publisher liability, thereby finding Prodigy liable for the defamation as if it were the original poster on the bulletin board.[84] Unlike ISP CompuServe, which did not filter any of its posters' commentary,[85] ISP Prodigy filtered content in accordance with its family-focused goals.[86] Oddly enough, Prodigy's content filtering subjected Prodigy to harsher publisher liability, while CompuServe's complete lack of filtering is what enabled it to be saved by the more lenient standard of distributor liability instead.[87]

This new era of common law decisions regarding ISP liability created a perverse incentive: if ISPs did not moderate anything their users posted, they would be less likely to face harsh publisher's liability.[88] The practical effects of a *Stratton Oakmont*-influenced Internet landscape could result in a largely un-moderated Internet, as it incentivized web hosts to refuse to edit or modify any of their users' statements to avoid getting sued under publisher liability.[89]

In light of the *Cubby* and *Stratton Oakmont* decisions, ISPs petitioned Congress to enact legislation—the Communications Decency Act (CDA)—shielding them from such widespread publisher liability under a *Stratton Oakmont* regime.[90] The CDA's sponsors feared that imposing common law publisher liability on ISPs would incentivize ISPs to refuse to moderate content whatsoever in order to avoid liability, resulting in a lawless Internet filled with pornography and other obscenities that children could accidentally view.[91] Consequently, the CDA was marketed as a way to prevent obscene materials online from getting into the hands of children without chilling free speech.[92] Proponents of the bill recognized that it would be impossible for the government to completely censor obscene material, noting that ". . . the Internet operates worldwide, and not

84. *Id.* at *4.

85. *Cubby*, 776 F. Supp. at 140.

86. *Stratton Oakmont*, 1995 WL 323710, at *2; *see* Rustad & Koenig, *supra* note 77, at 367.

87. *See* Rustad & Koenig, *supra* note 77, at 367.

88. *Id.*

89. *Id.*

90. *Id.* at 368.

91. H.R. REP. NO. 104-458, at 194 (1996) ("One of the specific purposes of this section is to overrule Stratton-Oakmont v. Prodigy and any other similar decisions which have treated such providers and users as publishers or speakers of content that is not their own because they have restricted access to objectionable material. The conferees believe that such decisions create serious obstacles to the important federal policy of empowering parents to determine the content of communications their children receive through interactive computer services.")

92. 47 U.S.C. § 230(b)(4) (2012) ("It is the policy of the United States . . . to remove disincentives for the development and utilization of blocking and filtering technologies that empower parents to restrict their children's access to objectionable and inappropriate material online.").

even a Federal Internet censorship army would give our Government the power to keep offensive materials out of the hands of the children who use the new interactive media."[93]

The CDA's sponsors emphasized that private citizens—not the government—would be best suited to create a "21st century policy" for the Internet.[94] In order to achieve the crucial balance of filtering obscenity with preserving unfettered speech online, the CDA included a "safe harbor" provision for ISPs in Section 230(c).[95] This section, entitled "Protection for 'Good Samaritan' blocking and screening of offensive materials," explicitly exempted ISPs from being subject to publisher liability or treated as publishers, stating that "[n]o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."[96] Section 230(c)(2) also explicitly protected ISPs attempting to remove objectionable content from exposure to civil liability.[97]

One of the first cases to apply Section 230(c), *Zeran v. America Online, Inc.*, involved a post on an AOL bulletin board directing phone calls for purchasing t-shirts with incendiary slogans relating to the 1995 Oklahoma City bombing to Zeran's phone number.[98] Zeran, who had nothing to do with selling such t-shirts, received a barrage of angry phone calls, some of which included death threats.[99] Although AOL took down the initial post, subsequent similar postings appeared, and Zeran was again bombarded with threats.[100] Zeran sued AOL, alleging that AOL both failed to remove the postings in a timely manner after Zeran had given notice and failed to screen for further defamatory postings.[101] Zeran further argued that Section 230(c) mentioned nothing explicitly about extending a "safe harbor" to

93. 141 CONG. REC. H8472 (1995) (statement of Rep. Goodlatte) ("[The CDA] allows parents to make the important decisions with regard to what their children can access, not the government. It doesn't violate free speech or the right of adults to communicate with each other.").

94. 141 CONG. REC. H8470 (statement of Mr. Wyden) ("[T]he new media is simply different. We have the opportunity to build a 21st century policy for the Internet employing the technologies and the creativity designed by the private sector.").

95. 47 U.S.C. § 230(c). It is also important to note that another section of the CDA, 47 U.S.C. § 223 (2012), outlawed the transmission of "patently offensive communications" via interactive computer services. This statute was deemed unconstitutional in *Reno v. American Civil Liberties Union*, 521 U.S. 844, 849 (1997), for being overbroad and not sufficiently tailored to achieve the government's interest in protecting children from obscenities online.

96. 47 U.S.C. § 230(c)(1).

97. *Id.*

98. Zeran v. Am. Online, Inc., 129 F.3d 327, 329 (4th Cir. 1997).

99. *Id.*

100. *Id.*

101. *Id.* at 330.

distributors; as such, Zeran argued he should be able to sue AOL under a theory of distributor liability.[102] The Fourth Circuit rejected Zeran's argument, noting that:

> . . . notice cannot transform one from an original publisher to a distributor. . . . [O]nce a computer service provider receives notice of a potentially defamatory posting, it is thrust into the role of a traditional publisher. The computer service provider must decide whether to publish, edit, or withdraw the posting.[103]

Consequently, distributors were entitled to the same level of deference under Section 230(c) as publishers.[104]

The wide-reaching effects of *Zeran* were amplified by the 1998 District Court for the District of Columbia's decision in *Blumenthal v. Drudge* in which plaintiffs brought a defamation suit against "Drudge Report" creator Matt Drudge and AOL for an article that claimed that Blumenthal, a White House staffer, abused his spouse.[105] Drudge had a licensing agreement with AOL, making the Drudge Report available to all AOL members.[106] Drudge was able to create, edit, and "otherwise manage" content, but AOL retained the right to remove content that violated its terms of service.[107] Despite AOL's exercise of some editorial control over Drudge's work and its profiting from the benefits of the CDA immunity scheme without "accepting any of the burdens that Congress intended," Section 230(c) rendered AOL immune from suit.[108] Importantly, in *Blumenthal*, AOL actively promoted its online gossip column, a marked departure from its role as mere passive host to bulletin boards in *Zeran*.[109]

*B. The Post-*Zeran *Landscape: Broad ISP Immunity*

After *Zeran* and *Blumenthal*, courts have generally recognized three requisite criteria in order for Section 230 immunity to apply to a defendant ISP.[110] First, a defendant must be the provider or user of

---

102. *Id.* at 331.

103. *Id.*

104. *Id.*

105. *See* Blumenthal v. Drudge, 992 F. Supp. 44, 46–48 (D.D.C. 1998).

106. *Id.*

107. *Id.*

108. *Id.* at 52–53.

109. *Compare Zeran*, 129 F.3d at 329 (ISP AOL immune from liability under Section 230 when serving as forum for users to make postings on online bulletin boards), *with Blumenthal*, 992 F. Supp. at 47 (ISP AOL immune from liability under Section 230 when retaining ability to remove certain content created by third party contractor and promoting gossip column).

110. *See* Shiamili v. Real Estate Grp. of N.Y., Inc, 952 N.E.2d 1011, 1015 (2011) ("A defendant is therefore immune from state law liability if (1) it is a 'provider or user of an interactive computer service'; (2) the complaint seeks to hold the defendant liable as a 'publisher or speaker'; and (3) the action is based on 'information provided by another information content provider.'"); Nemet Chevrolet, Ltd. v. Consumeraffairs.com, Inc., 564 F. Supp. 2d 544, 548 (E.D.

an "interactive computer service."[111] The CDA defines an "interactive computer service" as "any information service, system, or access software provider that provides or enables computer access to the Internet and such systems operated or services offered by libraries or educational institutions."[112] Many types of online services have qualified as interactive computer services, including classified ad sites such as Craigslist, dating websites, and social media platforms such as MySpace and Facebook.[113] Second, the claim must treat the defendant interactive computer service as the publisher or speaker of the information.[114] Third, another user, separate from the ISP, must have created the disputed content.[115] This third criterion has raised many questions in its application; if an ISP crosses the line from existing as a passive "interactive computer service" to serving as an "information content provider" pursuant to the CDA, the ISP may not qualify for Section 230 immunity.[116] In other words, ISPs are generally not liable for hosting third-party tortious content. As such, web hosts have a marked interest in ensuring that they take the appropriate steps to avoid crossing over to the side of the CDA that exposes them to liability.

Following in the ISP-deferential steps of *Zeran* and *Blumenthal*, however, courts have been reluctant to deem ISPs "information content providers" when they are performing editorial functions, choosing to remove or add content, or making minor adjustments to third-party content. For example, in *Ben Ezra, Weinstein, and Co. v. America Online, Inc.*, the Tenth Circuit found

---

Va. 2008) ("Courts engage in a three-part inquiry when determining the attachment of immunity under the CDA."); Schneider v. Amazon.com, Inc., 31 P.3d 37, 39–40 (2001).

111. *See Shiamili*, 952 N.E.2d at 1015; *Nemet*, 564 F. Supp. 2d. at 548; *Schneider*, 31 P.3d at 39–40.

112. 47 U.S.C. § 230(f)(2) (2012).

113. *See* Doe v. MySpace Inc., 528 F.3d 413, 420 (5th Cir. 2008) (social networking website qualifies as interactive computer service); Chi. Lawyers' Comm. for Civil Rights Under Law, Inc. v. Craigslist, Inc., 519 F.3d 666, 671 (7th Cir. 2008) (online classified ad service qualifies as interactive computer service); Carafano v. Metrosplash.com, Inc., 339 F.3d 1119, 1124 (9th Cir. 2003) (online dating service qualifies as interactive computer service).

114. *See Shiamili*, 952 N.E.2d at 1015; *Nemet*, 564 F. Supp. 2d. at 548; *Schneider*, 31 P.3d at 39–40.

115. *See Shiamili*, 952 N.E.2d at 1015; *Nemet*, 564 F. Supp. 2d. at 548; *Schneider*, 31 P.3d at 39–40.

116. An "information content provider" is "any person or entity that is responsible, in whole or in part, for the creation or development of information provided through the Internet or any other interactive computer service." 47 U.S.C. § 230(f)(3). The distinction between "interactive computer service" and "information content provider" has been discussed at length. *See, e.g.*, Fair Hous. Council v. Roommates.com, LLC., 521 F.3d 1157, 1163–74 (9th Cir. 2008); *Carafano*, 339 F.3d at 1123 ("Under the statutory scheme, an 'interactive computer service' qualifies for immunity so long as it does not also function as an 'information content provider' for the portion of the statement or publication at issue.").

that AOL's minor edits of stock quotations provided by a third party or deletion of incorrect stock information provided by a third party did not turn AOL into an "information content provider" of that content.[117] In *Batzel v. Smith*, the Ninth Circuit found that an ISP that made minor alterations to an allegedly defamatory third-party email before replicating the email on a listserv also did not count as an "information content provider" of that content.[118] In *Carafano v. Metrosplash*, an anonymous harasser created a fake profile for actress Christine Carafano on a dating website and posted Carafano's home address and phone number.[119] After receiving numerous threats, Carafano sued the dating service, Matchmaker.com, alleging invasion of privacy, misappropriation of the right of publicity, defamation, and negligence.[120] The Ninth Circuit deemed Matchmaker.com's Section 230(c) defense as valid despite Matchmaker.com contributing to the creation of dating profiles with questionnaires designed to format users' profiles, noting that "[u]nder Section 230(c) . . . so long as a third party willingly provides the essential published content, the interactive service provider receives full immunity regardless of the specific editing or selection process."[121]

While the Ninth Circuit took a stand in 2008 by denying Section 230 immunity to ISPs who "materially contribute" to alleged illegal conduct,[122] the trend of wide judicial deference to ISPs under the Section 230 framework has otherwise largely continued. In 2014, the Sixth Circuit held in *Jones v. Dirty World Entertainment* that a web host does not forfeit Section 230 immunity even when that host encourages or otherwise ratifies third-party tortious content.[123] *Jones* involved controversial site TheDirty.com, a non-celebrity gossip site, where users could post about members of the community without fact-checks or other verifications for accuracy.[124] Site creator Nik Richie would then comment on the users' postings, often encouraging his "dirty Army" to find more information on its targets.[125] The district

117. *See* Ben Ezra, Weinstein, & Co. v. Am. Online, Inc., 206 F.3d 980, 985–86 (10th Cir. 2000).

118. *See* Batzel v. Smith, 333 F.3d 1018, 1030–31 (9th Cir. 2003).

119. *Carafano*, 339 F.3d at 1121.

120. *Id.* at 1122.

121. *Id.* at 1124.

122. *See* Fair Hous. Council v. Roommates.com, LLC, 521 F.3d 1157, 1167–70 (9th Cir. 2008) (finding ISP that designed website intended to solicit and enforce allegedly illegal housing preferences "materially contributed" to the illegality and was therefore not entitled to Section 230 immunity).

123. Jones v. Dirty World Entm't Recordings LLC, 755 F.3d 398, 401–03 (6th Cir. 2014).

124. *Id.*

125. Jones v. Dirty World Entm't Recordings, LLC, 965 F. Supp. 2d 818, 823 (E.D. Ky. 2013).

court ruled that TheDirty.com forfeited its Section 230 immunity due to this editorial commentary encouraging users to find more "dirt" on others, ruling that

> a website owner [that] intentionally encourages illegal or actionable third-party postings to which he adds his own comments ratifying or adopting the posts becomes a 'creator' or 'developer' of that content and is not entitled to immunity.[126]

The Sixth Circuit reversed, noting that an "encouragement" test was not proper for determining whether an ISP lost immunity under the CDA.[127] Since plaintiff Sarah Jones was suing because of the defamatory third-party statements posted on TheDirty.com, and not based upon the subsequent editorial commentary of site creator Nik Richie, Section 230 prevented Jones from bringing her defamation suit against Richie.[128] The Sixth Circuit further noted that the underlying purpose of the CDA was to promote a free and open Internet, and the CDA should be read in accordance with this principle.[129]

C. Scholarly Critiques of the CDA Incentive Framework

ISP-deferential judicial interpretations of the CDA have left many plaintiffs in online defamation cases exasperated and largely without remedy. Many legal scholars have criticized both the inadequate incentive scheme for ISPs created by the structure of Section 230 as well as the broad judicial deference ISPs receive in litigation involving application of Section 230.[130] Courts have also taken note of the odd incentive schemes created by the CDA framework; in *Doe v. GTE*, a 2003 Seventh Circuit case that involved a defendant ISP that raised a Section 230 defense, Judge Easterbrook noted:

---

126. *Id.* at 821.

127. *Jones*, 755 F.3d at 414.

128. *Id.* at 415–16.

129. *Id.* at 415.

130. *See* Heather Saint, Note, *Section 230 of the Communications Decency Act: The True Culprit of Internet Defamation*, 36 LOY. L.A. ENT. L. REV. 39, 41 (2014) (noting sweeping judicial deference given to ISPs under the CDA); Patricia Spiccia, Note, *The Best Things in Life Are Not Free: Why Immunity Under Section 230 of the Communications Decency Act Should Be Earned and Not Freely Given*, 48 VAL. U. L. REV. 369, 370 (2013) (noting that the CDA does not incentivize ISPs to assist victims of defamation).

> . . . [Section] 230(c) as a whole makes ISPs indifferent to the content of information they host or transmit: whether they do . . . or do not . . . take precautions, there is no liability under state or federal law. As precautions are costly, not only in direct outlay but also in lost revenue from the filtered customers, ISPs may be expected to take the do-nothing option and enjoy immunity under Section 230(c)(1).[131]

Among proposed changes to the current CDA framework, some have advocated for the addition of notice and takedown procedures for defamatory materials posted online, modeled after the current notice and takedown provisions of the Digital Millennium Copyright Act.[132] This would serve to balance the goals of the CDA with the need for victims of online attacks to remove defamatory content.[133] Others have called for amending the CDA to deny immunity to ISPs that make editorial publication decisions and therefore serve as more than mere conduits to third-party content.[134] More drastically, others have suggested repealing the CDA in its entirety and returning to the use of notice-based liability for third-party tortious content imposed upon ISPs that was used before the CDA's passage.[135]

While imperfect, the CDA has arguably fostered the free growth and expansion of the Internet; one scholar notes that "Section 230 is the breathing space for the Internet's extraordinarily free expression."[136] Others feel that reverting back to common law notice-based liability for ISPs for distributing tortious content would irreparably stall the free-flowing nature of information online due to an increase in the marginal cost of each users' postings.[137] The Electronic Frontier Foundation notes that Section 230 is "perhaps the most influential law to protect the kind of innovation that has allowed

---

131. Doe v. GTE Corp., 347 F.3d 655, 660 (7th Cir. 2003).

132. *See* Olivera Medenica & Kaiser Wahab, *Does Liability Enhance Credibility? Lessons from the DMCA Applied to Online Defamation*, 25 CARDOZO ARTS & ENT. L.J. 237, 239 (2007) (advocating for DMCA-styled notice and takedown provisions to be added to Section 230 of the CDA).

133. *Id.*

134. *See* Saint, *supra* note 130, at 66.

135. *See* Matthew G. Jeweler, Note, *The Communications Decency Act of 1996: Why § 230 Is Outdated and Publisher Liability for Defamation Should Be Reinstated Against Internet Service Providers*, 8 U. PITT. J. TECH. L. & POL'Y 3, 3 (2007) ("[W]ith the Internet being the dominant medium that it is, the CDA is outdated and unfair, and should be amended or repealed in favor of the common law framework for publisher liability in defamation.").

136. William H. Freivogel, *Does the Communications Decency Act Foster Indecency?*, 16 COMM. L. & POL'Y 17, 20 (2011).

137. *See* Cecelia Ziniti, Note, *The Optimal Liability System for Online Service Providers: How* Zeran v. Am, Online *Got It Right and Web 2.0 Proves It*, 23 BERKELEY TECH. L.J. 583, 600–01 (2008).

the Internet to thrive since 1996."[138] However, despite its existing issues, the CDA is unlikely to be amended in an age of Congressional gridlock. Consequently, the CDA, despite its controversies, must be considered when determining the extent to which ISPs can act in reducing offensive third-party content online.

III. HOW WEB HOSTS CAN AND SHOULD UTILIZE THE EXISTING SECTION 230(C) FRAMEWORK TO TAME THE MOB

The CDA's framework, by both its design and judicial interpretations, grants ISPs the ability to develop web-hosting policies of their choosing and moderate content without the fear of facing liability for publishing or distributing potentially tortious content created by third parties.[139] ISPs enjoy this broad immunity for the editorial choices they do or do not make in removing objectionable content under the CDA. ISPs, therefore, have untapped potential under the CDA to curb the prevalence of online shaming. This Note suggests that ISPs can achieve this goal by altering the choice architecture that users engage with to inform users about the dangers of shaming before engaging in shaming activity. Good choice architecture ensures that a signal is consistent with a desired subsequent action.[140] Implementing a shaming "warning"—a signal—would prompt an individual to engage in a desired behavior, such as declining to engage in shaming activity. Prompting users to consider the far-reaching consequences of online shaming beforehand is a feasible way to lessen the prevalence of shaming without unduly impinging upon users' free speech. Curbing the prevalence and severity of online public shaming goes to the very heart of the CDA's purpose: to reduce the prevalence of offensive content online.

---

138. *CDA 230: The Most Important Law in Protecting Internet Speech*, ELECTRONIC FRONTIER FOUND., https://www.eff.org/issues/cda230 [https://perma.cc/E9YV-3NZ7] (last visited Feb. 7, 2017).

139. Many judicial interpretations of Section 230 have granted immunity to ISPs that exert editorial control over third-party content. *See* Batzel v. Smith, 333 F.3d 1018, 1031 (9th Cir. 2003) (holding that the ISP was not liable as an information content provider for allowing potentially defamatory content to be posted on its listserv because the ISP did nothing more than make minor alterations to original defamatory content); Ben Ezra, Weinstein, & Co. v. Am. Online, Inc., 206 F.3d 980, 985 (9th Cir. 2000) (holding that ISP was not liable as an information content provider when ISP altered and edited stock quotations provided by third party); *see supra* Section II.A.

140. *See* RICHARD H. THALER & CASS R. SUNSTEIN, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS 83–85 (2008).

A. Altering Choice Architecture to Discourage Shaming

One powerful tool that ISPs currently have at their disposal to diminish the prevalence of online public shaming is employing choice architecture that would result in users declining to engage in shaming activity.[141] The idea behind choice architecture is based in "stimulus response compatibility"—a signal that an individual receives should be consistent with the desired outcome.[142] An example of good choice architecture is where large door pulls—a stimulus—would result in a door opening inward, not outward, when the handles are pulled.[143] When a stimulus is inconsistent with the desired outcome—a red stop sign that has the word "GO" printed on it, for example—people make errors, choosing to stop at the sign instead of proceeding with its printed instructions.[144] Choice architects thereby indirectly influence the choices of others by altering available choices and cues.[145] Another example of good choice architecture is capitalizing on the understanding that people making a decision will choose the path of least resistance by defaulting to the optimal option.[146] For example, magazine franchises can increase their rates of subscription renewal simply by defaulting renewal and requiring that readers opt *out* of the subscription renewal if they would like to stop receiving the magazine.[147] Based on this choice architecture, more subscribers will take the path of least resistance and keep subscribing to the magazine instead of taking the necessary steps to unsubscribe by opting out.[148]

ISPs have taken advantage of their roles as choice architects; for example, Google added a "Forgotten Attachment Detector" in 2008 that alerted users who mentioned the term "attachment" in the body of an email but forgot to actually attach any files.[149] ISPs could take advantage of choice architecture to reduce shaming by prompting users to be more contemplative before engaging in often mindless shaming activity online. By prompting users with a "shaming" warning, those who are contemplating engaging in an act of shaming may become aware of the wide array of damaging consequences their

141. *Id.*
142. *Id.*
143. *Id.*
144. *Id.*
145. *Id.*
146. *Id.*
147. *Id.* at 87.
148. *Id.* at 85–87.
149. Jon Kotker, *New in Labs: Forgotten Attachment Detector*, GOOGLE BLOG (Sept. 15, 2008), https://gmail.googleblog.com/2008/09/new-in-labs-handy-intern-tweaks.html [https://perma.cc/BC35-25XQ].

shaming behavior could inflict upon a stranger,[150] consequently lowering the prevalence of shaming online. Forcing an individual to confront the harms he may inflict upon another could serve as a deterrence function, which is especially important in an age where the ability to engage in shaming activity is tantalizingly easy.

What would such a "shaming" warning look like in practice? An ISP could use algorithms and other filters to detect certain keywords or usernames that indicate that users are contemplating engaging in shaming activities. Upon detection, the ISP would send an automated warning to the user before completing the shaming post. The user would be confronted with the decision to post or not to post, but in either scenario would be advised of the possible long-term or even unintended ramifications of shaming before actually posting the content. Importantly, a user would have the option of posting the shaming content nonetheless—he would simply be prompted with a warning beforehand.

Figure 1 demonstrates what a "shaming" warning could look like. The warning would remind individuals of the long-term and often disproportionate effects of online public shaming and offer users a chance to view the norm violator's public apology, if one has been issued.[151] A user could then proceed with her post, or choose to edit or delete her post, in recognition of the warning.

> *Hi, [user]. It looks like your post is directed towards [norm violator]. Shaming can impose permanent and disproportionate consequences upon this individual. You can read [norm violator]'s public apology here. Please consider these implications before posting.*

FIGURE 1

How would this work in practice? A user attempting to engage in an act of shaming—making a shaming post on Twitter that reads "Can't believe a mother was so irresponsible and careless at the Cincinnati Zoo today!"—would be prompted with an iteration of the shaming warning as described in Figure 1. The user would subsequently be confronted with the choice of whether to proceed with the shaming activity in light of the warning or decline. Due to the often mindless nature of online shaming, the implementation of a warning could prompt individuals to contemplate the harms of shaming that may never have occurred to them.

---

150. *See* discussion of the consequences of shaming, *supra* Section I.B.

151. Other legal scholars have discussed the use of public apologies to combat the prevalence of shaming. *See* Klonick, *supra* note 37, at 1063–64.

B. Anticipated Objections

1. ISPs Lack the Incentive to Develop Anti-Shaming Policies

One anticipated objection to the proposed solution of ISPs implementing shaming warnings is that the CDA immunizes ISPs regardless of whether they do *or do not* take measures to screen offensive content.[152] As many other scholars and courts have noted in the context of dealing with third-party defamation, many ISPs would choose to enjoy the benefits of immunity under the CDA scheme without spending resources on developing new anti-shaming policies.[153] However, while the CDA's incentive scheme may have contributed to the "do nothing" approach that some ISPs have adopted, that certainly does not mean ISPs should be written off as a means to reduce the prevalence of online shaming. Online public shaming is a relatively recent development that ISPs may not fully understand or realize needs remedying. Arguably, in light of increased societal awareness about online shaming, prudent or progressive ISPs could investigate the potential to develop new policies to enhance their users' experiences on their platforms. In other words, just because the CDA structure does not *require* ISPs to take action to curb shaming does not mean that ISPs should not be *encouraged* to do so.

Progressive ISPs that adopt anti-shaming policies could highlight their efforts and distinguish their services to gain more users and obtain a competitive advantage. Major companies, such as Twitter, are already coming into the spotlight for failures in implementing better anti-harassment policies for their users.[154] Twitter CEO Jack Dorsey recently stated that "harassment has no place on Twitter" after a summer of controversies, including the suspension of Breitbart News Technology Editor Milo Yiannopolous's Twitter account, as well as actress Leslie Jones denouncing Twitter harassment after facing a spate of racist and sexist tweets.[155] While acknowledging the prevalence of online harassment and calling for

---

152. *See* 47 U.S.C. §§ 230(c)(1), (c)(2) (2012).

153. *See* Doe v. GTE Corp., 347 F.3d 655, 660 (7th Cir. 2003) ("[Section] 230(c) as a whole makes ISPs indifferent to the content of information they host or transmit: whether they do (subsection (c)(2)) or do not (subsection (c)(1)) take precautions, there is no liability under state or federal law. As precautions are costly, not only in direct outlay but also in lost revenue from the filtered customers, ISPs may be expected to take the do-nothing option and enjoy immunity under § 230(c)(1).").

154. *See* Christine Yang, *Jack Dorsey Said Online Harassment 'Has No Place on Twitter'*, CNBC (July 26, 2016), http://www.cnbc.com/2016/07/26/jack-dorsey-said-online-harassment-has-no-place-on-twitter.html [https://perma.cc/27X5-NS8E].

155. *Id.*

reform is a much-needed positive step, anti-shaming policies should become an addition to the array of harassment policies that major Internet companies are currently developing, and it could serve to differentiate companies competing for growth online.

2. Free Speech Concerns

Free speech concerns arise when considering the extent to which ISPs should filter offensive online content. As some judges and legal scholars have noted, implementing notice-based liability for ISPs—what existed before the advent of the CDA— could lead ISPs to remove any flagged content without properly vetting it for fear of facing distributor liability, leading to a considerable chilling of free speech online.[156] Consequently, a successful solution must require an ISP to weigh an online poster's right to free speech with the right of a norm violator to be free from the permanent effects of shaming.

The proposed "shaming warning" achieves this balance. Under the proposed solution, an individual would merely have to respond to an additional prompt in order to complete her post.[157] Altering the choice architecture involved in posting shaming content by adding a shaming warning merely alters the decision making framework that a user engages with before posting. A shaming warning would ideally require a user to reflect upon the nature of her post, as well as the deleterious effects that shaming can have on an individual, and potentially deter the shamer from engaging in the harmful activity in the first place.

IV. CONCLUSION

Online public shaming has exploded in popularity in the age of the Internet. Due to the ease of creating shaming content online, shaming's disproportionate harms, and the lack of legal recourse available to victims, web hosts should take advantage of the sweeping protections that the Communications Decency Act offers and begin implementing policies to lessen the prevalence of online public shaming. One way that web hosts can achieve this goal is by altering the choice architecture that users engage with before posting shaming

156. *See* Zeran v. Am. Online, Inc., 129 F.3d 327, 331 (4th Cir. 1997) ("The specter of tort liability in an area of such prolific speech would have an obvious chilling effect. It would be impossible for service providers to screen each of their millions of postings for possible problems. Faced with potential liability for each message republished by their services, interactive computer service providers might choose to severely restrict the number and type of messages posted."); *see also* Ziniti, *supra* note 137, at 600–01.

157. *See supra* Figure 1.

material to force users to consider the effects of it before engaging in shaming activity. The deeper message, applicable to all web users in 2016: think before you post.

*Kristine L. Gallardo**

* J.D. Candidate, 2017, Vanderbilt University Law School; B.A., 2014, University of Arizona. I would like to thank Kate Klonick for her valuable feedback and insight during the drafting process. I would also like to extend a special thanks to the VANDERBILT JOURNAL OF ENTERTAINMENT AND TECHNOLOGY LAW editors, particularly Sarah C. Dotzel, Jennifer Blasco, Natalie M. Gabrenya, and Laura E. Powell for their patience, support, and encouragement throughout this process.