

*Social Justice and Social Media: Responding to the Enforcement of Justice by the Internet
Community*
by
Gallant Fish

Submitted in partial fulfillment of the requirements of the
King Scholar Program
Michigan State University College of Law
under the direction of
Professor Michael Lawrence
Spring 2017

SOCIAL JUSTICE AND SOCIAL MEDIA: RESPONDING TO THE ENFORCEMENT OF JUSTICE BY THE INTERNET COMMUNITY

INTRODUCTION: THE ENFORCEMENT OF JUSTICE BY THE INTERNET COMMUNITY

The internet has reshaped how the community participates in the work of enforcing justice. The community has always played an important role in enforcing social norms through tools such as shaming and ostracization. In the past, norm enforcement by the community played a primary role in sanctioning deviant behavior and maintaining social order. Even after the formal legal system took control of criminal punishment, the community continued to play an important role. As communities grew larger and diversified, the role of the community in punishing deviant behavior decreased. Recently, however, the advent of the internet and social media has allowed community norm enforcement to again have a central role in enforcing justice by both expanding the tools of enforcement available to the community and increasing the effectiveness of these tools.

Scholars and journalists have given this enhanced community norm enforcement a number of labels—online shaming,¹ digital vigilantism,² or internet outrage.³ None of these definitions, however, fully capture the unique nature of this phenomenon. These scenarios often begin with shaming, but they lead to and incorporate tangible consequences, such as the loss of a job or even death threats that necessitate relocation. In addition, the entire community is involved in the norm enforcement, so “vigilantism” is an improper label. Finally, “internet outrage” partially describes the phenomenon but does not take into account how the outrage is directed toward a specific individual to condemn a discrete action. Because this behavior is a community endeavor with the purpose of reinforcing norms and punishing unethical behavior, this Note refers to the phenomenon as “the enforcement of justice by the internet community,” abbreviated “internet justice.”

Internet justice is unique from other, harmful online behaviors, such as internet harassment⁴ or cyber bullying.⁵ While internet justice uses some of the same tools as cyber harassment and bullying, internet justice has several distinguishing characteristics. First, internet

¹ JON RONSON, *SO YOU'VE BEEN PUBLICLY SHAMED* 292-94 (2015).

² Bianca Bosker, *Why We Should All Fear the Righteous Online Mob*, HUFFINGTON POST (Dec. 27, 2013, 1:29 PM), http://www.huffingtonpost.com/2013/12/27/justin-sacco-online-vigilantism_n_4505452.html; Bronwen Clune, *Digital Vigilantism: Think Before Putting Pictures of 'Wrongdoing' Online*, THE GUARDIAN (Nov. 28, 2013, 7:22 PM), <https://www.theguardian.com/commentisfree/2013/nov/29/digital-vigilantism-think-before-pictures-of-wrongdoing-online>.

³ Spencer Kornhaber, *Ending the Internet Outrage Cycle*, THE ATLANTIC (Apr. 3, 2015), <https://www.theatlantic.com/entertainment/archive/2015/04/so-youve-been-publicly-shamed-trevor-noah-and-why-everyone-needs-to-quiet-down-about-everything/389360/>.

⁴ Cyber harassment includes acts such as “trolling” and “revenge porn.” Trolling targets vulnerable populations such as women, racial minorities, or homosexuals and involves the posting of hateful or violent messages directed at individuals in these groups. See Joel Stein, *How Trolls Are Ruining the Internet*, TIME (Aug. 18, 2016), <http://time.com/4457110/internet-trolls/>. Revenge porn involves the posting online of sexual explicit photos of a former intimate partner in order to humiliate him or her. Paul J. Larkin, Jr., *Revenge Porn, State Law, and Free Speech*, 48 LOY. L.A.L. REV. 57, 63-64 (2014).

⁵ Cyberbullying involves the use of social media to humiliate classmates in a variety of ways. See *State Cyberstalking, Cyberharassment and Cyberbullying Laws*, NATIONAL CONFERENCE OF STATE LEGISLATURES, <http://ssl.csg.org/dockets/2012cycle/32B/32Bdocmins/Cyberstalking,%20Cyberharassment%20and%20Cyberbullying%20Laws.pdf> (last updated Dec. 20, 2010) (providing definitions of cyberstalking, cyberharassment, and cyberbullying).

justice typically focuses on issues currently in the social justice spotlight, such as sexism, racism, rape, or animal cruelty.⁶ Second, where perpetrators of cyber harassment or bullying typically target those in positions of vulnerability, internet justice targets individuals in perceived positions of privilege.⁷ Finally, while acts of cyber harassment or bullying are usually condemned by the larger community, acts of internet justice are typically accepted or lauded as justified.⁸

While negative internet behaviors such as cyber harassment raise difficult legal questions of how to best punish the behavior without violating the First Amendment, most in the legal community would agree that cyber harassment is negative and should be punished to the extent possible. With internet justice, however, the discussion is not so simple. Community norm enforcement, although secondary to the judicial system in sanctioning deviant behavior, plays an important role in society, a role that the judicial system cannot fill. Because it does not have the checks and balances of the judicial system, however, community norm enforcement can apply sanctions that are overly harsh or undeserved entirely. Internet justice, community norm enforcement as evolved in the internet age, reflects in amplified form both the positive and negative aspects of community norm enforcement.

This Note formulates a suggested response to internet justice by the legal community. First, this Note explores the positive aspects of internet justice, arguing that it has significant value in supplementing the judicial system in the pursuit of a just society. Second, this Note also proposes a solution to the need to regulate the negative aspects of internet justice without unduly impeding the positive role that it plays. This solution will require drawing a theoretical line where an instance of internet justice shifts from beneficial to dangerous, which this Note argues is when the shaming begins to resemble cyberstalking, which places the target of the shaming in physical danger. Having identified this line, the next step should be crafting a regulatory framework requiring internet service providers to remove certain posted content that places individuals targeted by shaming in danger. Before exploring such a solution, however, an understanding of the role that norm enforcement plays in society is necessary.

I. A BRIEF HISTORY OF COMMUNITY NORM ENFORCEMENT

Law is not the only force that governs the actions of individuals. Communities also regulate the actions of community members by enforcing norms through informal sanctions. In his field-defining work *Order Without Law: How Neighbors Settle Disputes*, Robert C. Ellickson defines norms as standards of behavior that are both descriptive and prescriptive; that is, norms are behavior patterns that people normally follow and that most people ought to adhere to.⁹ When individuals violate these norms, the community applies sanctions to correct the behavior.¹⁰ Similarly, the community applies rewards to encourage behavior that complies with the norms.¹¹

⁶ See *infra* Part II.

⁷ See *id.*

⁸ See *id.*

⁹ ROBERT C. ELICKSON, *ORDER WITHOUT LAW: HOW NEIGHBORS SETTLE DISPUTES* 126 (1994).

¹⁰ *Id.* at 124-26.

¹¹ *Id.* Ellickson defines rewards as “goods, services, or obligations that a person would assign a positive monetary value.” *Id.* at 124. Conversely, “punishments are goods, services, obligations that a person would pay to be rid of.” *Id.*

Ellickson recognizes five “controllers” that enforce sanctions to govern behavior. First, individuals self-regulate their own behavior according to personal ethics.¹² Second, parties to a contract regulate behavior by applying rewards (reciprocating goods or services) or punishments (not reciprocating) in order to compel the other party’s performance of the contract terms.¹³ Third, nongovernmental organizations apply organizational rules to govern behavior of organization members.¹⁴ Fourth, governmental actors apply legal sanctions to compel obedience to the law.¹⁵ Finally, the community as a whole, which Ellickson labels “social forces,” enforces norms by means of informal methods of control.¹⁶

These community norm enforcement sanctions can be incredibly powerful. In *Conditions of Successful Degradation of Ceremonies*, sociologist Harold Garfinkel describes the effect of the most severe type of sanction—the public denunciation.¹⁷ Garfinkel describes the successful public denunciation as resulting in a complete destruction and recreation of the offender’s identity.¹⁸ The offender is transformed from someone fundamentally alike to the rest of the society’s members to someone who is in essence an offender, i.e. a “murderer,” “liar,” or “traitor.”¹⁹

In the past, and presently in some areas of the world, tribal communities have used informal, shame-based methods of norm enforcement as the main means of regulating the behavior of community members.²⁰ In colonial America, where communities were small and intimate, the formal justice system used similar shame-based sanctions as the main method of justice enforcement for much of America’s early history.²¹ As society expanded and diversified, shame-based sanctions became less effective.²² As a result, the justice system largely abandoned shame-based sanctions in favor of incarceration, although some judges have experimented with shaming sanctions in recent years.²³ Today, however, the internet has again created a community where informal norm enforcement can effectively conscript behavior on a large scale.

A. Norm Enforcement Amongst Tribal Communities

Although all cultures use informal community norm enforcement on some level, some tribal communities have relied upon informal shame-based sanctions as the primary method of maintaining group order.²⁴ For example, anthropologists living among Inuit tribes in the early 20th century noted that the tribal members relied upon “song duels” to resolve conflicts between community members.²⁵ If one community member accused another of wrongdoing, that wronged community member would publicly challenge the alleged wrongdoer and, through a chant,

¹² *Id.* at 131.

¹³ *Id.*

¹⁴ *Id.*

¹⁵ *Id.*

¹⁶ *Id.*

¹⁷ Harold Garfinkel, *Conditions of Successful Degradation Ceremonies*, 61 AM. J. SOC. 420 (1956).

¹⁸ *Id.* at 421.

¹⁹ *Id.* at 421-22.

²⁰ *See infra* Subsection I.A.

²¹ *See infra* Subsection I.B.

²² *See infra* Subsection I.C.

²³ *See id.*

²⁴ *See* Toni M. Massaro, *Shame, Culture, and American Criminal Law*, 89 MICH. L. REV. 1880, 1991-12 (1991).

²⁵ *Id.* at 1991.

would seek to shame him or her.²⁶ Through the chant, the accuser would list the alleged wrongdoer's offenses before the community.²⁷ In this way, the tribes were able to resolve conflicts and maintain social order without the use of formal legal process.²⁸

B. The Use of Community Shame-Based Sanctions in Colonial America

In colonial America, local communities used similar shame-based sanctions via both the formal legal system and informal methods as the primary means of punishing wrong behavior and regulating group conduct.²⁹ The formal legal system used a variety of shaming techniques ranging from coerced public repentance to physically and emotionally severe public humiliation.³⁰ On one end of the spectrum, the local government officials would use an "admonition," which would involve a public recitation of the wrongdoer's offenses, followed by an opportunity for the offender to confess and seek the community's forgiveness.³¹ More severe shaming punishments included the stocks and the pillory, where the offender would be displayed and humiliated in a public location and potentially subjected to whipping as well.³² The most severe shame punishment inflicted was banishment, devastating because of both the emotional weight of total rejection by the community and the practical difficulty of building a new life in another community.³³

Colonial communities, primarily through the church, would also utilize shame-based norm enforcement outside of the legal system. Puritanism shaped colonial society in many ways, particularly how communities would enforce norms.³⁴ At an early age, parents would use harsh discipline and the doctrine of original sin to teach children to view all willfulness and rebellion as manifestations of sin.³⁵ As a result, community members were raised with a high sensitivity to shame.³⁶ Thus, church communities could effectively sanction behavior through practices such as forced confession and repentance before the congregation.³⁷ In this way, communities could enforce norms through not only formal, legal means, but also through extra-legal, community-based methods.

²⁶ *Id.*

²⁷ *Id.*

²⁸ *Id.* ("Using these song duels, the Eskimos sought to preserve the social order without violence or other formal coercion.").

²⁹ Ryan K. Williams, *Shaming and Shunning*, in *THE SOCIAL HISTORY OF CRIME AND PUNISHMENT IN AMERICA: AN ENCYCLOPEDIA 1657* (Wilber R. Miller, ed. 2012).

³⁰ *Id.* at 1657-59 (describing the various shaming punishments used by the colonist communities).

³¹ See Massaro, *supra* note 24, at 1912.

³² See Williams, *supra* note 29, at 1658-59.

³³ See *id.*

³⁴ See Massaro, *supra* note 24, at 1914 ("Puritan culture, especially the belief in the doctrine of predestination, may help to explain the seemingly unfeeling tone, and the specific methods, of Puritan punishment"); see also Adam J. Hirsch, *From Pillory to Penitentiary: The Rise of Criminal Incarceration in Early Massachusetts*, 80 MICH. L. REV. 1179, 1221-23 (1982) (describing the effects of puritanism on the colonial approach to crime).

³⁵ See Massaro, *supra* note 24, at 1915.

³⁶ See *id.*

³⁷ See *id.* at 1913 ("The offender often was forced to confess publicly to her congregation, sometimes dressed in a white cloth, and beg their forgiveness. This forgiveness, or redemption, effectively drew the offender back into the fold and further reinforced the moral order.").

C. Moving Away from Shame-Based Sanctions

In both tribal communities and in colonial American, similar factors allowed for the success of shame-based and community-based sanctions in regulating group behavior. The primary factor in both cases was the small size of these communities.³⁸ Colonial communities most often had less than two thousand inhabitants.³⁹ This small size resulted in communities where most inhabitants knew each other and would experience great shame if disgraced publicly before other community members.⁴⁰

Several other factors, however, also facilitated the success of shame-based punishments. For example, in tribal and colonial communities, inhabitants relied heavily upon each other, so sanctions that resulted in reintegration into society, as opposed to removal from society through incarceration, were preferable.⁴¹ In addition, because of this mutual interdependence, community members had incentive to be involved in the lives of others in the community instead of remaining distant and uninvolved.⁴² Community members also could not easily leave their community and move to a new community to escape the pain of a shaming. Finally, colonial communities in particular had a clearly defined set of shared moral beliefs that allowed communities to apply shame-based punishments that would be widely effective.⁴³

As societies grew and diversified, these factors faded, which resulted in the decreasing effectiveness of both informal and formal shame-based social control. Communities grew in size, and transitions between communities become much more common.⁴⁴ Industrialization decreased the need for reliance on fellow community members. Because of this breakdown in community cohesion, shame sanctions no longer carried the same weight, either of emotional pain or practical effect.⁴⁵

Crime rates spiked as previously standard shame-based sanctions lost their effectiveness.⁴⁶ As a result, during this time period, the formal legal system turned to incarceration as the standard sanction for wrong behavior.⁴⁷ Today, although some judges have experimented with shame-based sanctions in recent years,⁴⁸ incarceration remains the main tool of punishment used by the formal legal system. In recent years, however, with the rise of social media, the internet has opened the door for effective social control through informal means of community norm enforcement.⁴⁹

³⁸ *See id.* at 1916; Hirsch, *supra* note 34, at 1223.

³⁹ Hirsch, *supra* note 34, at 1223.

⁴⁰ Massaro, *supra* note 24, at 1912 (“Moreover, the social intimacy of colonial communities meant that criminal offenders typically were known members of the group, not transient outsiders. Thus, the fear of disgrace before the community was considerable.”).

⁴¹ *See id.* at 1916 (“If [the community member] must depend greatly on the group for social, economic, or political support, or cannot leave the group easily, then a social sanction will have a tremendous impact.”).

⁴² *See id.*

⁴³ *See id.* (“These cultures have widely shared moral or behavior expectations of their citizens, which are public.”); *see also supra* notes 34-36 and accompanying text.

⁴⁴ *See* Hirsch, *supra* note 34, at 1223, 1228 (“As late as 1690, no town in Massachusetts, save Boston and Salem, held more than 2,000 persons. . . . By 1765, some thirty towns contained more than 2,000 residents apiece, and almost half of all towns within the province had swelled beyond the 1,000 person mark.”).

⁴⁵ *See id.* at 1229-32.

⁴⁶ *Id.* at 1229-31.

⁴⁷ *See id.* at 1234-42.

⁴⁸ *See* Massaro, *supra* note 24, at 1884-90 (describing recent examples of modern shaming punishments).

⁴⁹ *See infra* Section I.D.

D. The Reemergence of Community Norm Enforcement with the Internet

The internet has created an online community with characteristics that allow for the effective use of informal shame-based norm enforcement.⁵⁰ Social media is a public forum where a perceived offender can be brought before the community and shamed. Like in the colonial communities, leaving the community to escape the pain of the shame punishment is difficult, if not impossible. While an individual can leave the social media forum, the shaming will continue in the individual's absence and affect the individual's friends, family, and co-workers.

In addition to bearing similarities to colonial and tribal communities that allowed for the effectiveness of shaming sanctions, the social media community also has differences that both increase the effectiveness of shaming penalties and remove safeguards present in colonial and tribal communities. The social media community can enforce penalties that are more severe than those used in colonial times because of the size of the community and the permanence of the shaming. In tribal and colonial times, the small size of the communities increased the effectiveness of shamings.⁵¹ Because the internet creates a forum visible by millions simultaneously, however, the immense size of the social media community only intensifies the severity of a shaming, as opposed to detracting from it.

Because everything on the internet is preserved permanently, the effects of a shamings also have the potential to last much longer than in colonial times.⁵² In colonial times, the shaming would end when the individual was removed from the stocks and released. On social media, however, the shaming can continue indefinitely, as a google search of the shamed individual's name can bring back the same Tweets, photos, and news articles that resulted from the shaming.⁵³ Shamings on the internet, then, have the potential to reach a larger audience and to last much longer than punishments inflicted in colonial times.

Not only are the penalties potentially more severe, however, but social media shamings also lack the restraints, both legal and societal, that tempered shaming sanctions in the past. Societally, shamings in colonial times were restrained by the mutual interdependence of community members and the resulting need to reintegrate offenders back into the community.⁵⁴

⁵⁰ See Jon Ronson, SO YOU'VE BEEN PUBLICLY SHAMED (2015) ("We have always had some influence over the justice system, but for the first time in 180 years—since the stocks and pillory were outlawed—we have the power to determine the severity of some punishment."); see also Lauren M. Goldman, *Trending Now: The Use of Social Media Websites in Public Shaming Punishments*, 52 AM. CRIM. L. REV. 415 (2015) ("It appears that an offender's community via online social media platforms contains the cultural conditions that existed during the colonial period, which fostered an environment in which public shaming thrived.").

⁵¹ See *supra* notes 38-40 and accompanying text.

⁵² See Cole Stryker, *The Problem with Public Shaming*, THE NATION (Apr. 24, 2013), <https://www.thenation.com/article/problem-public-shaming/> ("Then there is the permanence problem. Once embarrassing information about a person is online, it's never going to go away. Imagine, thirty years from now, some potential employer evaluating a candidate based on a thoughtless remark she made as a teenager. The permanence of uploaded information ensures that modern shamings, while obviously milder in severity, can far exceed the scope of the scarlet letter, the most extreme manifestation of which was at least branded on the chest, where it could be covered.").

⁵³ See Ronson, *supra* note 50, at 211-13 (discussing the fear for one particular individual subjected to shaming that a new employer or friend will find evidence online of her shaming that took place a year earlier).

⁵⁴ See Massaro, *supra* note 24, at 1916.

In the online community, however, little to no interdependence exists and, as a result, little incentive exists to forgive and reintegrate the offender.⁵⁵

Perhaps more importantly than the lack of societal restraints, online shamings also do not have the protections of legal due process. While due process has evolved considerably since colonial times, the colonial legal system still protected community members from entirely arbitrary shaming punishments.⁵⁶ Internet shamings have no such safeguards.⁵⁷ So then, the internet has created a forum for shaming sanctions that are both potentially more severe than those in colonial American and not restrained by the social and legal forces present in colonial communities. An examination of how social media shamings have operated so far reveals that these theoretical concerns have played out in reality.⁵⁸

II. COMMUNITY NORM ENFORCEMENT ON THE INTERNET

History has shown that community norm enforcement through shaming sanctions has the potential to be sufficiently effective to operate as the main method of maintaining social order.⁵⁹ That the internet has created the conditions for community norm enforcement to again operate effectively, but without the legal and social safeguards of the past, raises concerns. Numerous examples of social media shamings over the past several years illustrate that community norm enforcement on the internet has indeed effected severe results in the lives of perceived offenders targeted by shamings.

Social media shamings have generally fallen into three categories. First, the internet community may initiate a shaming in response to online speech, such as a Tweet or a picture posted on Facebook.⁶⁰ Second, the internet community may impose online shaming sanctions in response to actions taken offline, actions that may or may not legally constitute a crime.⁶¹ Third, and finally, the internet may react with a shaming to a supposed norm violation that never actually occurred.⁶²

Regardless of which of these forms an online shaming takes, the shaming usually follows a common pattern.⁶³ Understanding this pattern is essential to effectively regulating internet justice without unduly hindering the positive functions that it can perform. Effective regulation

⁵⁵ See Ronson, *supra* note 50, at 306-08 (describing the angry online response the author received after stating online that he felt badly for a particular victim of online shaming).

⁵⁶ Although the community could apply shame-based sanctions outside of legal community, these were limited to lighter sanctions, such as forced repentance in front of the church congregation, as opposed to sanctions such as the stocks or pillories. See Massaro, *supra* note 24, at 1913.

⁵⁷ Journalists outside the legal community have recognized the dangers of online shamings because of the lack of any due process, further evidencing the need for the legal community to respond to internet justice. See Erik Kain, *Internet Mob Justice Isn't Justice at All*, Forbes (May 14, 2015, 4:51 PM), <https://www.forbes.com/forbes/welcome/?toURL=https://www.forbes.com/sites/erikkain/2015/05/14/internet-mob-justice-isnt-justice-at-all/&refURL=https://www.google.com/&referrer=https://www.google.com/> (“Sometimes the people being shamed are bad people. But always the shaming circumvents due process [and] precedes true justice. . . . Innocent until proven guilty doesn't matter at all to the outraged internet.”).

⁵⁸ See *infra* Part II.

⁵⁹ See *supra* Part I.

⁶⁰ See *infra* Subsection II.A.1.

⁶¹ See *infra* Subsection II.A.2.

⁶² See *infra* Subsection II.A.3.

⁶³ See *infra* Subsection II.B.

will require pinpointing when in the internet shaming process the primary effect shifts from benefit to harm.

A. Archetype Examples of Internet Justice

1. *When the Internet Sanctions Online Speech*

Likely the most common internet shaming scenario is a response to norm-violating internet speech. One of the most powerful norms today is the norm against prejudice, particularly from a position of privilege. Because prejudice can be expressed through speech as easily as through actions, social media operates as a medium where this norm can be both violated and enforced.⁶⁴ In one of the most well-known instances of online shaming, Justine Sacco experienced first-hand the full force of the internet community's ability to enforce the norm against prejudice.⁶⁵

With one Tweet,⁶⁶ Justice Sacco inadvertently sparked an outrage that illustrates the norm enforcement power of the internet community.⁶⁷ Before her Tweet, Sacco worked at an internet marketing corporation and had a minimal internet presence with only 170 Twitter followers.⁶⁸ Then, while on a trip to Africa, Sacco sent out the following Tweet while waiting in the airport: "Going to Africa. Hope I don't get AIDS. Just kidding. I'm white!"⁶⁹ When Sacco checked her phone before boarding the plane, her Tweet had received no responses.⁷⁰ By the time the plane landed eleven hours later, Sacco had become the number one worldwide trending topic on Twitter.⁷¹

While on the plane, one of Sacco's Twitter followers had sent the Tweet to Sam Biddle, a writer at Gawker with fifteen thousand Twitter followers.⁷² Biddle shared the Tweet with his

⁶⁴ In an interested illustration of this, in late 2014 Slate published an article highlighting an instance of internet outrage for every day of 2014. Julia Turner et al., *The Year of Outrage*, SLATE (Dec. 17, 2014 11:48 PM), http://www.slate.com/articles/life/culturebox/2014/12/the_year_of_outrage_2014_everything_you_were_angry_about_on_social_media.html. The topics of outrage ranged from environmentalism to politics to international relations. *See id.* Of the 365 instances of outrage highlighted, roughly 112 were responses to perceived prejudice against women or racial or ethnic minorities. *See id.* While Slate has a fairly liberal reputation, *see* Aaron Blake, *Ranking the Media from Liberal to Conservative, Based on Their Audiences*, WASH. POST (Oct. 21, 2014), https://www.washingtonpost.com/news/the-fix/wp/2014/10/21/lets-rank-the-media-from-liberal-to-conservative-based-on-their-audiences/?utm_term=.039c3b181f47, this ratio is still an interesting indication of the importance of the norm against prejudice because the article at least attempts to highlight instances of outrage across the spectrum of political views.

⁶⁵ The internet community's self-reflection on Justine Sacco began when the New York Times published an article by journalist Jon Ronson detailing Justine's story of how online shaming drastically affected her life. *See* 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>. Jon Ronson later continued the conversation concerning Justine Sacco and internet shaming in general with his book *SO YOU'VE BEEN PUBLICLY SHAMED?*. *See* RONSON, *supra* note 50.

⁶⁶ A Tweet is a message posted by a Twitter user on his or her Twitter page. The user's followers on Twitter can then view the Tweet on their Twitter pages.

⁶⁷ *Id.*; *see also* Jon Ronson, *When Online Shaming Goes Too Far*, TED (June 2015), https://www.ted.com/talks/jon_ronson_what_happens_when_online_shaming_spirals_out_of_control#t-285789.

⁶⁸ Ronson, *One Stupid Tweet*, *supra* note 65.

⁶⁹ *Id.*

⁷⁰ *Id.*

⁷¹ *Id.*

⁷² *Id.*

followers and wrote an accompanying piece on Gawker.⁷³ Immediately, the internet community began to react, sharing the Tweet and writing responses such as “How did @JustineSacco get a PR job?! Her level of racist ignorance belongs on Fox News. #AIDS can affect anyone!” and “All I want for Christmas is to see @JustineSacco’s face when her plane lands and she checks her inbox/voicemail.”⁷⁴ Twitter users waited anxiously for Sacco to land, and one Twitter user even went to the airport to take a photo of Sacco as she arrived at the airport and posted the photo on Twitter.⁷⁵ By the time Sacco realized this fallout and deleted both her Tweet and her Twitter account, an internet community that stretched across the world had placed Sacco in the internet’s equivalent of the community forum and subjected her to a shaming punishment unlike anything colonial America could have envisioned.⁷⁶

For Sacco, the consequences extended beyond mere embarrassment. Her employer posted its disapproval of the Tweet and promptly fired Sacco.⁷⁷ Her family told her that she had tarnished the family name.⁷⁸ She ended her vacation early, in part because the workers at the hotels where she planned on staying threatened to strike if she arrived.⁷⁹ Several months after the initial Tweet, Sacco shared with journalist Ron Jonson that she still experienced shaming from the internet community and struggled to engage in activities as normal as dating, for fear her date would Google her name.⁸⁰

Since Justice Sacco, the internet has continued to effectively use shaming as a tool to sanction norm-violating online speech. Online speech, however, is not the only behavior that the internet community can reach through shame-based sanctions. The internet community has shown itself effective at sanctioning norm-violating behavior outside the internet as well.

2. When the Internet Sanctions an Individual’s Actions Outside the Internet

Although the legal system now punishes most actions that the community categorizes as antinormative, the internet community may use shaming to punish actions when the community believes that the justice system has failed adequately to do so. In some instances, the internet community may utilize shaming when the justice system has applied a punishment not sufficient in relation to the action. In other cases, the internet community may use norm enforcement to punish actions that the justice system does not recognize or punish as criminal.

Brock Turner experienced the ability of the internet community to apply its own set of sanctions when the justice system applies a punishment that the community views as inadequate. In June of 2016, Turner, a student at Stanford University, was convicted of assaulting an unconscious woman after a party.⁸¹ While sexual assault is unfortunately not uncommon, the

⁷³ *Id.*

⁷⁴ *Id.*

⁷⁵ *Id.*

⁷⁶ *Id.*

⁷⁷ *Id.*

⁷⁸ *Id.* (“[Justine’s] extended family in South Africa were African National Congress supporters — the party of Nelson Mandela. They were longtime activists for racial equality. When Justine arrived at the family home from the airport, one of the first things her aunt said to her was: ‘This is not what our family stands for. And now, by association, you’ve almost tarnished the family.’”).

⁷⁹ *Id.*

⁸⁰ *Id.*

⁸¹ Liam Stack, *Light Sentence for Brock Turner in Stanford Rape Case Draws Outrage*, N.Y. TIMES (June 6, 2016), <https://www.nytimes.com/2016/06/07/us/outrage-in-stanford-rape-case-over-dueling-statements-of-victim-and-attackers-father.html>.

internet community responded uniquely to this case for at least two reasons. First, at the trial, the victim of the assault read an articulate, moving account of her experience since the assault, which BuzzFeed published online the next day.⁸² This statement gave words to a growing awareness of campus sexual assault and was viewed five million times within the first three days after it was published.⁸³ Second, the judge in the case sentenced Turner to only six months imprisonment, which the internet community found insufficient and unacceptable.⁸⁴

In response, the internet community applied shaming sanctions where it perceived that the justice system had failed. Individuals on social media Tweeted and posted extensively about Turner, and major media responded by publishing innumerable pieces on the case.⁸⁵ Images such as those pictured below inundated the internet.



Three months later, after Turner was released for good behavior, the sanctions applied and encouraged by the internet community continued past articles and posts about Turner. Protestors gathered outside the jail on the day of Turner's release in order to speak out against

⁸² See Katie J.M. Baker, *Here Is the Powerful Letter the Stanford Victim Read Aloud to Her Attacker*, BUZZFEED (June 3, 2016, 4:17 PM), https://www.buzzfeed.com/katiejmbaker/heres-the-powerful-letter-the-stanford-victim-read-to-her-ra?utm_term=.dkX1R1ENv#.ny22p2aYr.

⁸³ Stack, *supra* note 81 (“The statement, a 7,244-word cri de coeur against the role of privilege in the trial and the way the legal system deals with sexual assault, was provided by the victim and has since gone viral. By Monday, it had been viewed more than five million times on the BuzzFeed site.”).

⁸⁴ *See id.*

⁸⁵ See, e.g., Ashley Fantz, *Outrage Over 6-Month Sentence for Brock Turner in Stanford Rape Case*, CNN (June 7, 2016, 8:45 AM), <http://www.cnn.com/2016/06/06/us/sexual-assault-brock-turner-stanford/>; Marina Koren, *Why the Stanford Judge Gave Brock Turner Six Months*, THE ATLANTIC (June 17, 2016), <https://www.theatlantic.com/news/archive/2016/06/stanford-rape-case-judge/487415/>; Sam Levin, *Brock Turner Laughed After Bystanders Stopped Stanford Sex Assault, Files Show*, THE GUARDIAN (Aug. 26, 2016, 3:43 PM), <https://www.theguardian.com/society/2016/aug/26/brock-turner-stanford-sexual-assault-victim-testimony-laugh>; Gadi Schwartz & Erik Ortiz, *Brock Turner, Convicted Sexual Assault Offender, Released From Jail After 3 Months*, NBC NEWS (Sept. 2, 2016, 7:33 PM), <http://www.nbcnews.com/news/us-news/brock-turner-convicted-sexual-assault-offender-released-jail-after-3-n641981>; Bill Chappell, *Brock Turner Freed From Jail After Serving Half of 6-Month Sentence*, NPR (Sept. 2, 2016, 10:42 PM), <http://www.npr.org/sections/thetwo-way/2016/09/02/492390163/brock-turner-freed-from-jail-after-serving-half-of-6-month-sentence>; *Brock Turner Registers As Sex Offender After Controversial Jail Stint*, FOX NEWS (EPT. 6, 2016), <http://www.foxnews.com/us/2016/09/06/brock-turner-registers-as-sex-offender-after-controversial-jail-stint.html>.

⁸⁶ Available at <http://imgur.com/gallery/NOCxy>.

his early release.⁸⁷ USA Swimming publicly denounced Turner and permanently banned him from membership in the organization.⁸⁸ When he returned home, armed protesters gathered outside his house carrying signs that carried threatening messages, such as “Castrate All Rapists” and “If I Rape Brock Turner Will I Only Do 3 Months?”⁸⁹ Turner’s case, then, illustrates not only the ability of the internet community to apply sanctions when it perceives the formal justice system to have failed but also the range of sanctions that the internet community can impose.

3. *When the Internet Sanctions an Individual in Response to False Accusations*

Unfortunately, the internet community can also use these powerful tools in response to allegations that are later revealed to be mistaken or fabricated. Because anyone can upload a statement or photo making an accusation and the internet community can respond almost immediately, instances of internet justice based on false or mistaken information are likely inevitable. While the ease of internet communication that allows for the spreading of false accusations can also facilitate the communication of the truth, the accused individuals may still suffer the consequences of sanctions imposed before the truth comes to light.

The case of the “Darth Vader Selfie Dad” illustrates the ability of the internet community to quickly spread both false accusations and corrections.⁹⁰ In 2015, a woman at a mall in Australia took a picture of a stranger whom she thought had just taken a picture of her children.⁹¹ After telling security, who contacted the police, the woman posted the picture of the man on Facebook, along with the warning “Ok people, take a look at this creep . . . [B]e safe with your kids.”⁹² In reality, the man had not taken pictures of her children but had only taken a selfie in front of a Darth Vader cut-out to send to his children.⁹³ Soon, the post had been shared over 20,000 times, and the man began receiving death threats.⁹⁴ As soon as the man learned of the post, he contacted the police to explain what had really happened.⁹⁵

Because of how widely circulated the original post had been, the revelation of the actual story also spread quickly. The Daily Mail of Australia published the man’s story, and the woman apologized publicly and expressed her regret over sharing the post.⁹⁶ Despite the public recognition of his innocence, the accused man expressed the devastating effect of the shaming on

⁸⁷ See Emanuella Grinberg & Catherine E. Shoichet, *Brock Turner Released from Jail After Serving 3 Months for Sexual Assault*, CNN (Sept. 2, 2016, 8:52 PM), <http://www.cnn.com/2016/09/02/us/brock-turner-release-jail/>.

⁸⁸ Ray Sanchez, *USA Swimming Bans Brock Turner for Life*, CNN (June 10, 2016, 11:40 AM), <http://www.cnn.com/2016/06/10/us/sexual-assault-brock-turner-swimming/index.html>. Even though the internet community did not directly apply this sanction, the publicity of Turner’s action is attributable to the internet community, which likely motivated USA Swimming to take this action, or at least to do so publicly.

⁸⁹ See Morgan Winsor, *Protesters Gather Outside Brock Turner’s Family Home in Ohio*, ABC NEWS (Sept. 3, 2016, 3:36 PM), <http://abcnews.go.com/US/protesters-gather-brock-turners-family-home-ohio/story?id=41846386>.

⁹⁰ See Ashley Welch, *Woman Apologizes for Mistakenly Shaming Man on Facebook*, CBS NEWS (Ma 11, 2015, 1:43 PM), <http://www.cbsnews.com/news/woman-apologizes-after-mistakenly-shaming-man-on-facebook/>.

⁹¹ *Id.*

⁹² Sarah Michael, *‘OK People, Take a Look at This Creep!’: Man Who Mum Shamed on Facebook Because She Thought He Was Taking Photos of Her Kids . . . Was Just Taking A Selfie in Front of a Darth Vader Display to Show HIS Children*, DAILY MAIL AUSTRALIA (May 7, 2015, 11:07 PM), <http://www.dailymail.co.uk/news/article-3073095/Mother-mistakenly-shames-dad-thought-taking-photos-kids-Facebook-post-shared-hundreds-actually-taking-selfie-Star-Wars-display-children.html>.

⁹³ *Id.*

⁹⁴ *Id.*

⁹⁵ *Id.*

⁹⁶ See *id.*; Welch, *supra* note 90.

him and his children.⁹⁷ Thus, while the internet community can correct itself when it shames someone based on false or mistaken information, the shaming sanctions can still have a severe effect on the accused individual.

B. The Lifecycle of Internet Justice

Understanding the power of internet justice to impose effective sanctions, even in response to false accusations, highlights the need to understand the phenomenon and formulate an effective solution. Crafting such a solution requires understanding where in the lifecycle of internet justice the primary effect changes from norm enforcement, typically a positive social function, to danger to the individual, a negative social result. An examination of the standard stages of internet justice will help determine where in the process the law should step in and regulate.

1. *Initial Recognition*

Internet justice typically begins with initial recognition of an antinormative act by an publisher with an established internet following. For example, Justine Sacco's shaming began after Gawker editor Sam Biddle published an article highlighting Sacco's tweet.⁹⁸ Similarly, the internet community targeted Brock Turner after BuzzFeed published the witness statement from the victim that retold her experiences after the sexual assault.⁹⁹ In both these instances, the antinormative behavior would have likely never come to the public's attention without the initial recognition by an online publisher.

2. *Mutually Reinforcing Publicity*

After the initial recognition by an online publisher, the antinormative action typically gains widespread exposure through mutually reinforcing publicity between private individuals and major media sources. This begins with extensive sharing of the initial article that highlighted the antinormative behavior. Once the article becomes popular to certain extent, major media begins publishing articles discussing the popularity of the topic, including a discussion of the original action that caused the uproar.¹⁰⁰ These major media articles serve to further extend the widespread recognition of the offense.

⁹⁷ Michael, *supra* note 92.

⁹⁸ See *supra* notes 72-74 and accompanying text.

⁹⁹ See *supra* notes 82-83 and accompanying text.

¹⁰⁰ For examples of major media articles highlighting the outrage over the Brock Turner case, see *Brock Turner Sentence Causes Emotions to Run High in His Hometown*, CBS NEWS (June 11, 2016, 6:09 AM), <http://www.cbsnews.com/news/brock-turner-stanford-sexual-assault-hometown/>; Marina Koren, *Why the Stanford Judge Gave Brock Turner Six Months*, THE ATLANTIC (June 17, 2016), <https://www.theatlantic.com/news/archive/2016/06/stanford-rape-case-judge/487415/>; Curt Mills, *Sentence in Mass. Sexual Assault Case Draws Outrage*, U.S. NEWS & WORLD REPORTS (Aug. 25, 2016, 1:54 PM), <https://www.usnews.com/news/articles/2016-08-25/david-beckers-sexual-assault-sentence-draws-brock-turner-comparisons>.

3. *Online Shaming*

Simultaneously with the growing recognition of the offense begins the shaming of the targeted individual. Shaming partially consists of simply spreading the publicity of the offense by sharing articles discussing the offense, but it also includes statements made by individual social media users discussing the offense and criticizing the offender. The shaming can also incorporate mediums other than written posts, such as the images used to label Brock Turner as a rapist.¹⁰¹ Once the shaming reaches its peak, the ultimate goal becomes “effect[ing] the ritual destruction of the person denounced.”¹⁰² Before the targeted individual experiences any tangible consequences of internet justice, then, the individual suffers the reconstitution of their identity into an offender.

4. *Consequences for the Individual*

In most cases, the targeted individual does experience real-life consequences as well. Justine Sacco lost her job almost immediately as a result of the internet justice response to her antinormative speech, and she was forced to cancel her trip because of safety concerns and employee boycotts at the hotels on her itinerary.¹⁰³ Similarly, Brock Turner permanently lost his membership of USA Swimming soon after being targeted by internet justice.¹⁰⁴ As internet justice progresses, then, it applies tangible sanctions on those targeted beyond simply labeling the individual as an offender.

5. *Learning the Physical Location of the Individual*

Internet justice grows more dangerous as the internet community learns the physical location of the targeted individual. In some cases, the targeted individual’s general location may be obvious from the beginning, such as the widely publicized fact that Justine Sacco was on a plane to Africa while she was being shamed online.¹⁰⁵ In other cases, however, members of the internet community may intentionally seek out and publicize the targeted individual’s address or phone number.¹⁰⁶ Additionally, members of the internet community may seek to determine the exact location of the targeted individual, utilizing methods similar to stalking.¹⁰⁷ Although learning the physical location of the targeted individual may occur simultaneously with the rest of the shaming, it adds an important layer to the sanctions, and potential danger, placed upon the individual.

¹⁰¹ See *supra* Subsection II.A.2.

¹⁰² Garfinkel, *supra* note 17, at 421.

¹⁰³ See *supra* Subsection I.A.1.

¹⁰⁴ See *supra* note 88 and accompanying text.

¹⁰⁵ See Ronson, *One Stupid Tweet*, *supra* note 65.

¹⁰⁶ See generally Emily Bazelon, *The Online Avengers*, N.Y. TIMES (Jan. 15, 2014), <https://www.nytimes.com/2014/01/19/magazine/the-online-avengers.html> (discussing and interviewing a group of hackers that punish cyberbullies by exposing online their names, addresses, and other personal information).

¹⁰⁷ For example, in Sacco’s case, a Twitter user waited at the airport for her to land, took a picture of her when he found her, and posted the picture on Twitter, allowing the internet community to know her exact location. See Ronson, *One Stupid Tweet*, *supra* note 65.

6. Physical Danger to the Individual

Although learning the physical location of the targeted individual does not always signify danger to the individual, it allows physical safety to become a legitimate concern. If the location of the victim is known, then violent and threatening posts or messages demand careful consideration because a member of the internet community could conceivably respond by carrying out the advocated-for violence. Indeed, one of the main reasons Justice Sacco was forced to abandon her trip in Africa was the fact that local officials could not guarantee her safety.¹⁰⁸ Similarly, although Brock Turner did not experience actual violence, he was placed in fear of violence by the armed protestors who surrounded his house.¹⁰⁹

Internet justice may not always follow this progression exactly, but analyzing separately each aspect of internet justice assists in determining how to appropriately regulate it. The legal community's response to internet justice should regulate internet justice without impeding the positive role that internet justice can play. In light of the stages in the lifecycle of internet justice, regulation of internet justice should operate to remove fear of physical danger for the targeted individual to the extent possible. The legal solution can accomplish this by removing threatening and stalking material that places the individual in fear of harm. Before formulating such a regulatory solution, however, the legal community must consider whether internet justice should be allowed to exist at all.

III. HOW THE LEGAL SYSTEM BENEFITS FROM INTERNET JUSTICE

Looking at the examples of internet justice explored in the previous section, one could easily determine that online shaming should have no place in society. Shamings on social media are effective in ways beyond what was possible in the past, while at the same time more unhindered than at any point in history. Examples of internet shaming confirm that social media indeed can apply sanctions that are either too severe or unwarranted entirely. Taking these factors into consideration, some may conclude that the legal community should respond to internet justice by limiting it to greatest extent possible. Despite the concerns raised by internet justice, however, it can also play a significant, complementary role to the formal justice system that the legal community should recognize in formulating a response to internet justices.

A. Why Recognize Any Benefits to Internet Justice?

Several authors have discussed the dangers of both social media shamings and shaming sanctions generally. Dan Markel, in *Are Shaming Punishments Beautifully Retributive?*, argues that shaming sanctions are dangerous, in part because they bear more commonalities to revenge than to retribution.¹¹⁰ First, Markel draws attention to “the potentially unconstrained scope and intensity of shaming punishments,” closer to lynch mobs than to justly retributive punishment.¹¹¹

¹⁰⁸ See *id.*

¹⁰⁹ See *supra* note 87 and accompanying text.

¹¹⁰ Dan Markel, *Are Shaming Punishments Beautifully Retributive? Retributivism and the Implications for the Alternative Sanctions Debate*, 54 VAND. L. REV. 2157, 2216-22 (2001). In his article, Markel focuses on the use of shaming sanctions by the formal legal system, however, because his arguments focus on the dangers of community shaming punishments generally, his points are relevant as well to internet shaming sanctions outside the formal legal system. See *id.*

¹¹¹ See *id.* at 2217.

Second, Markel argues that shame-based punishments improperly place authority to punish into the hands of the public, thus denying the judiciary its proper role as the sole arbiter of punishment.¹¹² Finally, Markel looks to the effect of shaming punishments on both the offenders and the punishers. Shaming punishments dehumanize both the offender and the punishers as the community participates in stripping away the dignity of the offender.¹¹³

Similarly, discussing social media shaming specifically, Kate Klonick compares shamings to typical, state-enforced sanctions and concludes that, unlike state-enforced sanctions, social media shamings are indeterminate, uncalibrated, and inaccurate.¹¹⁴ Klonick explains that acts usually included in online shaming, such as re-Tweeting a photo of the offender, have indeterminate social meaning because the act could be motivated by a desire to reinforce a social norm or, alternatively, by boredom or even sadism.¹¹⁵ As noted by Markel, shamings are also uncalibrated because no governmental body restrains the punishments inflicted by the community.¹¹⁶ But Klonick points out that social media shamings amplify this concern because the internet allows for the community to impose shamings punishments much more easily than communities in the past.¹¹⁷ Finally, Klonick expresses the concern that social media shamings are often inaccurate, as illustrated by multiple examples of social media shamings that were triggered by a mistake or a lie.¹¹⁸

Despite these arguments illustrating the dangers of shaming, there are several reasons why the legal community can and should recognize the positive role that social media shamings can play. First, as discussed earlier, community norm enforcement through shame-based sanctions has played a healthy and effective role in society in the past.¹¹⁹ Viewed through the lens of this history, while the internet opens the door for new dangers associated with online shamings, it also creates a new opportunity for shamings to play a unique and beneficial role in society.

Second, because the legal system cannot realistically end online shaming entirely,¹²⁰ a response to internet justice from the legal community should recognize the potential positive roles that online shaming can play. Because speech is the principal tool of internet justice, the

¹¹² See *id.* at 2218. Markel asserts that, because the judiciary has been vested with sole authority to enact punishments, community enforcements of shaming sanctions “disrupts the notion of the state serving as the public’s agent of retribution when it delegates that power back to the public at large again.” *Id.*

¹¹³ See *id.* at 2219.

¹¹⁴ See Kate Klonick, *Re-Shaming the Debate: Social Norms, Shame, and Regulation in an Internet Age*, 75 MARYLAND L. REV. 1029, 1015 (2016).

¹¹⁵ See *id.* at 1053. By comparison, Klonick explains, the judicial system applies punishments with determinate meanings because the punishments are clearly intended to punish criminal actions. *Id.* at 1051.

¹¹⁶ See *id.* at 1053-54.

¹¹⁷ See *id.* (“[W]hile the public was always armed with its own tools and means of norm enforcement, the ready availability, low-cost, and anonymity of modern technology and the Internet has drastically changed whatever natural limits formerly conscribed shaming.”).

¹¹⁸ See *id.* at 1054-55.

¹¹⁹ See *supra* Part I.

¹²⁰ See Klonick, *supra* note 113, at 1063 (“Though there are not currently any absolute solutions to curtailing online shaming, we can learn much by considering online shaming within the framework of social norms and the shame debate, and the lessons of those who have been shamed.”). Klonick also recognizes that an absolute conscription of online shaming would not be desirable, not necessarily because of any valuable role that online shaming can play, but because of the competing normative value of free speech. See *id.* at 1062 (“[T]here is now a growing sense that online shaming is something we want to control; but any regulation must be balanced against competing normative and legal values like free speech.”).

First Amendment prevents the legal community from altogether eliminating internet justice.¹²¹ In light of this, recognizing the benefits of internet justice will allow the legal community to craft a solution that will regulate internet justice without hindering its beneficial aspects.

Third, and most importantly, the legal community should recognize the value of internet justice because internet justice has recognizable potential to play a significant role complementary to the formal legal system. Two examples will illustrate the ways that internet justice can play an important role in complementing the formal legal system.¹²²

B. The Complementary Role of Internet Justice

Most significantly, internet justice can complement the formal legal system by conscripting antinormative behavior that law enforcement cannot or will not punish. Informal norm enforcement operates to sanction behavior that the community recognizes as negative, but that the formal justice system does not reach, either because of constitutional limitations or because of societal consensus that legal liability is not appropriate.¹²³ As an enhanced form of norm enforcement, internet justice can allow a large, yet cohesive community to effectively express condemnation of and sanction negative behavior on a scale not before possible. Two areas of potential positive applications of internet justice are antinormative speech and unethical business practices.

1. Repugnant Speech

Internet justice presents a potential solution to the tension between the desire to protect speech from governmental regulation and the societal disapproval of speech that is harmful or offensive.¹²⁴ Although the First Amendment allows for regulation of some speech that involves a threat to another or spreads false accusations against another, the First Amendment prohibits the government from punishing speech solely because of its offensive nature.¹²⁵ Through internet justice, however, the internet community can express disapproval of repugnant speech and apply sanctions to deter it. Although the sanctions applied by the internet community may not always be proportionate to the offense, such as was arguably the case with Justice Sacco, internet justice has the potential to accomplish the good of reducing hateful speech, while avoiding the evil of governmental regulation of speech.¹²⁶

¹²¹ See U.S. CONST. amend. I.

¹²² See *infra* Section III.B.

¹²³ See generally Ellickson, *supra* note 9, at 4-6.

¹²⁴ See generally *Whitney v. California*, 274 U.S. 357, 374 (1927) (“[A] State is, ordinarily, denied the power to prohibit dissemination of social, economic and political doctrine which a vast majority of its citizens believes to be false and fraught with evil consequence.”).

¹²⁵ See *Virginia v. Black*, 538 U.S. 343, 358-59 (2003).

¹²⁶ There are some indications that the internet community is beginning to self-regulate the degree of punishment imposed on antinormative speech. For example, a year after Justice Sacco’s shaming, Sam Biddle, the Gawker editor who initially brought attention to Sacco’s Tweet, published a piece recognizing the “incredibly disproportionate” punishment that Sacco received. See Sam Biddle, *Justine Sacco Is Good at Her Job, and How I Came to Peace with Her*, GAWKER (Dec. 20, 2014, 9:00 AM), <http://gawker.com/justine-sacco-is-good-at-her-job-and-how-i-came-to-pea-1653022326>.

2. Ethical Business Practice

Another difficult area for the law to regulate that internet justice could address is sanctioning unethical business practices. For some forms of unethical business practices, the law may impose liability, such as knowingly utilizing forced labor in a company's supply chain.¹²⁷ Investigation and prosecution is difficult in these cases, however, and will likely only succeed in the most unambiguous cases and, even when successful, may not succeed in deterring corporate behavior.¹²⁸ Other business practices, such as poor working conditions or wages for workers overseas, may not implicate legal liability at all but may still be considered unethical by members of society.

Through internet justice, the internet community can sanction both forms of unethical business practices when the law cannot. The internet provides both the tools to learn about a business's practices¹²⁹ and the platform to call attention to and sanction businesses that violate ethical norms. Because online shaming focuses on assigning the label of "offender" to the target of the shaming, online shaming can be especially effective in deterring negative business practices, as businesses spend billions each year to maintain positive name recognition among consumers.¹³⁰ In addition, the internet community can impose further sanctions beyond shaming, such as coordinating consumer boycotts of offending companies.¹³¹ Through these tools, the internet community can complement the formal legal system by holding publicly accountable companies that engage in unethical practices that the law does not reach.

¹²⁷ See 18 U.S.C. § 1593A (2008) (assigning liability to any individuals or corporations who knowingly benefit from forced labor).

¹²⁸ See Jonathan Todres, *The Private Sector's Pivotal Role in Combating Human Trafficking*, 3 CAL. L. REV. CIR. 80, 91 (2012) ("[A] natural tendency would be to suggest corporate liability as a strategy for ensuring the private sector takes action to combat human trafficking, and it has a role to play in such instances. Criminal liability for corporate entities in this context is likely, however, to be utilized only in the most egregious cases (for example, a hotel in a red light district that offers rooms by the hour for commercial sex) and to provide little incentive for legitimate businesses that benefit indirectly from human trafficking to take ameliorative measures.").

¹²⁹ See Kathleen Wong, *5 Tools to Check If Your Clothing Is Ethically Made*, MASHABLE (Apr. 24, 2015), <http://mashable.com/2015/04/24/ethical-fashion-tools/#nABUq3PA.uqY> (discussing five organizations that provide information of the business practices of major corporations).

¹³⁰ See Lara O'Reilly, *These Are the 10 Companies That Spend the Most on Advertising*, BUSINESS INSIDER (July 6, 2015, 10:21 AM), <http://www.businessinsider.com/10-biggest-advertising-spenders-in-the-us-2015-7/#10-walt-disney-company-ad-spend-up-74-to-21-billion-the-company-spent-big-on-marketing-its-box-office-phenomenon-frozen-with-fresh-content-and-merchandise-long-after-its-2013-opening-1> (reporting that thirty-eight companies spent more than one billion dollars on advertising in 2014).

¹³¹ See Bruce Watson, *Do Boycotts Really Work?*, THE GUARDIAN (Jan. 6, 2015, 10:14 AM), <https://www.theguardian.com/vital-signs/2015/jan/06/boycotts-shopping-protests-activists-consumers> (discussing the ability of consumer boycotts to open policy discussions between the public and corporations).

IV. REGULATING INTERNET JUSTICE

Having established that internet justice has the potential to play a beneficial, complementary role to the formal legal system, the next step for a holistic response from the legal community is establishing a solution to the dangers of internet justice. As discussed above, the solution should focus on how to limit online shaming when it progresses beyond harassment and begins approaching cyberstalking. Potential solutions exist in private action,¹³² tort law,¹³³ and in criminal statutes.¹³⁴ Each of these solutions, however, has significant limitations and cannot be relied upon as a comprehensive solution to internet justice. While not a simple solution, the most effective solution for protecting against the dangers of internet justice is a regulatory scheme that places responsibility on website operators for removing threatening material from their webpages.¹³⁵

A. The Range of Potential Solutions

1. Private Action

Private action responses to social media shamings can take at least three different forms. First, victims of an unjustified online shaming can make an appeal to the internet community to “clear their names” and end the shaming.¹³⁶ This approach has had some limited success in the past in cases where a shaming was based upon mistaken or false information.¹³⁷ Appeals from a target of online shaming will not always succeed, however. And even when they do succeed, the process for the victim may be long and painful.¹³⁸ As a result, this solution is not ideal.

Second, those with an influential voice in the internet community can appeal to the community to change shaming patterns. Numerous authors and journalists have made such appeals, seeking to persuade the internet community to stop online shamings, or at least to learn more of the facts before engaging in online shaming.¹³⁹ These calls for reform, however, can

¹³² See *infra* Subsection IV.A.1.

¹³³ See *infra* Subsection IV.A.2.

¹³⁴ See *infra* Subsection IV.A.3.

¹³⁵ See *infra* Section IV.B.

¹³⁶ See Klonick, *supra* note 113, at 1062 (“Some of the private efforts of shamed individuals to reclaim their reputation and end their shaming demonstrate the power of private action to conscript shaming or cyber harassment.”).

¹³⁷ *Id.* Klonick uses the story of Jill Filipovic to illustrate the potential success of this type of action. See *id.* Filipovic became the subject of an online discussion board discussing her rape and murder, most likely in response to her feminist writings on her blog. *Id.* at 1049-50; see also Jill Filipovic, *Let’s Be Real: Online Harassment Isn’t ‘Virtual’ For Women*, TALKINGPOINTSMEMO.COM (Jan. 10, 2014 6:00 AM), <http://talkingpointsmemo.com/cafe/let-s-be-real-online-harassment-isn-t-virtual-for-women>. These threads became so viral that the first result in a Google search of Filipovic’s name would be from this message board. See Klonick, *supra* note 113, at 1050. At first, Filipovic withdrew from social networks, both online and in real life, in order to avoid further humiliation. See *id.* Eventually, however, Filipovic chose to push back, sending out a message to followers of her blog, asking them to share positive links about her online to drown out the Google results from the message board. *Id.* This tactic worked, and the primary Google results for Filipovic changed to her blog or social media profiles. *Id.* While this was an instance of pure internet harassment and not internet justice because its purpose was to lash out at Filipovic for her feminist writing, it still illustrates the potential success of private action in combating an undeserved shaming.

¹³⁸ See *id.* at 1062.

¹³⁹ See, e.g., Joel Stein, *How Trolls Are Ruining the Internet*, TIME (Aug. 18, 2016), <http://time.com/4457110/internet-trolls/>; Scott Sauls, *Internet Outrage, Public Shaming and Modern Day Pharisees*,

send conflicting messages, which may detract from their effectiveness.¹⁴⁰ In addition, based upon the continued prevalence of online shaming, these calls to action have not fully accomplished their goal.

Third, influential corporations connected to social media can develop solutions to identifying and eliminating online shaming that devolves into harassment. Google has led in this area. Through a subsidiary named Jigsaw, Google has been developing an algorithm to detect and censor harassment online.¹⁴¹ Jigsaw has designed the algorithm, named Conversation AI, to recognize language that a human moderator would label harassment.¹⁴² Already, companies such as The New York Times and the Wikimedia Foundation have partnered with Jigsaw to test Conversation AI in the comments sections of their websites, with some success.¹⁴³

While technology-driven solutions like this are promising, they also raise concerns. On one hand, the technology is not perfect at recognizing harassing language.¹⁴⁴ On the other hand, if the technology becomes too effective, it may censor too much content, thus hindering political speech on the internet.¹⁴⁵ Finally, even if the algorithm works perfectly to identify harassment, websites still have no obligation to use the algorithm.

Ultimately, while private action solutions are promising and should be encouraged by the legal community, they are not solutions that can be effected by the legal system. The legal system plays an important role in applying coercive force to parties who refuse to change their conduct in response to private appeals or to utilize technological solutions. As a result, a holistic response from the legal community should include a proposed legal solution.

RELEVANT MAGAZINE (June 10, 2015), <http://archives.relevantmagazine.com/culture/internet-outrage-public-shaming-and-modern-day-pharisees/>; Ryan Holiday, *This Is the Hollowed-Out World That Outrage Culture Has Created*, OBSERVER (Jan. 13, 2016 11:30 AM), <http://observer.com/2016/01/this-is-the-hollowed-out-world-that-outrage-culture-has-created/>; RONSON, *supra* note 50.

¹⁴⁰ Compare Kain, *supra* note 57 (calling for self-reflection and restraint before engaging in online shaming) and Ronson, *supra* note 50, at 282-87 (placing the responsibility for tempering online shaming on the social justice community that has abused shaming), with Jim Geraghty, *Trolls Haven't Ruined the Internet*, NATIONAL REVIEW (Aug. 20, 2016 4:00 AM), <http://www.nationalreview.com/article/439198/internet-trolls-can-be-avoided> (“It is easy enough to wall yourself off from people you find too toxic to abide, just as you would in real life. Or, if things get really bad, you could just log off and go enjoy the sunshine. Would that be so bad?”) and Spencer Kornhaber, *Ending the Internet Outrage Cycle*, THE ATLANTIC (Apr. 3, 2015), <https://www.theatlantic.com/entertainment/archive/2015/04/so-youve-been-publicly-shamed-trevor-noah-and-why-everyone-needs-to-quiet-down-about-everything/389360/> (calling for social media users to consider more carefully what they post online in order to avoid shaming).

¹⁴¹ See Andy Greenberg, *Inside Google's Internet Justice League and Its AI Powered War on Trolls*, WIRED (Nov. 19, 2016 6:55 AM), <https://www.wired.com/2016/09/inside-googles-internet-justice-league-ai-powered-war-trolls/>.

¹⁴² See *id.* (“In fact, by some measures Jigsaw has now trained Conversation AI to spot toxic language with impressive accuracy. Feed a string of text into its Wikipedia harassment-detection engine and it can, with what Google describes as more than 92 percent certainty and a 10 percent false-positive rate, come up with a judgment that matches a human test panel as to whether that line represents an attack. For now the tool looks only at the content of that single string of text. But Green says Jigsaw has also looked into detecting methods of mass harassment based on the volume of messages and other long-term patterns.”).

¹⁴³ See *id.*

¹⁴⁴ See *id.* (explaining that the algorithm does not recognize some especially violent or degrading language as harassment if the language is communicated to a general audience, rather than directed to a specific individual).

¹⁴⁵ See *id.* (explaining that a more political phrase, such as “Donald Trump is a moron,” rates at 99 out of 100 on the harassment indicator scale for Conversation AI).

2. Solutions from Tort Law

Private law presents a potential solution to some aspects of internet justice. Three areas of tort law could potentially assist in preventing social media shamings from devolving into cyberstalking by giving targeted individuals recourse against those who posted threatening material online. Although each of these could potentially provide a remedy, however, tort law is an inadequate tool for regulating internet justice.

a. Defamation

Defamation could provide a remedy to some individuals targeted by harmful forms of internet justice, particularly those individuals targeted based solely on false accusations. To prevail in an action for defamation, a plaintiff must show that the defendant (1) made a false and damaging statement, (2) distributed the statement to a third party, (3) acted at least negligently in doing so, and (4) caused harm to the plaintiff.¹⁴⁶ Importantly, in many of jurisdictions, a plaintiff can hold the defendant liable for special harm caused by the repetition of the defamatory statement by third-parties if the defendant intended the repetition or should have reasonably expected it.¹⁴⁷ In conducting this analysis, courts generally take into account “if the originator has himself widely disseminated the defamation . . . in determining whether there were grounds to expect the further dissemination of the slander.”¹⁴⁸ As a result, an individual targeted by harmful internet justice can likely recover for harm caused not only by an original statement online, but also for harm caused by the widespread sharing of the statement.

The availability of defamation, however, will be limited by the requirement that the defamatory statement be false.¹⁴⁹ Because of this requirement, an individual such as Justine Sacco would not be able to recover under defamation for harmful or dangerous statements leveled against her in relation to true accusations based on her tweet.¹⁵⁰ A individuals falsely accused and shamed, however, such as the Darth Vader Selfie Dad, may have a cause of action under defamation to recover damages for harm caused by the false statement and its widespread circulation. Defamation, then, could provide a remedy to the subset of individuals targeted by internet justice who experienced shaming sanctions based on false information, but not to those who received threats based on true accusations.

b. IIED

Individuals targeted by internet justice that crosses the threshold from primarily beneficial to primarily dangerous could also potentially find success in tort actions for intentional infliction of emotional distress (IIED). Under the doctrine of IIED, “[a]n actor who by extreme and outrageous conduct intentionally or recklessly causes severe emotional harm to another is subject to liability for that emotional harm and, if the emotional harm causes bodily harm, also

¹⁴⁶ See RESTATEMENT (SECOND) OF TORTS § 558-59 (2017).

¹⁴⁷ See *id.* § 576.

¹⁴⁸ *Id.* cmt. d.

¹⁴⁹ See *id.* § 581A (“One who publishes a defamatory statement of fact is not subject to liability for defamation if the statement is true.”).

¹⁵⁰ See Klonick, *supra* note 114, at 1060 (“Moreover, as was the case with Sacco and Richards, it is often the offenders’ true, self-publicized actions or words that lead to the shaming, making them ineligible for any kind of defamation relief.”).

for the bodily harm.”¹⁵¹ “Extreme and outrageous conduct” sets a high bar that requires conduct which “goes beyond the bounds of human decency such that it would be regarded as intolerable in a civilized community.”¹⁵² In addition, the emotional harm suffered must be serious in order for IIED to provide a remedy.¹⁵³ As a result, a plaintiff seeking damages under a theory of IIED faces a heavy burden.

In cases where internet justice leads to instances of location sharing and threats, however, targeted individuals may have successful claims for IIED. For example, in *Delfino v. Agilent Technologies Inc.*, the California Court of Appeals held that repeated graphic threats of violence on an internet messaging board could fulfill the “extreme and outrageous” element of IIED.¹⁵⁴ Similarly, in *Kleopfel v. Bokor* the Court of Appeals of Washington held that stalking and threats constituted “extreme and outrageous conduct.”¹⁵⁵ When internet justice begins to incorporate stalking and violent threats, then, the targeted individuals may find success in bringing claims for IIED.¹⁵⁶

c. Privacy Torts¹⁵⁷

Finally, in some instances of harmful internet justice, the privacy torts may provide a remedy for targeted individuals. The privacy torts include (1) intrusion, (2) appropriation of another’s name or likeness, (3) publicity given to another’s private life, and (4) false light portrayal.¹⁵⁸ As with defamation, the tort of false light portrayal may provide a remedy where an individual is shamed based on false information.¹⁵⁹ Additionally, in some cases of internet justice that cross the threshold into cyberstalking, the torts of intrusion and publicity may provide a remedy. Intrusion involves the unwanted prying into the private life of another, such as taking photos of the inside of another’s house or continual harassment through phone calls.¹⁶⁰ Publicity given to private life, on the other hand, involves the offensive publication of a private matter that is not of legitimate concern to the public.¹⁶¹ Intrusion and publicity given to private life may give plaintiffs a remedy who experience stalking, such as photos taken of them in the privacy of their home or possibly repeated harassing messages online, as a result of internet justice. The

¹⁵¹ RESTATEMENT (THIRD) OF TORTS: PHYSICAL AND EMOTIONAL HARM § 46 (AM. LAW INST. 2012).

¹⁵² See *id.* cmt. d.; see also *Miller v. Currie*, 50 F.3d 373, 378 (6th Cir. 1995).

¹⁵³ See *Yeager v. Local Union 20, Teamsters, Chauffeurs, Warehousemen, & Helpers of Am.*, 453 N.W.2d 666, 671 (Ohio 1983) (“Thus, we hold that in order to state a claim alleging the intentional infliction of emotional distress, the emotional distress alleged must be serious.”).

¹⁵⁴ 52 Cal. Rptr. 3d 376, 392 (Cal. Ct. App. 2006).

¹⁵⁵ 110 Wash. App. 1059 (2002) (unpublished opinion).

¹⁵⁶ For further arguments that IIED could provide an effective solution to online harassment, see Victoria McIntyre, Comment, “Do(x) You Really Want to Hurt Me?”: *Adapting IIED as a Solution to Doxing by Reshaping Intent*, 19 Tul. J. Tech. & Intell. Prop. 111, 125-33 (2016) (arguing that courts should begin to apply a totality of the circumstances test in IIED cases to better allow recovery for plaintiffs in cases of “doxing,” the online sharing of personally identifiable information of another person).

¹⁵⁷ See, e.g., Larkin, *supra* note 4, at 76.

¹⁵⁸ RESTATEMENT (SECOND) OF TORTS § 652A (2017).

¹⁵⁹ See *id.* § 652E. False light portrayal requires that the defendant “places the other before the public in a false light” and “(a) the false light in which the other was placed would be highly offensive to a reasonable person, and (b) the [defendant] had knowledge of or acted in reckless disregard as to the falsity of the publicized matter and the false light in which the other would be placed.” *Id.* Again, these elements would likely be met in the case of the Darth Vader Selfie Dad, which involved the public, highly offensive, and false accusation of being a pedophile.

¹⁶⁰ *Id.* § 652B; *id.* cmt. d.

¹⁶¹ *Id.* § 652D.

availability of this remedy, however, will be limited by the requirement that the conduct be “highly offensive” and intrude upon an area that the plaintiff has made private, such the interior of the home.¹⁶²

Although each of these areas of tort law may provide a remedy to targets of harmful internet justice, tort law remedies are not an ideal solution to controlling internet justice for several reasons. First, the costs of trial, both civil and financial, will likely be prohibitive for many potential plaintiffs, especially when relief is uncertain because of the novelty of the claims.¹⁶³ Second, because internet justice is a group effort, selecting a particular defendant may be a difficult or impossible task, especially when many of the group participants may be anonymous users of an online forum.¹⁶⁴ Finally, and perhaps most importantly to those targeted by internet justice, seeking a legal judgment against individuals will likely serve to exacerbate the shaming they are already experiencing online. Consider, for example, if Brock Turner had brought suit against the armed protesters outside his house for IIED. While a few members of the internet community may have supported a lawsuit, the majority of the internet community would likely respond with reinvigorated contempt and shaming. For this reason, in addition to the other difficulties associated with bringing tort actions, tort law does not provide an adequate solution to regulating internet justice.

3. *Statutory Solutions*

Finally, criminal law provides a third potential solution to protecting against the dangers of internet justice. The U.S. Code already has a provision, 18 U.S.C. § 875(c), that prohibits the interstate communication of a threat.¹⁶⁵ Because the solution to regulating internet justice should aim to remove threatening material that creates danger for the targeted individual, § 875(c) facially appears to be an ideal remedy. Under § 875(c), or equivalent state statutes, individuals who post threats or calls to violence against targets of internet justice could face criminal liability, the strongest possible deterrence against posting violent material.¹⁶⁶

Relying upon a criminal statute, however, also has several disadvantages that make it a poor solution to controlling internet justice. Most significantly, criminal law requires reliance upon law enforcement officials to identify and respond to threatening online material. Law enforcement may have neither the training necessary to recognize online threats nor the

¹⁶² See *id.* §§ 652B, 652D.

¹⁶³ See Klonick, *supra* note 114, at 1060 (“But litigation is a costly and timely remedy. Years can—and often do—pass before a case is settled or tried. In the meantime, the shaming can continue, or even accelerate.”).

¹⁶⁴ See Nicole P. Grant, *Mean Girls and Boys: The Intersection of Cyberbullying and Privacy Laws and Its Social-Political Implications*, 56 *How. L.J.* 169, 198-200 (2012) (discussing the challenges associated with bringing tort actions against anonymous internet users in cases of cyberbullying); Emily Poole, Note, *Hey Girls, Did You Know? Slut-Shaming on the Internet Needs to Stop*, 48 *U.S.F. L. REV.* 221, 259 (2013) (“Additionally, cyberbullying often takes place at the hands of numerous bullies. There is a mob mentality on the Internet—one person starts bullying and another jumps on board, and then another, and another. Victims may have a hard time singling out a single attacker or figuring out against whom to bring a claim. It is logistically and jurisdictionally difficult to sue a large group of bullies who are spread out across the country or the world.”).

¹⁶⁵ 18 U.S.C. § 875(c) (1994). Violation of the provision is punishable by fine or by imprisonment of not more than five years. *Id.*

¹⁶⁶ Federal prosecutors have brought charges under § 875(c) for violent and threatening online posts in the past. See, e.g., *Elonis v. United States*, 135 S. Ct. 2001 (2015); see also *infra* Subsection IV.B.2. (discussing *Elonis* as an illustration of social media speech constituting a true threat).

resources to investigate every potential case.¹⁶⁷ If law enforcement chooses not to respond, then the targeted individual is left exposed to danger. Additionally, identifying the perpetrators in such cases may be difficult, requiring investigative efforts to determine the individuals behind anonymous usernames and further disincentivising involvement by law enforcement. Because of the necessity and difficulty of securing law enforcement involvement, criminal law does not provide a reliable solution to protecting individuals against the potential harms of internet justice.

B. Suggestions for a Regulatory Solution

To holistically protect against the dangers of internet justice, while not hindering the beneficial role of online community norm enforcement, Congress should utilize regulatory law to craft a solution to internet justice. Through regulatory law, an agency can place responsibility for the removal of threatening material directly on website operators, thus removing the requirement that the targeted individual bring suit against anonymous internet users to find relief through tort law. Additionally, under an administrative scheme, the responsibility for imposing liability would rest with the agency, a body typically equipped with greater expertise in the area than law enforcement. Because regulatory law avoids the difficulty of placing responsibility on either the targeted individual to bring suit or law enforcement to prosecute, an administrative solution would best protect targeted individuals against the dangers of internet justice.

To accomplish this solution, Congress should enact a statute giving the Federal Communications Commission (FCC) authority to promulgate regulations for when internet service providers must remove cyberstalking material. In order to give this authority to the FCC, however, Congress must first amend the Communication Decency Act, which has already come under severe criticism for not evolving in tandem with the rapidly changing internet landscape. Additionally, because Congress will be granting the FCC authority to regulate online speech, the statute will need to comply with the requirements of the First Amendment, which is possible under the True Threat doctrine elucidated in *United States v. Watts* and in *Virginia v. Black*. Having passed through these hurdles, Congress will be able to authorize the FCC to promulgate rules for internet service providers to follow regarding when cyberstalking material must be removed.

1. Amending the Communications Decency Act

Before Congress can give any agency authority to impose requirements on internet service providers regarding content published by third-parties, Congress first must amend § 230(c) of the Communications Decency Act (CDA).¹⁶⁸ Section 230(c)(1) currently protects internet service providers (ISPs) from liability for content posted by third-parties.¹⁶⁹ It does so by directing that ISPs cannot be treated as publishers of content posted by a third party.¹⁷⁰ As non-

¹⁶⁷ See McIntyre, *supra* note 156, at 123-24 (“Criminal penalties, in addition to presenting potential proof and application issues, depend upon law enforcement and prosecutors to take the victim’s complaints seriously in order to succeed.”).

¹⁶⁸ 47 U.S.C. § 230 (2012).

¹⁶⁹ See *id.* § 230(c)(1).

¹⁷⁰ See *id.* § 230(c)(1) (“No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.”).

publishers, ISPs cannot face publisher liability for offensive or illegal content posted on their websites.¹⁷¹

Ironically, Congress originally intended the CDA to protect internet service providers from liability for good faith attempts to remove harassing or offensive content.¹⁷² Indeed, § 230(c) is entitled “Protection for ‘Good Samaritan’ blocking and screening of offensive material” and § 230(c)(2) protects ISPs from liability for removal of content that the ISP deems offensive, violent, or harassing.¹⁷³ Instead of encouraging proactive removal of offensive content, however, the statute has primarily operated to protect the continued publication of harassing or harmful content online because internet service providers cannot face liability for any material posted on their websites.¹⁷⁴ Numerous authors have advocated for an amendment to the Communications Decency Act in light of this unintended consequence.¹⁷⁵

Congress should now revise the CDA to allow ISPs to face publisher liability in certain circumstances. Elimination of § 230(c)(1), removing all protection from liability for content posted by third-parties, would likely be too drastic a revision, both in terms of its abruptness and its breadth. Instead, Congress should amend the CDA by including a provision giving the FCC authority to enact regulations determining when an ISP may be held liable as a publisher of third-party content. Determining the appropriate rules for the FCC to enact to protect against dangerous forms of internet justice first requires an examination of the First Amendment.

2. True Threat as a First Amendment Exception for Cyberstalking-Related Speech

Even if Congress were to amend the Communications Decency Act to allow for the regulation of internet service providers, Congress would still need to navigate the First Amendment in crafting a statute that would require the removal of cyberstalking speech online. As Supreme Court has repeatedly held, the First Amendment does not prohibit all regulation of

¹⁷¹ See Vanessa S. Browne-Barbour, *Losing Their License to Libel: Revisiting § 230 Immunity*, 30 BERKELEY TECH. L.J. 1505, 1527 (2015) (“A substantial majority of the United States Courts of Appeal have determined that § 230 immunizes ISPs from liability for online content provided by a third party.”). As interpreted by federal courts, § 230 provides immunity to website operators, such as Facebook or Craigslist, as well as internet service providers, such as AOL or Comcast. According to its technical definition, “internet service provider” refers to companies to provide internet access, such as AT&T or Verizon. See Tim Fisher, *Internet Service Provider (ISP)*, LIFEWIRE (Mar. 9, 2017), <https://www.lifewire.com/internet-service-provider-isp-2625924>. Federal circuit court decisions have used “internet service provider to refer to website operators, as well as to technical service providers. See, e.g., *Chi. Lawyers’ Comm. For Civil Rights Under Law, Inc. v. Craigslist, Inc.*, 519 F.3d 666 (7th Cir. 2008) (referring to craigslist as an internet service provider); *Jane Doe No. 1 v. Backpage.com, LLC*, 817 F.3d 12 (1st Cir. 2016) (holding that § 230 immunity protects Backpage.com).

¹⁷² See *id.* § 230(c)(2); see also Larkin, *supra* note 4, at 67.

¹⁷³ See § 230(c).

¹⁷⁴ Larkin, *supra* note 4, at 67 (“Ironically, Congress enacted section 230 in order to encourage the free exchange of ideas over the Internet and to encourage websites voluntarily to monitor and *delete* obscene or offensive material, not to shelter items such as revenge porn. Section 230 therefore serves as a classic example of the law of unintended consequences.”).

¹⁷⁵ See, e.g., Browne-Barbour, *supra* note 171 (amending the CDA to allow for publisher liability for defamatory content published online); Poole, *supra* note 164, at 248-51 (amending the Section 230 to allow for prevention of cyber-bullying); Stephanie Silvano, Note, *Fighting a Losing Battle to Win the War: Can States Combat Domestic Minor Sex Trafficking Despite CDA Preemption?*, 83 FORDHAM L. REV. 375, (2014) (amending the CDA to allow for prosecution of internet service providers who turn a blind eye to human trafficking on their websites).

speech.¹⁷⁶ Instead, the First Amendment allows for the prohibition of certain categories of speech that are “of such slight social value as a step to truth that any benefit that may be derived from them is clearly outweighed by the social interest in order and morality.”¹⁷⁷ These proscribable categories include speech designed to cause an immediate breach of the peace,¹⁷⁸ defamation,¹⁷⁹ and intimidation.¹⁸⁰ For the purpose of regulating speech related to cyberstalking, the most relevant category of proscribable speech is true threat, first recognized in *Watts v. United States*.

In *Watts*, the Supreme Court first recognized that the government may prohibit and punish speech intended to communicate a threat.¹⁸¹ Since *Watts*, the Court has refined the elements of what constitutes a true threat. Under the current true threat doctrine, a true threat is a statement that “the speaker means to communicate a serious expression of an intent to commit an act of unlawful violence to a particular individual or group of individuals.”¹⁸² Whether a statement constitutes a true threat does not depend on the speaker’s intent to carry out the threat but upon the speaker’s intent to place the subject of the threat in fear of bodily harm.¹⁸³ The purpose of the true threat doctrine is to protect individuals from the fear of violence.¹⁸⁴

Recently, in *Elonis v. United States*, the Supreme Court implicitly accepted that violent internet posts could fall under the true threat exception.¹⁸⁵ In *Elonis*, the government brought charges under 18 U.S.C. § 875(c), which prohibits the interstate communication of threats.¹⁸⁶ *Elonis* had posted online violent writings, discussing acts of violence against his wife and against police.¹⁸⁷ These posts were not sent, or even addressed, directly to the subjects of his violent statements.¹⁸⁸ The Supreme Court ultimately reversed *Elonis*’s conviction based upon a holding that the lower court had incorrectly interpreted the *mens rea* requirement of § 875(c).¹⁸⁹

¹⁷⁶ See *Virginia v. Black*, 538 U.S. 343, 358 (2003) (“The protections afforded by the First Amendment, however, are not absolute, and we have long recognized that the government may regulate certain categories of expression consistent with the Constitution.”).

¹⁷⁷ *Chaplinsky v. State of New Hampshire*, 315 U.S. 568, 572 (1942)

¹⁷⁸ See *id.* at 571-72 (“There are certain well-defined and narrowly limited classes of speech, the prevention and punishment of which have never been thought to raise any Constitutional problem. These include the lewd and obscene, the profane, the libelous, and the insulting or ‘fighting’ words—those which by their very utterance inflict injury or tend to incite an immediate breach of the peace.”).

¹⁷⁹ See *R.A.V. v. City of St. Paul*, 505 U.S. 377, 383 (1992).

¹⁸⁰ See *Black*, 538 U.S. at 362.

¹⁸¹ 394 U.S. 705, 708 (1969).

¹⁸² *Black*, 538 U.S. at 359.

¹⁸³ *Id.* at 359-60.

¹⁸⁴ *Id.* (citing *R.A.V.*, 505 U.S. at 388).

¹⁸⁵ See *Elonis v. United States*, 135 S. Ct. 2001 (2015).

¹⁸⁶ *Id.*; 18 U.S.C. § 875(c). In order to meet the requirements of the First Amendment, a violation of § 875(c) must constitute a true threat. See, e.g., *United States v. Landham*, 251 F.3d 1072, 1087 (6th Cir. 2001) (dismissing a conviction under § 875 where the communication did not constitute a true threat).

¹⁸⁷ *Elonis*, 135 S. Ct. at 2005-07.

¹⁸⁸ *Id.*

¹⁸⁹ *Id.* at 2012-13. The issue in the case concerned *Elonis*’s intent in making the statements. The government argued that § 875(c) did not contain a *mens rea* requirement beyond a minimal understanding of the “content and context” of the words used. *Id.* at 2011. *Elonis*, on other hand, argued that the statute required an intent to communicate a threat and that he did not intend his online posts to contain threats. *Id.* at 2008. The Court, while declining to specify the necessary *mens rea* under the statute, held that the statute does contain a *mens rea* requirement and that *Elonis*’s conviction, based solely on how a reasonable person would receive the statements, could not stand. *Id.* at 2012. Because the Court’s ruling focused solely on the *mens rea* requirement of § 875(c), *Elonis* did not affect the interpretation of true threat from a First Amendment perspective. Courts are divided on the specific intent required

However, neither the parties nor the Court argued that *Elonis*'s conviction raised First Amendment issues based upon the statements not constituting true threats, indicating that online posts advocating violence, such as those in this case, can constitute true threats.

Because the goal of cyberstalking is to place the target in fear of harm, a regulatory rule requiring the removal of cyberstalking speech would fall under the true threat doctrine. Examples of cyberstalking speech that could constitute a true threat include: a picture posted online of a targeted individual with a caption calling for violence against that individual, a message sent directly to a targeted individual threatening violence, or a post sharing the address of a targeted individual with the intent to place the individual in fear of violence. Because true threat does not require intent to actually carry out the violence, these statements still constitute true threats, even if the individual making the statements is unable to actually carry them out. Furthermore, these statements can create a reasonable fear of actual violence for the targeted individual because the call to violence is widely visible to those who can carry out violence against the targeted individual. As a result, under the true threat doctrine, Congress could enact a statute requiring the FCC to promulgate rules requiring the removal of cyberstalking speech that passes scrutiny under the First Amendment. The last step, then, is crafting regulations that would remain within the strictures of the true threat doctrine and would reasonably and effectively require internet service providers to remove cyberstalking material.¹⁹⁰

3. Suggested Rules for Internet Service Providers

To accomplish the goal of protecting individuals from dangerous forms of internet speech while staying within the strictures of the First Amendment, the regulations enacted by the FCC should utilize a notice-removal requirement and provide guidelines for when an online post constitutes a true threat. Liability for internet service providers for illegal speech of which they have notice is not an original proposition. Legal scholars have suggested a notice-removal scheme for regulating other harmful forms of internet speech, such as cyberbullying or defamation.¹⁹¹ Additionally, Congress has already utilized a notice-removal requirement under the Digital Millennium Copyright Act (DMCA) to prevent copyright violations.¹⁹²

Under the DMCA, certain website operators are liable for copyright infringement when they are put on notice of the infringement by the copyright holder.¹⁹³ The DMCA requires operators of websites that allow for the public uploading and access of content to designate on the website a contact-person for complaints of copyright infringement.¹⁹⁴ To effectively put a website operator on notice of copyright infringement, the copyright holder must follow specific requirements, such as identifying the infringing material and providing a statement that the

for a true threat. *See generally* Jing Xun Quek, *Elonis v. United States: The Next Twelve Years*, 31 BERKELEY TECH. L.J. 1109 (2016) (discussing the remaining confusion over the intent requirement for a true threat after *Elonis*).

¹⁹⁰ *See infra* Subsection IV.B.3.

¹⁹¹ *See* Browne-Barbour, *supra* note 171, at 1553-54 (suggesting an amendment to the CDA that replaces absolute immunity with immunity from liability only for defamatory content of which the ISP was not on notice); Poole, *supra* note 164, at 248-51 (suggesting a notice liability scheme to prevent cyberbullying).

¹⁹² 17 U.S.C. § 512 (2006).

¹⁹³ *See id.* § 512(b)-(c). This assignment of liability is possible under the CDA because the CDA specifies that “[n]othing in this section shall be construed to limit or expand any law pertaining to intellectual property.” 17 U.S.C. § 230(e)(2).

¹⁹⁴ *See* § 512(c); *see also* Browne-Barbour, *supra* note 171, at 1549-50.

complainant has a good faith belief that the material infringes a copyright.¹⁹⁵ If, after receiving notice, the website operator fails to remove the infringing material, the copyright holder can bring suit against the website operator.¹⁹⁶

Similarly, a notice-removal rule could impose liability on website operators who are put on notice of threatening or stalking material uploaded to their websites. The notice-removal rule for threatening material, however, should be different from the DMCA notice-removal requirement in two respects. First, the rule for removal of threatening material should utilize a less stringent notice requirement than the DCMA. Under the DCMA, the class of individuals with knowledge of the copyright infringement is limited to the copyright holders themselves. In contrast, any individual can identify a post that calls for violence against a specific individual. Allowing anyone to file a notification of a threatening post could accomplish the removal of dangerous posts that a targeted individual may not have yet seen but that still place the individual in danger. Similar to the DMCA, then, notice should require specific identification of the threatening material, along with a description of why the post constitutes a threat. Notice, however, should not require complaints to identify themselves as the subjects of the threatening material.

Second, the remedy for violation of the notice-removal requirement should be a fine imposed by the FCC, rather than civil liability. Requiring the targeted individual to bring suit to impose liability on the website operator would raise the same concerns as reliance upon tort law—the high cost of bringing suit and the exacerbation of the targeted individual’s shaming. Instead, the FCC should impose fines upon website operators who fail to comply with the notice-removal requirement.

To determine whether website operators are complying with the requirement, the FCC could utilize both a private complaint procedure and a requirement that website operators issue a report of their procedures for receiving and responding to notifications. For websites operators that in good faith attempt to respond to notifications, but consistently fail to recognize and take down posts that constitute true threats, the FCC could issue opinion letter to urge better practices. Conversely, for website operators who do not make a good faith attempt to respond to notifications, the FCC should impose coercive penalties, such as fines. In this way, encouraging the removal of threatening material does not require lengthy litigation that would often yield only a small amount of damages.

To assist website operators in identifying material that requires removal, the FCC should issue guidelines for when online material constitutes a true threat. These guidelines could list a number of factors, such as whether the post identifies an individual by name, shares the location of the individual, or advocates violence. The guidelines could also include context-based indicators for when a post may or may not constitute a true threat that requires removal under the rule. These guidelines would both assist website operators in removing dangerous material and provide the FCC with a benchmark for whether websites are complying substantially with the notice-removal requirement of the rule.

Some may argue that a rule requiring removal of material determined by website operators to be a true threat may chill the free speech rights of internet users. A subset of legal

¹⁹⁵ See § 512(c)(3). Notice also requires the signature of the copyright holder (or representative), identification of the copyrighted work, contact information for the complainant, and a statement made under penalty of perjury that the complainant is authorized to act for the copyright owner. *Id.*

¹⁹⁶ See § 512(c).

scholars have criticized the notice-removal provision of the DCMA for this reason.¹⁹⁷ The notice-removal rule should not raise concerns for the chilling of speech for several reasons. First, a requirement placed upon the website operator to remove threatening speech does chill the speech of internet users because it does not subject internet users to liability for their speech. While the notice-removal requirement may result in the removal of some material, it will not place individual internet users in danger of liability. Second, under the CDA, website operators are already encouraged to remove harassing or offensive speech by having a shield from liability for removal of such speech.¹⁹⁸ Finally, a notice-removal requirement for threatening speech raises even less of a free speech concern than the DMCA because it is focused solely on low-value speech. While the DMCA requires removal of material that could be of social or political value aside from the copyright infringement, the notice-removal requirement proposed here would focus solely on the removal of threatening material. Even if a website removes material that does not constitute a true threat under Supreme Court precedent, it will still be removing low value speech that should not raise the same level of free speech concerns as the DMCA may raise. For these reasons, concerns of chilling online speech should not dissuade Congress and the FCC from the use of a notice-removal rule to require the deletion of true threats.

CONCLUSION

Internet justice is a new phenomenon with significant repercussions for the legal community. While shame-based community norm enforcement played an important role in tribal and colonial communities, the modern criminal justice system has abandoned community shaming punishments for exclusively government-imposed incarceration. The internet, however, has allowed shamed-based community norm enforcement to again emerge as a significant force in sanctioning antinormative behavior, in many ways even more effectively than was possible in the past. Because of the potential for internet justice to both cause substantial harm to individuals and complement the legal system in positive ways, internet justice requires a holistic response from the legal community. Through a notice-removal rule that requires website operators to take down threatening online speech that places targeted individuals in danger, the legal community can limit the potential dangers of internet justice while also allowing internet justice to fulfill its complementary role. Alternatively, by failing to respond to internet justice, the legal community will lose the opportunity to guide internet justice in fulfilling its potentially significant roles and to protect targeted individuals from its potential harms.

¹⁹⁷ See, e.g., Amanda Beshears Cook, *Copyright and Freedom of Expression: Saving Free Speech from Advancing Legislation*, 12 CHI.-KENT J. INTELL. PROP. 1, 26 (2013) (“The DMCA takedown notice procedure has as much ‘chilling effect’ on free speech as a prior restraint, and should be struck down as unconstitutionally restrictive.”); Rebecca Alderfer Rock, Comment, *Fair Use Analysis in DMCA Takedown Notices: Necessary or Noxious*, 86 Temp. L. Rev. 691, 700 (2014) (arguing that the DMCA takedown notice provision raises free speech concerns).

¹⁹⁸ See 18 U.S.C. 230(c)(2) (2006).