

Article

***61 CYBER CIVIL RIGHTS**

[Danielle Keats Citron \[FN1\]](#)

* * *

I. Anonymous Mobs of the Twenty-First Century

* * *

A. The Destructive Nature of Online Mobs

Online assaults exist along several interconnected dimensions. [\[FN34\]](#) First, attacks involve threats of physical violence. Death and rape threats are legion on the Web. [\[FN35\]](#) The threats may foreshadow offline stalking and physical violence. [\[FN36\]](#) They often include references to victims' home addresses and personal information, suggesting attackers' familiarity with them, and the attackers encourage readers to physically assault the victims, putting them in fear of genuine danger.

In response, victims stop blogging and participating in online forums. [\[FN37\]](#) A Pew Internet and American Life Project study attributed a nine percent decline ***70** in women's use of chat rooms to menacing sexual comments. [\[FN38\]](#) Victims may also make their sites private or assume pseudonyms to mask their identity. [\[FN39\]](#) As one victim explains, it does not take many rape threats to “make women want to lay low.” [\[FN40\]](#)

Second, assaults invade victims' privacy. Attackers hack into victims' computers and e-mail accounts to obtain personal information, such as Social Security numbers, driver's license information, and confidential medical data. [\[FN41\]](#) The stolen information is then posted online. [\[FN42\]](#) Disclosing such personal information poses imminent risks, such as the threat of identity theft, employment discrimination, and online or offline stalking. [\[FN43\]](#) It also inflicts harm in the longer term. Victims feel a sustained loss of personal security and regularly dismantle their online presence to avoid further devastation of their privacy. [\[FN44\]](#)

Third, assaults can involve statements that damage reputations and interfere with victims' economic opportunities. [\[FN45\]](#) Online comments may assert that individuals suffer from mental illnesses. [\[FN46\]](#) They may claim individuals have sexually transmitted diseases. [\[FN47\]](#) Attackers sometimes publish doctored ***71** photographs of victims. [\[FN48\]](#) In addition, attackers send damaging statements about victims to their employers and manipulate search engines to reproduce the damaging statements and pictures for others to see, [\[FN49\]](#) creating digital “scarlet letters” that destroy reputations. [\[FN50\]](#)

Fourth, some assaults do not involve online postings at all. Instead, attackers use technology to force victims offline. Groups coordinate denial-of-service attacks [\[FN51\]](#) and “image reaping” campaigns to shut down sites and blogs. [\[FN52\]](#) While the other types of assaults silence victims indirectly with fear and humiliation, this fourth type of assault muzzles them directly.

Groups commonly wield all four of these tools in their attacks against individuals. Some attacks originate online and continue offline, while others move

in the opposite direction. [FN53] For example, in 2007, the social networking site AutoAdmit hosted a pattern of attacks on female law students. [FN54] Thirty-*72 nine posters targeted named students on the site's message board. [FN55] The posters, writing under pseudonyms, generated hundreds of threatening, sexually-explicit, and allegedly defamatory comments about the victims. [FN56]

Posters threatened female law students with violence. One poster asserted that a named female student "should be raped." [FN57] That remark begat dozens of more threats. For instance, a poster promised: "I'll force myself on [the identified student]" and "sodomize" her "repeatedly." [FN58] Another said the student "deserves to be raped so that her little fantasy world can be shattered by real life." [FN59]

Discussion threads suggested the posters had physical access to the female students. A poster described a student's recent attire at the law school gym. [FN60] Posts mentioned meeting targeted women and described what they looked like and where they spent their summer. [FN61] Posters urged site members to follow a woman to the gym, take her picture, and post it on AutoAdmit. [FN62] Others provided updates on sightings of a particular woman. [FN63] Another poster provided the e-mail address of a female law student under a thread entitled "Mad at [named individual]? E-mail her . . ." [FN64]

Posters also asserted damaging statements about the women. One asserted that a female student spent time in a drug rehabilitation center. [FN65] Another claimed the student had a lesbian affair with a law school administrator. [FN66] *73 Others remarked that the student appeared in Playboy. [FN67] Posters claimed that another female student had a sexually transmitted disease. [FN68] Others provided her purported "sub-par" LSAT score. [FN69] The victims asserted that these were lies. [FN70]

In addition to publishing the alleged lies online, posters spread them offline to undermine the victims' job opportunities. One poster urged the group to tell top law firms about the female student's LSAT score "before she gets an offer." [FN71] Posters e-mailed their attacks to the student's former employer, recommending that the employer show it to its clients, who would "not want to be represented by someone who is not of the highest character value." [FN72]

Another poster sent an e-mail to a particular female law student's faculty asserting that her father had a criminal record. [FN73] The poster displayed the e-mail on AutoAdmit before sending it, explaining: "I've assembled a spreadsheet with [faculty e-mail] addresses and every single one of them will be notified about what our darling [named student] has done. I post this here as a warning to all those who would try to regulate the more antisocial posters - we have the power now." [FN74]

Site members applauded the e-mail and rallied around the sender. For instance, a poster stated that the e-mail sender should be awarded a "Congressional medal." [FN75] Others recommended sending the e-mail from a public computer and a "hushmail account," or with anonymizing software. [FN76]

The attackers waged a "Google-bombing" campaign that would ensure the prominence of offensive threads in searches of the female students' names. [FN77] Posters made plain the goal of their Google-bombing campaign: "We're not going to let that bitch have her own blog be the first result from googling her name!" [FN78] An individual writing under the pseudonym "leaf" detailed the steps *74 AutoAdmit posters would have to take to engage in Google-bombing. [FN79] Leaf explained that posts should include the adjective "big-titted" next to the woman's name. [FN80] "Big-titted [name of female student]'s name is never to be used in parts - it must always be [name of student] at the least, and 'big-titted [name of the student]' ideally" with pictures of her accompanying the thread. [FN81] This would work because search engine algorithms assign a high rank to a Web page if sites linking to that page use consistent anchor text.

[\[FN82\]](#)

Posters admitted their desire to intimidate and harm the female students. After one of the women did not get a summer job, a poster asked if the other “bitch got what she deserved too?” [\[FN83\]](#) Another said: “I’m doing cartwheels knowing this stupid Jew bitch is getting her self esteem raped.” [\[FN84\]](#) A poster explained that the women were targeted “just for being women.” [\[FN85\]](#)

A lawsuit filed by two of the women alleged the AutoAdmit site managers refused to remove the offensive threads even though the women told them that the messages caused them severe emotional distress. [\[FN86\]](#) On March 15, 2007, a site manager asserted that he would not remove the offensive threads until the female students apologized for threatening litigation and until *75 ReputationDefender, a group assisting the women, acknowledged the mistakes the manager alleged the group had made. [\[FN87\]](#)

In a similar vein, a group called Anonymous has devoted itself to terrorizing and silencing hundreds of women writing on the Web. [\[FN88\]](#) For instance, in 2007, Anonymous used message boards and wikis [\[FN89\]](#) to plan an attack on a nineteen-year-old woman who maintained a video blog about Japanese language and video games. [\[FN90\]](#) Group members hacked her online accounts, including her YouTube blog account, e-mail, Facebook profile, and MySpace page, to obtain her personal information. [\[FN91\]](#) They published her account passwords and private medical history on various sites. [\[FN92\]](#) Postings disclosed her full name, home address, and her mother's e-mail address. [\[FN93\]](#) Group members sent messages from the woman's e-mail account to her loved ones. [\[FN94\]](#) They claimed the woman had committed suicide on various message boards. [\[FN95\]](#)

Members of Anonymous posted doctored photographs of the woman including one picture that featured the woman's head atop naked bodies. [\[FN96\]](#) Next to her picture appeared the promise that group members would rape her *76 “at full force in her vagina, mouth, and ass.” [\[FN97\]](#) A drawing depicted men brutally raping the woman. [\[FN98\]](#)

Anonymous urged its members to “seek and destroy” the woman's online identity. [\[FN99\]](#) Group members saturated her video blog with sexually violent material. [\[FN100\]](#) They took down her videos. [\[FN101\]](#) Anonymous updated its members on the status of her sites. [\[FN102\]](#) When her live journal or video blog reappeared, Anonymous urged members to “rape” and “nuke[] [her sites] from orbit.” [\[FN103\]](#)

Anonymous similarly attacked a journalist writing under the pseudonym “Heart” who maintained a blog and discussion forum about women's issues. [\[FN104\]](#) Group members pieced together her identity from her postings and revealed her name and home address on her discussion forum. [\[FN105\]](#) They made death threats and sexually menacing comments on her blog. [\[FN106\]](#) Anonymous urged members to engage in “image reaping” to shut down her site. [\[FN107\]](#) The group succeeded in overloading and closing Heart's website during the summer of 2007. [\[FN108\]](#) In August 2007, Heart closed her blog and website. [\[FN109\]](#)

Anonymous maintains a list of sites and blogs addressing women's issues that it claims to have forced offline. [\[FN110\]](#) The list includes the names of shuttered sites with a line crossed through them and the accompanying message: “Down due to excessive bandwidth - great success!” [\[FN111\]](#) When a site reappears online, *77 Anonymous tells its members: “It's back! Show no mercy.” [\[FN112\]](#) The group takes credit for closing over 100 feminist sites and blogs. [\[FN113\]](#) Anonymous has also targeted journalists, such as Anna Greer, who have reported on the group's attacks. The group published Ms. Greer's home and e-mail addresses with instructions to “choke a bitch.” [\[FN114\]](#)

Targeted female bloggers and website operators confirm the group's claims

of attacks. [\[FN115\]](#) They describe the denial-of-service attacks and “image reaping” campaigns that have shut down their sites. [\[FN116\]](#) A victim explained: “Being silenced for over two weeks felt infuriating, stifling, imprisoned by gang rapists [sic] just waiting for me to try to get up from underneath their weight so they could stomp me down again.” [\[FN117\]](#) Victimized website operators and bloggers have asked Anonymous in vain to stop its attacks. [\[FN118\]](#)

Groups attack women on the website JuicyCampus with threats of violence, and their posts have generated offline stalking. [\[FN119\]](#) For instance, anonymous posters disclosed a woman's cell phone and dorm address with instructions that she was available for sex. [\[FN120\]](#) After the posts appeared, strange men started knocking on the woman's door at night. [\[FN121\]](#)

Online mobs have targeted African-American and Hispanic women. [\[FN122\]](#) As blogger “La Chola” explains, women-of-color bloggers have consistently *78 received horrific, vile e-mails and comments threatening violent sexual assault, death, and attacks against family members. [\[FN123\]](#) After the author of the blog “Ask This Black Woman” posted commentary about the Resident Evil 5 video game, anonymous posters attacked her on her blog and other sites. [\[FN124\]](#) She received death threats. [\[FN125\]](#) Posters told her to “[g]et back into the cotton fields, you filthy [n****r]” [\[FN126\]](#) and threatened to over-run her blog. [\[FN127\]](#)

Posters on a white supremacist website targeted Bonnie Jouhari, the mother of a biracial girl. [\[FN128\]](#) The site posted an image of Jouhari's workplace burning in flames with a caption that read “race traitor . . . beware, for in our day, they will be hung from the neck from the nearest tree or lamp post.” [\[FN129\]](#) The site included a picture of Jouhari's child and an image of her burning office with bomb-making instructions posted beneath it. [\[FN130\]](#) Ms. Jouhari and her daughter received harassing phone calls at home and at work. [\[FN131\]](#)

Other people of color have faced similar attacks. [\[FN132\]](#) An Asian-American columnist who writes a blog called “Yellow Peril” explained that a group of individuals attacked her online after she wrote about a hate crimes march. [\[FN133\]](#) The group posted a picture of her on a white supremacist watch list, which included her phone number and address, and its members sent threatening e-mails to her. [\[FN134\]](#) College students wrote racially threatening messages on a *79 Hispanic student's Facebook profile, [\[FN135\]](#) promising to “come find you and drag you behind my (expletive) car.”

[Ed. Note: Examples of online attacks against religious minorities and gays and lesbians listed also but deleted for space.]

* * *

The harm online mobs inflict is potent. The threats and privacy intrusions produce damage in numerous ways. Publishing a woman's home address alongside the suggestion that she should be raped or is interested in sex raises the risk that readers of the post will stalk her or commit physical violence against her. Posting a person's Social Security number increases the chance that she will be subject to identity theft. Victims fear that threats or identity theft will be realized: the Internet's anonymity disaggregates the threats from their social context, eliminating cues that might signal the extent of peril. Online anonymity also may prevent an effective law enforcement response. A victim's feeling that she is “being watched” also may stifle her creativity and sense of well-being. [\[FN151\]](#)

Victims may lose job opportunities due to damaging statements and threats posted online. Employers often review Google search results before interviewing and hiring candidates. [\[FN152\]](#) The damaging statements and threats may raise doubts about the victim's competence, or suggest the victim attracts unwanted controversy, causing the employer to hire someone else. When victims stop blogging because of threats, they lose opportunities to establish their online

presence in a manner that could enhance their careers and attract clients. [\[FN153\]](#)

If online groups select victims for abuse based on their race, ethnicity, gender, or religion, they perpetrate invidious discrimination. Important parallels exist between the harm inflicted by prior centuries' mobs and this *81 century's destructive online crowds. Much like their offline counterparts, online hate mobs deprive vulnerable individuals of their equal right to participate in economic, political, and social life. They silence victims and stifle public discourse. [\[FN154\]](#) Although online mobs do not engage in lynching and physical beatings, their attacks produce serious individual and societal harm that should not be ignored.

B. The Dynamics of Mob Behavior

These destructive crowds continue a disturbing pattern from the past, when anonymous groups such as the anti-immigrant mobs of the nineteenth century and the Ku Klux Klan inflicted serious harm on their victims. [\[FN155\]](#) Social scientists have identified four factors that influence the potential dangerousness of a group. [\[FN156\]](#)

First, groups with homogeneous views tend to become more extreme when they deliberate. [\[FN157\]](#) Group members' interactions tend to reinforce preexisting views as members offer a disproportionately large number of arguments supporting their views and only a small number of arguments tilting the other way. [\[FN158\]](#) Hearing agreement from others bolsters group members' confidence, entrenching and radicalizing their views. [\[FN159\]](#)

Second, a group member's deindividuation encourages the member to act on destructive impulses. [\[FN160\]](#) According to one school of thought, people in groups fail to see themselves as distinct individuals and lose a sense of personal responsibility for their destructive acts. [\[FN161\]](#) Another school of thought attributes deindividuation to anonymity rather than an individual's immersion in a group. *82 This account explains that people behave aggressively when they believe that they cannot be observed and caught. [\[FN162\]](#)

Third, groups are more destructive when they dehumanize their victims. [\[FN163\]](#) By viewing victims as devoid of humanity and personal identity, group members feel free to attack without regret. [\[FN164\]](#) Groups rarely target those who are important to their personal well-being. [\[FN165\]](#) For instance, the incidence of lynching in the South similarly tracked the degree of interdependence between victims and the violent crowd, with black newcomers more vulnerable to violence than black employees who worked for the white community. [\[FN166\]](#)

Lastly, group members are more aggressive if they sense that authority figures support their efforts. Social scientists emphasize a perceived leader's role in accelerating dangerous group behavior. [\[FN167\]](#) As recently as the early 1900s, Southern newspapers explicitly "legitimated mob violence" by reporting that lynch mobs included prominent members of the white *83 community. [\[FN168\]](#) As legal historian Robert Kaczorowski explains, federal authorities implicitly encouraged the Klan by failing to enforce civil rights laws. [\[FN169\]](#)

The Internet magnifies the dangerousness of group behavior in each of these respects. Web 2.0 platforms create a feeling of closeness among like-minded individuals. [\[FN170\]](#) Online groups affirm each other's negative views, which become more extreme and destructive. [\[FN171\]](#) Individuals say and do things online they would never consider saying or doing offline because they feel anonymous, even if they write under their real names. [\[FN172\]](#) Because group members often shroud themselves in pseudonyms, they have little fear that victims will retaliate against them or that they will suffer social stigma for their

abusive conduct. Online groups also perceive their victims as “images” and thus feel free to do anything they want to them. [\[FN173\]](#)

* * *

III. Protecting Online Dialogue

* * *

A. Online Mobs and Individual Autonomy

One of free speech's most important functions is promoting individual autonomy. [\[FN263\]](#) This view urges that people be free to choose their own path. [\[FN264\]](#) Free speech facilitates self-mastery, allowing people to author their own narratives. [\[FN265\]](#) Commentators characterize respect for autonomy of speech and thought as necessary for legitimate government. [\[FN266\]](#) For some, freedom from ^{*98} any form of coercion is paramount for autonomy and dignity. [\[FN267\]](#) Others argue that autonomy and dignity require equitable and effective participation in political self-government, and thus the regulation of certain speech, such as racist and sexist speech, may be an essential prerequisite to secure equal citizenship. [\[FN268\]](#)

Restraining a mob's most destructive assaults is essential to defending the expressive autonomy and equality of its victims. [\[FN269\]](#) Preventing mobs from driving vulnerable people offline would “advance the reasons why we protect free speech in the first place,” even though it would inevitably chill some speech of online mobs. [\[FN270\]](#) Free from mob attacks, victims might continue to blog, join online discussions, and generally express themselves on public issues. Protecting them from grotesque defamation, threats, invasions of privacy, and technological attacks would allow them to be candid about their ideas. [\[FN271\]](#)

Although online mobs express themselves and their autonomy through their assaults, their actions also implicate their victims' autonomy and ability to participate in political and social discourse. [\[FN272\]](#) Self-expression should receive no protection if its sole purpose is to extinguish the self-expression of another. [\[FN273\]](#) As Owen Fiss argues, sometimes we must lower the voices of some to permit the self-expression of others. [\[FN274\]](#) Similarly, Cass Sunstein contends that threats, libel, and sexual and racial harassment constitute low-value speech of little First Amendment consequence. [\[FN275\]](#) Rarely is that more true than when one group of voices consciously exploits the Internet's aggregating power to silence others and its disaggregative power to escape social responsibility for the group's actions.

^{*99} B. Civil Rights and the Theory of Free Speech Online

The importance of free speech warrants vigilance against threats that weaken public discourse. [\[FN276\]](#) These concerns, however, are not absolute. [\[FN277\]](#) Our society permits restrictions on speech that is “of such slight social value as a step to truth that any benefit that may be derived from [it] is clearly outweighed by the social interest in order and morality.” [\[FN278\]](#) The Internet poses several challenges to striking that balance.

This Section explores the challenges of applying First Amendment theory to cyberspace. Subsection 1 identifies the special problem of distinguishing protected expressions from unprotected actions in a medium that functions exclusively by transmitting packets of data. Subsection 2 considers online mobs in the context of some prominent First Amendment theories. Subsection 3 assesses

whether private correction of online mobs' abuses might obviate the need for a legal response. Finally, Subsection 4 explores the extent to which online mobs' protected speech might be curtailed or chilled by enforcement of the doctrines advocated here.

1. The Expression-Action Distinction on the Internet

A core problem in theorizing the First Amendment is distinguishing expressions from actions. This speech-conduct dichotomy pervades free speech discourse. [\[FN279\]](#) Advances in law and technology, however, complicate this distinction as they make more actions achievable through “mere” words. Indeed, the Internet's very essence is to aggregate expressions so as to convert them into actions. Some Internet behaviors that are akin to the offline crimes of breaking and entering and vandalism - hacking and denial-of-service attacks - are accomplished by sending communications to other computers. Moreover, the Internet's powerful aggregative capacity converts seemingly individual expressions (e.g., visiting a website or sending an e-mail) into criminal acts through their repetition (e.g., denial-of-service attacks and image ***100** reaping). The Internet also routinely allows individuals to aggregate their efforts with strangers. Thus, the fact that someone may not know the identity of a thief or rapist who uses posted personal information does not eliminate the danger, because the poster knows that such predators may put the information to malicious use.

The Internet may also disaggregate communications into components that operate as actions. For example, some online rape threats engender serious fear that they will be carried out offline because they arrive without cues - such as the identity or location of the person who made the threat or a joking tone of voice - that might diminish the nature of the threat. [\[FN280\]](#) The person's refusal to leave cues that would mitigate the victim's fear arguably demonstrates that person's intent to terrorize the victim. This can convert expression into criminal conduct. In short, because everything that happens on the Internet ultimately takes the form of 1s and 0s does not mean that it is all the expression of ideas.

The expression-action distinction for cyberspace is elusive. [\[FN281\]](#) One could argue that hacking and denial-of-service attacks work first on computers and only indirectly on the computers' owners. The same, however, could be said of responding to an online poll or visiting a site solely to enhance its hit count. Alternatively, one could categorize online activity based on its offline analogues. Unfortunately, this would lead to difficult debates over the strength of competing analogies. Even more importantly, this approach ignores the ways in which the Internet's aggregative and disaggregative character fundamentally transforms online activity. One might ask which characteristic - expression or action - dominates the activity. [\[FN282\]](#) But this question may be difficult to answer, as behavior is often equal parts expression and action. [\[FN283\]](#) For instance, the picture with Ms. Sierra being suffocated by lingerie [\[FN284\]](#) ***101** arguably constitutes both action meant to terrorize her and expression designed to communicate feelings of hatred and misogyny.

A final option would be to treat online behavior as conduct if a reasonable person would expect or intend it to have offline effects independent of the expression of ideas. Thus, threats that frighten recipients and disclosures of personal information that empower identity thieves to obtain victims' money could both be regarded as criminal conduct, like denial-of-service attacks and hacking. This principle would compel some judgment calls, although these would be broadly similar to those the Court's Confrontation Clause doctrine requires in determining the admissibility of statements of witnesses who are unavailable for cross-examination. [\[FN285\]](#) It would, however, take full account of the funda-

mental changes in our modes of both expression and action wrought by the Internet.

2. The Values the First Amendment Protects

Freedom of expression serves several important purposes. [\[FN286\]](#) Limiting online mobs' abuses as proposed above would not threaten any of their core values. Freedom of expression facilitates deliberation about public issues and hence promotes democratic governance. [\[FN287\]](#) Under this view, expression deserves protection if it promotes ideas and information necessary for a self-governing citizenry to make decisions about what kind of life it wishes to live. [\[FN288\]](#) A mob's online attacks do not involve discourse on political issues. Quite the contrary, the attacks deprive vulnerable individuals of their right to engage in political discourse. The threats, lies, and damaging photographs generate a fear of physical violence, exclusion, and subordination that may propel victims offline. [\[FN289\]](#)

Democratic culture theorists argue that freedom of expression promotes “democracy in the widest possible sense, not merely at the level of governance, or at the level of deliberation, but at the level of culture” where we interact, ***102** create, build communities, and build ourselves. [\[FN290\]](#) Free speech permits innovation in a networked age where people aggregate their ideas with those of others, create works of art, gossip, and parody, and thus continually add to the cultural mix in which they live. [\[FN291\]](#) It enables individuals to participate in creating culture on equal terms. [\[FN292\]](#) Free speech also dissolves unjust social barriers of rank and privilege. [\[FN293\]](#) In this vein, Diane Zimmerman highlights the role of gossip as generating intimacy and a sense of community among disparate groups. [\[FN294\]](#) Gossip provides people a way to learn about social groups to which they do not belong and fosters relationships by giving strangers the means to bridge awkward silences when thrown together in social situations. [\[FN295\]](#)

Online mobs do indeed engage in gossip. Sites such as JuicyCampus promote themselves as gossip facilitators. But the attacks perpetrated by online mobs have little to do with building bonds among disparate communities. Rape threats, lies, damaging photographs, and denial-of-service attacks not only preclude any connection with differently-minded group members, but they also sever the victim's connections with her own community. The attacks inflict serious social harm rather than generating ideas in popular culture or enforcing positive social norms. Defeating such discrimination outweighs the imperceptible contribution that online mobs make to our cultural interaction and exchange.

Still others focus on free speech as an engine promoting truth. [\[FN296\]](#) In this view, any silencing of speech prevents us from better understanding the world in which we live. [\[FN297\]](#) Justice Holmes drew from this theory when he articulated the notion of the marketplace of ideas: “that the best test of truth is the power of the thought to get itself accepted in the competition of the market.” [\[FN298\]](#) The ***103** marketplace-of-ideas metaphor places no special premium on political discussion. [\[FN299\]](#) Instead, it captures the idea that “truth must be experimentally determined from the properties of the experience itself.” [\[FN300\]](#)

An extreme version of truth-seeking theory might insist the market should sort out online mobs' deceptions. Though to do so, the theory would have to consider a Social Security number a truthful fact, disclosure of which contributes to an understanding of that person. [\[FN301\]](#) A more plausible vision of truth-seeking theory, however, is not served with the disclosure of a person's personal identifying information. [\[FN302\]](#) Rather than revealing a fact to be tested in the marketplace, a Social Security number is simply a key to a person's

credit and bank accounts. [\[FN303\]](#) In this context, it is a weapon, not a truth or half-truth to be tested in the marketplace. Rape and death threats similarly tell us nothing about the victims - no truths are contested there. This is equally true of denial-of-service attacks and “image reaping.” Even where online mobs make factual assertions, features of the Internet prevent the marketplace of ideas from performing its intended curative function, as the next Subsection shows. Moreover, as Daniel Solove notes, “truth isn't the only value at stake.” [\[FN304\]](#)

3. The Inadequacy of Private Responses

One common offline response to some kinds of unpleasant speech is exclusion of the speaker. Someone who disrupts a private party or meeting may not receive an invitation back. A proliferation of annoying sound trucks may yield time, place, and manner restrictions, such as a noise ordinance. [\[FN305\]](#) *104 Stations that broadcast obscenity when children often listen may lose their licenses. [\[FN306\]](#) Although exclusion can be ineffective against many forms of offensive expression offline, it is particularly ineffective online where individuals can easily frustrate any exclusion by disaggregating their on- and offline identities: an individual ejected from a website under one screen name could promptly return under another.

Nonetheless, some may argue that private responses obviate the need for criminal, tort, and civil rights remedies. In particular, they may contend that victims can defeat online crowds without this proposal by recruiting advocacy groups to defend them. Women's groups could coordinate efforts to rebuild a victim's reputation online. They could engage in “Google-bombing” to optimize positive posts on a search of a victim's name. Groups like ReputationDefender have helped victims establish their online presence to offset destructive postings. [\[FN307\]](#)

Such a response, however, would be inadequate. First, it would not remove the threats and lies that produce emotional distress and fear. It would not restore the confidentiality of the victim's Social Security number and other sensitive information. Even in the case of “merely” defamatory attacks, it is inconceivable that all damage will be restored. [\[FN308\]](#) Because so many people will see the material, some will inevitably miss the victim's response while others will not believe, or only partially believe, it. In the diffuse world online, the shortcomings of any response will be much, much worse.

When issues are being debated, the failure of a point to connect with its counterpoint is less of a concern: some people's views may be skewed by seeing only one side of an argument, while others' ideas may reflect disproportionate exposure to the opposing side. In the end, society can hope the two roughly balance. When dealing with attacks on someone's character, however, the victim does not have an affirmative case she is trying to convey - she is only seeking to dispel the harm from the mob's attack. People seeing a disproportionate number of her rebuttals will not counterbalance those who have seen none.

Second, the efforts of advocacy groups may be unable to drown out the assaults of cyber mobs. Consider the case of Nicole Catsouras, who died in a horrific car crash. [\[FN309\]](#) Gruesome photographs of the carnage appeared on the Internet, spreading to over 1500 sites. [\[FN310\]](#) Posters urged cohorts to harass her *105 family and facilitated this harassment by providing the family's home address. [\[FN311\]](#) The woman's family asked sites to remove the pictures but to no avail. [\[FN312\]](#) Tracking down the anonymous posters proved impossible for the family, and the pictures remained online. [\[FN313\]](#)

Third, instead of slowing down an online mob, counter-measures may sustain the life of the attacks. The very purpose of many online attacks is to force

victims off the net; the mobs are likely to respond with particular venom against a victim who not only stays online but tries to fight back. A victim may plausibly conclude that more people will see the defamatory or private material if she responds than if she does not.

Fourth, online attacks are vastly more numerous and easier to launch than defenses. The online advocacy groups are hopelessly outnumbered and out-matched, and basic collective action theory says they will remain so. [\[FN314\]](#) Few free or inexpensive resources are available for defending one's online reputation, and the services of groups like ReputationDefender are expensive and beyond the means of many victims. [\[FN315\]](#) Even if a victim could afford such assistance, anticipating that cost could discourage an individual from expressing herself online. Thus, the fact that some victims of mobs may be able to enlist allies does not justify limiting or denying relief to the many who cannot.

Finally, this view ignores the social harm resulting from attacks by online mobs. If expressing opinions online subjects someone to the risk of assault, even if the damage is only temporary, the result will change the kinds of people who participate in online discourse. If we believe the Internet is, and should remain, a wild west with incivility and brutality as the norm, then those who are impervious to such conduct will remain online while the vulnerable may not. To that end, we may get more bull-headed posters and fewer thoughtful ones. An online discourse which systematically under-represents people - particularly women and people of color - cannot effectively process our various attitudes and convert them into truly democratic decisions.

***106** 4. The Extent of Interference with Protected Expression

In considering restrictions on the time, place, and manner of expression, the Supreme Court has emphasized the availability of alternative avenues for expression. [\[FN316\]](#) Although defamation, true threats, and online mobs' other unlawful actions are subject to far more extensive regulation than merely time, place, and manner, this reasoning is nonetheless instructive. None of the criminal statutes, tort laws, and civil rights theories discussed here impede or even chill the mobs' expression of their core ideas, whether they be disagreement with their targets' ideas, hatred for their targets, or even hatred of women or other classes of people. In addition, no statutes limit the vehemence of those expressions. They instead further important interests unrelated to the suppression of hateful, racist, or sexist speech.

Although the Court has upheld excluding from public property fully protected expressions of political and religious views in ways that sharply narrowed their potential audiences, [\[FN317\]](#) none of the remedies proposed here would curtail in any way the audience for the mobs' expressions of disagreement. At the margins, of course, some may be uncertain as to whether a particular threat will be considered sufficiently severe to qualify as a true threat or whether particular abuse is sufficiently outrageous to be considered an intentional infliction of emotional distress. These problems, however, existed in the analog world, and they affect only a tiny fraction of the ways in which an idea might be expressed. Protecting the civil rights of online mobs' victims comes at an extremely small cost to legitimate expression.

C. First Amendment Doctrine

1. Criminal and Tort Law

Threats fall outside the First Amendment's protection if speakers mean to

communicate a serious intention to commit an act of unlawful violence against particular individuals. [\[FN318\]](#) The speaker need not actually intend to commit a violent act because the prohibition of “true threats” protects individuals from the fear of violence and the disruption that such fear engenders. [\[FN319\]](#) Once a statement meets the “true threat” standard, it no longer qualifies as protected ***107** speech because it “is so intertwined with violent action that it has essentially become conduct.” [\[FN320\]](#)

A “true threat” determination typically depends upon whether a reasonable person would consider the statement a serious and unconditional expression of intent to inflict bodily harm and not mere hyperbole. [\[FN321\]](#) A person, however, cannot escape responsibility merely by combining the threatening language with an issue of public concern. [\[FN322\]](#) Courts have upheld online threats as unprotected “true threats” even though the defendants never sent the messages directly to the recipients. [\[FN323\]](#) Whether statements constitute “true threats” is a ***108** jury question unless no reasonable jury could find that they amounted to “true threats.” [\[FN324\]](#)

Similarly, First Amendment doctrine offers little protection to defamatory statements because they offer “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.” [\[FN325\]](#) Statements do, however, enjoy immunity from defamation liability if they do not assert or imply verifiable facts. [\[FN326\]](#) To that end, courts prohibit defamation actions based on loose, figurative language that no reasonable person in that context would believe presented facts. [\[FN327\]](#) For example, criticizing another's views in an online debate has been understood as constituting privileged opinion and not verifiable facts. [\[FN328\]](#)

Nonetheless, anonymous message-board postings are not immune from defamation liability simply because they are too outrageous to be believed. [\[FN329\]](#) A California court explained:

***109** Even if the exchange that takes place on these message boards is typically freewheeling and irreverent, we do not agree that it is exempt from established legal and social norms. . . . We would be doing a great disservice to the Internet audience if we were to conclude that all speech on Internet bulletin boards was so suspect that it could not be defamatory as a matter of law. [\[FN330\]](#)

If these damaging statements here were indeed false, many would not enjoy immunity from liability - they could reasonably be understood as asserting verifiable facts. For instance, statements concerning a victim's specific actions or conditions, such as a stay in a drug rehabilitation center, an infectious disease, or a specific LSAT score, seem factual, and thus, a plaintiff could prove them true or false for defamation purposes. [\[FN331\]](#)

Moreover, plaintiffs need not prove “actual malice” if the alleged defamation involves the personal affairs of private individuals. [\[FN332\]](#) First Amendment doctrine requires plaintiffs to show “actual malice” only if the plaintiffs are “public figures” [\[FN333\]](#) or if the statements concern matters on which the public has a justified and important interest. [\[FN334\]](#) For instance, in *Time, Inc. v. Firestone*, the Court found no need for proof of “actual malice” in a defamation case concerning the plaintiff's divorce because the dissolution of a marriage does not involve a matter of public interest, even though the marital difficulties of wealthy individuals might be of some interest to the public, and because the plaintiff did not freely choose to publicize issues related to her married life. [\[FN335\]](#)

Few of the targets of online mobs are likely to be “public figures” even for special purposes: their influence is simply too minimal to suggest they “have ***110** assumed roles of especial prominence in the affairs of society.” [\[FN336\]](#) A person whose published writings reach a relatively small category of people in a

particular field is not a public figure. [\[FN337\]](#) Nor do the public controversies that surround attacks, and victims' attempts to defend themselves, render them public figures: “[T]hose charged with defamation cannot, by their own conduct, create their own defense by making the claimant a public figure.” [\[FN338\]](#) Moreover, the assaults often involve highly personal matters on which the public lacks an important interest. If the public does not have an important interest in learning about the divorce of a wealthy couple, it surely has no interest in charges of private individuals' sexually transmitted diseases or mental illnesses. Thus, victims such as students and bloggers would not have to prove actual malice to pursue defamation claims.

Free speech doctrine would also not limit emotional distress claims, as the attacks here mainly involve private individuals, rather than public figures. Only in cases involving “public debate about public figures” does First Amendment doctrine require proof that “the publication contains a false statement of fact which was made with ‘actual malice.’” [\[FN339\]](#) As with defamation claims, the actual malice standard does not apply to a private person whose emotional distress concerns personal matters. [\[FN340\]](#) Many of the women targeted online, such as the female law students, would not be considered public figures because they never sought to attract the public's attention. Although some victims, such as Kathy Sierra, might be considered public figures due to their especial prominence, the assaults do not address ***111** issues of public concern and thus may be actionable without showing actual malice.

2. Civil Rights Law

Civil rights violations have a dual character. On one hand, they single out people of color, religious minorities, women, and other traditionally subjugated groups for abuse that wreaks special harm on the victims and their communities. On the other hand, they explicitly or implicitly communicate a racist or otherwise bigoted viewpoint. The Court has made clear that the First Amendment poses no obstacle to punishing a defendant for his decision to target vulnerable individuals for abuse because of their gender or race, and for the grave harm the targeting of vulnerable individuals inflicts. Moreover, the Court has refused to allow perpetrators to immunize actions by adding some explicit expressions. The Court has, however, rejected attempts to proscribe abusive expressions solely because their content may be more offensive to vulnerable people. This viewpoint discrimination restriction poses little obstacle to the pursuit of federal and state antidiscrimination actions against online mobs.

The two leading cases in this area are *R.A.V. v. City of St. Paul* [\[FN341\]](#) and *Wisconsin v. Mitchell*. [\[FN342\]](#) In *R.A.V.*, the city criminalized conduct that an individual “knows or has reasonable grounds to know arouses anger, alarm or resentment in others on the basis of race, color, creed, religion, or gender.” [\[FN343\]](#) The Court held that this ordinance unconstitutionally discriminated on the basis of the content of certain offensive expressions - expressions that offensively demonstrated bigoted ideas were proscribed by the ordinance, yet those that gave offense in other ways were not. [\[FN344\]](#) The Court explained:

Displays containing abusive invective, no matter how vicious or severe, are permissible unless they are addressed to one of the specified disfavored topics. Those who wish to use “fighting words” in connection with other ideas - to express hostility, for example, on the basis of political affiliation, union membership, or homosexuality - are not covered. The First Amendment does not permit St. Paul to impose special prohibitions on those speakers who express views on disfavored subjects.

[\[FN345\]](#)

The government cannot discriminate on the basis of the ideas expressed, even within categories of speech and conduct that the First Amendment does ***112** not protect independently. [\[FN346\]](#) Thus, it can prohibit all obscene or defamatory statements, or all fighting words, but not only those conveying a particular type of message independent of their obscene, defamatory, or incendiary character. [\[FN347\]](#)

The Court emphasized the narrowness of its findings by explicitly upholding several civil rights laws prohibiting sexual harassment in the workplace, even though that type of harassment is commonly accomplished through expressions that are far more likely to be misogynistic than feminist. [\[FN348\]](#) It explained that Congress directed Title VII's prohibition on "sexually derogatory 'fighting words'" at conduct and thus its "content-based subcategory of a proscribable class of speech [is] swept up incidentally within the reach of a statute directed at conduct rather than speech." [\[FN349\]](#) Also acceptable are laws that focus on those expressions most likely to cause harm, so long as those laws do not define harmfulness in terms of the viewpoint expressed. [\[FN350\]](#) More generally, the Court explained that "[w]here the government does not target conduct on the basis of its expressive content, acts are not shielded from regulation merely because they express a discriminatory idea or philosophy." [\[FN351\]](#)

A year later, in *Mitchell v. Wisconsin*, a unanimous Court confirmed that *R.A.V.*'s holding was narrow indeed. *Mitchell* involved a Wisconsin statute allowing harsher sentences for certain crimes if the perpetrator selected his or her victim "because of . . . race, religion, color, disability, sexual orientation, national origin or ancestry." [\[FN352\]](#) A unanimous Court rejected the defendant's claim that the statute discriminated against him on the basis of his racist views. [\[FN353\]](#) It noted that the additional penalties attached to the defendant's discriminatory intent because of his conduct, not his bigoted ideas. [\[FN354\]](#)

The Court analogized the Wisconsin statute to federal and state antidiscrimination laws, which, it explained, were immune from First Amendment challenges. [\[FN355\]](#) It pointed to Title VII of the Civil Rights Act of 1964 and [42 U.S.C. § 1981](#) as examples from civil rights law of "permissible content-neutral regulation of conduct." [\[FN356\]](#) The Court noted that "whereas the ordinance struck down in *R.A.V.* was explicitly directed at expression (i.e., 'speech' or 'messages'), the statute in this case is aimed at conduct unprotected ***113** by the First Amendment." [\[FN357\]](#) It found Wisconsin was justified in singling out "bias-inspired conduct because this conduct is thought to inflict greater individual and societal harm. . . . The State's desire to redress these perceived harms provides an adequate explanation for its penalty-enhancement provision over and above mere disagreement with offenders' beliefs or biases." [\[FN358\]](#) The Court underscored that "[t]he First Amendment . . . does not prohibit the evidentiary use of speech to establish the elements of a crime or to prove motive or intent." [\[FN359\]](#)

Applying civil rights statutes to the attacks of cyber mobs falls clearly on the *Mitchell* side of this line. The statutes' proscriptions turn on an online mob's discriminatory choice of victim and the distinct harm to victims and society that the defendant's abusive conduct produces, rather than on the opinions that either the victims or the attackers express. [\[FN360\]](#) Seeking to prevent a woman from maintaining an income-generating blog through threats and denial-of-service attacks because she is a woman is equally offensive, and equally proscribed, no matter the perpetrator's specific views. Aiming to prevent a person of color from securing gainful employment because of her race is no more or less offensive depending on the nature of the lies or the private information disseminated. Many online attacks have included racist, sexist, or other bigoted language; others have not. When the law punishes online attackers due to the special severity

of the social harm produced by targeting these classes of victims on bases of gender or race, and not due to the particular opinions the victims express, no First Amendment values are implicated.

R.A.V. confirmed “that nonverbal expressive activity can be banned because of the action it entails, but not because of the ideas it expresses.” [\[FN361\]](#) The application of civil rights laws to online mobs clearly targets actions - the interference with job opportunities through threats, damaging statements, and technological attacks - and is indifferent to the mobs' ideas. Indeed, even if these laws did single out some sub-types of proscribed speech, such as severe threats or especially injurious defamation, the Court noted that it would raise no First Amendment concerns. The Court explained:

When the basis for the content discrimination consists entirely of the very reason the entire class of speech at issue is proscribable, no significant danger of idea or viewpoint discrimination exists. Such a reason, having been adjudged neutral enough to support exclusion of the entire class of ***114** speech from First Amendment protection, is also neutral enough to form the basis of distinction within the class. [\[FN362\]](#)

Virginia v. Black [\[FN363\]](#) provides further support for this reading of R.A.V. In Black, the Court held that a state “may ban cross burning carried out with the intent to intimidate,” but struck down a provision in Virginia's statute that treated all cross-burnings as prima facie evidence of intent to intimidate. [\[FN364\]](#) The Court explicitly reaffirmed R.A.V.'s holding that the government may prohibit low-value speech across-the-board, but not specific speech that discriminates on the basis of message. [\[FN365\]](#) The Court distinguished cross-burning with the intent to intimidate, which it deemed a proscribable “true threat,” from cross-burning for other purposes, which it held constituted a protected expression of a viewpoint. [\[FN366\]](#) Thus, far from being immune from scrutiny, Black confirms that actors' motives may be decisive in the classification of their actions. [\[FN367\]](#) This is precisely the point Mitchell made.

Attempting to prevent anyone from making a living is offensive. By contrast, attempting to prevent someone from making a living because of her race is a civil rights violation. Federal and state antidiscrimination laws focus on the perpetrator's discriminatory intent in targeting the victim and the special harm that results, not on any views that either the perpetrator or victim might have, and thus the laws' application here would not offend the First Amendment. As such, their application to online mobs poses no First Amendment problems under current doctrine.