Article

Disinformation, Democracy, and the Social Costs of Identity-Based Attacks Online

October 22, 2019

Editor's Note: The following article quotes and references toxic and hateful speech, including attacks based on gender, ethnicity, religion, and other markers of identity. It also includes threats or descriptions of sexual assault and other references to violence.

The Ugliness of Strangers

In July, US president Donald Trump posted a now-infamous thread on Twitter: "So interesting to see 'Progressive' Democrat Congresswomen, who originally came from countries whose governments are a complete and total catastrophe, the worst, most corrupt and inept anywhere in the world (if they even have a functioning government at all), now loudly and viciously telling the people of the United States, the greatest and most powerful Nation on earth, how our government is to be run," he tweeted. "Why don't they go back and help fix the totally broken and crime infested places from which they came. Then come back and show us how it is done."

As has become characteristic of American politics, a media battle erupted in the wake of these remarks, conservatives and progressives alike inflamed with moral indignation. Many pundits and political figures condemned the comments as racist, xenophobic, and anti-American, while others defended them vociferously—describing critics as the true racists, shamelessly playing the victim for political benefit. While unbiased observers might have made a range of interpretations of the president's original tweets, the subsequent chants of "Send her back!" that were left unchallenged at a Trump rally reduced the ambiguity of his missive. News of the chants fanned the fire on talk radio, on cable news analysis shows, and across social media, disturbing many in the US who have been made to feel disparaged or unwelcome at some point in their lives on the basis of their race, ethnicity, or national origin.

Ilhan Omar, Democratic representative from Minnesota and one of the presumptive targets of Trump's remarks, took to social media in response, tweeting a segment of Maya Angelou's defiant 1978 poem, *Still I Rise*:

You may shoot me with your words,

You may cut me with your eyes,

You may kill me with your hatefulness,

But still, like air, I'll rise.

Omar used Angelou's powerful message to assert her own resilience, but her public display of backbone led others to test it. Although Omar was applauded by many, public life online for women—particularly women from historically underrepresented groups—is often marred by the ugliness of strangers. Omar's experience was no exception. Amid the virtual fist-bumps and standing ovations, the civil arguments about immigration, and the contradictory assertions about the true nature of patriotism were a stunning number of hate-based barbs from those who saw Omar's declaration of determination as an open invitation to threaten, harass, and malign her. A few examples of @mentions directed at her in the subsequent week:

Replying to @IlhanMN

Hahah bitch as if! you belong in fucking Somalia 😂 fucking lying cunt. Omg You, @AOC and the other gorrilla looking bitch are making every minority in the US look like a privilege cry baby asshole and we all see through your ploy. Fucking antifa is your base you scum.

1:51 AM · Jul 18, 2019 · Twitter for iPhone

Replying to @IlhanMN

The noose is tightening. Tick tock

8:43 PM · Jul 23, 2019 · Twitter Web App

Replying to @IlhanMN

You need to leave my fucking country before your people end up dead over here. We will not succumb to your demands. We will not give up our guns, our flag, or our pork. Just get the fuck out you stupid cunt!!!!!! I do speak for the American people, you do not!!

8:50 PM · Jul 18, 2019 · Twitter for Android

Replying to @IlhanMN

Shut your cock holster you racist terrorist. Go eat some bacon.

12:16 AM · Jul 22, 2019 · Twitter for iPhone

This backlash should give us pause because of its xenophobia, misogyny, and violence. It should give us pause because this kind of rhetoric has become tediously common (Awan 2016; Madden et al. 2018; Farrell et al. 2019; Schäfer and Schadauer 2018; Sobieraj and Merchant, forthcoming). It should give us pause because it threatens the lives and liberties of those it targets. And it should give us pause because it has deleterious consequences for the strength of our communities and the health of our democracy. Finally, we must pause to attend to the important but unacknowledged links between identity-based abuse and what Wardle and Derakhshan (2017) refer to as information disorder. As I will show, attackers weaponize disinformation and misinformation as part and parcel of their abuse, making it simultaneously a product of, and a resource for, digital hate.

This essay will address the uneven distribution of identity-based hate online, explain what these patterns reveal, address the threats to democracy posed by this kind of behavior, unravel the symbiotic connections between disinformation and digital abuse, and argue that platforms and policymakers have been inexcusably slow to respond. My hope is that the "pauses" referenced above will create the space necessary to begin to address this digital hate in a more meaningful fashion.

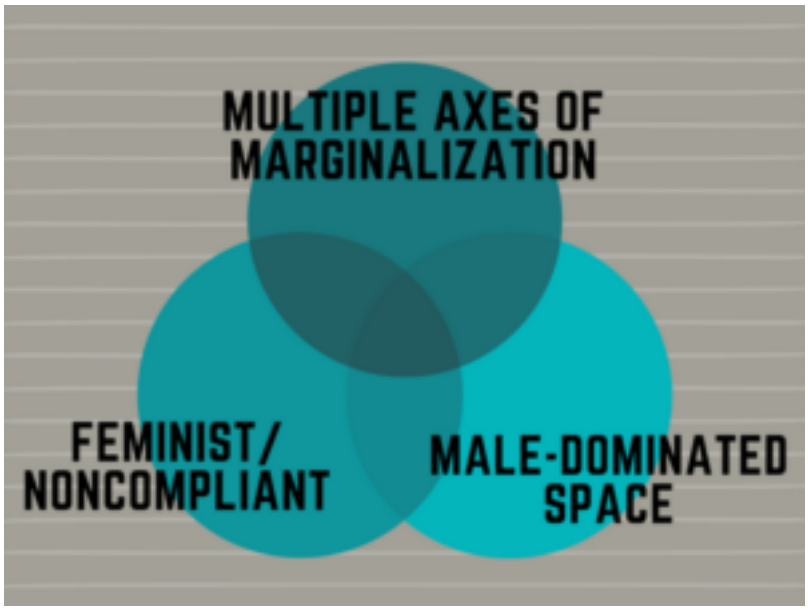
Patterned Resistance

Communications scholar Joseph Reagle notes that most theories of digital toxicity boil down to accounts of "good people acting badly"—their behavior altered by the unique characteristics of life online (e.g., anonymity, social distance)—and those of "bad people acting out," which suggest that a subset of people with particular personality profiles are to blame (Reagle 2015, 94–97). But if these were adequate explanations, we would expect abusers and targets to be randomly distributed. They aren't. Instead we find a preponderance of men lashing out at women, particularly women of color and queer women of all races (Amnesty International 2018; Phillips 2015; Citron 2014; Herring 2002; Herring et al. 2002). What's more, the *substance* of these attacks is patterned. Targets are frequently attacked on the grounds that their very identities are unacceptable, with their perceived race, gender, ethnicity, religion, sexual orientation, and/or social class as the central basis for condemnation. Ilhan Omar is challenged for her stance on immigration, for example, but her national origin, religion, and gender are invoked in the process. Twitter @mentions use gendered epithets such as "cunt" and "bitch," insist that she is not American and belongs in Somalia, and denigrate her as a Muslim (e.g., "go eat some bacon" and "may the good lord rain down blood of your Muslim brothers and sisters"). Other "feedback" uses terms such as "sand whore" and "sand nigger," mocks her hijab, and sexualizes her in humiliating ways. Not exactly the kind of political discourse toward which democratic theorists might have us aspire.

One of the peculiar things about identity-based abuse is that it is intimate and ad hominem, yet simultaneously generic. The attacks are largely *im*personal, brushed across anyone with a similar social location, such that the slurs and threats seem nearly interchangeable. This generic quality is parodied by scholars Emma Jane and Nicole Vincent, who designed a "Random Rape Threat Generator." The generator is a digital slot machine with three wheels that spin independently, each containing misogynistic commentary culled from an archive of actual attacks against women online. When users press "play," the spinning wheels combine to yield a three-part compound insult (e.g., "Hope you get raped to death" + "you PC" + "cumdumpster"). Jane and Vincent's tool illustrates an important point: the hate directed at women online *feels* intimate and deeply personal, but most of the missives contain nonspecific misogyny, divorced from the behaviors, ideas, or attributes of any particular woman. This rubberstamp quality is revealing. It tells us that although identity-based abuse can feel and look like interpersonal bullying, the rage is more structural, rooted in hostility toward the voice and visibility of individual speakers as representatives of *specific groups of people*.

Behind this brand of digital toxicity is a struggle to control who is allowed to hold sway in public discourse. Recognizing it as such helps explain why the attacks are unevenly distributed—even among women. As I have noted elsewhere (Sobieraj 2018; Sobieraj and Merchant, forthcoming), while women receive far more abuse than men in general, the abuse is unusually burdensome for women who are members of multiple marginalized groups, such as women of color. They receive racialized gender-based attacks and race-based attacks in addition to the kinds of harassment that white women experience (Gray 2012, 2014). Queer women of all races are also frequent targets (Citron 2009; Finn 2004), as are women from ethnic and religious minority groups, women with disabilities, and poor women. There is also an increased amount of hostility directed toward women who challenge (whether intentionally or by their mere presence) the norms and practices in male-dominated arenas, such as science, technology, gaming, sports, politics, and the military. Finally, feminist women and those who are otherwise noncompliant with traditional gender expectations are also subject to particularly vicious attacks. By "noncompliant" I mean anyone who disturbs the gender order, such as women who hold positions of power, who are open about enjoying sex, or who celebrate their "unacceptable" bodies—unashamed of their gender nonconformity, cellulite, body hair, or abundant curves (Sobieraj 2018; Sobieraj, forthcoming). And, of course, as figure 1 illustrates, many women fall into more than one of these categories. Representative Ilhan Omar sits precisely in the eye of this storm (as do Alexandria Ocasio-Cortez [D-NY], Ayanna Pressley [D-MA], and Rashida Tlaib [D-MI], the other three members of the so-called Squad).

Figure 1:



All three categories of women represented in figure 1 can be understood as destabilizers, whether or not they intend to be. They challenge gender norms (and often norms around race, sexuality, ability, etc.) simply by entering digital publics to address matters of social or

political concern. It is helpful, then, to resee the hatred directed at Ilhan Omar (or that directed at Anita Sarkeesian of Gamergate, Parkland survivor and gun control advocate Emma Gonzalez, climate activist Greta Thunberg, and comedian Leslie Jones, among others) as a visceral response. The hatred budding from an unsettling—and often subconscious—sense that long-taken-for-granted social hierarchies may be shifting, upsetting the footing of those who are currently advantaged—in many cases, white men. For those who benefit from inequality (especially those who perceive the social order as meritocratic), it is little surprise that shifts toward inclusion, access, diversity, and meaningful equal opportunity feel remarkably like *in*justice, as I and others have written elsewhere (Sobieraj, Berry, and Connors 2013; Kimmel 2013).

What's more, understanding digital attacks as patterned resistance to underrepresented voices explains why race, sexuality, gender, and the like are so prominent in the substance of the abuse. It is not simply that these "destabilizers" (most often women, but also men of color, queer people, etc.) are attacked more often or more intensely, but that they are regularly attacked in ways that center their identities. Calling Omar a "sand nigger" or an "incestuous cunt" asks others to disregard her ideas and opinions on the basis of her national origin, ethnic identity, and gender. In a cultural context where these attributes are devalued and people who possess them are seen as uppity, ignorant, disgusting, or dangerous, *who Ilhan Omar is* proves easier to assail than *what she has to say*. The not-so-subtle subtexts of these attacks exclaim, "This person does not have the authority to speak"; "Do not listen to her"; "How dare this brown Muslim immigrant woman criticize my country? Who does she think she is?" The epithets, stereotypes, and patronizing missives are a gut-punch reminder to targets that they should stay in their lane.

The Costs of Identity-Based Attacks Online

In my research with women who have been on the receiving end of identity-based attacks online from strangers, I have heard many accounts that lay bare the impact of digital toxicity. Although the fallout varies, women's experiences have been financially draining, sapped scarce time and energy, and come at tremendous professional, personal, and psychological cost—amounting to a rarely acknowledged form of persistent gender inequality, one that reflects intersectional oppressions and falls more heavily on some shoulders than others.

In response to this climate, some women leave their preferred social media platforms, take breaks to recover from the fear and frustration, write under pseudonyms, or use identity-disguising avatars. Many turn to private digital spaces where they are surrounded by like-minded and supportive peers. These spaces are deeply meaningful and may themselves

foment great political dialogue, energy, and involvement, but to the extent that these enclaves become alternatives to participation in broader digital publics, rather than supplements to it, identity-based attacks diminish the vitality of our public discourse. And, based on the uneven distribution of vitriol, it is likely that those who leave are disproportionately women of color, queer women, and women with other "unpopular" identities such as Muslim women, immigrant women, and women who are poor. The most underrepresented voices and perspectives are quite likely to be the first pushed out.

Not all women are attacked online, but digital abuse is powerful even in its absence. When women see a peer deluged with rape threats, they recognize themselves as potential targets. The looming threat of being attacked serves as a cautionary tale—a deterrent—that inhibits other women from speaking freely about social and political issues. Like the female jogger who avoids running alone in the forest for fear of attack but who would have most likely been safe, stories of digital backlash have a chilling effect strong enough to discourage some women from speaking about public issues in digital spaces, even if their participation might have been uneventful. This is especially true if the issues they wish to address are particularly controversial, or their point of view unpopular. Robust democracies are built on political discourse in which people—including those without power—discuss even the most sensitive topics (e.g., abortion, guns, immigration) and share their ideas, experiences, and opinions without fear. Whether or not we share Ilhan Omar's preferences or opinions, her voice is a valuable part of our national conversation. And for young people, particularly young Muslim women, seeing a Muslim woman on the public stage tells them that their ideas and opinions matter. Such models of political voice, involvement, and impact are critical to political socialization and the development of the sense of efficacy. When we lose women's voices, the breadth and depth of the information we encounter online withers, leaving gaps in the perspectives, information, and experiences to which we have access that are decidedly not random.

Identity-based attacks deal a second blow to information quality by trading freely and enthusiastically in misinformation and disinformation. The importance of misinformation and disinformation in identity-based attacks is visible in the backlash directed toward Ilhan Omar. Among the tweets are myriad accusations that she married her brother, an unfounded conspiracy theory that has been circulating since 2016. There were dozens of references to the alleged incest in the replies to Ilhan Omar's Maya Angelou tweet alone. The replies also included assertions that Omar herself is a terrorist, a theory proffered by Glenn Beck that caught fire online. Consider, for example, *Front Page*'s not-so-subtle coverage of the Beck theory, entitled "CONGRESSWOMAN ILHAN OMAR: TERRORIST," which was shared almost 58,000 times on Facebook alone. The shadows of such misinformation simultaneously fuel the rage directed at Omar and help propagate the falsehoods. It seems fitting that *Still I Rise* begins with a reference to this kind of

defamation. Angelou's opening stanza reads:

You may write me down in history

With your bitter, twisted lies,

You may trod me in the very dirt

But still, like dust, I'll rise.

At times the disinformation echoes the arguments and language used by those in talk media, cable news, and ideologically driven "news" sources. Sometimes it parrots back remarks and innuendoes made by other political figures, and at other times the digital attacks introduce new or lesser-heard "facts" into the conversation. In the Omar case, some of these include debunked rumors such as the story that Omar insisted she would not pledge allegiance to America. The repetition of such narratives matters; the more a piece of misinformation is repeated, the more likely we are to believe that it is true (Claypool et al. 2004; Fazio, Rand, and Pennycook 2019; Foster et al. 2012; Pennycook, Cannon, and Rand 2018). Other attackers introduce more outlandish assertions, such as this reply to one of Omar's earlier tweets:

A dispassionate observer might think the suggestion that a member of the US House of Representatives tortures, guts, and eats children is not technically disinformation, because it is far too outrageous for anyone to believe. This is not the case, especially in the context of other well-circulated "facts" that claim Omar was personally involved in the massacre of Somalis or that she is a fake citizen sent to start a war to destroy the United States.

This kind of content has consequences for those who are defamed. Omar has had to address these issues in the media ad nauseum. In an interview with the *Star Tribune* during the 2018 election season, she said, "It's really strange, right, to prove a negative," in reference to the incest allegations. "If someone was asking me, do I have a brother by that name, I

don’t. If someone was asking … are there court documents that are false … there is no truth to that.” And yet, reminiscent of the false accusations leveled against Barack Obama by the Birthers, questions persist. Reputation management takes time and energy, and the rumors likely distract constituents, colleagues, and journalists from focusing on issues that Omar feels are more salient. The fear-mongering may ultimately cost her the next election. But this is far from a personal problem. Disinformation and misinformation are antithetical to a healthy information environment, as elections are only meaningful if the citizenry has adequate information to make informed decisions on their own behalf when they enter the voting booth.

The ability to obtain “adequate information” is hampered not only by the falsehoods deployed vis-à-vis identity-based attacks but also by the way misinformation, disinformation, propaganda, and fake news have increased skepticism and contributed to a generalized distrust of journalism. This is compounded, of course, by the historically familiar epistemological costs of patterned exclusion from other spheres of public discourse. Histories of white, land-owning, male deliberation have already shown how severely the products of social and political discourse are distorted when the things we come to believe as truths or facts or reality are by definition incomplete, because members of specific groups of society have been systematically pushed out of public conversations.

Not Enough

Ilhan Omar has been the subject of a stunning amount of digital abuse, but she is not a peculiarity. In August 2018, amid an unprecedented number of women running for public office in the US, of whom Omar was one, the *New York Times* released a video of current and former female candidates talking about their experiences with harassment and sexism, much via social media, during their campaigns (Kerr, Tiefenthäler, and Fineman 2018). Throughout the video, the women describe abuse that is gendered and racialized. Among the speakers in the video is Iowa Democrat Kim Weaver, who pulled out of her 2016 congressional race amid a torrent of sexist and anti-Semitic abuse (Astor 2018). My own research has shown that this abuse does not necessarily subside once the targets are in office (Sobieraj and Merchant, forthcoming). Focusing on politics highlights the costs to democracy most effectively, but these patterns of digital abuse have been documented in arenas ranging from gaming (Gray 2012; Fox and Tang 2017) to academia (Ferber 2018; Veletsianos et al. 2018) to journalism (Gardiner 2018; Adams 2018; Chen et al. 2018). In talking about the challenge of keeping people safe online, Del Harvey, vice president for trust and safety at Twitter herself, said, “Given the scale that Twitter is at, a one-in-a-million chance happens 500 times a day. It’s the same for other companies dealing at this sort of

scale. For us, edge cases, those rare situations that are unlikely to occur, are more like norms." In other words, there is a lot of digital toxicity. And why not? Users are operating in a nearly consequence-free environment.

Those targeted need recourse and, in spite of critics' fervor over free speech, there is no legal reason to tolerate this abuse. Civil rights abuses are not protected by the First Amendment to the US Constitution. As Citron explains, "When law punishes online attackers due to their singling out victims for online abuse due to their gender, race, sexual orientation, or other protected characteristics and the special severity of the harm produced, and not due to the particular opinions that the attackers or victims express, its application does not transgress the First Amendment" (2014, 221). What's more, platforms are not government actors and are free to make whatever rules they like for the speech that takes place on their sites.

At first blush, it appears as though opportunities for recourse are plentiful. There are several potential pathways for legal action, from criminal charges for stalking and harassment or threats of violence to civil claims for infractions such as defamation and intentional infliction of emotional distress. Civil rights laws, which are designed to punish abuse motivated by race, national origin, religion, and (in some states) gender and sexual orientation, have been infrequently embraced as a form of redress, according to Citron, but to the extent that online harassment threatens to drive people from protected groups offline, they have a great deal of potential (2014, 126–29). Unfortunately, when the abuse comes in the form of a million paper cuts—most of which individually do not meet the threshold for legal action—and when harassers whose efforts break the law prove hard to identify, these forms of legal recourse are largely inaccessible. Citron concludes, "Law cannot communicate norms, deter unlawful activity, or remedy injuries if defendants cannot be found. Perpetrators can be hard to identify if they use anonymizing technologies or post on sites that do not collect IP addresses. Because the law's efficacy depends on having defendants to penalize, legal reform should include, but not focus exclusively on, harassers." (2014, 143). In other words, the most effective path toward eliminating digital abuse is unlikely to be one that hinges on catching individual attackers.

This demands we take the ecosystem in which attackers flourish seriously. Prominent platforms such as Twitter, Facebook, and YouTube all have terms of service that reject this kind of digital abuse. For example, Twitter does not allow "hateful conduct," which includes, among other things: (1) targeting individuals with content intended to incite fear or spread fearful stereotypes about a protected category, including asserting that members of a protected category are more likely to take part in dangerous or illegal activities, and (2) targeting individuals with repeated slurs, tropes, or other content that intends to dehumanize, degrade, or reinforce negative or harmful stereotypes about a protected

category (https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy). But such prohibitions are ineffective without meaningful enforcement. As the Omar backlash illustrates, platform enforcement is inadequate, making the policies feel more akin to guidelines. This is not to say that platforms have ignored enforcement thus far. Facebook, for example, uses artificial intelligence, community flagging features, and employs 15,000 content reviewers globally (counting full- and part-time employees as well as those working via third-party contractors), a number that has tripled since 2017 (Madrigal 2018; Wong 2019). And in the first half of 2018 alone, Twitter reportedly took action against nearly 300,000 accounts in response to user reports (Chen 2019). But their efforts are simply not enough. This is particularly clear in light of recent research that demonstrates the way platform affordances shape influence levels of incivility and expressions of hatred in the moment and in future interactions (Jaidka, Zhou, and Lelkes 2019; Matamoros-Fernández 2017; Merrill and Åkerlund 2018; Shmargad and Klar 2019).

Until meaningful reform takes place, employers and managers who incentivize or require their subordinates to promote themselves or their work on social media or to engage with constituents, clients, customers, or readers in these contexts must be aware that they are asking women (and men from historically marginalized groups) to enter a hostile speaking environment. Those who have known only the luxury of a relatively comfortable digital life—such as the *New York Times*' Bret Stephens, who was recently outraged after being called a "bedbug" on Twitter—may find the notion that these are often abusive spaces hard to fathom. But this hostility is now well-documented, and responsible employers will need to adjust their expectations, reward structures, and support systems accordingly.

Finding a way to mete out consequences is not easy or inexpensive, but as communications scholar and moderation expert Tarleton Gillespie argues, "When an intermediary grows this large, this entwined with the institutions of public discourse, this crucial, it has an *implicit contract* with the public…The primary and secondary effects these platforms have on essential aspects of public life, as they become apparent, now lie at their doorstep" (2018, 208). If platforms such as Twitter, Facebook, and YouTube are unable or unwilling to manage the emergent threats posed by harassment, such as civil rights abuses, reductions in information diversity, and the circulation of disinformation, then it is time to acknowledge that industry self-regulation—the de facto norm in the United States as a result of Section 230 of the Communications Decency Act of 1996—is no longer in the public interest.

Expert Reflections are submitted by members of the MediaWell Advisory Board, a diverse group of prominent researchers and scholars who guide the project's approach. In these essays, they discuss recent political developments, offer their predictions for the near future, and suggest concrete policy recommendations informed by their own research. Their opinions are their own, with minor edits by MediaWell staff for style and clarity. You can

find other Expert Reflections here.

Note: An earlier version of this essay incorrectly named the act that includes Section 230.

[workscited]

Adams, Catherine. 2018. "'They Go for Gender First.'" *Journalism Practice* 12 (7): 850–69. https://doi.org/10.1080/17512786.2017.1350115.

Amnesty International. 2018. "Troll Patrol." https://decoders.amnesty.org/projects/troll-patrol/findings#what_did_we_find_container.

Astor, Maggie. 2018. "For Female Candidates, Harassment and Threats Come Every Day." *The New York Times*, August 24, 2018. https://www.nytimes.com/2018/08/24/us/politics/women-harassment-elections.html.

Awan, Imran. 2016. "Virtual Islamophobia: The Eight Faces of Anti-Muslim Trolls on Twitter." In *Islamophobia in Cyberspace*, edited by Imran Awan, 37–54. London: Routledge.

Claypool, Heather M., Diane M. Mackie, Teresa Garcia-Marques, Ashley McIntosh, and Ashton Udall. 2004. "The Effects of Personal Relevance and Repetition on Persuasive Processing." *Social Cognition* 22(3): 310–35. https://doi.org/10.1521/soco.22.3.310.35970.

Chen, Gina Masullo, Paromita Pain, Victoria Y Chen, Madlin Mekelburg, Nina Springer, and Franziska Troger. 2018. "'You Really Have to Have a Thick Skin': A Cross-Cultural Perspective on How Online Harassment Influences Female Journalists." *Journalism*, April. https://doi.org/10.1177/1464884918768500.

Chen, Rachel. 2019. "Social Media Is Broken, But You Should Still Report Hate." *Vice* (blog). January 23, 2019. https://www.vice.com/en_us/article/d3mzqx/social-media-is-broken-but-you-should-still-report-hate.

Citron, Danielle Keats. 2009. "Law's Expressive Value in Combating Cyber Gender Harassment." *Michigan Law Review* 108 (3): 373–415. https://papers.ssrn.com/abstract=1352442.

———. 2014. *Hate Crimes in Cyberspace*. Cambridge, MA: Harvard University Press.

Farrell, Tracie, Miriam Fernandez, Jakub Novotny, and Harith Alani. 2019. "Exploring Misogyny across the Manosphere in Reddit." In *WebSci '19 Proceedings of the 10th ACM Conference on Web Science*, 87–96. Boston, MA, June 30–July 3, 2019.https://doi.org/10.1145/3292522.3326045.

Fazio, Lisa K., David G. Rand, and Gordon Pennycook. 2019. “Repetition Increases Perceived Truth Equally for Plausible and Implausible Statements.” *Psychonomic Bulletin & Review*, 1–6. https://doi.org/10.3758/s13423-019-01651-4.

Ferber, Abby L. 2018. “‘Are You Willing to Die for This Work?’ Public Targeted Online Harassment in Higher Education: SWS Presidential Address.” *Gender & Society* 32 (3): 301–20. https://doi.org/10.1177/0891243218766831.

Finn, Jerry. 2004. “A Survey of Online Harassment at a University Campus.” *Journal of Interpersonal Violence* 19 (4): 468–83. https://doi.org/10.1177/0886260503262083.

Foster, Jeffrey L., Thomas Huthwaite, Julia A. Yesberg, Maryanne Garry, and Elizabeth F. Loftus. 2012. “Repetition, Not Number of Sources, Increases Both Susceptibility to Misinformation and Confidence in the Accuracy of Eyewitnesses.” *Acta Psychologica* 139(2): 320–26. https://doi.org/10.1016/j.actpsy.2011.12.004.

Fox, Jesse, and Wai Yen Tang. 2017. “Women’s Experiences with General and Sexual Harassment in Online Video Games: Rumination, Organizational Responsiveness, Withdrawal, and Coping Strategies.” *New Media & Society*19 (8): 1290–1307. https://doi.org/10.1177/1461444816635778.

Gardiner, Becky. 2018. “‘It’s a Terrible Way to Go to Work:’ What 70 Million Readers’ Comments on the Guardian Revealed about Hostility to Women and Minorities Online.” *Feminist Media Studies*18 (4): 592–608. https://doi.org/10.1080/14680777.2018.1447334.

Gillespie, Tarleton. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven, CT: Yale University Press.

Gray, Kishonna L. 2012. “Intersecting Oppressions and Online Communities.” *Information, Communication & Society*15 (3): 411–28. https://doi.org/10.1080/1369118X.2011.642401.

———. 2014. *Race, Gender, and Deviance in Xbox Live: Theoretical Perspectives from the Virtual Margins*. New York: Routledge.

Herring, Susan C. 2002. “Computer-Mediated Communication on the Internet.” *Annual Review of Information Science and Technology* 36(1): 109–168. https://doi.org/10.1002/aris.1440360104.

Herring, Susan, Kirk Job-Sluder, Rebecca Scheckler, and Sasha Barab. 2002. “Searching for Safety Online: Managing ‘Trolling’ in a Feminist Forum.” *The Information Society* 18(5), 371–84. https://doi.org/10.1080/01972240290108186.

Jaidka, Kokil, Alvin Zhou, and Yphtach Lelkes. 2019. “Brevity Is the Soul of Twitter: The Constraint Affordance and Political Discussion.” *Journal of Communication* 69(4): 345–72. https://doi.org/10.1093/joc/jqz023.

Kerr, Sarah Stein, Ainara Tiefenthäler, and Nicole Fineman. 2018. “‘Where’s Your Husband?’ What Female Candidates Hear on the Trail.” Video, 4:13. *The New York Times*, August 24, 2018. https://www.nytimes.com/video/us/politics/100000006027375/women-politics-harassment.html.

Kimmel, Michael. 2013. *Angry White Men: American Masculinity at the End of an Era*. New York: Nation Books.

Madden, Stephanie, Melissa Janoske, Rowena Briones Winkler, and Amanda Nell Edgar. 2018. “Mediated Misogynoir: Intersecting Race and Gender in Online Harassment.” In *Mediating Misogyny*, edited by Jacqueline Ryan Vickery and Tracy Everbach, 71–90. Cham, Switzerland: Palgrave Macmillan. https://doi.org/10.1007/978-3-319-72917-6_4

Madrigal, Alexis C. 2018. “Inside Facebook’s Fast-Growing Content-Moderation Effort.” *The Atlantic*, February 7, 2018. https://www.theatlantic.com/technology/archive/2018/02/what-facebook-told-insiders-about-how-it-moderates-posts/552632/.

Matamoros-Fernández, Ariadna. 2017. “Platformed Racism: The Mediation and Circulation of an Australian Race-Based Controversy on Twitter, Facebook and YouTube.” *Information, Communication & Society* 20(6): 930–46. https://doi.org/10.1080/1369118X.2017.1293130.

Merrill, Samuel, and Mathilda Åkerlund. 2018. “Standing Up for Sweden? The Racist Discourses, Architectures and Affordances of an Anti-Immigration Facebook Group.” *Journal of Computer-Mediated Communication* 23(6): 332–53. https://doi.org/10.1093/jcmc/zmy018.

Pennycook, Gordon, Tyrone D. Cannon, and David G. Rand. 2018. “Prior Exposure Increases Perceived Accuracy of Fake News.” *Journal of Experimental Psychology: General* 147(12): 1865–80. http://dx.doi.org/10.1037/xge0000465.

Phillips, Whitney. 2015. *This Is Why We Can’t Have Nice Things: Mapping the Relationship between Online Trolling and Mainstream Culture*. Cambridge, MA: MIT Press.

Reagle, Joseph M. Jr. 2015. *Reading the Comments: Likers, Haters, and Manipulators at the Bottom of the Web*. Cambridge, MA: MIT Press.

Schäfer, Claudia, and Andreas Schadauer. 2018. “Online Fake News, Hateful Posts Against

Refugees, and a Surge in Xenophobia and Hate Crimes in Austria." In *Refugee News, Refugee Politics: Journalism, Public Opinion and Policymaking in Europe*, edited by Giovanna Dell'Orto and Irmgard Wetzstein. New York: Routledge.

Shmargad, Yotam, and Samara Klar. 2019. "How Partisan Online Environments Shape Communication with Political Outgroups." *International Journal of Communication* 13: 2287–2313.

Sobieraj, Sarah. 2018. "Bitch, Slut, Skank, Cunt: Patterned Resistance to Women's Visibility in Digital Publics." *Information, Communication & Society* 21 (11): 1700–1714. https://doi.org/10.1080/1369118X.2017.1348535.

———. Forthcoming. *Credible Threat: Attacks Against Women Online and the Future of Democratic Discourse.* Oxford University Press.

Sobieraj, Sarah, Jeffrey Berry, and Amy Connors. 2013. "Outrageous Political Opinion and Political Anxiety in the U.S." *Poetics* 41 (5): 407–432. https://doi.org/10.1016/j.poetic.2013.06.001.

Sobieraj, Sarah., and S. Merchant, forthcoming. "Gender and Race in the Digital Town Hall: Identity-Based Attacks Against US Legislators on Twitter." In *Social Media and Social Order*, edited by D. Herbert and S. Fisher-Høyrem. Berlin: De Gruyter.

Veletsianos, George, Shandell Houlden, Jaigris Hodson, and Chandell Gosse. 2018. "Women Scholars' Experiences with Online Harassment and Abuse: Self-Protection, Resistance, Acceptance, and Self-Blame." *New Media & Society* 20 (12): 4689–4708. https://doi.org/10.1177/1461444818781324.

Wardle, Claire, and Hossein Derakhshan. 2017. *Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking.* Strasbourg: Council of Europe.

Wong, Queenie. 2019. "Facebook Content Moderation Is an Ugly Business. Here's Who Does It." *CNET*, June 19, 2019. https://www.cnet.com/news/facebook-content-moderation-is-an-ugly-business-heres-who-does-it/.

[/workscited]