

“Men Are Scum”: Self-Regulation, Hate Speech, and Gender-Based Censorship on Facebook

CHLOÉ NURIK¹

University of Pennsylvania, USA

Because social media sites are self-regulating, each site has developed its own community standards, which serve as regulatory tools. However, the processes of content moderation are often unclear, subjective, and discriminatory. Drawing from a series of interviews with individuals in the “Men Are Scum” movement, this article describes the experiences of women who have been censored on Facebook and explores whether self-regulatory processes on this platform are distinctly gendered. It asserts that both explicit censorship (e.g., limited displays of the body) and implicit censorship (e.g., rampant and unchecked hate speech silencing women’s voices) are operative on Facebook, limiting women’s expressive potentiality. Thus, this article proposes the term “gender-based censorship” as a lens through which to understand women’s experiences on Facebook. These findings help reveal the pitfalls of industry self-regulation in which profit motives are prioritized over protection of users (especially those who may be marginalized offline).

Keywords: Facebook, self-regulation, gender-based censorship, content moderation, social media

“Facebook has more power in determining who can speak and who can be heard around the globe than any Supreme Court Justice, any king or any president.”

—George Washington University law professor Jeffrey Rosen
(as cited in Helft, 2010, para. 7)

In the contemporary moment, personal expression on social media is used for identity formation, civic engagement, and political participation (Gil de Zúñiga, Jung, & Valenzuela, 2012). However, individual posts on social media platforms are shaped by each site’s community standards, which serve as regulatory tools. Although these explicit guidelines are listed on the website of each social media network, the processes of content moderation are vague and subjective, and users are often unsure what

Chloé Nurik: chloe.nurik@asc.upenn.edu

Date submitted: 2018–05–05

¹ I would like to thank Drs. Monroe Price, Victor Pickard, Jessa Lingel, and Kenneth Winneg for their feedback, support, and guidance. Additionally, I am appreciative of the reviewers who offered helpful comments that were used to deepen the article’s analysis. Thank you to the interviewees (especially Kayla Avery) who made this project possible.

Copyright © 2019 (Chloé Lynn Nurik). Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at <http://ijoc.org>.

material is acceptable to post. Further, the rules may be applied disparately in the absence of external oversight, leading to inconsistent enforcement of the site's guidelines. In such a regulatory schema, there is the potential for voices to be weighed unequally and for material to be selectively restricted. Drawing from a series of interviews with individuals in the "Men Are Scum" movement, this article describes the experiences of women who claim to have been censored and harassed on Facebook and explores whether the platform's self-regulatory systems are distinctly gendered. Although problems with content moderation, hate speech, and threats have also been impactful for individuals identifying as nonbinary, this study largely focuses on the accounts of users who self-identify as women (a group that comprises the plurality of Facebook users). Gender-based censorship is important to examine because an investigation of the socially constructed allocation of speech and the implicit biases inherent in self-regulatory processes on social media exposes the fundamental limitations of commercial social media for providing equal access to speech for all users.

Contextualizing Facebook's Self-Regulation

Numerous social media sites, including Facebook, are based in the United States (U.S.) and have relied on American legal norms and frameworks to construct their policies and standards (Klonick, 2018). The Communications Decency Act of 1996 (CDA) in the U.S. both restricts liability for user-generated content on interactive computer services and permits these entities to develop their own regulatory schemas. Because interactive computer services such as platforms are not currently categorized as publishers, they are treated in a different manner than traditional content providers who do not receive the same degree of immunity. Section 230 of the CDA establishes that neither users nor providers of interactive computer services are liable for

any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected. § 230 (c)(2)(a)

Herein, the phrase "any action voluntarily taken" (§ 230) highlights the voluntary rather than obligatory nature of content moderation, which does not incentivize social media sites to be transparent, accountable, or equitable. This lax framework has been criticized by many scholars, including Georgetown law professor Rebecca Tushnet (2008), who expressed concern that platforms have been granted "power without responsibility" (p. 986), a scenario that enables the digital proliferation of sexism, hate speech, and threats on a global scale.

The devolvement of regulation onto platforms may have been appropriate in the mid-1990s (Citron, 2014), but the sheer size, number of users, and international reach of social media sites have problematized existing legal frameworks. In contrast to the U.S.'s laissez-faire approach toward platforms, several European nations have adopted stronger regulatory measures. The European Union (EU) has motivated social media and technology companies to adopt a voluntary code of conduct countering illegal hate speech (which focuses primarily on racism and xenophobia). Further, the EU passed the sweeping General Data Protection Regulation, which strengthens user protections (e.g.,

making data collection transparent and enabling users to demand that certain posts be removed from platforms). Additionally, countries such as Germany have passed laws requiring social media sites to quickly remove “clearly illegal” material, including hate speech, or face steep fines (“Germany Starts,” 2018). Although these policy initiatives may provide inroads into the problems of self-regulation, the power of social media sites to determine the parameters of speech is still vast and largely unchecked. In fact, Facebook and other social media sites have already devised ways to bypass these requirements and reduce compliance costs.² Therefore, although recent regulatory efforts hold promise, the lax nature of U.S. provisions (such as § 230) is still globally impactful.

Processes and Complications of Self-Regulation

Given its lack of regulatory oversight, Facebook established terms of service and community guidelines for users. Facebook’s stated goal is as follows: “We want people to feel safe when using Facebook” (“Community Standards,” n.d.). Consequently, the platform created several provisions, which are intended to protect users. With regard to hate speech (the focus of this article), Facebook prohibits hateful language and categorizes verbal attacks as falling into one of three tiers—dehumanizing speech, individual attacks, and calls for segregation or exclusion (“Hate Speech,” n.d.). Facebook also indicates that users cannot threaten either a private or public figure (“Harassment,” n.d.).

Facebook works to enforce the same standards across the globe. This approach can be problematic because the content users post may be evaluated based on the norms of another culture. Further, the notion of “one Facebook” (Ammori, 2014, p. 2281) is complicated by the fact that users’ experiences on the site are not monolithic, but instead are deeply shaped by intersectional identifiers such as race, gender, and nationality, making the push for a consistent and fair policy difficult to enact and yet critically important to realize.

To uphold its unitary standards, Facebook relies on individual users as well as a team of content moderators. The platform delegates aspects of content moderation to users by asking them to report others’ posts for violating communal guidelines. Even though the site has more than 2.2 billion users and reviews more than 100 million pieces of content each month (Wagner & Swisher, 2017), to date it employs only 7,500 censors whose demographics are not released by the site (Madrigal, 2018). Predominantly outsourced and underpaid (Roberts, 2016), censors are entrusted with deciding whether reported content complies with the community guidelines. Facebook “largely allows the companies that hire the moderators to police themselves” (Beirne, 2018, para. 16), revealing the subjectivity and lack of accountability threaded throughout this process. The vast number of users and the equivocal nature of Facebook’s policies endows this platform (and its censors) with tremendous clout to influence cultural standards around the world. Susan Benesch, director of the Dangerous Speech Project, explained, “Facebook is regulating more human speech than any government does now or ever has. . . . They are like a de facto

² For example, although Facebook users outside the U.S. and Canada are currently governed by terms of service under the site’s international headquarters in Ireland, the platform plans to have only users in Europe follow these terms of service so as to avoid applying General Data Protection Regulation provisions globally (Ingram, 2018).

body of law, yet that law is a secret" (as cited in Jan & Dwoskin, 2017, para. 15). This relatively unchecked power (along with Facebook's lack of transparency) contributes to abuse and discrimination on the platform (especially against users from historically marginalized groups).

Literature Review and Research Questions

Given the prevalence of social media sites and the importance of content regulation, there has been a recent growth in scholarship on this topic. This article draws upon two bodies of literature: studies of social media content moderation and works analyzing women's interactions online.

A major area of study pertains to the legal frameworks, technical affordances, social implications, and labor practices of content moderation. Scholars have examined how self-regulatory content moderation emerged in response to U.S. laws such as the CDA (e.g., Citron, 2014; Klonick, 2018). Although flagging practices differ across social media sites, researchers have expressed great concern about censorial practices overall (e.g., Crawford & Gillespie, 2016). Describing platforms' immense power to control speech, Gillespie (2018) argued that content moderation is the central task of social media sites rather than a peripheral function. The institutional study of social media regulation has also been advanced by scholars such as Sarah T. Roberts, who researches the labor processes of commercial content moderation. After conducting interviews with censors on social media sites, Roberts (2016) explained that these individuals who "act as digital gatekeepers for a platform" (p. 148) must balance competing demands such as free expression, profit motives, and brand protection when deciding on the acceptability of content. Roberts further described how the opacity of this process contributes to normalization of sexist, racist, and homophobic content on social media, a claim that will be considered in this article.

Studies that examine women's participation on the Internet reveal the deeply gendered nature of online interactions. Collectively, this scholarship demonstrates that although early visions of social media (and the Internet in general) were suffused with technological utopianism and views of an enriched public sphere (Papacharissi, 2002), neither technologies nor experiences of using the Internet are gender neutral or democratic (Dubrofsky & Magnet, 2015; Eckert, 2018; Noble, 2018; Turton-Turner, 2013). Instead, "misogynistic hostility . . . has become a *lingua franca* in many sectors of the cybersphere" (Jane, 2014, p. 558). The anonymity of the Internet and the perceived divide between online speech and off-line effects have fostered an explosion of sexist hashtags (Fox, Cruz, & Lee, 2015), gender-based trolling (Mantilla, 2015), rape threats (Filipovic, 2007; Hess, 2014), revenge porn (Abah, 2016), hate speech (Bernstein, 2014; Citron, 2014; Eckert, 2018; Massanari, 2017; Turton-Turner, 2013), and misogyny online (Banet-Weiser & Miltner, 2016; Jane, 2016). As aptly stated by researcher Emma Jane (2017), "while the Internet did not invent sexism, it *is* amplifying it in unprecedented ways" (p. 3). In combination, these sources reveal how pernicious Internet practices limit women's voices and agency, illustrating the intimate and mutually constitutive relationship between online interactions and offline marginalization.

Building on previous scholarship, this work makes several unique contributions. First, by interviewing women who have experienced censorship on Facebook, this article helps to shed light on how users conceptualize and respond to online content moderation. Second, the study's focus on enforcement as a locus of inequality and discrimination deepens existing accounts of sexism online and

provides further evidence of the ideological imperatives threaded throughout content moderation. Drawing from works that describe how technical features and the absence of protective policies on platforms hurt women (e.g., Massanari, 2017), this article argues that the way existing policies are administered may be problematic as well. In particular, Facebook's self-regulatory governance has given rise to enforcement asymmetry between men and women, as subjective content moderation decisions adjudicate not only questions of free speech but also social value. This article proposes the term "gender-based censorship" as a lens through which to understand women's experiences on Facebook. Gender-based censorship may be understood as encompassing explicit guidelines, implicit regulatory methods, and unregulated hate speech and threats against women, factors that are exacerbated by the unfettered control and guiding profit motive of social media companies.

This article seeks to answer three central questions:

- RQ1: How do women interface with Facebook's community standards, and what forms of censorship do they experience?*
- RQ2: How does the discriminatory enforcement of Facebook's community standards disenfranchise women and replicate societal inequities?*
- RQ3: What do women's interactions on Facebook reveal about the impact of self-regulation on the expressive potentiality of historically marginalized groups?*

Method and Theory

I chose to investigate Facebook for this case study because of the platform's vast size, its hypercurated nature (in contrast to more decentralized networks), and recent vociferous protests surrounding this site. Since 2016, Facebook has come under increased scrutiny for a series of scandals, including its possible facilitation of Russian interference in the U.S. 2016 presidential election, its privacy leaks, and its transgression of user rights and research ethics (e.g., Cambridge Analytica). Consequently, there have been calls for individuals to leave the platform as a means of protecting themselves and their data, affecting the company's bottom line (Dwoskin, 2018). Given the greater critical attention paid to this site, it is important to study the gendered nature of content moderation on the platform.

Interviews were selected as the data source for this study because of methodological limitations (caused by Facebook's lack of transparency) as well as my desire to honor the subjective experiences of users. Social media sites operate in an opaque manner, withholding details about the content they take down, the censors they employ (including demographics such as age, gender, and race), and the ideological directives they champion. As a result, it is necessary to rely on users to obtain information about the effects of content moderation. Interviews were also an appropriate method to offer users the opportunity to vocalize their experiences, which are often erased from the platform. My interviewing style was informed by feminist approaches of developing trust with interviewees, respecting their choices and narratives, and participating in the interview as opposed to adopting a stance of detached neutrality

(Oakley, 1981). With these guidelines in mind, I tried to cultivate a nonhierarchical relationship with interviewees and to let them codetermine the content and context of the interview.

Interviewing began with members of the Facebook Jailed community and "Men Are Scum" movement (to be described in the upcoming analysis). Chain referral sampling (Penrod, Preston, Cain, & Starks, 2003) was used to broaden the interview base as interviewees opened up their networks of friends with similar experiences on Facebook. Additionally, key informants (including researchers, lawyers, and individuals from nonprofit organizations) were contacted.³ In this type of purposive sampling, I targeted "individuals who are particularly knowledgeable about the issues under investigation" (Schutt, 2014, p. 171). Interviewing continued until the point of saturation (Small, 2009) was reached with the initial community of interest.

I conducted 13 interviews from January to February 2018 over the phone, through Skype, via FaceTime, through e-mail, or face-to-face, depending on the availability and preference of the interviewee.⁴ Interviews were audio recorded, uncompensated, conducted in English, and lasted approximately 35 minutes. Table 1 lists the pseudonym, occupation, location, self-identified gender, race/ethnicity, and age of participants as well as the format and date of the interview.⁵

Interview questions related to experiences of content removal on Facebook, awareness of community guidelines, and opinions about how content should be moderated on social media sites. Sample questions included the following: "How did you become involved with the 'Men Are Scum' movement?" "Have your posts ever been removed from Facebook?" "What is your understanding of Facebook's community standards?" Academics/activists were asked about the legal dimensions of content regulation, the complications of self-regulation by social media sites, and trends related to the flagging of specific content categories. Sample questions included the following: "How did you become involved in advocacy work related to Facebook's content moderation standards?" "Have you noticed any gender-based patterns in the reports you receive from users about hate speech?" After transcribing the interviews, I used open coding (by hand) to analyze the transcripts (Corbin & Strauss, 2008). Within open coding, a constant comparative method was used to make connections across the interviews. Coding topics consisted of censorial practices (i.e., unequal regulation and the proliferation of hate speech) and issues related to Facebook's operational style (such as lack of accountability, transparency, and accessibility).

³ Despite repeated attempts, I did not hear back from a Facebook representative regarding an interview, highlighting the way in which the site distances itself from users and thwarts efforts to promote transparency.

⁴ Influenced by feminist research methods (Oakley, 1981), I prioritized the comfort of interviewees by enabling them to select the interview mode, resulting in varied formats.

⁵ Although interviewees from several countries were included in this study, I do not make regional comparisons. The selection of interviewees was based primarily on connection to the community of interest and/or expertise in this topic. Future research is needed to tease apart how content moderation may be influenced by (and in turn influence) geographical contexts.

Table 1. Participants.

Interviewee	Gender	Race/ Ethnicity	Age	Occupation	Location	Interview Format	Date
<i>Kayla Avery</i> ⁶	Female	Caucasian	30s	Comedian/ Paralegal	United States	Phone call	1/26/18
<i>Tami Smith</i>	Female	Mixed race	30	Comedian	United States	Phone call	1/29/18
<i>Liz Wilson</i>	Female	Black	30	Comedian/ Activist	United States	Phone call	1/29/18
<i>Jessica Jones</i>	Female	Black	26	Comedian	United States	Phone call	1/29/18
<i>Abby Wood</i>	Female	Caucasian	31	Business owner	United States	Phone call	1/30/18
<i>Catherine Perry</i>	Gender fluid/ Female	Caucasian	39	Artist	Australia	Skype	1/30/18
<i>Rós Helgadóttir</i>	Female	Caucasian	37	Activist	Iceland	E-mail	1/31/18
<i>Maria Miller</i>	Female	Caucasian	36	Researcher/ Activist	Germany	Skype	2/8/18
<i>Olivia Lee</i>	Female	Caucasian	26	Photographer	Australia	FaceTime	2/8/18
<i>Sommer Cooper</i>	Female	Multiethnic/ Multiracial	52	Writer/ Activist	United States	Phone call	2/9/18
<i>Ryan Lewis</i>	Male	Caucasian	50s	Lawyer/ Professor	United States	Face-to- face	2/13/18
<i>Ava Gagnon</i>	Female	Caucasian	30s	Photographer/ Doula	Canada	Phone call	2/14/18
<i>Alana Johnson</i>	Female	Caucasian	29	Comedian	United States	Phone call	2/23/18

The article links the economic directive of Facebook, a publicly owned corporation, to Gramsci's theory of hegemony. This theory suggests that leaders enforce conformity to the existing social order by convincing individuals that their interests are prioritized while masking forces of domination and coercion (Gramsci, 1929/1995). As social media companies are primarily beholden to shareholders, they seek to selectively limit speech that may offend dominant interests (along with advertisers) and threaten profits. The desire to expunge such content maintains prevailing ideologies as the hegemonic viewpoint has been naturalized by social processes and may therefore appear less problematic to censors. The theory of hegemony is applicable to social media self-regulation because even though this structure upholds dominant interests at the expense of marginal viewpoints, it has won the support of users who report each other for posting certain content. These users participate in the system without fully realizing that social media's egalitarian appeals to self-regulation

⁶ Given Kayla's central role in the movement and her creation of Facebook Jailed, I decided (with permission) to forgo a pseudonym to give Kayla credit for her work (Bruckman, Luther, & Fiesler, 2015).

and equal speech obscure the subjective decisions that undemocratically appointed censors make regarding content.

This article also applies feminist theorizations of censorship. Although content removal based on Facebook's community guidelines may be considered an explicit form of censorship, implicit censorship (i.e., "powerful operations of censorship that are not based in explicit state policy or regulation"; Butler, 1998, p. 249) also occurs on Facebook. Implicit censorship, which "may be more efficacious than explicit forms in rendering certain kinds of speech unspeakable" (Butler, 1998, p. 250), is operative on Facebook in two ways. First, when content moderators make subjective, personalized decisions about permissible content, these choices are often guided by implicit biases and discriminatory beliefs that reflect the dominant social order; consequently, the expression of marginalized groups can be censored without a formal directive (revealed through the site's enforcement asymmetry). Second, hate speech and threats against women (which the site fails to adequately control) can be seen as another manifestation of implicit censorship. Hate speech that seeks to delegitimize women may be regarded as illocutionary disablement (Langton, 1998). Langton (1998) outlined the power dynamics undergirding illocutionary disablement: "One's power to do illocutionary things with words can sometimes be constrained and circumscribed by someone else's speech—that some speech builds a space that makes other speech possible and still other speech impossible" (p. 274). Illocutionary disablement is applicable to a gendered study of social media because Facebook's lack of protection for women perpetuates a climate of misogyny, hate speech, and threats, which renders women's speech less effective and undermines their expressive potential on the platform. Although women are also purveyors of hate speech, a recent survey revealed that they are twice as likely to be harassed online because of their gender and that they tend to regard such abuse as more problematic than men do (Duggan, 2017). Thus, Facebook's lax self-regulation and its deprioritization of fair standards for content moderation enable sexist speech and illocutionary disablement to flourish, disempowering women and silencing them.

Facebook Jailed, Community Standards, and the Perpetuation of Hegemonic Norms

This project was born when I came across Facebook Jailed (facebookjailed.com), a website dedicated to "exposing Facebook's double standard with regard to monitoring hate speech" (para. 1). Although the website's gallery of misogynistic posts (submitted by users) claims that the platform's monitoring process is deeply flawed and gendered, it is important to contextualize the creation of this site.

In October 2017, Alana, a feminist comedian and interviewee, became fed up with the relentless harassment and sexist speech directed at her friend on Facebook. After her friend posted 211 screenshots of sexist comments made by men (that were not removed by censors), Alana posted "Men Are Scum" on her friend's photo album. The comedian was subsequently reported by another user for perpetuating hate speech and was banned for 30 days from the platform. This punitive action stands in sharp contrast to Facebook's nonresponsive approach to users who continued to hurl sexist insults at Alana's friend even after she filed a series of complaints (revealing disparate enforcement). Rallying behind Alana, on November 24, 2017, 500 comedians (several of whom were interviewed for this project) posted variations of the phrase "Men Are Scum" on Facebook to protest the company's arbitrary censorship. Almost every one of these women experienced a ban from the site as a result (Lorenz, 2017). In response, Facebook Jailed was created by Kayla Avery to draw attention to this issue, which is ongoing, as another comedian was banned for a

similar post 10 months after Alana's initial remark (Harding, 2018), revealing the enduring nature of this concern.

As the creation of Facebook Jailed indicates, Facebook's enforcement of community standards is fraught with inconsistency and subjectivity, disproportionately affecting those who are socially marginalized. The site's hate speech policies and capricious application of guidelines reveal the problematic nature of social media self-regulation that weighs the expression of some groups over others. Facebook's censors monitor posts made against protected categories (i.e., on the basis of race, sex, gender, religion, nationality, ethnicity, sexual orientation, and disability) but do not block comments directed at subsets of protected categories (Angwin & Grassegger, 2017). For example, groups such as White men are protected because both race and gender are defended, but subgroups such as female drivers or Muslim children are not shielded from hate speech because one of their characteristics (i.e., drivers or children) is not protected (Angwin & Grassegger, 2017). Such a policy indicates Facebook's piecemeal rather than holistic approach to curbing hate speech and demonstrates a loophole by which those seeking to deliver hate speech can circumvent detection by censors. Further, because content moderation is executed overwhelmingly by humans (rather than by artificial intelligence), subjective opinions govern this process. Under such conditions, there is a tendency for conventional views to be upheld, as hegemonic norms "mark the boundaries of permissible discourse" (Lears, 1985, p. 570), rendering hate speech directed at dominant groups outside the bounds of acceptable discussion. On the other hand, hate speech directed at vulnerable subsets of users may not appear problematic to censors, leaving these groups with little protection or recourse and facilitating hateful language that may lead to illocutionary disablement.

When interviewed, writer/activist Sommer described how the sheer volume of posts may contribute to inconsistent content moderation:

They are processing millions and millions and millions of reports a day. They do it in 30-second increments, with very little context, with no jurisdictional clarity, with cultural and sociocultural issues that they cannot ever possibly take into account. (personal communication, February 9, 2018)

When censors are pressed to make quick decisions using vague guidelines, they are likely to rely on conventional views of social dynamics to avoid alienating dominant interests. Although moderators are provided with some level of sensitivity training (Levine, 2013), interviewee Alana expressed concern that "human biases are not really evaluated or checked" (personal communication, February 23, 2018). Thus, Facebook's self-regulation (which is relatively unsupervised to date) props up the hegemonic social order, which limits the expression, comfort, and safety of certain groups. As a result, this platform replicates problematic offline social dynamics, placing women at risk instead of adequately protecting them.

Unequal Enforcement: Adjudicating Speech, Adjudicating Social Value

Interviewees supporting the "Men Are Scum" movement expressed concerns that hate speech directed at men and women is adjudicated differently (especially for women who may be marginalized on

the basis of their race or other intersectional identifiers).⁷ Business owner Abby complained that the site consistently grants men latitude when making sexist comments about women, but restricts women who fire back in kind:

When the roles are reversed and you report something that is sexist and degrading toward women, it never gets taken down, and Facebook sends you a cute little message back saying we reviewed this, and it doesn't violate our standards. (personal communication, January 30, 2018)

This sense of inequity fuels resentment over Facebook's regulatory standards, which reinforce the hegemonic social structure by blocking criticisms of men. Selectively restricting the speech of women illustrates implicit (and gendered) censorship on the site.

Unequal enforcement is grounded in contemporary societal polarization. The #MeToo zeitgeist promoted open discussion of sexual abuse, efforts to hold harassers responsible, and highly visible (and viral) networked social movements. While #MeToo and other hashtags have given a powerful voice to feminist protest, there have also been competing calls to silence women as a means of maintaining men's dominant social position. To this point, several participants in this study believe Facebook's discriminatory approach provides evidence of a growing backlash against women. Comedian Jessica linked the censorship of women on Facebook to recent sociopolitical developments:

It seems like a very organized effort is happening, maybe in response to the political climate and #MeToo and men getting called out. I don't know what's going on, but it just seems like a woman can't write anything mean about a man, even as a joke, not any women I know, without it getting taken down. (personal communication, January 29, 2018)

In this manner, the subjectivity of the content moderation process may be harnessed to further the interests of the patriarchal social order, harming women who require equal access to speech platforms.

When evaluating the concerns of the "Men Are Scum" movement, it is important to consider the complexity of content moderation on a platform as widely used as Facebook. Because the surrounding context of a post is removed when sent to censors for review, moderators have difficulty interpreting humor or irony, especially as they race to delete an average of 288,000 posts a month that are deemed hate speech (Jan & Dwoskin, 2017). In an article featured in *The Daily Beast*, Taylor Lorenz (2017) explained, "When moderators can't make this distinction they punish innocent parties and embolden trolls" (para. 29). However, interviewees regarded power dynamics undergirding posts as an important element shaping the classification of hate speech.

Notably, participants did not conceptualize their comments as hate speech, but rather considered posts such as "Men Are Scum" to be a means of "punching up" and "critiquing the power structure" (Tami,

⁷ There have also been claims of racial bias in Facebook's content moderation (Jan & Dwoskin, 2017).

personal communication, January 29, 2018). Respondents insisted that their posts challenge the patriarchal order they have been socialized into: "There is not an equal balance of power between men and women for you to even consider it hate speech" (Kayla, personal communication, January 26, 2018). In this manner, they connected their posts to offline power dynamics, suggesting the site's refusal to consider societal hierarchies protects hegemonic interests and subordinates women. Alana asserted that the site's decontextualized approach fuels discrimination toward women because the platform has established that "misogyny and racism are equal to misandry and 'reverse-racism' and they're ignoring the power structure" (as cited in Reghay, 2017, para. 22). Although masking the context of posts is meant to protect users, women would benefit from content moderators evaluating their material in relation to other responses. This information would give greater insight into the possibility of women using such posts as a form of counterspeech and self-defense.

Several participants linked discriminatory censorship to the demographics of Facebook employees. In 2017, 65% of Facebook's employees worldwide were men ("Distribution of Facebook Employees Worldwide," 2018), and 89% of employees in the U.S. (where the company is based) were either White or Asian ("Distribution of Facebook Employees in the United States," 2018). Interviewees argued that stark imbalances in race and gender within the company affect the way Facebook conceptualizes and responds to hate speech and threats, as the platform has largely failed to consider the needs of women and the power dynamics surrounding their posts. Several participants stressed the importance of making the demographics of content moderators transparent, which are currently hidden by the platform. Facebook's decision to keep this information hidden is a strategic choice that enables the site to obscure the degree of editorial control it exercises and the processes it uses to filter content, which are at odds with its projected image as a vehicle of free speech (Gillespie, 2018). Concerned with the opaque nature of this site's content moderation system, comedian/paralegal Kayla stated, "If we don't know who is flagging and removing posts and censoring them, then how are we to trust them?" (personal communication, January 26, 2018). Because the demographics of Facebook's content moderators play a key role in shaping the gender dynamics of the platform and the way hate speech is evaluated, it is problematic that the site has not been forthcoming about this information, keeping users in the dark and perpetuating harmful practices.

Facebook as a Faceless Entity

A recurrent theme in the interviews was frustration about Facebook's lack of accountability, accessibility, and transparency. Although Facebook's mission statement is to "give people the power to build community and bring the world closer together" ("About: Business Info," n.d.), the platform distances itself from users, intentionally making itself unreachable.

Difficulty connecting to Facebook is particularly problematic when issues related to hate speech and censorship arise. Interviewees were exasperated about their inability to reach Facebook representatives when appealing a ban or gathering information about the moderation process. In fact, only two of the interviewees (the creator of Facebook Jailed and the originator of the "Men Are Scum" movement) were successful in getting a response from the site (likely because these individuals were able to drum up attention from the press). Australian photographer Olivia described Facebook's inaccessibility:

The thing is like you get banned on Facebook, you can't contact Facebook. So there is nobody to talk to. There is no human to talk to. You can't e-mail them. Like there is absolutely nothing; you're just banned. And once you're banned once, and you're banned again, your bans just keep adding. (personal communication, February 8, 2018)

That a site predicated on fostering social connection fails to respond to users in addressing their concerns is not only ironic but also troubling, as it exposes the profit-driven nature of the company. Ava, a Canadian photographer/doula, considered Facebook's lack of responsiveness to be "disheartening and discouraging" (personal communication, February 14, 2018). Interviewees, like Ava, who rely on Facebook for business, were especially frustrated when they could not reach the company to overturn their ban and were therefore unable to use the platform to contact clients. Although men may also experience difficulty reaching the platform, Facebook's lack of regard for women compounds existing biases on the site as well as societal power dynamics.

To this point, interviewees linked their positionality and the gendered nature of their concerns to Facebook's lack of responsiveness. Several women felt disregarded by the site and believed the platform's inaccessibility indicated a broader pattern, as reflected by the following statements: "I realized nothing that is racist or sexist against marginalized people matters to Facebook" (Jessica, personal communication, January 29, 2018); "I think the overarching theme is Facebook does not really seem to care at the end of the day about anyone who is not a White male" (Abby, personal communication, January 30, 2018); and "I think it's about time that Facebook starts listening to their female customers. Women have been pointing this problem out for years, and Facebook has always gotten away with ignoring them" (Rós, personal communication, January 31, 2018). Facebook's refusal to acknowledge the voice of women when they speak up about concerns is particularly troubling and constitutes an implicit form of gender-based censorship on this platform.

Facebook Weaponized by Misogynists: Underenforcement of Threats

In addition to enforcement asymmetry about hate speech, women are also discriminated against on the site through the underpolicing of threats (which are formally prohibited). The platform's lack of protection mirrors the long-standing disregard of crimes against women (Citron, 2014), highlighting how societal biases are woven throughout the very fabric of content moderation. Interviewees recalled countless stories of rape and death threats on Facebook and their inability to elicit a response from the site regarding these circumstances. After harassers photoshopped the face of Alana's acquaintance to look like she had been battered, she was unable to get the site to remove this disturbing image until she involved the press (personal communication, February 23, 2018). Similarly, Icelandic activist Rós described how she reposted screenshots of threats against women that were not taken off the platform. This outing of misogyny led to pushback from other users as Rós subsequently received death threats via personal phone calls and experienced self-reported symptoms of posttraumatic stress disorder. She declared, "I don't understand a world where men can get away with threatening to rape women, but women are punished for reposting such threats" (Rós, personal communication, January 31, 2018). This statement reveals a powerful inequity in which women are penalized for reporting what men said to them while men are protected for stepping outside the bounds of civil discourse. Such enforcement decisions illustrate a tendency to silence women who try to assert themselves, thereby preserving unequal gender relations and abusive power dynamics.

By failing to adequately respond to threats, Facebook implicitly sends the message that its values are aligned with those of misogynists, prioritizing their free speech over the safety of women on the platform. From the perspective of interviewees, censors not only harm women by suppressing their speech but also by deleting "warning posts about a specific person who is being predatory" (Kayla, personal communication, January 26, 2018). Removing these posts undermines community activism and endangers women. Although a friend of comedian/paralegal Kayla had been advised by local police to post a warning about a man who threatened her, this attempt to protect other women backfired as Facebook deleted the post and reportedly reversed her attempt to block the man from seeing her profile. The way the platform handled this incident directly contradicts its stated goal of facilitating community formation, indicting a contradiction between the lofty aspirations of the site and the lived reality of users.

Facebook's inaction with regard to threats against women may be linked to the site's difficulty evaluating credible threats as well as the gendered nature of its policies. In its review of Facebook's internal documents, *The Guardian* revealed that the platform did not consider the following posts to be credible threats: "To snap a bitch's neck, make sure to apply all your pressure to the middle of her throat," and "Little girl needs to keep to herself before daddy breaks her face" (Hopkins, 2017, para. 12, 20). While these posts do not threaten a specific person, it is notable that in both instances the targeted individual was clearly a woman, demonstrating the site's tolerance for violent rhetoric against women. It may be the case that threats against women are so commonplace (and thus naturalized) that these statements are not deemed problematic enough to be removed from the site. In this manner, offline tendencies to belittle women's experiences are replicated and exacerbated online.

As a result of the site's repeated failure to respond to complaints regarding inconsistencies, inequities, and threats, some women have resourcefully circumvented Facebook to force change. The creation of Facebook Jailed is one type of protest and workaround system where women found a way to harness the power of the Internet to rally against the abuses of social media platforms. In 2013, activists called attention to violent imagery and graphic language on Facebook by affecting the site's advertising profits, which totaled 84% of its global revenue at the time (Fuchs, 2014). Interviewee Sommer was involved in the 2013 petition for Facebook to adopt a stricter stance in blocking threats against women (Levine, 2013). After she "tried working with Facebook for about nine months politely and diligently and that went nowhere," Sommer initiated "a campaign that bypassed Facebook" (personal communication, February 9, 2018). Along with other activists, she sent companies screenshots of their ads, which were situated next to abusive imagery of women on the site. A photo featured in this campaign is shown below:

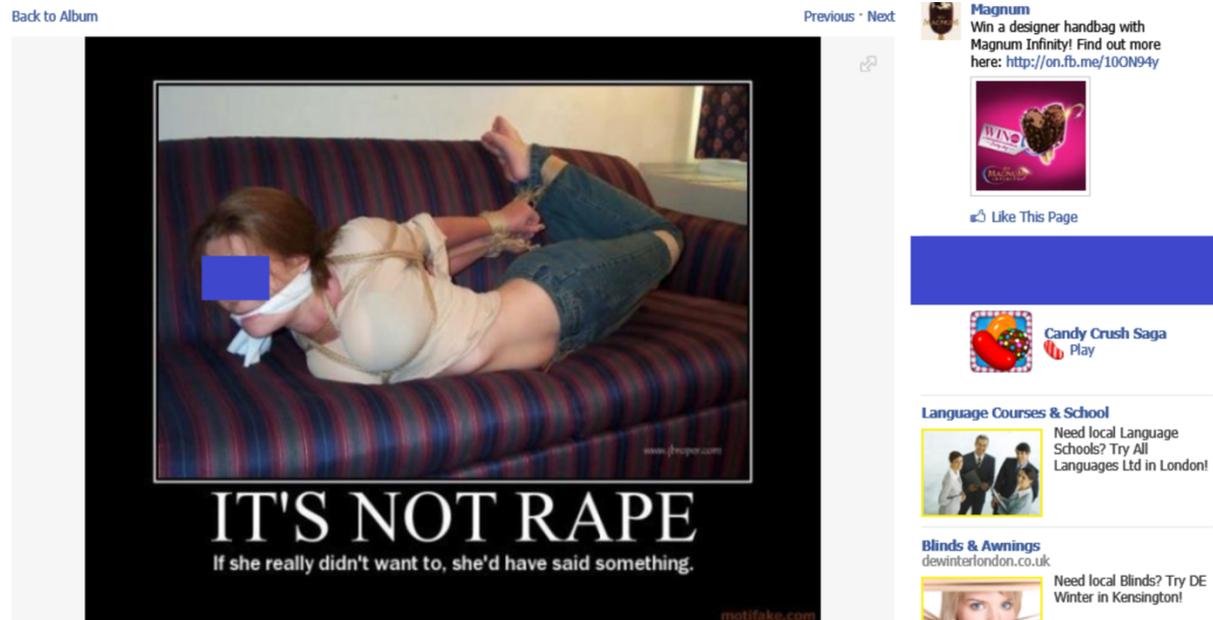


Figure 1. This image, which was sent to advertisers and companies, shows corporations' ads next to imagery promoting rape culture ("Examples of Gender-Based Hate Speech on Facebook," n.d.).

As advertising revenue dropped off (from companies such as Nissan), Facebook engaged in negotiations with consumer advocacy groups, leading to outcomes such as greater involvement by activists in the moderation process, increased training of censors, and more frequently updated community standards. However, Facebook's responsiveness to advertisers rather than users reveals the company's continued focus on its bottom line, demonstrating one of the structural problems of weakly enforced self-regulation. Although the 2013 campaign initially brought much needed reform, interviewees attest that sexism, threats, and degradation of women persist on the site. The continuation of Facebook's ineffectual response to gendered threats has led women to fear for their safety on the platform, as illustrated by the following statements: "Facebook is not a place that I can safely express myself on" (Jessica, personal communication, January 29, 2018), and "I, as a female, very rarely feel safe on Facebook" (Abby, personal communication, January 30, 2018). When Facebook refuses to adequately police threats against women, it implicitly sanctions the censorship of women and increases their vulnerability. Facebook's silence speaks volumes about the disparity between its publicly stated values (i.e., equality) and the priorities it demonstrates in practice (i.e., protection of dominant interests and groups).

Censorship and Speech: Silencing Voices

Facebook's censorship of women reveals and reinforces the gendered nature of speech allocation. In her interview, Sommer explained that self-regulation of social media content presupposes "some sort of utopian space where everybody has equal access to speech, which of course we know is not the case" (personal communication, February 9, 2018). According to Sommer and several other interviewees, the

idealistic rhetoric of egalitarian speech on social media sites masks power dynamics underlying these networks:

So to think that there isn't already socially constructed allocation of speech and its power is false. We know that speech allocations are socially constructed depending on where you stand in the hierarchy of power that is related to money or resources or technology. . . . Your speech is going to be more or less limited. (personal communication, February 9, 2018)

It should be noted that gender-based censorship is not a new problem for women but rather a long-standing pattern that predates the introduction of social media. Since ancient times, the right to speak and the ability to be heard have been predicated on gender as public spaces have been closed off to women, reinforcing their limited sphere of influence and subordinate social position (Lane, 2015).

Facebook's policies compound transhistorical inequities surrounding public speech. The site's toleration of hateful language against women operates as a mechanism of repression because "abuse and harassment diminish free speech" (Bernstein, 2014, p. 1). Threats against women and gendered hate speech on Facebook constitute a form of censorship and illocutionary disablement. When women are forced to worry about the implications of their statements and the harm they may receive, they are likely to alter their speech or refuse to speak altogether. In this manner, pernicious Internet practices are part of a broader tendency to censor women:

Gendertrolling, then, follows in the long historical tradition of attempting to control women's access to public and professional spaces, using abuse, harassment, and threats in an effort to silence women's speech and to intimidate women from fully participating in public discourse. (Mantilla, 2015, p. 141)

Therefore, Facebook's regulatory schema and inconsistent policing exacerbate an underlying and enduring pattern of blocking women's civic engagement, which is particularly troubling when one considers the clout, reach, and global presence of the platform.

Facebook's tendency to weigh voices unequally through biased content moderation decisions is of great concern in the current sociocultural climate. Comedian/activist Liz pointed out how public speech represents invaluable currency in the #MeToo moment: "As you see the pushback, you see who is getting silenced" (personal communication, January 29, 2018). Silencing thus operates as a way to perpetuate the status quo, hegemonic norms, and entrenched gender dynamics. Muting women's voices and stifling their ability to press for change render women especially vulnerable to acts of violence, sexual aggression, and economic disparities. Comedian Alana spoke to this point: "You're censoring the people who need that platform the most and giving more of a voice to the people who already have it" (personal communication, February 23, 2018). Consequently, Facebook's problematic regulatory practices are weaponized to subordinate women and deny their political agency, thereby reinforcing the dominant social order.

Conclusion

This case study of gender-based censorship on Facebook reveals some of the fundamental flaws of commercial social media that can never be neutral, equitable, or nondiscriminatory, particularly when left to self-regulate. Motivated by profit generation, platforms are incentivized to segment users and treat them differently based on their perceived social or economic value to the site. This type of segmenting and disparate treatment leaves historically marginalized groups vulnerable to the prejudicial policies and administrative strategies of social media sites. Although platforms nominally promise equal treatment, there is a fundamental disjuncture between the way policies are written and how they are implemented (and experienced). Enforcement asymmetry between men and women and the nonpolicing of threats against women on Facebook represent loci of inequality that while currently underexplored in the existing literature are impactful and consequential. Therefore, it is important to not only analyze explicit policies and the speech they prohibit but also to evaluate how these policies are executed and how gender-based censorship is produced, parsing out hidden layers of bias and discrimination endemic to social media.

In the current moment, there is great potential for women to call out inequities and force changes on social media platforms. The 2013 campaign to send screenshots to advertisers and the creation of Facebook Jailed provide evidence of women coming together and collectively protesting platform policies. These occurrences demonstrate how "online spaces remain a double-edged sword for women" (Eckert, 2018, p. 1284) as social media sites leave women vulnerable to threats, harassment, and hate speech while facilitating group membership and collective action. As the plurality of Facebook users (Nakamura, 2015), women gross substantial revenue for the platform through their regular exposure to advertisements. Given their collective clout, women can engage in communal efforts to document and protest both sexist speech on the site and gender-biased enforcement procedures. Women may also opt to leave Facebook (as well as other commercial social media sites) and/or create alternative (noncommercial) platforms that better honor the needs of users by prioritizing equal access to communication over profit generation. However, because commercial social media sites have become embedded in everyday life, women and other marginalized groups "are the least able to 'quit Facebook'" (Nakamura, 2015, p. 223). Consequently, even when they are aware of inequitable practices, women may choose to protest sites from within rather than leave them altogether.

In addition to continued pressure on platforms through activism, external oversight should be explored as a necessary alternative to self-regulation. To date, the lack of outside regulation has exacerbated many underlying problems, as social media sites have been given a free pass to create arbitrary standards, to adjudicate them unfairly and inconsistently, and to mask their operations from both the public and lawmakers. The capitalist directive of platforms explains their lack of incentive to enact change, making it unlikely that substantial adjustments will be made by sites themselves. Consequently, there is a great need to impose meaningful regulations to curb some of the excesses of weakly enforced self-regulation that serves to reinforce gender disparities, social biases, and hegemonic worldviews (Nurik, 2018). The combined approach of ongoing communal activism and legislative reform is therefore essential to reel in the unchecked power of platforms so that women are protected from online abuse, systematic silencing, and being "jailed" for defending themselves.

References

- Abah, A. L. (2016). Legal regulation of CSR: The case of social media and gender-based harassment. *University of Baltimore Journal of Media Law & Ethics*, 5, 38–55.
- About: Business info. (n.d.). *Facebook*. Retrieved from https://www.facebook.com/pg/facebook/about/?ref=page_internal
- Ammori, M. (2014). The “new” *New York Times*: Free speech lawyering in the age of Google and Twitter. *Harvard Law Review*, 127, 2259–2295.
- Angwin, J., & Grassegger, H. (2017, June 28). Facebook’s secret censorship rules protect White men from hate speech but not Black children. *ProPublica*. Retrieved from <https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms>
- Banet-Weiser, S., & Miltner, K. M. (2016). #MasculinitySoFragile: Culture, structure, and networked misogyny. *Feminist Media Studies*, 16(1), 171–174.
- Beirne, A. (2018, December 27). 5 takeaways from Facebook’s leaked moderation documents. *The New York Times*. Retrieved from <https://www.nytimes.com/2018/12/27/world/facebook-moderators-takeaways.html>
- Bernstein, A. (2014). Abuse and harassment diminish free speech. *Pace Law Review*, 35, 1–29.
- Bruckman, A., Luther, K., & Fiesler, C. (2015). When should we use real names in published accounts of Internet research? In E. Hargittai & C. Sandvig (Eds.), *Digital research confidential: The secrets of studying behavior online* (pp. 243–258). Cambridge, MA: MIT Press.
- Butler, J. (1998). Ruled out: Vocabularies of the censor. In R. Post (Ed.), *Censorship and silencing: Practices of cultural regulation* (pp. 247–260). Los Angeles, CA: The Getty Research Institute for the History of Art and the Humanities.
- Citron, D. K. (2014). *Hate crimes in cyberspace*. Cambridge, MA: Harvard University Press.
- Communications Decency Act, 47 U.S.C. §230 (1996).
- Community Standards. (n.d.). *Facebook*. Retrieved from <https://www.facebook.com/communitystandards>
- Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Thousand Oaks, CA: SAGE Publications.

Crawford, K., & Gillespie, T. (2016). What is a flag for? Social media reporting tools and the vocabulary of complaint. *New Media & Society, 18*(3), 410–428.

Distribution of Facebook employees in the United States from 2014 to 2017, by ethnicity. (2018). *Statista*. Retrieved from <https://www.statista.com/statistics/311847/facebook-employee-ethnicity-us/>

Distribution of Facebook employees worldwide from 2014 to 2017, by gender. (2018). *Statista*. Retrieved from <https://www.statista.com/statistics/311827/facebook-employee-gender-global/>

Dubrofsky, R. E., & Magnet, S. A. (Eds.). (2015). *Feminist surveillance studies*. Durham, NC: Duke University Press.

Duggan, M. (2017, July 14). Men, women experience and view online harassment differently. *Pew Research Center*. Retrieved from <http://www.pewresearch.org/fact-tank/2017/07/14/men-women-experience-and-view-online-harassment-differently/>

Dwoskin, E. (2018, November 13). Quitting Instagram: She's one of the millions disillusioned with social media. But she also helped create it. *The Washington Post*. Retrieved from https://www.washingtonpost.com/technology/2018/11/14/quitting-instagram-shes-one-millions-disillusioned-with-social-media-she-also-helped-create-it/?noredirect=on&utm_term=.1290b8ffe07d

Eckert, S. (2018). Fighting for recognition: Online abuse of women bloggers in Germany, Switzerland, the United Kingdom, and the United States. *New Media & Society, 20*(4), 1282–1302.

Examples of gender-based hate speech on Facebook. (n.d.). *Women, Action, & The Media*. Retrieved from <http://womenactionmedia.org/examples-of-gender-based-hate-speech-on-facebook/>

Filipovic, J. (2007). Blogging while female: How Internet misogyny parallels real-world harassment. *Yale Journal of Law & Feminism, 19*, 295–304.

Fox, J., Cruz, C., & Lee, J. Y. (2015). Perpetuating online sexism offline: Anonymity, interactivity, and the effects of sexist hashtags on social media. *Computers in Human Behavior, 52*, 436–442.

Fuchs, C. (2014). *Social media: A critical introduction*. London, UK: SAGE Publications.

Germany starts enforcing hate speech law. (2018, January 1). *BBC*. Retrieved from <https://www.bbc.com/news/technology-42510868>

Gil de Zúñiga, H., Jung, N., & Valenzuela, S. (2012). Social media use for news and individuals' social capital, civic engagement and political participation. *Journal of Computer-Mediated Communication, 17*(3), 319–336.

Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. New Haven, CT: Yale University Press.

Gramsci, A. (1995). *Further selections from the prison notebooks*. Minneapolis, MN: University of Minnesota Press. (Original work published 1929)

Harassment. (n.d.). *Facebook*. Retrieved from <https://www.facebook.com/communitystandards/harassment>

Harding, X. (2018, August 20). Here's why Facebook is quick to ban users who say "men are trash," but took longer to ban Alex Jones. *Mic*. Retrieved from <https://mic.com/articles/190835/heres-why-facebook-is-quick-to-ban-users-who-say-men-are-trash-but-took-longer-to-ban-alex-jones#.SPEE5K7rl>

Hate speech. (n.d.). *Facebook*. Retrieved from https://www.facebook.com/communitystandards/hate_speech

Helft, M. (2010, December 13). Facebook wrestles with free speech and civility. *The New York Times*. Retrieved from <http://www.nytimes.com/2010/12/13/technology/13facebook.html>

Hess, A. (2014, January 6). Why women aren't welcome on the Internet. *Pacific Standard*. Retrieved from <https://psmag.com/social-justice/women-arent-welcome-internet-72170>

Hopkins, N. (2017, May 21). Revealed: Facebook's internal rulebook on sex, terrorism and violence. *The Guardian*. Retrieved from <https://www.theguardian.com/news/2017/may/21/revealed-facebook-internal-rulebook-sex-terrorism-violence>

Ingram, D. (2018, April 18). Exclusive: Facebook to put 1.5 billion users out of reach of new EU privacy law. *Reuters*. Retrieved from <https://www.reuters.com/article/us-facebook-privacy-eu-exclusive/exclusive-facebook-to-put-1-5-billion-users-out-of-reach-of-new-eu-privacy-law-idUSKBN1HQ00P>

Jan, T., & Dwoskin, E. (2017, July 31). A White man called her kids the n-word. Facebook stopped her from sharing it. *The Washington Post*. Retrieved from https://www.washingtonpost.com/business/economy/for-facebook-erasing-hate-speech-proves-a-daunting-challenge/2017/07/31/922d9bc6-6e3b-11e7-9c15-177740635e83_story.html?utm_term=.9e7941e033d4

Jane, E. A. (2014). "Back to the kitchen, cunt": Speaking the unspeakable about online misogyny. *Continuum*, 28(4), 558–570.

Jane, E. A. (2016). Online misogyny and feminist digilantism. *Continuum*, 30(3), 284–297.

- Jane, E.A. (2017). *Misogyny online: A short (and brutish) history*. Thousand Oaks, CA: SAGE Publications.
- Klonick, K. (2018). The new governors: The people, rules, and processes governing online speech. *Harvard Law Review*, 131, 1598–1670.
- Lane, L. (2015). Feminist rhetoric in the digital sphere: Digital interventions & the subversion of gendered cultural scripts. *Ada: A Journal of Gender, New Media, and Technology*, 8, 1–18.
- Langton, R. (1998). Subordination, silence, and pornography's authority. In R. Post (Ed.), *Censorship and silencing: Practices of cultural regulation* (pp. 261–284). Los Angeles, CA: The Getty Research Institute for the History of Art and the Humanities.
- Lears, T. J. (1985). The concept of cultural hegemony: Problems and possibilities. *The American Historical Review*, 90(3), 567–593.
- Levine, M. (2013, May 28). *Controversial, harmful and hateful speech on Facebook*. Retrieved from <https://www.facebook.com/notes/facebook-safety/controversial-harmful-and-hateful-speech-on-facebook/574430655911054/>
- Lorenz, T. (2017, December 4). Facebook is banning women for calling men "scum." *The Daily Beast*. Retrieved from <https://www.thedailybeast.com/women-are-getting-banned-from-facebook-for-calling-men-scum>
- Madrigal, A. (2018, February 7). Inside Facebook's fast-growing content-moderation effort. *The Atlantic*. Retrieved from <https://www.theatlantic.com/technology/archive/2018/02/what-facebook-told-insiders-about-how-it-moderates-posts/552632/>
- Mantilla, K. (2015). *Gendertrolling: How misogyny went viral*. Santa Barbara, CA: Praeger.
- Massanari, A. (2017). #Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. *New Media & Society*, 19(3), 329–346.
- Nakamura, L. (2015). Afterword: Blaming, shaming, and the feminization of social media. In R. E. Dubrofsky & S. A. Magnet (Eds.), *Feminist surveillance studies* (pp. 221–228). Durham, NC: Duke University Press.
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York, NY: New York University Press.
- Nurik, C. (2018). 50 shades of film censorship: Gender bias from the Hays Code to MPAA ratings. *Communication Culture & Critique*, 11(4), 530–547.

- Oakley, A. (1981). Interviewing women: A contradiction in terms. In H. Roberts (Ed.), *Doing feminist research* (pp. 30–61). New York, NY: Routledge.
- Papacharissi, Z. (2002). The virtual sphere: The Internet as a public sphere. *New Media & Society*, 4(1), 9–27.
- Penrod, J., Preston, D. B., Cain, R. E., & Starks, M. T. (2003). A discussion of chain referral as a method of sampling hard-to-reach populations. *Journal of Transcultural Nursing*, 14(2), 100–107.
- Reghay, N. (2017, November 20). Are jokes about men “hate speech”? Facebook seems to think so. *The Daily Dot*. Retrieved from <https://www.dailydot.com/irl/facebook-women-hate-speech/>
- Roberts, S. T. (2016). Commercial content moderation: Digital laborers’ dirty work. In S. U. Noble & B. M. Tynes (Eds.), *The intersectional Internet: Race, sex, class, and culture online* (pp. 147–160). New York, NY: Peter Lang International Academic Publishers.
- Schutt, R. K. (2014). *Investigating the social world: The process and practice of research*. Thousand Oaks, CA: SAGE Publications.
- Small, M. L. (2009). “How many cases do I need?” On science and the logic of case selection in field-based research. *Ethnography*, 10(1), 5–38.
- Turton-Turner, P. (2013). Villainous avatars: The visual semiotics of misogyny and free speech in cyberspace. *Forum on Public Policy: A Journal of the Oxford Round Table*, 1, 1–18.
- Tushnet, R. (2008). Power without responsibility: Intermediaries and the First Amendment. *George Washington Law Review*, 76, 986–1016.
- Wagner, K., & Swisher, K. (2017, February 16). Read Mark Zuckerberg’s full 6,000-word letter on Facebook’s global ambitions. *Recode*. Retrieved from <https://www.recode.net/2017/2/16/14640460/mark-zuckerberg-facebook-manifesto-letter>