Sex Roles
https://doi.org/10.1007/s11199-019-01095-z

ORIGINAL ARTICLE

Sexist Slurs: Reinforcing Feminine Stereotypes Online

Diane Felmlee[1] · Paulina Inara Rodis[1] · Amy Zhang[2]

Abstract

Social media platforms are accused repeatedly of creating environments in which women are bullied and harassed. We argue that online aggression toward women aims to reinforce traditional feminine norms and stereotypes. In a mixed methods study, we find that this type of aggression on Twitter is common and extensive and that it can spread far beyond the original target. We locate over 2.9 million tweets in one week that contain instances of gendered insults (e.g., "bitch," "cunt," "slut," or "whore")—averaging 419,000 sexist slurs per day. The vast majority of these tweets are negative in sentiment. We analyze the social networks of the conversations that ensue in several cases and demonstrate how the use of "replies," "retweets," and "likes" can further victimize a target. Additionally, we develop a sentiment classifier that we use in a regression analysis to compare the negativity of sexist messages. We find that words in a message that reinforce feminine stereotypes inflate the negative sentiment of tweets to a significant and sizeable degree. These terms include those insulting someone's appearance (e.g., "ugly"), intellect (e.g., "stupid"), sexual experience (e.g., "promiscuous"), mental stability (e.g., "crazy"), and age ("old"). Messages enforcing beauty norms tend to be particularly negative. In sum, hostile, sexist tweets are strategic in nature. They aim to promote traditional, cultural beliefs about femininity, such as beauty ideals, and they shame victims by accusing them of falling short of these standards.

Keywords Stereotypes · Social media · Harassment · Beauty ideals · Social networks · Victimization · Online aggression · Hostility toward women

Harassment on social media constitutes an everyday, routine occurrence, with researchers finding 9,764,583 messages referencing bullying on Twitter over the span of two years (Bellmore et al. 2015). In other words, Twitter users post over 13,000 bullying-related messages on a daily basis. Forms of online aggression also carry with them serious, negative consequences. Repeated research documents that bullying victims suffer from a host of deleterious outcomes, such as low self-esteem (Hinduja and Patchin 2010), emotional and psychological distress (Ybarra et al. 2006), and negative emotions (Faris and Felmlee 2014; Juvonen and Gross 2008). Compared to those who have not been attacked, victims also tend to report more incidents of suicide ideation and attempted suicide (Hinduja and Patchin 2010). Several studies document that the targets of cyberbullying are disproportionately women (Backe et al. 2018; Felmlee and Faris 2016; Hinduja and Patchin 2010; Pew Research Center 2017), although there are exceptions depending on definitions and venues. Yet, we know little about the content or pattern of cyber aggression directed toward women in online forums. The purpose of the present research, therefore, is to examine in detail the practice of aggressive messaging that targets women and femininity within the social media venue of Twitter. Using both qualitative and quantitative analyses, we investigate the role of gender norm regulation in these patterns of cyber aggression.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11199-019-01095-z) contains supplementary material, which is available to authorized users.

✉ Diane Felmlee
dhf12@psu.edu

[1] Department of Sociology and Criminology, Pennsylvania State University, 206 Oswald Tower, University Park, PA 16802, USA

[2] Department of Statistics, Pennsylvania State University, University Park, PA 16802, USA

Cyber Aggression

Bullying represents a broad societal problem (Beauchere 2014; Faris and Felmlee 2011, 2014; Miller 2016) that now extends well beyond face-to-face interaction and instead tracks down its targets through their electronic devices (Xu

Published online: 28 November 2019

et al. 2012). *Cyber aggression*, which here refers to intentional electronic communication intended to insult or harm an individual, remains particularly problematic. This electronic form of aggression can provide perpetrators with a sense of anonymity, which makes it easier to engage in harmful communication without the fear of direct retribution (Bartlett et al. 2018). Forms of online harassment also carry the serious risk that an attack can spread far beyond the original incident, potentially reaching a broad network of social media users and multiplying the embarrassment and harm experienced by the victim.

Twitter represents a well-known and popular forum for online communication in today's world in which bullying occurs regularly (Xu et al. 2012). *Twitter* is a micro-blogging service in which people communicate online, using short messages or "tweets." In 2017, there were over 330 million active monthly Twitter users, with approximately 24% of online adults (or 21% of all Americans) using the site in 2016 (Pew Research Center 2016). However, this social media giant has come under repeated scrutiny due to its widely exposed public cases of bullying and harassment.

A group of women initiated a Twitter boycott (#WomenBoycottTwitter; https://www.cbsnews.com/news/harvey-weinstein-rose-mcgowan-women-boycott-twitter/) in 2017 that was triggered, in part, by the suspension of the Twitter account of actress Rose McGowan, who had posted allegations of sexual abuse by producer Harvey Weinstein. Following on the heels of the Women's Boycott, Twitter announced that they would begin implementing new "anti-abuse" rules meant to stem the prevalence of hate symbols, unwanted sexual advances, and other problematic or abusive content, and they introduced such changes in subsequent months. It remains to be seen how successful these alterations will be in stemming the tide of online abuse.

Drawing on theories of group processes (Homans 1950; Simmel 1950), we argue that two fundamental social mechanisms contribute to the development of cyber aggression: (a) the enforcement of social norms and (b) the establishment of social hierarchies (see also Faris and Felmlee 2011; Felmlee and Faris 2016). The first mechanism refers to the regulation of gender norms whereby demeaning online messages highlight the social expectations, or norms, that surround gendered behavior and reinforce traditional stereotypes. This type of gender norm enforcement is evident in multiple types of online discussions, such as those involving homophobic epithets (Pascoe and Diefendorf 2019) and hate speech (Wilhelm and Joeckel 2019). In addition, lesbian, gay, bisexual, intersex, transsexual, and others with non-traditional sexual and/or gender identities tend to be attacked online at a much higher rate (two to four times higher) than their heterosexual counterparts (Felmlee and Faris 2016; Hinduja and Patchin 2010; Schneider et al. 2012). These frequent attacks on individuals who challenge the norms of heterosexuality further demonstrate attempts at gender regulation (Hlavka 2014). Finally, tweets that insult women of color with messages that emphasize gendered racial/ethnic stereotypes also are common (Felmlee et al. 2018) and exhibit both sexist and racist norm enforcement.

A second mechanism involved in the development of sexist online aggression concerns social status. The maintenance of status represents a particularly powerful dimension of a range of human interaction, according to Ridgeway (2011) and others, and the motivation to increase one's social status emerges in forms of harassment and bullying (Miller 2016; Sijtsema et al. 2009). For instance, adolescents who are somewhat central within their peer groups engage in particularly high levels of social aggression, as compared to those already established at the top of the social hierarchy (Faris and Felmlee 2011, 2014). Perpetrators of derogatory Twitter communications toward women may believe, correctly or not, that engaging in such behavior will garner greater respect and status among their close associates or, at the very least, earn attention.

We argue here that aggressive tweets that target women attempt to reinforce traditional gendered norms, especially those relating to idealized forms of femininity. Furthermore, people disseminating such hostile messages are likely to be motivated by the goal of improving their social esteem and recognition among their group of supporters or a wider societal audience. By underscoring traditional gender expectations, perpetrators likely anticipate positive reverberations to their messages among those who continue to perpetuate long-held stereotypes of feminine behavior. Sexist online harassment does not occur at random, in other words, but it is strategic in nature and aimed at reinforcing gender inequality.

Gender and Feminine Stereotypes

Feminist theories suggest that hostility toward women arises within the maintenance of a patriarchal culture and its accompanying attitudes of sexism, misogyny, and objectification of women (e.g., Jeffreys 2005). Scholars note further that hostility emanates from women's oppression within larger societal structures of power that place men in the dominant position and women in the role of the subordinate. Both of these cultural and structural processes likely contribute to the flourishing of antagonism toward women more broadly (Hollander et al. 2011). Institutional sexism creates the social situations in which gendered stereotypes are established, and, in return, the perpetuation of stereotypes preserves and reinforces the social culture of sexism and dominant masculinity. At an individual level, these stereotypes can be used to reinforce and verify one's own position regarding acceptable norms or to increase one's social status and further denigrate others' non-normative behavior.

One prominent example of a traditional feminine norm that evolves from these processes is that women and girls should be physically attractive. This gender role expectation derives from Western beauty ideals directed toward women. Female beauty ideals (e.g., facial symmetry and appeal; BMI or level of thinness, height, and body shape) convey powerful messages about the necessity and consequences of physical attractiveness to young girls and adults alike (Bailey et al. 2013; Jeffreys 2005; Wolf 2002). Research illustrates that individual support of beauty ideals is highly related to sexist attitudes as well as to acts of hostility toward women (Forbes et al. 2007; Swami et al. 2010). Furthermore, the presentation of an appropriately "gendered self," with just the right amount of feminine appearance and sexual appeal is key to a positive and socially engaging online presence (Bailey et al. 2013).

The classic feminine personality stereotype involves two major components: one associated with warmth and being nice and the other with a lack of competence (e.g., passive, emotional, illogical) (Ellemers 2018). This overarching idealized femininity comprises traits typically believed to be the opposite of those associated with masculinity (Ellemers 2018). Other feminine stereotypes noted in the literature include those of being more family- or community-oriented and embodying a sense of social purity. Rather than having selfish intentions or actions, therefore, women are expected to be moral paragons and keepers of virtue, which also extends to requirements that they be "virginal" or relatively sexually inexperienced (Valenti 2009). Thus, central themes in gendered norms regarding women in our society revolve around beauty as well as niceness, morality, and sexual inexperience.

In the present paper, we argue that acts of aggressive behavior oriented toward women on Twitter tap into many of these same core themes in feminine stereotypes. Such behavior could reflect either explicit or implicit gender biases. Harassers reinforce traditional stereotypes, consciously or not, by attempting to shame women with labels that counter these normative expectations. Several typical labels include the word "bitch" (i.e., "a malicious, spiteful, or overbearing woman"), the terms "slut" and "whore" (i.e., offensive terms for a "promiscuous woman"), and the word "cunt" (i.e., a "disparaging, and obscene" reference to a woman) (https://www.merriam-webster.com/dictionary/). Calling a woman a "bitch" in an aggressive message, for example, implies that she is not conforming to the expectations that she should be "sugar and spice and all things nice." Identifying a woman as a "whore" or "slut" means that she falls far out of line with the norm of sexual inexperience. Labelling a woman a "cunt," on the other hand—one of the most hateful words in the English language according to one urban dictionary—implies that she has absolutely no redeeming features, feminine or otherwise.

Research Goals

The purpose of our research is to investigate instances of aggression oriented toward women on Twitter and to examine the regulation of traditional feminine norms in these messages. One of our goals is to highlight online hostility toward women as a social problem and to bring to light several of its characteristics. At the same time, we do not claim that online aggression solely targets women, as opposed to men, transgender individuals, and others. In such cases, the insulting terms often differ from those oriented toward women (e.g., bitch, slut), however, and consist instead of homophobic slurs (e.g., fag, homo) (Sterner and Felmlee 2017). Men vilified online appear to be labelled a "fag," for example, more frequently than they are called a "slut" or "cunt." On the other hand, even if a man or a transgender person is the target of a tweet using a female-oriented slur, the use of terms and messages enforcing and regulating stereotypical femininity itself highlights a broader phenomenon of sexism.

Note, too, that we are not able to identify the demographic characteristics of the perpetrators—those who send hostile tweets in our data set could be men, women, or those of other gender identities. Men generally engage in aggressive behavior toward women more frequently than toward men (Anderson and Anderson 2008), and yet women, who operate within the same dominant culture, also can exhibit hostility toward other women (Loya et al. 2006; Wilhelm and Joeckel 2019).

In a mixed methods format, we begin by describing the frequency with which tweets occur that use one of four key insulting female terms, and we provide several illustrations of messages that include one or more of these keywords. We also note that a unique quality of online aggression is that a hostile message can spread readily beyond the initial episode to involve multiple individuals and even extend to reach very large groups. Thus, we illustrate the spread of cyber aggression by investigating several case studies of networks of conversations that follow from a hostile tweet. (We display graphs of such networks in an online supplement; see S1, Figs. 1s and 2s.)

In our final set of analyses, we use sentiment analysis to investigate the degree of negativity of tweets that contain our key terms and those that include common negative adjectives that challenge femininity (e.g., ugly, stupid, fat, skinny, crazy, old, promiscuous). We expect that insulting adjectives will increase the overall negativity of the tweets, and we explore whether those that contradict ideal beauty standards (e.g., overall appearance: ugly, weight-related: skinny or fat) will be particularly negative in sentiment. We use multivariate regression analysis to investigate these issues.

Method

Procedure and Measures

We collected data from Twitter for approximately 2½ years, beginning in 2016 and through the beginning of 2019. During much of this time, users were limited to messages with 140 characters or fewer (prior to the change in Twitter policies to allow users to write longer messages). Institutional Review Board (IRB) approval was obtained prior to data collection.

We gathered Twitter messages in two phases. First, we utilized keyword searches in NodeXL (Smith et al. 2010) to scrape recent tweets containing specified terms. From these searches, a research team spent 2½ years exploring patterns of sexist and misogynistic language on Twitter, following both current events and through the exploration of specific insults and gendered slurs (Baker et al. 2019; Lawson et al. 2017; Shartle et al. 2016; Stumm et al. 2016). During this process, the authors compiled a list of commonly used insults that made use of negative concepts and stereotypes directed toward women on Twitter. In total, these searches collected over 50,000 tweets, and the messages encompassed a broad variety of gendered derogatory language (e.g., "butterface": slang for a woman described as attractive except for her face); "ho[e]": slang for "whore"). For this project, we focus on the four terms ("bitch," "cunt," "slut," and "whore") that were the most common. These four words also represent frequent curse words on Twitter, with "bitch" ranked number four and the others within the top 20 of all types of curse words used (Wang et al. 2014).

In our second phase of data collection, we scraped data directly from Twitter's Streaming API. The scraped data set was collected in four separate streams filtered to find tweets including at least one of the four key terms we noted. Gathering data directly from Twitter allowed us to collect tweets as they were published on the platform. The dataset was collected over a period of one week in June 2017 (June 4–June 11). We chose a week during the year in which there was no major U.S. holiday nor was there an obvious political or news event that might have dominated the searches for sexist slurs. Next, we processed the tweets for greater analysis efficiency and anonymity, replacing Twitter handles with "@USER#," for example, and website links with "URL." The tweets were further filtered to remove those by accounts that were highly likely to be bots, that is, tweets spread by automated software (Davis et al. 2016). After processing, the direct streams collected 2.9 million tweets across four categories, each downloading tweets with one of the four terms. Of the total sample, 87% came from the stream collecting the keyword, "bitch" (which is equivalent to 2,530,832 tweets), 5% "cunt" (155,059 tweets), 4% "slut" (131,155 tweets), and 4% "whore" (110,307 tweets).

Reinforcing Feminine Stereotypes

Along with searching for tweets that contained the four key derogatory terms, we examined the degree of negativity of tweets that appeared to endorse feminine stereotypes through the inclusion of derogatory adjectives. In total, we targeted seven groups of adjectives that were synonyms of, or closely related to, the following terms: ugly, overweight, stupid, underweight, crazy, old, and promiscuous. Each of these adjective groups was chosen to highlight insults that focus on a different stereotypical aspect of femininity. The terms ugly, overweight, and underweight (and their synonyms), for example, can be used to emphasize traditional feminine appearance stereotypes suggesting that a woman should be beautiful and slender (slim, but not too skinny). The adjectives stupid and crazy raise concerns that a woman is either not intelligent or is too irrational and emotional (as compared to the rationality and intelligence of men). Finally, the terms old and promiscuous contradict two ideals of femininity, those of youth and sexual inexperience. Although these categories of adjectives do not necessarily represent an exhaustive list of adjectives associated with feminine stereotypes, they were chosen because they relate to several prominent themes in the stereotype literature (e.g., Ellemers 2018).

Re-appropriation

One issue we confronted in our sentiment analysis concerned the re-appropriation, or reclaiming, of certain negative terms by women themselves. In reclaiming a word, oppressed groups attempt to make a formally pejorative word used by the dominant culture more acceptable (e.g., Bianchi 2014). The term "bitch," for example, can be used in positive ways between women in a joking manner or as a way to take the sting out of the word more broadly. In our individual searches, we could avoid positive uses of our keywords and adjectives, but this was not always the case for our sentiment analysis of millions of tweets. One of the reasons for discrepancies between human coders and our classifier in the final sentiment score of tweets was the occasional practice of re-appropriation of negative words. Nevertheless, our classifier outperformed many common approaches, as we discuss in the following.

Social Network Analysis

We examine the social network of Twitter "conversations" that develop out of messages that contain our keywords, with a focus on tweets located within a network of users. We investigate the network spread that results from a negative tweet, in which the network consists of "retweets," "replies," and "likes" that arise following an initial insulting tweet. Next, we examine the social roles involved in the Twitter interchange in our smaller networks and identify those of the "perpetrator," the "victim," the "reinforcer," and the "defender." We also note when there appears to be a "bystander," which refers to someone who was involved in a Twitter conversation prior to the derogatory tweet, but then who does not respond in either a supportive or a critical manner. We chose examples that represent the use of each of our key terms and that illustrate the various roles in which individuals can participate during an online Twitter interchange. (In an online supplement, S1, Figs. 1s and 2s, we include visualizations of these four cases.)

When describing the smaller social networks of everyday hostile networks, we indicate both the roles of the users in the network and the tweets within the hostile conversation. First, we identify *aggressors or perpetrators* who represent the user responsible for the highlighted aggressive message. Second, we spot *reinforcers*, that is, users who directly support, retweet, or like the aggressive message. Third, we locate *victims* or users who are attacked or otherwise victimized by the aggressive message. Fourth, we label *defenders* as individuals who support or otherwise positively respond to the victim's tweets. Finally, we regard *bystanders* as those individuals who participate in the Twitter conversation but do not respond to the aggressive message.

Sentiment Analysis

Assessing Tweet Sentiment

We developed a classifier that we use in a sentiment analysis of the content of our sample of tweets (Zhang and Felmlee 2017). Sentiment classifiers leverage dictionaries of words, called lexicons, which are then manually annotated by several research assistants as carrying positive or negative sentiment. The optimal choice of lexicon is dependent on the data. The sentiment classifier in this project utilizes the lexicon created for the VADER (Valence Aware Dictionary and sEntiment Reasoner) classifier. The VADER classifier is particularly useful here because VADER is a lexicon approach designed to assess sentiment in social media (fully open-sourced under the [MIT License] at https://github.com/cjhutto/vaderSentiment). We updated VADER's lexicon to include the derogatory and targeted terms toward women that we found in our manual examination of tweets and translated the score into a − 4 (most negative) to +4 (most positive) scale.

To measure sentiment toward the keyword in the tweet, we adjusted the individual sentiment scores for the words in each tweet by their distance from the key term, as done by Flores (2017). We diverge from Flores by dividing the sentiment score of a word by the natural log of its distance from the keyword, which we found to be more accurate in our data than using the absolute distance. In other words, an adjective that is two places away from the target will have its sentiment halved under previous approaches, but with the natural log, the sentiment will largely be preserved and decay more gradually. We then take the sum of the distance-adjusted sentiment scores and use it as the independent variable in a regression analysis. This procedure improved greatly on using VADER's default classifier, which is designed to measure the overall sentiment in a tweet.

To further increase the overall accuracy of the sentiment score, we compared the performance of using VADER's lexicon with several of the most commonly used lexicons. In the end, we used a combination of scores from the top three lexicons in ensemble, which included VADER and two others: AFINN (Nielsen 2011) and Bing (Hu and Liu 2004). Our final classifier takes into account not only the relative negativity and positivity of the individual words included in the tweets, but also the presence of other emotional cues and the respective linguistic distance between gendered slurs and emotional intensifiers. Thus, our sentiment analyses include supplementary tests to help determine the sentiment of tweets beyond the mere presence or absence of positive and negative words. Finally, we compared the sentiment scores gathered from four human coders on a test set of 400 tweets, with the scores obtained by our final classifier using 10-fold cross-validation. Our sentiment classifier performed quite well, with overall F1 scores of .640 (micro) and .640 (macro). These F1 scores represented an improvement over scores from VADER's default classifier alone or from other combinations of common classifiers.

Analyzing the Effect of Stereotypes on Message Sentiment

To understand the specific influence of stereotypes in messages utilizing hostility toward women, we also study how the inclusion of specific insults modifies tweet sentiment. We use ordinary least squares (OLS) regression models in which the sentiment score (−4 to +4) is the dependent measure, and each independent variable consists of the presence of an insulting adjective within one of the seven categories of stereotypical language. To test the effect of these adjective groups on the sentiment score, we employ three regression models. We analyze one model each for the keywords "bitch" and "slut" and a third in which we collapse tweets with the key terms "slut" and "whore" into one analysis, due to their relatively small sample sizes and synonymous meaning.

Results

Descriptive Findings

Frequency of Key Terms

Using the data collected directly from the Twitter API via four streams, we begin by comparing the frequency of each of the four keywords in tweets over the week. We scraped tweets over one week starting on a Saturday. Across all four streams, some days produced more tweets than others.

Thus, there appears to be one day of the week that spikes in finding messages and that day differs for each of the terms (e.g., Day 4 for the "bitch" and "cunt" streams, Day 3 for "slut," and Day 1 for "whore"). In other words, these four negative terms were not concentrated within one particular day of the week, but rather appeared across the week. Note, too, that there were numerous messages gathered on each of these terms on any given day. The minimum number of tweets in the streams on any given day that contained one of these terms was around 10,000 tweets for the keywords "cunt" (on Day 1) and "whore" (on Day 2), whereas for the more common term, "bitch," the lowest number of messages obtained was approximately 30 times that size (300,000) on Day 6. Altogether, this analysis resulted in over 418,000 tweets per day that included at least one sexist term. Although we calculate an overall frequency of tweets scraped from the API over one week, the prevalence of such tweets cannot be inferred because Twitter does not make public the number of tweets created on any given day nor the number of tweets released publicly for download.

Illustrations of Hostile Tweets

Examples of negative aggressive messages oriented toward women that contain one or more of the four keywords are included in Table 1. For example, all the tweets with multiple key terms were classified as being particularly negative in content, with a maximum negative score of −4. Several of the examples included in this table were racist as well as sexist in content. For example, both the first and third tweets in Table 1 appear to harass women from a specific ethno-racial group, with the first employing not only the word "bitch," but also a common Hispanic slur, and the third adding an insult aimed at those of Chinese heritage.

Networks of Cyber Aggression

In collecting the network data on cyber aggression, we identified two types of "conversation" networks. First, a set of very large network discussions about targeted women emerged, many of which contained positive as well as negative content. These extensive Twitter networks tended to involve widely

Table 1 Examples of aggressive messages by key term

Keyword in Tweet	Tweet Content
Bitch	@USER WHAT DO U MEAN FACTS BITCH U BLOCKED ME U DUMB HOE UR MAD PRICK ALL UR DIRTY SPICK FRIENDS WANNA GET INVOLVED SO NOW U TALKING
	Then go buy ur chinese food elsewhere u skanky old bitch
	@USER ching chong bitch
	[a]Heard @USER had a mob of people protecting her because she's a scared little (aka fat) bitch that can't fight her own battles.
	You look ugly as fuck you fucking cunt nigger kill yourself please do us a favour KKK
Cunt	it's even cuter that I was fucking your boyfriend just days before y'all got together. *kissy emoji* enjoy, ya fat cunt.
	Don't rt. my tweet I'm mocking you you psychotic cunt
	@USER @USER ur name looks retarded u anorexic cunt
	@USER You ignorant fucking cunt, you are a waste of skin!!! Fuck you and the muslims!!!
Slut	[a] we get it you work out you stupid slut
	@USER @USER How about you shut the fuck you stupid slut
	@USER no one asked for your damn opinion ignorant slut
Whore	Maybe if you had some self respect & didn't whore yourself out & lie/manipulate people you wouldn't be so depressed☺
	There are only 2 genders. Climate change is not affected by humans. Your mom's a fat whore. You suck and everyone hates you. #DealWithIt
	[a]You're literally a whore shut your mouth
	@USER How about you go burn in hell you stupid whore, there is nothing wrong with @USER. Just because you don't like him..whore
Multiple Terms	Fuck you, #You dumb cunt ass bitch
	@USER You're a slut, kill yourself you old cunt
	@USER @USER Fuckin bitch,if I ever see you.I'm going to kill you.nasty ugly slut

Data collected from NodeXL Twitter Searches and Directly from Twitter API

[a] Aggressive Tweet Content Described in Illustrations of Everyday, Hostile Networks

known individuals such as celebrities or politicians and, in these cases, the messages appeared to be exchanged largely among strangers who were interested in the same general conversation. Typically, a tweet that contained an insult set off a chain of interconnected messages, some of which reinforced the harmful content in retweets to additional people and others that attempted to defend the victim. The second type of cyber aggression network represented smaller sets of conversations that used derogatory language toward women in everyday conversations. These types of interchanges were common among users who seemed to know one another and who interacted more frequently. In addition, there was a third category of negative messages that consisted of singular tweets that received no visible online response and failed to generate a network of interchanges. Such seemingly aborted interchanges were common in our searches and may represent the fact that Twitter only releases a small portion of their data for public study at any one time.

Illustration of Celebrity Network

We include one case study of a large Twitter network that emerged during a battle between an African American actress and a British commentator that spread widely within the "Twittersphere" in 2016. During this exchange, the actress announced she was planning to leave Twitter due to the large accumulation of hateful and offensive comments she received following her appearance in the remake of a popular movie franchise. Twitter responded by stating that they had not done enough to enforce their regulations against abusive tweeting and that they would work more consistently to review and enforce their policies (Altman 2016). The British political commentator reacted by declaring that Twitter's announcement was tantamount to a war on free speech. He then responded to the actress's original tweets in a manner that crossed Twitter's policy on abusive content, which led to his account being suspended. The interaction over time among the actress, the political commentator, and Twitter led to a great deal of public debate both on and off Twitter about the role and nature of Twitter as a social media platform, as well as the acceptability of online hate and aggression.

A network analysis of this case study illustrates that this type of an attack on a well-known celebrity can become quite large and spread far (see the online supplement S1, Fig. 1s). Several months following the initial attack, a high density of tweets emerges in the conversation network that is directed toward the actress, which displays the intense concentration of discussion regarding the actress. In addition, although a number of messages in this Twitter interchange are positive and supportive in content, derogatory aggressive tweets continue to appear in the virtual network. For example, one message insults the actress, as follows: "@USER1 @USER2 [Actress] is a gorilla." Another hostile tweet attacks one of her online supporters, calling the person an "incestuous child rapist," and then describes the African American actress herself as a "thin-skinned racist." From analyzing the aggressive tweets, the messages are likely not sent by the original British commentator, but by users who are unconnected to the celebrity they attacked here.

Illustrations of Everyday Hostile Networks

In addition to large network conversations about specific women or groups of women, we found many instances of smaller everyday conversations that contained hostile messages. The texts of several tweets containing such aggressive messages are depicted in Table 1. We describe in detail three such examples that include at least one of our four key slurs. (For displays of the resulting Twitter networks of conversation in these three cases see the online supplement S1, Fig. 2s.)

The first example of an everyday network contains an aggressive tweet using the key term, "bitch," in the following message: "Heard @USER had a mob of people protecting her because she's a scared little (aka fat) bitch that can't fight her own battles." In the text, the bully directly mentions the victim, connecting the two accounts directly. The aggressive message also includes an adjective from the Overweight category, suggesting that the victim is not just a "scared little" bitch, but also too heavy; the use of this extra insulting term seems designed to "add insult to injury" in the attack. In addition to the bully and victim, six other individuals like or retweet the bully's original message, and they therefore act as reinforcers of the negative content. In a subsequent reply to the original aggressive tweet, the bully also mentions another user who never responds. We label this last user—who is not attacked nor responds to any of the conversation—as a bystander. A bystander represents a person who probably was aware of an aggressive message, but who neither comes to the defense of the victim nor joins in on the attack.

The second network contains twelve individuals, including one who publishes an aggressive message and four individuals who like or retweet the original aggressive tweet. The tweet content, "You're literally a whore shut your mouth" is an insult using the keyword, "whore." While short, this tweet represents a sizeable number of tweets in the sample that use aggressive slurs in a concise, derogatory strike. Note that the target of this message did not appear in our dataset, perhaps because the victim's account was private or was removed after the conflictual interaction. After publishing the original tweet, nevertheless, there are three individuals who condone the aggressive message. Seven bystanders also are present in the conversation network, that is, individuals who do not directly support nor condemn the offensive content of the tweet.

The final example contains an illustration of an aggressive message using the key term, "slut." The message is as follows: "we get it you work out you stupid slut." This tweet is part of a longer conversation that involves several users discussing their experiences watching and judging others at a gym. Despite a seemingly innocuous start to this conversation, the thread quickly moves to a discussion about women going to the gym "to be seen," and thus the identified tweet concerns (inappropriate) appearance. In addition, this tweet uses an adjective from the Unintelligent group, "stupid," presumably to amplify the negative intent.

Of these final three network cases, this last network is the largest with 30 individuals involved. Whereas two people act in a supportive capacity toward the victim, there are six others who initially support the aggressor, but then later retweet or like a message in defense of the original victim. In this case, the actors appear to be interconnected. Multiple tweets directed between network members suggest some familiarity and association, which may help to explain apparent shifting loyalties. This case also demonstrates that not all instances of sexist language emanate from anonymous Twitter accounts or from strangers at a distance (such as in the Celebrity example), but instead involve likely acquaintances and friends.

Descriptive Statistics from Sentiment Analyses

Sentiment of Tweets with Key Terms and Adjectives

Next, we conducted a sentiment analysis of our sample of tweets. As expected, messages that include any one of the key terms, "bitch," "cunt," "slut," and "whore," tend to be negative in sentiment. The average sentiment score ranged from −.38 for messages with "slut" to −.52 for those containing the term "cunt," on a scale of −4 to 4, with an overall, sample average of −.43 (see first row of Table 2). When tweets include an adjective that reinforces a feminine stereotype (e.g., "ugly," "fat"), the average sentiment score more than doubles in negativity (by a factor of 2.6), from −.39 for messages with no insulting adjectives to −1.03 for messages with such an adjective (last column of Table 2). For example, the average sentiment score for tweets with the term "bitch" changes from −.39 to −.95 when it includes a negative adjective (first column of Table 2). The shift in sentiment is even more notable among the "whore" messages," in which case the sentiment increases by a factor of over 6.5, from an average of −.32 to −2.11 with the addition of an insulting adjective (fourth column of Table 2).

Most Negative Words

Tweet sentiment was particularly negative when uncomplimentary adjectives referring to physical attributes were included (see Table 3). For example, within "bitch" messages, the use of "fat" (or its synonyms in the Overweight category) increased the negativity of the message from an average of −.43 to −1.15. For the "slut" sample, the most negative tweets were those in the Underweight category (e.g, skinny). Adjectives in the Ugly category, on the other hand, were associated with the highest negative sentiment score for the "cunt" and "whore" messages. Referring to an "ugly whore" in a tweet, for example, multiplied the degree of negative sentiment by more than six times, on average, as compared to a message lacking the descriptor, ugly.

The Effect of Insulting Adjectives on Sentiment

Ordinary Least Squares Regression (OLS)

As shown in the preceding, certain adjectives are associated with more negative tweet sentiment than others, particularly adjectives referring to physical appearance. Next, we examine whether such tendencies are statistically significant by testing whether the inclusion of an adjective in one of the seven categories significantly alters the degree of overall message negativity. Given our distance weighting, the addition of a negative adjective should only decrease the sentiment score of a tweet if it is close to the target or if the tweet commonly contains other negative words.

Table 2 Description of tweets by keyword and mean score

	Bitch Tweets		Cunt Tweets		Slut Tweets		Whore Tweets		Total Tweets	
	Tweet Count	*M* (*SD*)	Tweet Count	*M* (*SD*)	Tweet Count	*M* (*SD*)	Tweet Count	*M* (*SD*)	Tweet Count	*M* (*SD*)
All Tweets	2,530,832	−.43 (.72)	155,059	−.52 (.76)	131,155	−.38 (.62)	110,307	−.43 (.84)	2,927,353	−.43 (.72)
Without Adjs	2,349,978	−.39 (.70)	138,769	−.44 (.72)	125,537	−.35 (.60)	99,498	−.32 (.62)	2,713,782	−.39 (.70)
With Adjs	180,854	−.95 (.72)	16,290	−1.21 (.74)	5618	−.92 (.77)	10,810	−2.11 (1.06)	213,572	−1.03 (.78)

Tweets are marked as either containing an adjective or not; this refers to the specifically identified adjectives studied in the present paper. Tweets in the present sample contain many different adjectives, but here we focus on a small subset of adjectives relating to stereotypes regarding femininity. "Without Adjs" refers to the sample of tweets that do not include a specified adjective pertaining to feminine stereotypes and "With Adjs" refers to tweets containing an identified adjective. Tweets were scored within a range of −4.00 (most negative) to 4.00 (most positive)

Table 3 Sentiment scores of tweets and frequency of most negative adjectives within tweets

Keyword in Tweet	Mean Sentiment Score of Tweets with Keyword	Within Tweets with Specified Keyword	
		Most Negative Adjective Used (Adjective Group)	Mean Score of Tweets Containing Key Term + Most Negative Adjective
Bitch	−.43	Fat (Overweight)	−1.15
Cunt	−.52	Ugly (Ugly)	−1.29
Slut	−.39	Skinny (Underweight)	−1.02
Whore	−.43	Ugly (Ugly)	−2.75

Data collected from Twitter API. Scores range from −4 (most negative) to +4 (most positive)

Robustness

In analyses not shown here, we compared results from this modelling approach to several others, such as proportional odds and partial proportional odds logistic regression, using both AIC and mean squared error (MSE) of fitted values. In all cases, OLS regression had lower MSE and AIC, indicating it was fitting better to the data despite assuming a continuous rather than ordinal response. When comparing the OLS results, we rounded the fitted values to the nearest integer to ensure the lower MSE is not a result of having a continuous response. In analyses not shown here, we also tried including controls for time of day and day of the week, but the results failed to show systematic differences and our conclusions were unaltered.

Comparing Sentiment Scores with Insulting Adjectives: OLS Results

In our analysis, we find support for the argument that the inclusion of insulting words that reinforce feminine stereotypes inflates the overall negative sentiment of a tweet significantly. These findings are depicted in Table 4, which gives the exact coefficients and standard errors. In the majority of cases (70%), the inclusion of one of the adjectives increases the negative sentiment of a tweet, suggesting that the word not only lowered sentiment, but that it was located near enough to the key term for its score to affect significantly the tweet's overall score. Furthermore, the inclusion of a normative adjective alone is enough to reduce the sentiment score of a message (i.e., make the tweet more negative) by 1.0 on average. Therefore, although the large sample size may influence the statistical significance of the results, the finding that the inclusion of adjectives results in sizable negative effects is particularly noteworthy.

Across each of the keywords separately, the addition of an adjective also significantly lowers tweet sentiment. For example, messages that contain only the key term "bitch" and no identified adjectives are predicted to have a negative sentiment (β = −.39) as depicted in the value of the Intercept (see the first column of Table 3). The inclusion of an adjective in one of the seven categories drops the sentiment score for a "bitch" tweet from between −.23 for words in the Old age group to

Table 4 Ordinary least squares regression results for each dataset

Predictors	Keyword					
	Bitch		Cunt		Slut/Whore	
	β (*SE*)	*p*	β (*SE*)	*p*	β (*SE*)	*p*
Intercept	−.39 (<.001)	<.001	−.44 (.002)	<.001	−.35 (.001)	<.001
Crazy	−.28 (.005)	<.001	−.46 (.037)	<.001	−.45 (.027)	<.001
Overweight	−.69 (.007)	<.001	−.65 (.012)	<.001	−.53 (.016)	<.001
Underweight	−.28 (.014)	<.001	−.42 (.042)	<.001	−.51 (.022)	<.001
Old	−.23 (.005)	<.001	−.75 (.014)	<.001	−.45 (.019)	<.001
Stupid	−.61 (.002)	<.001	−.73 (.008)	<.001	−.61 (.018)	<.001
Promiscuous	−.56 (.011)	<.001	−.57 (.040)	<.001	−.51 (.024)	<.001
Ugly	−.65 (.005)	<.001	−2.18 (.009)	<.001	−.50 (.028)	<.001
n	25,20,302		158,128		130,964	
R^2	.05		.31		.03	

Significant effects reported here indicate that each occurrence of the predictor within a tweet significantly increases the negativity of the tweet's sentiment a maximum of −.69 for terms in the Overweight category. For each Overweight type of adjective, in other words, the sentiment of a tweet with the word "bitch" decreases by an additional −.69 points.

Physical Appearance

In addition, we examine the possibility that tweets that demean a woman's appearance are particularly negative, as suggested by the prior descriptive analyses. As expected, messages containing "ugly" or one of its synonyms are consistently some of the most negative tweets across our keywords (see Tables 3 and 4). For instance, tweets that include an adjective in the Ugly category, when added to a tweet containing the key term "cunt," have the largest, and significant, increases in negativity of any adjective across all three models (see the cunt column of Table 4). For each adjective from the Ugly group, the sentiment score of a tweet with the term "cunt" is decreased further by over −2.0 points. The size of this effect is quite large given that the range of possible scores ranges from −4 to +4. Tweets that include an adjective in the Overweight category also inflated the negative content of the message, particularly when paired with the keyword "bitch." We also consider the effect of adjectives in the Underweight category. For the majority of tweets (those containing either "bitch" or "cunt"), those with Underweight terms are not as negative as those from the Overweight category. The only exception occurs in messages containing "slut" or "whore," in which case the negative effects of Overweight and Underweight adjectives do not differ significantly.

For the keyword group of "slut/whore," the most adverse type of adjectives included those in the Stupid category rather than descriptors relating to physical appearance. One reason for the negativity of the word stupid and its synonyms may be due to the frequent use of these words in well-known offensive phrases, such as "stupid slut," and "ignorant slut." In sum, references to physical appearance significantly amplify the negative content of a sexist tweet, but other types of abusive language can do so as well depending on the specific phrase.

Discussion

Social media sources are repeatedly accused of providing venues for their users to treat women unjustly (e.g., Women's Boycott), and we find evidence that hostile Twitter messages aimed at women represent common everyday occurrences. We located 2.9 million tweets in 1 week, or approximately 419,000 per day, that used one of four key feminine slurs. Note that Twitter only makes a small portion of their data available for public download, suggesting that the problem of derogatory, female-oriented messages is even more extensive than we document. In addition, we find that these tweets were negative in sentiment, on average, and that the incorporation of adjectives that insulted feminine stereotypes amplified their overall negativity to a significant and notable degree. In some cases, the reach of a hostile message extended far beyond the original target to involve numerous other people, especially when concerning a female celebrity.

Findings from our mixed methods study show that these aggressive online messages frequently rely on language suggesting that the target fails to embody traditional feminine stereotypes and ideals, in particular those of physical attractiveness, niceness, and sexual purity. The implicit message, therefore, is that women *should* align themselves with traditional images of beauty, sweetness, and innocence—that is, there is a correct way to "do gender" (West and Zimmerman 1987). Messages attacking a woman's physical appearance are particularly negative, and they diminish message sentiment significantly more than other types of insults in a majority of our tweets. As noted elsewhere (Baker-Sperry and Grauerholz 2003; Mazur 1986; Wolf 2002), ideals of feminine beauty remain powerful in many societies in which social media networks thrive.

We find support for our theoretical arguments that norm enhancement and status processes contribute to cyber aggression within social media. Negative online communication in our Twitter data tends to reinforce existing social norms of femininity, not unlike findings of gendered norm enforcement in other genres of online interaction (Pascoe and Diefendorf 2019; Wilhelm and Joeckel 2019) and in schools (Felmlee and Faris 2016). The spread of derogatory tweets within networks of Twitter conversations, furthermore, suggests that enhancing one's status among one's peers constitutes another likely motivation behind damaging electronic communication toward women. In addition, these interactions occur within a cultural system of beliefs and a societal power structure that diminishes the status of women (Ridgeway 2011; Ridgeway and Correll 2004), all of which enhance the ability and motivation of individuals to demean women.

Limitations and Future Research Directions

Our research is not without limitations. Our sample is nonrandom and contains only a small portion of all tweets and adjectives that target women, which places limits on generalizability. Moreover, some portion of these messages might attack men and others. Previous research finds that sexual minorities and transgender individuals, for instance, remain frequent victims of social media aggression (Hinduja and Patchin 2010; Schneider et al. 2012; Sterner and Felmlee 2017). Some messages may be confronting gay (or straight) men rather than women, although research suggests that alternative homophobic slurs tend to be used instead in these instances (Sterner and Felmlee 2017). Moreover, a number of tweets in our dataset address women of color, often targeting women both on the basis of their gender and their ethno-racial identity. We provide examples of these "intersectional" cases, but this topic deserves additional attention.

We note, too, that tweets can be ambiguous in meaning and that some may represent the work of bots or electronic gaming. Although we took pains to circumvent such problems, and although our sentiment classifier performed quite well, we acknowledge that analyzing Twitter data remains challenging. Possible misinterpretations cast little doubt on our main conclusions, which are supported by extensive attention on the part of our in-depth human coding. However, they highlight the need for further research regarding the analysis of tweet sentiment. Alternative methodologies, such as Natural Language Processing Hierarchical Topic Modeling (Blei et al. 2003), also could be useful in the future to detect stereotype themes and subthemes in sexist Twitter content.

Practice Implications

Twitter and other types of social media constitute a routine part of numerous people's lives, especially young adults, and cyber aggression is a pervasive and problematic feature faced by many. The results provided herein can be used to inform counselors and educators to develop policies to address this societal problem. Those individuals who are victimized by sexist messages could benefit from being aware that many such messages are broadly sexist in nature, aiming to reinforce traditional femininity. Such knowledge could aid professionals in framing procedures to mitigate the negative ramifications of such incidents. Women whose self-worth depends on feedback from social media are particularly likely to be at risk for poor psychological well-being (Sabik et al. 2019), and these individuals are deserving of particular attention when facing electronic aggression. Finally, since the time of our data collection, Twitter instituted new procedures intended to reduce instances of abuse and harassment on its platform. Independent research finds that such efforts can modestly, but significantly, decrease the negative content of sexist and racist tweets (Felmlee et al. 2019). Such attempts by social media conglomerates to mitigate abuse on their platforms deserve continued serious attention.

Conclusion

Cyber aggression targets women on a day-to-day basis, with 419,000 tweets per day containing one of four common sexist slurs, and thus it constitutes a significant challenge to civil society. In its attempt to attack women on the basis of their femininity, electronic harassment implicitly reinforces cultural norms and stereotypical ideals of female appearance and behavior. One purpose of the present research is to give voice to this social problem. Perhaps learning that others also are vilified as lacking in appearance, character, and morality will lessen some of the pain experienced by victims in knowing that they are not alone in this experience. Finally, we further our understanding of this type of online behavior by providing evidence that the social process of gender norm regulation contributes to its occurrence. Aggression toward women does not represent simply individual random acts of violence, but instead it operates strategically and within the bounds of established group and societal processes.

Acknowledgments We are grateful to Evan Baker, Ying Cheng, Brendan Gard, Sara Francisco, Jordan Lawson, Kaitlin Shartle, and Tyler Stumm for their assistance with the project and data coding. Support for this research was provided by the National Science Foundation under Grant No. 1818497. We acknowledge administrative assistance by the Population Research Institute at Penn State University, which is supported by an infrastructure grant by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (P2CHD041025). Partial support also was provided by the National Science Foundation under IGERT grant DGE- 1144860, Big Data Social Science.

Compliance with Ethical Standards

Conflict of Interest None of the authors has potential conflicts of interest regarding this research.

Research Involving Human Participants and/or Animals This study was approved by the Pennsylvania State University Institutional Review Board (Study00004666). No animals were involved in this project.

Informed Consent No informed consent was necessary for this project.

References

Altman, J. (2016, July 25). The whole Leslie Jones twitter feud, explained. *USA TODAY College*. Retrieved from http://college.usatoday.com/2016/07/25/the-whole-leslie-jones-twitter-feud-explained/. Accessed 18 Sept 2019

Anderson, C. A., & Anderson, K. B. (2008). Men who target women: Specificity of target, generality of aggressive behavior. *Aggressive Behavior, 34*(6), 605–622. https://doi.org/10.1002/ab.20274.

Backe, E. L., Lilleston, P., & McCleary-Sills, J. (2018). Networked individuals, gendered violence: A literature review of cyberviolence. *Violence and Gender, 5*(3), 135–146. https://doi.org/10.1089/vio.2017.0056.

Bailey, J., Steeves, V., Burkell, J., & Regan, P. (2013). Negotiating with gender stereotypes on social networking sites: From "bicycle face" to Facebook. *Journal of Communication Inquiry, 37*(2), 91–112. https://doi.org/10.1177/0196859912473777.

Baker, E., Inara Rodis, P., & Felmlee, D. (2019, April). *Twitter: Fit or bitter?* Poster presented at The 2019 Undergraduate Research Exhibition, University Park, PA.

Baker-Sperry, L., & Grauerholz, L. (2003). The pervasiveness and persistence of the feminine beauty ideal in children's fairy tales. *Gender and Society, 17*(5), 711–726. https://doi.org/10.1177/0891243203255605.

Bartlett, C. P., DeWitt, C. C., Maronna, B., & Johnson, K. (2018). Social media use as a tool to facilitate or reduce cyberbullying perpetration: A review focusing on anonymous and nonanonymous social media platforms. *Violence and Gender, 5*(3), 147–152. https://doi.org/10.1089/vio.2017.0057.

Beauchere, J. F. (2014). Preventing online bullying: What companies and others can do. *International Journal of Technoethics, 5*(1), 69–77. https://doi.org/10.4018/ijt.2014010106.

Bellmore, A., Calvin, A. J., Xu, J., & Zhu, X. (2015). The five W's of "bullying" on twitter: Who, what, why, where, and when. *Computers in Human Behavior, 44*, 305-314. https://doi.org/10.1016/j.chb.2014.11.052 .

Bianchi, C. (2014). Slurs and appropriation: An echoic account. *Journal of Pragmatics, 66*, 35–44. https://doi.org/10.1016/j.pragma.2014.02.009.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*(Jan), 993–1022.

Davis, C. A., Varol, O., Ferrara, E., Flammini, A., & Menczer, F. (2016). Botornot: A system to evaluate social bots. *International Conference Companion on World Wide Web, 25*(1), 273–274. https://doi.org/10.1145/2872518.2889302.

Ellemers, N. (2018). Gender stereotypes. *Annual Review of Psychology, 69*(1), 275–298. https://doi.org/10.1146/annurev-psych-122216-011719.

Faris, R., & Felmlee, D. (2011). Status struggles: Network centrality and gender segregation in same- and cross-gender aggression. *American Sociological Review, 76*(1), 48–73. https://doi.org/10.1177/0003122410396196.

Faris, R., & Felmlee, D. (2014). Casualties of social combat: School networks of peer victimization and their consequences. *American Sociological Review, 79*(2), 228–257. https://doi.org/10.1177/0003122414524573.

Felmlee, D., & Faris, R. (2016). Toxic ties: Networks of friendship, dating, and cyber victimization. *Social Psychology Quarterly, 79*(3), 243–262. https://doi.org/10.1177/0190272516656585.

Felmlee, D., Inara Rodis, P., & Francisco, S. C. (2018). What a b!tch!: Cyber aggression toward women of color. In M. T. Segal & V. Demos (Eds.), *Gender and the media: Women's places* (Vol. 26, pp. 105–123). Bingley, UK: Emerald Publishing Limited. https://doi.org/10.1108/S1529-212620180000026008.

Felmlee, D. H., DellaPosta, D., Inara Rodis, P., & Matthews, S. A. (2019, August). *Cyber aggression on social media: A quasi-experimental study of policy on sexist and racist messages*. Paper presented at the 114th Annual Meeting of the American Sociological Association, New York, NY.

Flores, R. D. (2017). Do anti-immigrant laws shape public sentiment? A study of Arizona's SB1070 using Twitter data. *American Journal of Sociology, 123*(2), 333–384. https://doi.org/10.1086/692983.

Forbes, G. B., Collinsworth, L. L., Jobe, R. L., Braun, K. D., & Wise, L. M. (2007). Sexism, hostility toward women, and endorsement of beauty ideals and practices: Are beauty ideals associated with oppressive beliefs? *Sex Roles, 56*(5–6), 265–273. https://doi.org/10.1007/s11199-006-9161-5.

Hinduja, S., & Patchin, J. W. (2010). Bullying, cyberbullying, and suicide. *Archives of Suicide Research, 14*(3), 206–221. https://doi.org/10.1080/13811118.2010.494133.

Hlavka, H. R. (2014). Normalizing sexual violence: Young women account for harassment and abuse. *Gender and Society, 28*(3), 337–358. https://doi.org/10.1177/0891243214526468.

Hollander, J. A., Renfrow, D. G., & Howard, J. A. (2011). *Gendered situations, gendered selves* (2nd ed.). Plymouth, UK: Rowman & Littlefield Publishers.

Homans, G. C. (1950). *The human group*. New York: Harcourt, Brace & World.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 168–177). Seattle, WA: ACM. https://www.cs.uic.edu/~liub/publications/kdd04-revSummary.pdf.

Jeffreys, S. (2005). *Beauty and misogyny: Harmful cultural practices in the West*. New York: Routledge.

Juvonen, J., & Gross, E. F. (2008). Extending the school grounds?—Bullying experiences in cyberspace. *The Journal of School Health, 78*(9), 496–505. https://doi.org/10.1111/j.1746-1561.2008.00335.x.

Lawson, J., Rodis, P., & Felmlee, D. (2017). *Bigotry takes to Twitter: Cyberbullying towards African Americans.* Poster presented at the 2017 Undergraduate Exhibition, University Park, PA.

Loya, B. N., Cowan, G., & Walters, C. (2006). The role of social comparison and body consciousness in women's hostility toward women. *Sex Roles, 54*(7–8), 575–583. https://doi.org/10.1007/s11199-006-9024-0.

Mazur, A. (1986). U.S. trends in feminine beauty and overadaptation. *The Journal of Sex Research, 22*(3), 281–303. https://doi.org/10.1080/00224498609551309.

Miller, S. A. (2016). "How you bully a girl": Sexual drama and the negotiation of gendered sexuality in high school. *Gender & Society, 30*(5), 721–744. https://doi.org/10.1177/0891243216664723.

Nielsen, F. Å. (2011). *A new ANEW: Evaluation of a word list for sentiment analysis in microblogs.* Retrieved from http://arxiv.org/abs/1103.2903. Accessed 18 Sept 2019

Pascoe, C. J., & Diefendorf, S. (2019). No homo: Gendered dimensions of homophobic epithets online. *Sex Roles, 80*(3), 123–136. https://doi.org/10.1007/s11199-018-0926-4.

Pew Research Center. (2016). *Social media update 2016*. Retrieved from http://www.pewinternet.org/2016/11/11/social-media-update-2016/. Accessed 10 Dec 2016

Pew Research Center. (2017). *Online harassment 2017*. Retrieved from http://www.pewinternet.org/2017/07/11/online-harassment-2017/. Accessed 18 Sept 2019

Ridgeway, C. L. (2011). *Framed by gender: How gender inequality persists in the modern world*. New York: Oxford University Press.

Ridgeway, C. L., & Correll, S. J. (2004). Unpacking the gender system: A theoretical perspective on gender beliefs and social relations. *Gender and Society, 18*(4), 510–531. https://doi.org/10.1177/0891243204265269.

Sabik, N. J., Falat, J., & Magagnos, J. (2019). When self-worth depends on social media feedback: Associations with psychological well-being. *Sex Roles*. Advance online publication. https://doi.org/10.1007/s11199-019-01062-8.

Schneider, S. K., O'Donnell, L., Stueve, A., & Coulter, R. W. S. (2012). Cyberbullying, school bullying, and psychological distress: A regional census of high school students. *American Journal of Public Health, 102*(1), 171–177. https://doi.org/10.2105/AJPH.2011.300308.

Shartle, K., Stumm, T., Felmlee, D., & Sterner, G. (2016, April). *The social networks of gender related cyberbullying on Twitter.* Poster presented at The 2016 Undergraduate Exhibition, University Park, PA.

Sijtsema, J. J., Veenstra, R., Lindenberg, S., & Salmivalli, C. (2009). Empirical test of bullies' status goals: Assessing direct goals, aggression, and prestige. *Aggressive Behavior, 35*(1), 57–67. https://doi.org/10.1002/ab.20282.

Simmel, G. (1950). *The sociology of Georg Simmel*. (K. H. Wolff, Trans.). Glencoe, Ill.: Free Press.

Smith, M. A., Milic-Frayling, N., Shneiderman, B., Mendes Rodrigues, E., Leskovec, J., & Dunne, C. (2010). *NodeXL: A free and open network overview, discovery and exploration add-in for Excel 2007/2010*. Social Media Research Foundation. Retrieved from http://nodexl.codeplex.com/. Accessed 18 Sept 2019

Sterner, G., & Felmlee, D. (2017). The social networks of cyberbullying on Twitter. *International Journal of Technoethics, 8*(2), 1–15. https://doi.org/10.4018/IJT.2017070101.

Stumm, T., Shartle, K., Sterner, G., & Felmlee, D. (2016, April). *The social networks of racially specific cyberbullying on Twitter*. Poster presented at The 2016 Undergraduate Exhibition, University Park, PA.

Swami, V., Coles, R., Wyrozumska, K., Wilson, E., Salem, N., & Furnham, A. (2010). Oppressive beliefs at play: Associations among beauty ideals and practices and individual differences in sexism, objectification of others, and media exposure. *Psychology of Women Quarterly, 34*(3), 365–379. https://doi.org/10.1111/j.1471-6402.2010.01582.x.

Valenti, J. (2009). *The purity myth: How America's obsession with virginity is hurting young women*. Berkeley, CA: Seal Press.

Wang, W., Chen, L., Thirunarayan, K., & Sheth, A. P. (2014). Cursing in English on Twitter. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 415-424). Retrieved from http://corescholar.libraries.wright.edu/knoesis/590/. Accessed 18 Sept 2019

West, C., & Zimmerman, D. H. (1987). Doing gender. *Gender & Society, 1*(2), 125–151. https://doi.org/10.1177/0891243287001002002.

Wilhelm, C., & Joeckel, S. (2019). Gendered morality and backlash effects in online discussions: An experimental study on how users respond to hate speech comments against women and sexual minorities. *Sex Roles*. Advance online publication. https://doi.org/10.1007/s11199-018-0941-5.

Wolf, N. (2002). *The beauty myth: How images of beauty are used against women* ().New York: HarperCollins Perennial.

Xu, J.-M., Jun, K.-S., Zhu, X., & Bellmore, A. (2012). Learning from bullying traces in social media. *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human language technologies* (pp. 656–666). Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=2382139/. Accessed 18 Sept 2019

Ybarra, M. L., Mitchell, K. J., Wolak, J., & Finkelhor, D. (2006). Examining characteristics and associated distress related to internet harassment: Findings from the second youth internet safety survey. *Pediatrics, 118*(4), e1169–e1177. https://doi.org/10.1542/peds.2006-0815.

Zhang, A., & Felmlee, D. (2017). *You *&#*%!: Identifying bullying tweets*. Poster presented at The 2017 Graduate Exhibition, University Park, PA.