A WRETCHED HIVE OF SCUM AND VILLAINY: HOW TWITTER ENCOURAGES

HARASSMENT (AND HOW TO FIX IT)

by

TIMOTHY CONBERE

A THESIS

Presented to the Conflict and Dispute Resolution Program
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Master of Science

June 2019

THESIS APPROVAL PAGE

Student: Timothy Conbere

Title: A Wretched Hive of Scum and Villainy: How Twitter Encourages Harassment (and How to Fix It)

This thesis has been accepted and approved in partial fulfillment of the requirements for the Master of Science degree in the Conflict and Dispute Resolution Program by:

Elizabeth Tippett	Chairperson
Erik Girvan	Member

and

Janet Woodruff-Borden   Vice Provost and Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded June 2019.

THESIS ABSTRACT

Timothy Conbere

Master of Science

Conflict and Dispute Resolution Program

June 2019

Title: A Wretched Hive of Scum and Villainy: How Twitter Encourages Harassment (and How to Fix It)

In August 2014, the video games community on Twitter became embroiled in a months-long harassment campaign called Gamergate. The campaign initially targeted indie video game developer Zoë Quinn, but quickly spread to many women and feminist games developers and journalists, targeting them with slurs, sexualized epithets, threats of violence, and more. This thesis examines what Twitter has done to prevent the kind of harassment that occurred during Gamergate and explains the social-psychological principles behind Twitter's failure to curb the culture of abuse on its platform. There is a fundamental disconnect between Twitter's value of allowing its users to share information instantly without barriers and its desire to protect its users from harassment. In order to solve the problem of harassment, Twitter will need to shift its foundational value away from free speech and towards deliberate community building.

CURRICULUM VITAE

NAME OF AUTHOR: Timothy Conbere

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, Oregon
University of Oregon School of Law, Eugene, Oregon
University of Washington, Seattle, Washington

DEGREES AWARDED:

Master of Science, Conflict and Dispute Resolution, 2019, University of Oregon
Juris Doctor, 2019, University of Oregon School of Law
Bachelor of Arts, International Studies: Asia, 2012, University of Washington

AREAS OF SPECIAL INTEREST:

Alternative Dispute Resolution
Psychology of Conflict
Internet Communities
Anti-Harassment Strategies

GRANTS, AWARDS, AND HONORS:

Orlando J. Hollis Scholar Award, 2018

Kerry Lewiecki Memorial Award, 2017

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

INTRODUCTION

On October 10, 2014, Brianna Wu posted a screenshot of public messages she had received on Twitter.[1] The picture contained eight messages clustered together, sent within four minutes of each other, the first one reading, "You just made a shitty game nobody liked. That's it. Nobody [will care] when you die." The last message read, "Guess what bitch? I now know where you live," and proceeded to publicly post Brianna's home address. The message-sender's username read "Death to Brianna."

Ms. Wu received these threats during a harassment campaign called Gamergate that targeted female and feminist game developers and game critics that began in August 2014 and continued for the next several months.[2] Throughout this deluge of hate, Twitter's harassment and reporting policies failed to meaningfully protect the victims' safety or counteract the Gamergate movement.

The fact that the Internet contributes to harassment is a well-documented phenomenon. What is less explored is how and why harassment on social media sites like Twitter forms and propagates and what companies should do to manage it. This thesis will examine a famous example of harassment on Twitter, explore the history of Twitter's responses to harassment, analyze the ways Twitter contributes to harassment, and provide suggestions for Twitter to reduce harassment online.

---

[1] Brianna Wu (@BriannaWu), TWITTER (Oct. 10, 2014, 5:57 PM), https://twitter.com/BriannaWu/status/520739878993420290.

[2] Kyle Wagner, *The Future of the Culture Wars Is Here, And It's Gamergate*, DEADSPIN (Oct. 12, 2014, 3:46 PM), https://deadspin.com/the-future-of-the-culture-wars-is-here-and-its-gamerga-1646145844.

CHAPTER II

#GAMERGATE: A CASE STUDY IN ONLINE HARASSMENT

Gamergate began in 2014 as a harassment campaign targeting indie game developer Zoë Quinn. As the campaign continued, it spread its focus across the Twitter video game community and targeted other feminists who worked in or talked about video games.[3] Many women, including Quinn, received a deluge of slurs, derogatory comments, sexualized epithets, and death threats during the course of Gamergate. Many of those same women continue to receive such harassment to this day.

By 2014, attacks against women and feminists in the gaming sphere had become commonplace. In 2012, feminist critic Anita Sarkeesian announced crowdfunding for her project *Tropes vs. Women in Video Games*, a video series detailing the commonly used, sexist stereotypes for female characters in video games.[4] The crowdfunding campaign was a success, raising almost $160,000 in 30 days from roughly 7,000 backers. However, its success caught the attention of conservative commentators, who released a torrent of misogynist hate mail on Sarkeesian that persisted long after the crowdfunding concluded.[5] The backlash against *Tropes vs. Women* formed an ongoing conflict centered on two main parties—Sarkeesian's attackers (generally conservative white men) and her supporters (generally progressives, including women, feminists, and people of color)— and set the stage for Gamergate.

---

[3] *Id.*

[4] Anita Sarkeesian, *Tropes vs. Women in Video Games*, KICKSTARTER, https://www.kickstarter.com/projects/566429325/tropes-vs-women-in-video-games (last visited May 20, 2019).

[5] Angela Watercutter, *Feminist Take on Games Draws Crude Ridicule, Massive Support*, WIRED (June 14, 2012, 6:30 AM), https://www.wired.com/2012/06/anita-sarkeesian-feminist-games/.

Quinn entered the public eye in 2013, a year before Gamergate, when she designed an interactive fiction game called *Depression Quest*, which used narrative text and player choice to explain the effects of depression.[6] Shortly thereafter, Quinn started receiving hate mail from people who did not believe *Depression Quest* was a real game and reacted with incredulity to the game's generally favorable critical reception.[7]

Then, in August 2014, Quinn's ex-boyfriend wrote a lengthy blogpost revealing intimate details of their relationship and accusing her of emotional abuse and infidelity.[8] Afterwards, Quinn was bombarded with harassing, threatening, and often sexually explicit posts on Twitter.[9] Her personal contact information was made public (an act known as "doxxing"[10]) as threats continued, compelling her to leave her home.[11] News of the blogpost and subsequent fallout was picked up by conservative news site Breitbart and mentioned by actor Adam Baldwin,[12] who popularized and ongoing use of

[6] Simon Parkin, *Zoe Quinn's Depression Quest*, THE NEW YORKER, (Sept. 9, 2014), https://www.newyorker.com/tech/elements/zoe-quinns-depression-quest.

[7] *Id.*

[8] Wagner, *supra* note 2.

[9] Amanda Marcotte, *Gaming Misogyny Gets Infinite Lives: Zoe Quinn, Virtual Rape, and Sexism*, THE DAILY BEAST (Aug. 22, 2014, 5:45 AM), https://www.thedailybeast.com/gaming-misogyny-gets-infinite-lives-zoe-quinn-virtual-rape-and-sexism.

[10] The origin of the term doxxing comes from the early days of the internet, when hackers would post a rival's private documents. Nellie Bowles, *How 'Doxxing' Became a Mainstream Tool in the Culture Wars*, NY Times (Aug. 30, 2017), https://www.nytimes.com/2017/08/30/technology/doxxing-protests.html.

[11] Wagner, *supra* note 2.

[12] Adam Baldwin (@AdamBaldwin), TWITTER (Aug. 24, 2014, 6:22 PM), https://twitter.com/AdamBaldwin/status/504801169638567936 [https://web.archive.org/web/20140903012212/https://twitter.com/AdamBaldwin/status/504801169638567936]; Devin Faraci, *Joss Whedon On #GamerGate, Jurassic World, Adam Baldwin And Speaking Out*, BIRTH MOVIES DEATH (Apr. 12, 2015), https://birthmoviesdeath.com/2015/04/12/joss-whedon-on-gamergate-jurassic-world-adam-baldwin-and-speaking-out.

"#GamerGate" on Twitter.[13] Many using the Gamergate hashtag claimed that Quinn was faking the threats she had received in order to garner sympathy,[14] even as members of the movement escalated their attacks on Quinn through Twitter.

In addition to the original allegations of infidelity and abuse in the blogpost, Gamergaters began to allege that Quinn slept with male video game journalists in exchange for positive reviews and expanded coverage. In tandem with these accusations, the discourse of the Gamergate movement shifted to include the promotion of ethics in games journalism.[15] Despite the accusations lobbied against Quinn, Gamergate supporters found no credible evidence to suggest that Quinn engaged in any unethical behavior regarding the promotion of her games.[16] Critics of Gamergate accused the movement of using the ethics argument as a pretext to legitimize their harassment of Quinn, her supporters, and other feminists in the gaming field.

Gamergate supporters frequently claimed that Quinn, Sarkeesian, and others were merely whining, that they were not receiving more or substantially worse abuse than any other public figure might receive on the Internet, and that they were lying about receiving

---

[13] *Id.*

[14] MannoSlimmins, *Depression Quest Dev Claims Harassment and Misogyny. Facts Come Out Showing She's Lying*, REDDIT (Aug. 17, 2014, 9:44 AM), https://www.reddit.com/r/TumblrInAction/comments/2dt5hu/depression_quest_dev_claims_harassment_an d/; Joe Rogan Bible, *"GAMERGATE"! Zoe Quinn the Fake Victim*, YOUTUBE (Jun. 4, 2018), https://www.youtube.com/watch?v=KOtjR2GWQLQ; Chris Tognotti, *What Is "#Gamer Gate"? It's Misogyny, Under the Banner Of "Journalistic Integrity"*, BUSTLE (Sept. 5, 2014), https://www.bustle.com/articles/38742-what-is-gamer-gate-its-misogyny-under-the-banner-of-journalistic-integrity.

[15] Caitlin Dewey, *The Only Guide to Gamergate You Will Ever Need to Read*, WASH. POST (Oct. 14, 2014), https://www.washingtonpost.com/news/the-intersect/wp/2014/10/14/the-only-guide-to-gamergate-you-will-ever-need-to-read/.

[16] *Id.*

death threats.[17] Many Gamergate critics, however, claimed that they had experienced or observed an inordinate increase in the amount of harassment relating to gender and video games since the movement began.[18]

Gamergate supporters both identified as "gamers" and held conservative views of women and people of color, criticizing feminism and resisting appeals for greater diversity in popular media.[19] Critics of Gamergate held progressive values, supporting feminism, social justice, and diversity.[20] These differing cultural systems and values drove actors in the conflict to support one side or the other, and caused both sides to interpret the conflict differently.[21]

Gamergate supporters held vastly different beliefs about the narrative of the conflict than Gamergate critics. From the perspective of the movement, Zoë Quinn's *Depression Quest* lacked the necessary game mechanics to qualify it as a game, yet it inexplicably received attention and praise from game critics, even though it had almost no advertising.[22] She claimed to have progressive ideals, but was "proved" to have been unfaithful and emotionally abusive, highlighting her hypocrisy.[23] Gamergate supporters

---

[17] Wagner, *supra* note 2.

[18] *Id.*

[19] Jesse Singal, *Gamergate Should Stop Lying to Journalists—and Itself*, N.Y. MAG.: THE CUT (Oct. 20, 2014), https://www.thecut.com/2014/10/gamergate-should-stop-lying-to-itself.html.

[20] *Id.*

[21] Frank A. Dubinskas, *Culture and Conflict: The Cultural Roots of Discord*, *in* HIDDEN CONFLICT IN ORGANIZATIONS: UNCOVERING THE BEHIND-THE-SCENES DISPUTES 187 (Deborah M. Kolb & Jean M. Bartunek eds., 1992).

[22] Parkin, *supra* note 6.

[23] Jay Hathaway, *The Angry Ex Who Ignited Gamergate Has No Regrets*, GAWKER (Oct. 16, 2014, 1:40 PM), https://gawker.com/the-angry-ex-who-ignited-gamergate-has-no-regrets-1647186033.

took the blogpost from Quinn's ex-boyfriend at face value because they were motivated

to find fault in a public figure they disliked, making unsavory, intimate details about her

seem credible. People who jumped on the Gamergate hashtag had already been

suspicious of Quinn's popularity and legitimacy as a game developer. The blogpost

confirmed their suspicions. For those who had not heard of Quinn, the narrative of a

social-justice oriented woman who had received fame for an artsy piece of interactive

fiction by taking advantage of the liberal-leaning games media scene made sense. The

"ethics in games journalism" line arose from a feeling that games media was insular,

opaque, and incestuous. If Quinn could use her personal relationships with the games

media to advance her career, other people could also manipulate the system.

Gamergate critics, on the other hand, saw the conflict as part of a continuum of

harassment against women who tried to insert themselves into the male-dominated world

of video games. They viewed the blogpost with a critical eye because they were familiar

with misogynistic attacks on women by ex-boyfriends. The anti-Gamergate side was used

to supporting women against attacks from conservative critics just for making games or

creating feminist criticism. The fact that the accusations about Quinn using her contacts

for unfair promotion were never substantiated showed to the anti-Gamergate side that

there was no substance behind the "ethics is games journalism" line, and that Gamergate

was purely a movement to harass women for existing in the video games community.

The Internet has long been heralded as a public sphere, a space for open dialogue

where all are able to participate equally.[24] But equality is not just the ability to occupy

space; it is also the ability to influence and speak within that space without fear of

---

[24] Jessica Megarry, *Online Incivility or Sexual Harassment? Conceptualising Women's Experiences in the Digital Age*, 46 WOMEN'S STUD. INT'L F. 46 (2014).

harassment.[25] Harassment makes it much more difficult for women to navigate online spaces and make their voices heard. It has caused some women to leave Twitter altogether, providing evidence that harassment can shut people off from vital sources of information and public interaction.[26]

A recent study by the Data & Society Research Institute, in collaboration with the Center for Innovative Public Health Research, found that 47% of Internet users 15 years or older have experienced harassment online and that 72% of Internet users have witnessed harassment online.[27] Additionally, 36% of Internet users have experienced direct harassment, 30% of Internet users have experienced invasions of privacy, and 17% of Internet users have experienced denial of access.[28] While the study did not report that women received more harassment by volume than men, they did received a wider variety of abuse and more serious incidents of harassment.[29] Women were also nearly three times as likely to say an experience scared them and twice as likely to say an experience made them feel worried.[30] This could be explained by women receiving more serious harassment than men and could be compounded by social networking sites not feeling like safe spaces for women.

---

[25] *Id.*

[26] AMNESTY INT'L, #TOXICTWITTER: VIOLENCE AND ABUSE AGAINST WOMEN ONLINE 50 (2018), https://www.amnesty.org/download/Documents/ACT3080702018ENGLISH.pdf.

[27] AMANDA LENHART, MICHELE YBARRA, KATHRYN ZICKUHR & MYESHIA PRICE-FEENEY, ONLINE HARASSMENT, DIGITAL ABUSE, AND CYBERSTALKING IN AMERICA 3 (2016).

[28] *Id.* Denial of access is a misuse of a site's reporting tools to block users from the site or sending a large number of unwanted messages in a way that prevents other users from meaningfully using the site. *Id.*

[29] *Id.* at 4.

[30] *Id.*

In a report by Pew Research, 92% of respondents said that Internet spaces allow people to be more critical of others compared to in-person experiences, but 68% of respondents agreed that Internet spaces allow people to be more supportive of others.[31] The qualitative responses that Pew collected generally revealed that people expect others to "just [be] mean online sometimes."[32] These results are not contradictory. Anonymity and access to communicate with a wide range of people make harassment easy and immediate, so people using the Internet are opening themselves up to potential harassment. On the other hand, the Internet makes it much easier for people with like interests to meet and form supportive networks than it is offline. For example, if you are a fan of a particular TV show, it is much easier to go online and begin discussing that TV show with like-minded peers than it is to go out and find someone to talk to about the show in your offline life. These interest-connections create supportive networks that users often cannot find outside of the Internet.

The Pew Research study also analyzed the perception of friendliness of online platforms toward men and women. 75% of respondents perceived social networking sites to be equally welcoming toward men and women.[33] On the other hand, 44% of respondents perceived online gaming as being more welcoming toward men and only 51% of respondents perceived it to be equally welcoming to men and women.[34]

---

[31] Maeve Duggan, *Online Harassment*, PEW RESEARCH CENTER (Oct. 14, 2014), http://www.pewInternet.org/2014/10/22/online-harassment/.

[32] *Id.*

[33] *Id.*

[34] *Id.*

In Gamergate, both sides used moral discourse to prove that their side was more virtuous the other and put the other side in its place. Within this discourse, the two groups tried to shame each other. Shame is often used to enforce norms, a society's "primary social control mechanism."[35] Each side accused the other of violating certain social norms: Gamergaters criticized Quinn for her infidelity and spread the blogpost about her to show her perceived hypocrisy and illegitimacy,[36] while Gamergate critics quoted Gamergaters to their followers and called them out for their sexist attacks.[37]

While the vitriol surrounding Quinn's supposed acts of indecency and the fierce accusations of unethical practices in games media eventually subsided, Gamergate has had a lasting impact on the online gaming community.[38] Zoë Quinn, Anita Sarkeesian, and others all continue to receive critical, harassing, and sexualized messages. Gamergate brought conservative politics to the fore in the gaming community's cultural discourse, and gave misogynist, anti-feminist harassment a legitimate-sounding pretext that encourages the harassment of women on the Internet to this day.[39]

---

[35] Kate Klonick, *Re-Shaming the Debate: Social Norms, Shame, and Regulation in an Internet Age*, 75 MD. L. REV. 1029, 1044 (2016).

[36] Wagner, *supra* note 2. The widespread proliferation of the blogpost regarding Quinn shows that people often use the accessibility of the Internet to share embarrassing or damaging content because it tells an engaging story. *See* DANAH BOYD, IT'S COMPLICATED: THE SOCIAL LIVES OF NETWORKED TEENS 146 (2014).

[37] Wagner, *supra* note 2.

[38] Jessica Valenti, *Zoe Quinn: After Gamergate, Don't Cede the Internet to Whoever Screams the Loudest*, GUARDIAN (Sept. 24, 2017), https://www.theguardian.com/technology/2017/sep/24/zoe-quinn-gamergate-online-abuse; William Turton, *Gamergate is Never Going Away*, OUTLINE (Sept. 7, 2017), https://theoutline.com/post/2218/gamergate-will-never-die-alt-right-trump?zd=1&zi=rtxk4tdn.

[39] Jesse Singal, *Why the Video-Game Culture Wars Won't Die*, N.Y. MAG. (Sept. 30, 2016), http://nymag.com/selectall/2016/09/why-the-video-game-culture-wars-wont-die-two-years-later.html.

CHAPTER III

TWITTER'S RESPONSES TO HARASSMENT ON THE INTERNET

Twitter is a social networking site on which millions of users can post short, 280-character messages called "Tweets" that, by default, can be seen and interacted with by anyone in the world. Twitter was created in 2006 and has since become an increasingly popular communication tool, reaching 326 million active users in 2018.[40] Given the huge number of people communicating with each other on Twitter, it is not surprising that harassment and abuse occur. One Amnesty International study revealed that 73% of women in the US agree abuse and harassment of women online is common, and 33% of women in the US have experienced online abuse one or more times.[41]

The inadequacies of Twitter's response to harassment is rooted in its history. Twitter was created first and foremost as a communication tool, not as a protective and welcoming community. While in recent years Twitter has made notable changes to its anti-harassment policies, its policies have not always been so robust. In 2008, two years into Twitter's lifespan, co-founder Biz Stone responded to the abuse of an early female Twitter user with an announcement that "Twitter is a communication utility, not a mediator of content."[42] Before 2009, Twitter did not have a public-facing conduct policy

---

[40] Brett Molina, *Twitter Overcounted Active Users Since 2014, Shares Surge on Profit Hopes*, USA TODAY (Oct. 26, 2017, 8:39 AM), https://www.usatoday.com/story/tech/news/2017/10/26/twitter-overcounted-active-users-since-2014-shares-surge/801968001/.

[41] Azmina Dhrodia, *Unsocial Media: The Real Toll of Online Abuse against Women*, MEDIUM (Nov. 20, 2017), https://medium.com/amnesty-insights/unsocial-media-the-real-toll-of-online-abuse-against-women-37134ddab3f4.

[42] Betsy Schiffman, *Twitterer Takes on Twitter Harassment Policy*, WIRED (May 22, 2008, 5:16 PM), https://www.wired.com/2008/05/tweeter-takes-o/; Charlie Warzel, *"A Honeypot For Assholes": Inside Twitter's 10-Year Failure To Stop Harassment*, BUZZFEED NEWS (Aug. 11, 2016, 8:43 AM), https://www.buzzfeednews.com/article/charliewarzel/a-honeypot-for-assholes-inside-twitters-10-year-failure-to-s.

of any kind,[43] and the first publication of the Twitter Rules began with a paragraph

highlighting Twitter's commitment to free speech.[44] The rules at this time did not

mention abuse, harassment, or hateful conduct. The closest these early rules come to an

anti-harassment policy are restrictions on posting private information of other users;

posting direct, specific threats of violence; creating serial accounts for disruptive or

abusive purposes; and using Twitter for unlawful purposes.[45]

      These policies stayed unchanged for years. In 2011, after refusing to comply with

a government subpoena for data regarding user activity, Twitter executives published a

blogpost stating: "[W]e strive not to remove Tweets on the basis of their content."[46] In

2012, Twitter publicly referred to itself as "the free speech wing of the free speech

party."[47] Then, in 2013, feminist activist Caroline Criado-Perez and female U.K.

Parliamentarian Stella Creasy, among others, were hit with a deluge of misogynistic

harassment, which became one of the first high profile harassment campaigns on

---

[43] Sarah Jeong, *The History of Twitter's Rules*, MOTHERBOARD (Jan. 14 2016, 7:00 AM), https://motherboard.vice.com/en_us/article/z43xw3/the-history-of-twitters-rules.

[44] *The Twitter Rules*, TWITTER (Jan. 14, 2009), http://twitter.zendesk.com/forums/26257/entries/18311 [https://web.archive.org/web/20090118211301/http://twitter.zendesk.com/forums/26257/entries/18311].

> Our goal is to provide a service that allows you to discover and receive content from sources that interest you as well as to share your content with others. We respect the ownership of the content that users share and each user is responsible for the content he or she provides. Because of these principles, we do not actively monitor user's content and will not censor user content, except in limited circumstances described below.

Id.

[45] *Id.*

[46] Warzel, *supra* note 42.

[47] Josh Halliday, *Twitter's Tony Wang: 'We Are the Free Speech Wing of the Free Speech Party'*, GUARDIAN (Mar. 22, 2012, 11:57 AM), https://www.theguardian.com/media/2012/mar/22/twitter-tony-wang-free-speech.

Twitter.[48] The amount of media attention the abuse received revealed that Twitter was ill-equipped to deal with harassment and protect its users. In response, Twitter added a "Report Abuse" button to individual Tweets for the first time[49] and added language to the Twitter Rules specifically addressing abuse and harassment:

> You may not engage in targeted abuse or harassment. Some of the factors that we take into account when determining what conduct is considered to be targeted abuse or harassment are: if you are sending messages to a user from multiple accounts; if the sole purpose of your account is to send abusive messages to others; if the reported behavior is one-sided or includes threats.[50]

In 2014, Gamergate forced Twitter to have another reckoning. The harassment that spawned as a result of Gamergate made headlines and dominated conversation on Twitter for months.[51] Twitter improved its system for reporting abuse, and, in 2015, it published a policy updating its rules on violent threats[52] and prohibiting the "promot[ion of] violence against others… on the basis of race, ethnicity, national origin, religion, sexual orientation, gender, gender identity, age, or disability."[53] Accompanying these changes, Twitter's general counsel, Vijaya Gadde, wrote in a Washington Post op-ed that "Freedom of expression means little as our underlying philosophy if we continue to allow voices to be silenced because they are afraid to speak up. We need to do a better job

---

[48] Warzel, *supra* note 42; Jeong *supra* note 43.

[49] Next to the "Report Spam" button, which had already existed for years.

[50] Jeong, *supra* note 43.

[51] *Id.*

[52] Shreyas Doshi, *Policy and Product Updates Aimed at Combating Abuse*, TWITTER (Apr. 21, 2015), https://blog.twitter.com/official/en_us/a/2015/policy-and-product-updates-aimed-at-combating-abuse.html; Issie Lapowsky, *Why Twitter Is Finally Taking a Stand Against Trolls*, WIRED (Apr. 21, 2015, 2:14 PM), https://www.wired.com/2015/04/twitter-abuse/.

[53] Jeong, *supra* note 43.

combating abuse without chilling or silencing speech."[54] At the end of 2015, Twitter

consolidated these policy changes into the Twitter Rules[55] and updated its preamble,

which remains substantively the same to this day: "We believe that everyone should have

the power to create and share ideas and information instantly, without barriers. In order to

protect the experience and safety of people who use Twitter, there are some limitations

on the type of content and behavior that we allow."[56] Twitter still places its value of free

speech front and center, but there appears to be greater acknowledgment of the protective

actions that are necessary to ensure free speech.

Following this shift, in 2016, Twitter added the ability for third parties to report

abusive or hateful conduct and introduced options for muting keywords, phrases, and

conversations in which users have been tagged.[57] In 2017, Twitter added a feature that

automatically hides Tweets that are potentially abusive or "low quality."[58] Following a

public boycott of Twitter after its suspension of actress Rose McGowen's account in the

midst of her campaign against sexual harassment and abuse in Hollywood, Twitter

announced additional protections for non-consensual nudity and unwanted sexual

---

[54] Vijaya Gadde, *Twitter Executive: Here's How We're Trying to Stop Abuse while Preserving Free Speech*, WASH. POST (Apr. 16, 2015), https://www.washingtonpost.com/posteverything/wp/2015/04/16/twitter-executive-heres-how-were-trying-to-stop-abuse-while-preserving-free-speech/.

[55] Megan Cristina, *Fighting Abuse to Protect Freedom of Expression*, TWITTER (Dec. 30, 2015), https://blog.twitter.com/en_au/a/2015/fighting-abuse-to-protect-freedom-of-expression-au.html; Jeong, *supra* note 43.

[56] Jeong, *supra* note 43; *The Twitter Rules*, TWITTER, https://help.twitter.com/en/rules-and-policies/twitter-rules (last visited May 3, 2019).

[57] Ingrid Lundgren, *Twitter Updates Its Abuse Policy and Adds Muting and Reporting Tools to Combat Trolls*, TECHCRUNCH (Nov. 15, 2016), https://techcrunch.com/2016/11/15/twitter-updates-abuse-policy/.

[58] Sherisse Pham, *Twitter Tries New Measures in Crackdown on Harassment*, CNN (Feb. 7, 2017, 9:57 PM), https://money.cnn.com/2017/02/07/technology/twitter-combat-harassment-features/.

advances and policy changes that defined hateful imagery and hate symbols "sensitive media."[59] And in 2018, Twitter created new policies to combat dehumanizing language.[60] Now, in 2019, Twitter is continuing to improve the enforcement of its policies. Twitter recently announced that its moderation teams proactively found 38% of abusive content that was subject to enforcement, up from 0% in 2018.[61] Twitter CEO Jack Dorsey has stated that he hopes to improve upon Twitter's enforcement rate and proactivity by relying on machine learning so Twitter can entirely remove the burden of victims needing to report the harassment they experience.[62]

Today, Twitter's conduct policies forbid "abusive behavior" and "hateful conduct". Abusive behavior is "the targeted harassment of someone," or inciting other people to harass someone.[63] Twitter's examples of abusive behavior include wishing or hoping serious harm on a person or group of people, unwanted sexual advances, using

---

[59] Greg Evans, *Twitter Policy Changes Under Review: Glorification of Violence, Unwanted Sexual Advances Targeted*, DEADLINE (Oct. 18, 2017, 9:38 AM), https://deadline.com/2017/10/twitter-policy-harassment-hate-speech-jack-dorsey-rose-mcgowan-1202190479/.

[60] Vijaya Gadde & Del Harvey, *Creating New Policies Together*, TWITTER (Sept. 25, 2018), https://blog.twitter.com/official/en_us/topics/company/2018/Creating-new-policies-together.html; Louise Matsakis, *Twitter Releases New Policy on 'Dehumanizing Speech'*, WIRED (Sept. 25, 2018, 9:00 AM), https://www.wired.com/story/twitter-dehumanizing-speech-policy/.

[61] Donald Hicks & David Gasca, *A Healthier Twitter: Progress and More to Do*, TWITTER (Apr. 16, 2019), https://blog.twitter.com/en_us/topics/company/2019/health-update.html; Kurt Wagner, *Twitter Says It's Getting Better at Detecting Abusive Tweets without Your Help*, VOX (Apr. 16, 2019, 3:35 PM), https://www.vox.com/2019/4/16/18410931/twitter-abuse-update-health-technology-harassment.

[62] Briane Greene, *How Twitter Shapes Global Public Conversation: Jack Dorsey Speaks at TED2019*, TEDBLOG (Apr. 16, 2019, 2:41 PM), https://blog.ted.com/how-twitter-shapes-global-public-conversation-jack-dorsey-speaks-at-ted2019/.

[63] *Abusive Behavior*, TWITTER, https://help.twitter.com/en/rules-and-policies/abusive-behavior (last visited May 3, 2019).

aggressive insults with the purpose of harassing or intimidating others, and encouraging

or calling for others to harass an individual or group of people.[64]

Hateful conduct, on the other hand, is "promot[ing] violence against or directly

attack[ing] or threaten[ing] other people on the basis of race, ethnicity, national origin,

sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious

disease."[65] Examples of hateful conduct include violent threats against a protected group;

wishing, hoping, or calling for serious harm on a person or group of people; references to

mass murder, violent events, or specific means of violence where protected groups have

been the primary targets or victims; inciting fear about a protected category, repeated

and/or non-consensual slurs, epithets, racist and sexist tropes, or other content that

degrades someone; and hateful imagery.[66]

In order to address the issue of harassment on its site, Twitter has a reporting

process that allows recipients or observers of conduct that violates Twitter rules to inform

Twitter moderators of such behavior. To report a Tweet or Direct Message ("DM"), a

user must fill out an online form that explains why the communication is abusive or

harmful.[67] Twitter moderators review the report and decide whether it violates Twitter's

policies and, if it does, what level of enforcement to provide. Twitter moderators have a

range of enforcement options from which to choose: low-level, temporary options like

limiting the public visibility of individual Tweet to high-level, permanent options like

---

[64] *Id.*

[65] *Hateful Conduct Policy*, TWITTER, https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy (last visited May 3, 2019).

[66] *Id.*

[67] *Report Abusive Behavior*, TWITTER, https://help.twitter.com/en/safety-and-security/report-abusive-behavior (last visited May 3, 2019).

permanently suspending a user account.[68] Once the moderators complete their review, they notify the reporting user whether they took enforcement action.[69]

Unfortunately, the internal metrics for how Twitter decides to enforce its policies are still unclear.[70] Twitter's enforcement policies say that, when determining whether to take enforcement action, its moderators considers a number of factors, including whether "the behavior is directed at an individual, group, or protected category of people; the report has been filed by the target of the abuse or a bystander; the user has a history of violating [Twitter's] policies; the severity of the violation; [and] the content may be a topic of legitimate public interest."[71] But Twitter has not explained how these factors are used in actual practice or how they determine the level of enforcement moderators use. Instead, Twitter has stated that it gives its moderators broad discretion over its response to reports.[72] As a result, whether Twitter's policies are implemented properly is determined by its moderators on a case-by-case basis. Furthermore, Twitter has provided no data on the trainings that its moderators receive or its reporting process.[73] The public

---

[68] *Our Range of Enforcement Options*, TWITTER, https://help.twitter.com/en/rules-and-policies/enforcement-options (last visited May 3, 2019).

[69] *Id.*

[70] AMNESTY INT'L, *supra* note 26, at 36.

[71] *Id.* at 35.

[72] *Id.* at 70 ("Our review teams are empowered to use their judgement and take appropriate action on accounts that violate our rules.").

[73] *Id.* at 36. Twitter has recently provided its reporting response rate for January–June of 2018, but without the context behind why certain accounts where actioned upon and others were not, it is hard to know whether Twitter is implementing its policy effectively. *Twitter Rules Enforcement*, TWITTER, https://transparency.twitter.com/en/twitter-rules-enforcement.html#twitter-rules-enforcement-jan-jun-2018 (last visited May 3, 2019). During the recorded time period, there were 2,814,940 reported accounts of abuse and 2,698,613 reported accounts of hateful conduct, while 248,629 accounts received some form of enforcement action for abuse and 285,393 for hateful conduct. *Id.*

has no way of knowing whether Twitter is doing its due diligence in implementing its policies.

If Twitter suspends a user's account or temporarily prevents a user from posting pending the removal of a violative Tweet, that user can appeal the enforcement action.[74] However, Twitter has no policy on how to respond to appeals and provides no information on how the appeals process functions.

Twitter users are skeptical that Twitter is consistently enforcing its policies and often express dissatisfaction with its service. Amnesty International found that, among Twitter users, 22% of women in the U.S. and 43% of women in the U.K. believe that Twitter's response to abuse or harassment is inadequate.[75] In 2016, a Buzzfeed survey found that Twitter failed to take action on 92.8% of respondents' last report of abusive behavior.[76] This narrative has continued in recent years. In January 2018, U.K. journalist Ash Sarkar posted a screenshot of a Tweet she reported that Twitter found did not violate its abuse or hateful conduct policies.[77] Despite the use of derogatory language, gendered slurs, and insults targeted at Sarkar's membership in a category protected by Twitter's rules, Twitter moderators took no action on the Tweet. In her post containing the screenshot, Ms. Sarkar noted that "[t]his is the tip of the iceberg when it comes to racist abuse I've

---

[74] *Our Range of Enforcement Options*, TWITTER, https://help.twitter.com/en/rules-and-policies/enforcement-options (last visited May 3, 2019).

[75] AMNESTY INT'L, *supra* note 26, at 37.

[76] Charlie Warzel, *"It Only Adds to the Humiliation" — How Twitter Responds to Harassers*, BUZZFEED NEWS (Sept. 22, 2016, at 9:01 PM), https://www.buzzfeednews.com/article/charliewarzel/after-reporting-abuse-many-twitter-users-hear-silence-or-wor.

[77] Ash Sarkar (@AyoCaesar), TWITTER (Jan. 18, 2018, 3:58 AM) https://twitter.com/AyoCaesar/status/953959860403687424.

had since yesterday. Not one complaint has been upheld. Twitter needs to fix up and

protect its users."[78]

Figure 1: Tweet sent to Ash Sarkar

Additionally, some users have experienced Twitter taking action on posts they

made in response to harassment or in support of Twitter's protected categories.[79] In

August 2017, a prominent Twitter user was suspended for posting screenshots of an email

he received that contained harassing and threatening content.[80] In September 2018, an

MIT researcher was briefly suspended for quoting feminist academic research that was

critical of anti-women sentiment in STEM and the sexist gendering of scientific fields.[81]

In February 2019, users found themselves caught in the crossfire when Twitter attempted

---

[78] *Id.*

[79] *See* Monique Judge, *Twitter Has a Serious Harassment and Abuse Problem but Doesn't Seem to Want to Cure It*, ROOT (Oct. 30, 2017, 4:59 PM), https://www.theroot.com/twitter-has-a-serious-harassment-and-abuse-problem-but-1819979725.

[80] Rob Beschizza, *Popehat Suspended from Twitter for Sharing a Threat He Received*, BOINGBOING (Aug. 3, 2017, 5:26 PM), https://boingboing.net/2017/08/03/popehat-suspended-from-twitter.html; Ken White, *How I Got Barred from Posting on Twitter*, POPEHAT (Aug. 3, 2017), https://www.popehat.com/2017/08/03/how-i-got-barred-from-posting-on-twitter/.

[81] Chris Peterson, *Twitter Suspended Me for Tweeting Feminist Academic Research. Here's Why That's a Problem.*, CIVIC MEDIA (Sept. 29, 2018), https://civic.mit.edu/2018/09/29/twitter-suspended-me-for-tweeting-feminist-academic-research-heres-why-thats-a-problem/; Cory Doctorow, *Twitter Suspends Academic Who Quoted Feminist STEM Research*, BOINGBOING (Sept. 29, 2018, 12:06 PM), https://boingboing.net/2018/09/29/platform-censorship.html.

to crack down on a harassment campaign targeting journalists.[82] In response to a harassment campaign targeting recently laid off journalist, Twitter removed Tweets directed at journalists that contained the line "learn to code."[83] Users who were Tweeting the phrase at the harassers as a countermeasure also found themselves subject to account restrictions.[84] Twitter not only fails to consistently remove harassing content, but it also enforces the rules against users who are acting within its conduct policies. This unjust enforcement further erodes users' trust in Twitter's reporting system.

Not only is Twitter's enforcement of its rules uneven, but Twitter users with connections to Twitter senior staff have found that communicating with the staff directly about harassment more consistently results in Twitter dealing with the harassing behavior.[85] One prominent U.K. feminist noted: "When I reported things to Twitter, it very rarely resulted in anybody being suspended. But when I was put in touch with someone who was higher up in the company, they took action and removed the harassment."[86]

This discrepancy between the actions of the ground-level moderators and the actions of the "higher ups" suggests that the reporting process is not working. Contacting Twitter staff directly to address harassment is circumventing the intended reporting

---

[82] Amber Athey, *Twitter Restricts People Who Tweet "Learn to Code" — Even If They Aren't Engaging in Harassment*, DAILY CALLER (Feb. 13, 2019, 5:31 PM), https://dailycaller.com/2019/02/13/twitter-restricts-learn-to-code/; Jon Levine (@LevineJonathan), TWITTER (Jan. 28, 2018, 9:38 AM), https://twitter.com/LevineJonathan/status/1089940895418466305.

[83] Athey, *supra* note 82.

[84] *Id.*

[85] AMNESTY INT'L, *supra* note 26, at 40.

[86] *Id.*

process and creating an informal dispute resolution system. But only people with connections to staff can access this informal system. The result is that a small number of privileged users can expect Twitter to deal with the harassment they receive, thereby making them more comfortable to continue using Twitter, while normal users are more susceptible to being silenced by harassment and pushed off the service.

Without data on how Twitter responds to reports of abuse and hateful conduct, it is difficult to know the extent to which Twitter fails to uphold its policies. Regardless, the belief that Twitter cannot be trusted to respond to reports and uphold its harassment policies is pervasive and affects users' interactions with the reporting system.[87] Because the reporting system does not appear to result in responsive action, users are discouraged from utilizing the reporting system in the first place.[88] As a result, the reporting system helps fewer people and Twitter's anti-harassment policies become even less effective.

Despite Twitter's policy and reporting changes in recent years, the user experience on Twitter remains as frightful as ever, and users do not seem optimistic that their experience on Twitter will improve. While refining the reporting system has likely made life on Twitter marginally better, the reporting system will only ever be a band-aid

---

[87] Shérazade Faynel, *A Case of 'Double Standards' in Twitter's Enforcement Policy?*, JOURNAL INTERNATIONAL (Mar. 13, 2018), http://www.lejournalinternational.info/en/deux-poids-deux-mesures-lapplication-de-politique-de-twitter/; Charlie Warzel, *Twitter Is Still Dismissing Harassment Reports And Frustrating Victims*, BUZZFEED NEWS (July 18, 2017, at 4:39 PM), https://www.buzzfeednews.com/article/charliewarzel/twitter-is-still-dismissing-harassment-reports-and; AMNESTY INT'L, *supra* note 26, at 40–41; Jake New, *Twitter Still Isn't Doing Enough about Rape Threats against Women*, TEEN VOGUE (Apr. 20, 2017), https://www.teenvogue.com/story/twitter-petition-rape-threats.

[88] AMNESTY INT'L, *supra* note 26, at 44. Amnesty International quotes Twitter users expressing this narrative: "When possible, I report it. By possible, I mean if I have the time and the emotional bandwidth to be disappointed about the report." *Id.* "I don't bother reporting anymore for myself because it doesn't matter." *Id.* "[I]t didn't occur to me to report [the message] to Twitter itself because I know that there are journalists who have experienced online abuse and whenever they've reported things to Twitter they almost invariably don't get a positive response and the materials tends to stay up." *Id.*

for Twitter's toxic culture. A more robust dispute system cannot change the fact that Twitter was founded on and continues to run on the principle that "everyone should have the power to create and share ideas and information instantly, without barriers."

Community without barriers will always tend toward the loudest and angriest voices. What Twitter's inability to solve harassment through reporting and moderation shows us is that Twitter, as a communication tool, was not designed with human behavior in mind.

CHAPTER IV

HUMAN BEHAVIOR AND THE OBSTACLES TO

BUILDING COMMUNITY ON TWITTER

Twitter's early designers failed to realize a platform in which users can share

information instantly, without barriers, is not a design goal that creates a platform that

contains the kinds of experiences that users wish to receive. There are many

psychological principles that explain why people will tend toward competition rather than

cooperation when given a totally open platform where there are no barriers to

communication. In order to fix its toxic culture, Twitter needs to address these behavioral

obstacles to cooperation and begin thinking about how to build community on its

platform.

A. Community, Lifestyle Enclaves, and Intergroup Contact

Instead of a platform with no barriers, Twitter should be looking to create

community and enforce pro-social community norms. Community is a "network of

ongoing relatively stable relationships among people holding diverse views, but with at

least some base of shared values and ethical norms; some degree of caring, trust, and

collaborative activity; working through channels of communication; and carrying out

certain ritual-like activities that have the effect of affirming the relationships."[89]

Because Twitter is so large and the boundaries between users nonexistent, users

have minimal tools to create community for themselves. Social groups on Twitter are

cobbled together by loose connections of Twitter users who follow and support each

other. There is no defined boundary for what defines a group, and groups may

---

[89] Terry L. Cooper, *Building Ethical Community*, 41 AM. REV. FOR PUB. ADMIN. 3, 5 (2011).

temporarily overlap in response to certain events. Users may create hashtag movements to help define group identity and form community, like the #GirlsLikeUs movement, through which trans activists used hashtags to connect trans users across Twitter.[90] "This rapport building between members of the network, regardless of the status, does important community-building work, shaping cultural solidarity and providing important emotional and psychological support."[91]

Unfortunately, these communities are ephemeral, lasting only as long as the hashtag remains popular, and, more importantly, hashtags give members of the community no control over who engages with the community. Anyone on Twitter can add #GirlsLikeUs to their posts, which could include users who are not trans, who do not wish to be considered part of the Twitter trans community, or worse, could be actively attempting to abuse and harass trans users.

Because Twitter does not give users the ability to create explicitly defined, lasting communities, instead users often create "lifestyle enclaves," which are "fundamentally segmental and celebrate[] the narcissism of similarity[,] . . . usually explicitly involv[ing] a contrast with others who do not share one's lifestyle."[92] It is easy to create and maintain lifestyle enclaves on Twitter. All a user need do is follow people on Twitter who talk about a subject the user is interested in, and they will begin to receive messaging from those users that validate their lifestyle. For example, if a Twitter user likes video games

---

[90] Sarah J. Jackson, Moya Bailey & Brooke Foucault Welles, *#GirlsLikeUs: Trans Advocacy and Community Building Online*, 20 NEW MEDIA & SOC'Y 1868, 1876 (2017).

[91] Cooper, *supra* note 89 (citing Moya Bailey, *#transform(ing)DH Writing and Research: An Autoethnography of Digital Humanities and Feminist Ethics*, 9 DIGITAL HUMAN. Q. (2015)).

[92] *Id.* (citing ROBERT N. BELLAH, WILLIAM M. SULLIVAN, RICHARD MADSEN, ANN SWIDLER & STEVEN M. TIPTON, HABITS OF THE HEART: INDIVIDUALISM AND COMMITMENT IN AMERICAN LIFE 72–73 (1st ed. 1985)).

they might decide to follow popular Twitch streamer PewDiePie. PewDiePie, in turn, will post his own content and retweet content that he supports or enjoys, all of which shows up on the user's Twitter feed. As a result, that user might follow other accounts PewDiePie has retweeted and get additional video game content. The user is connecting with other people, but only among the demographics that enjoy PewDiePie. This form of connection does not require the user to step outside their narrow and discrete interest in PewDiePie-related video game content.

While lifestyle enclaves can be useful sources of support in one's private life, they do not help people bridge differences or create dialogue between disparate groups. Even worse, when lifestyle enclaves are not tempered by real community building, it can lead to the enclave lashing out against outgroup members, as seen in Gamergate.

In order to move from lifestyle enclaves to communities, Twitter should consider what kinds of processes lead to less conflictual communication. Despite Twitter's ability to instantly connect disparate groups from almost anywhere in the world, it is a poor tool for dialogue. The reason for this is in the kind of contact that Twitter users experience with each other. Intergroup contact theory suggests that increased contact with individuals from out-groups can result in decreased prejudice, but only in situations where five optimal conditions are met: (1) equal status of the groups in the situation, (2) common goals, (3) intergroup cooperation, (4) the support of authorities, law or custom, and (5) contact that is extensive and repeated in a way that allows individuals to become friends.[93] The participants' choice in engaging in contact is also a moderator to positive

---

[93] Thomas F. Pettigrew, Linda R. Tropp, Ulrich Wagner & Oliver Christ, *Recent Advances in Intergroup Contact Theory*, 35 INT'L J. OF INTERCULTURAL REL. 271, 273 (2011); Thomas F. Pettigrew, *Intergroup Contact Theory*, 49 ANN. REV. PSYCHOL. 65, 76 (1998).

intergroup contact.[94] However, when these conditions are not met, intergroup contact

increases prejudice rather than decreasing it.

Optimal intergroup contact requires parties to spend extended time with each

other, wherein intergroup contact participants go through a process of decategorization,

salient group categorization, and recategorization to generalize positive feelings about

specific outgroup members to all outgroup members.[95] That is, intergroup contact leads to

reduced prejudice more frequently when intergroup members first meet in conditions

where group saliency is low (decategorization).[96] This allows intergroup members to

form personal bonds that are tied interests rather than group membership.[97] Once

personal bonds are formed, effective intergroup contact requires members to see

themselves as representatives for their group (salient group categorization).[98] With time,

intergroup members begin to perceive their interactions from a larger group perspective

(recategorization).[99]

In the case of Gamergate, even though the Twitter provided Gamergate supporters

and critics convenient methods for communicating with one another, it did not reduce the

prejudice between the two groups. Prejudice remained high because the optimal

conditions for intergroup contact had not been met. First, the groups did not have equal

status in the situation. When a single person gets a deluge of hateful messages day-in and

---

[94] Pettigrew et al., *supra* note 93, at 275–76.

[95] Pettigrew, *supra* note 93, at 77 fig.2.

[96] *Id.* at 74.

[97] *Id.*

[98] *Id.*

[99] *Id.* at 75.

day-out from many anonymous sources, the single person does not have the same power and control over the situation that the harassers do. A small number of concerted harassers can consume their target's day and make them fear for their life, while the target can do little in response.[100]

Second, while the two sides may have had some common goals regarding the enjoyment of video games and support for the industry, they perceived their salient goals as incompatible. Gamergate critics vocally supported increased diversity for game developers and character representation in games, while Gamergate supporters were frequently dismissive of the need to intentionally support diverse creators and create diverse characters. Gamergate supporters wanted to see games media make dramatic changes to how they dealt with and disclosed ethical concerns, while Gamergate critics dismissed the need for a serious focus on the personal relationships between game developers and game journalists.

Third, there was no opportunity for intergroup cooperation. When creators are putting their work on the Internet they are not working with commenters on the creation of a project; they are not building something tangible together. Harassers take this a step further by showing either negative or no interest in the creator's work. When they are fighting against the perceived injustice perpetrated by their enemy, people are not interested in cooperation. Social media does little in the way of providing social groups opportunities to cooperate. Twitter, where most of Gamergate conflict took place,

---

[100] While users can control who can directly message them, this does not ameliorate the flood of public comments they may receive. *About Direct Messages*, TWITTER, https://help.twitter.com/en/using-twitter/direct-messages#basics (last visited May 15, 2019). Currently, users cannot control who responds to their public posts without blocking individual users or setting their account to private.

abbreviates language and encourages bold statements, but does not provide systems through which users can meaningfully collaborate.

Fourth, the targets of harassment were given no choice as to whether they wished to engage in contact with their harassers. Many targets of harassment on Twitter use the service to publicize their work, which leaves them vulnerable to attack. They cannot make their accounts private without losing publicity and money, but by remaining public they cannot easily stop other Twitter users from sending them messages. This lack of choice, combined with the overwhelming deluge of harassment Quinn and Sarkeesian received, put them in an emotionally vulnerable place that heightened their anxiety and reduced their empathy for Gamergate supporters and, in turn, increased their prejudice against their attackers.

Finally, Twitter does not provide a space for members of different groups to become friends. Twitter is often a hostile place, with little chance of connection between people who do not already agree with each other. It is difficult to engage in the kind of communication that follows the proper sequence of a decategorization response, salient group categorization, and recategorization response. Instead, communication on Twitter often frontloads salient group categorization and, as such, there is little opportunity to develop interpersonal bonds based on common interests before salient group dynamics produce intergroup conflict.

Because Twitter has prioritized free speech over community since its conception, it has created a platform without addressing the socio-psychological principles that lead to intergroup conflict. The next section examines the psychology behind intergroup

conflict on Twitter and uses Gamergate to show how Twitter's failure to address these principles has exacerbated intergroup conflict and neglected community-building.

B. Why Intergroup Contact Has Failed to Reduce Prejudice on Twitter

One of the reasons the extensive intergroup contact on Twitter is often conflictual because social groups tend to vie for status as if it were a limited resource. Social Identity Theory (SIT) posits that people try to maintain positive self-concept through their affiliation with social groups, which can be achieved through positive comparisons to their ingroup or negative comparisons to an outgroup.[101] SIT also predicts that if someone has a negative affiliation with their ingroup, they will either seek to distance themselves from that ingroup identity or work to improve their image of their ingroup.[102] SIT is predicated on the idea that human beings perceive status as a limited resource and engage in conflict to improve status, just as they engage in conflict to secure physical resources.[103]

As stated in in section I, Gamergate supporters had a strong sense of group identity associated with being "gamers." Attached to this identity was the idea that gamers as a social group are looked down on, that they had hollowed out a niche hobby for themselves that was uncool and derided by popular culture.[104] To maintain positive

---

[101] Rupert Brown, *Social Identity Theory: Past Achievements, Current Problems and Future Challenges*, 30 EUR. J. SOC. PSYCHOL. 745, 747 (2000).

[102] *Id.*

[103] *Id.* at 748.

[104] David Wong, *5 Reasons It's Still Not Cool to Admit You're a Gamer*, CRACKED (May 24, 2010), https://www.cracked.com/article_18571_5-reasons-its-still-not-cool-to-admit-youre-gamer.html ("I spent years putting up with the 'gamers are pale loners crouched in the dark among Mountain Dew bottles and pizza boxes' stereotype[.]"); Ricardomz, *Why Are Videogames So 'Uncool' on the Social Media?*, GAMESPOT, https://www.gamespot.com/forums/offtopic-discussion-314159273/why-are-videogames-so-uncool-on-the-social-media-29357431/ (last visited May 14, 2019) ("I think that everyone has the vision that people who play videogames have terrible social lives [sic] . . . ."); MDK12345, *Afraid of Getting*

28

self-concept, self-identified gamers were highly motivated to improve the image of their ingroup. Through Gamergate, they attempted to do so by creating negative comparisons to an outgroup: women.

The increased diversification of video games caused gamers to view women as an outgroup with which gamers needed to contend for social status. As video games became more popular and made more money, their audience expanded. Women started playing video games in increasingly higher percentages and became more involved with video game development and critique. However, popular understanding of video game culture still, including that of gamers, held to the idea that mostly men play video games, and big budget games were still primarily targeted at a male demographic. This disconnect created the sense that women were outsiders intruding in the gamer social sphere, so their interest in games, even the same games as men, did not make them part of the gamer ingroup.

Therefore, games like Quinn's *Depression Quest* were seen as not only a violation of a social norm, but as a status threat from an outside group encroaching on gamer territory. This threat was then compounded by the game's support from games media. The criticism of *Depression Quest* can also be analyzed as the gamer ingroup degrading the work of an outgroup member to maintain positive self-concept regarding the games that they like and think are legitimate.

Gamergate further exacerbated group conflict by creating a widespread story from which it was easy to make negative outgroup comparisons. Gamers had an incentive to

Judged for Being a Gamer, GAMESPOT, https://www.gamespot.com/forums/games-discussion-1000000/afraid-of-getting-judged-for-being-a-gamer-31167494/ (last visited May 14, 2019) ("I'm going to be 30 sooner or later and I really worry about what girls think of me if they were to see my gaming collection.").

share the story of Quinn's alleged infidelity because they could use the story reduce the status of an individual outgroup member and, by extension, the entire group. The ensuing conflict occurred as a result of gamers struggling to improve their perceived status.

Gamergate supporters' response to feminist criticism of video games, like Sarkeesian's, can also be seen as an attempt to maintain positive self-concept by attacking the legitimacy of an outgroup member's criticism of their hobby. As mentioned above, gamers have a strong identity associated with what games gamers like and what kinds of games are seen as legitimate. By criticizing popular games, gamers saw Sarkeesian's videos as a status threat in need of strong response in order to maintain their sense of quality for popular games and, by extension, the quality of the hobby itself. Ingroup members used the degradation of Quinn, Sarkeesian, and other women to bolster their own sense of positive self-concept.

Gamergate's social identity conflict was intensified by the historical divisions of men and women in video game culture. Women tend to receive more harassment than men in video game culture because women are seen as intruding on a traditionally male space.[105] Expectation states theory predicts that, because video games are a male-dominated space, expectations for masculine behavior would be dominant.[106] Men are expected to fulfill a public role while women are expected to fulfill a private, domestic role.[107] Masculinity is correlated to favoring sexual harassment and disfavoring gender

---

[105] Jesse Fox & Wai Yen Tang, *Sexism in Online Video Games: The Role of Conformity to Masculine Norms and Social Dominance Orientation*, 33 COMPUTERS HUM. BEHAV. 315 (2014).

[106] *Id.*

[107] Megarry, *supra* note 24, at 48.

equality.[108] Therefore, when women wish to make their voices heard, it is considered transgressive behavior and is subject to attack and harassing behavior.[109] Social Dominance Orientation (SDO) is a measure of a person's endorsement of inequality, leading to prejudice and the belief of one group's superiority over others.[110] Because "men often outnumber women in networked video games and that masculine behavior is typically rewarded, this may provide men with the opportunity to express social dominance in the virtual world in a way they cannot in the physical world."[111]

The social identity model of deinvidiuation effects (SIDE) predicts that people lose their sense of personal identity in anonymous spaces and defer to salient group membership.[112] The salient group membership in video games is typically male and masculine, which further encourages masculine behavior. The loss of personal identity also enables group members to engage in anti-normative behavior.[113] Twitter retains the anonymity and lack of individual identity from online gaming and creates spaces where group membership can be very important. Without community norms enforcing pro-social behavior, SIDE could lead to enclaves creating and supporting anti-normative behavior.

Once the two sides in the Gamergate conflict had solidified, the two groups were psychologically predisposed to find their own positions reasonable and well-supported

---

[108] Fox & Tang, *supra* note 105, at 315.

[109] Megarry, *supra* note 24, at 48.

[110] Fox & Tang, *supra* note 105, at 316.

[111] *Id.* at 318.

[112] *Id.* at 316.

[113] *Id.*

and the other side's positions unreasonable and biased. The inability for each group to see

the other side's perspective is a psychological process called naïve realism. Naïve realism

is the theory that people believe they see the world objectively and that, therefore,

reasonable other people will see and respond to similar circumstances similarly.[114]

Studies have shown that people are likely to think their own beliefs are supported by

objective considerations, such as the use of logic and reason, and think that beliefs that

differ from their own are attributable to non-objective factors or biases.[115]

When other people do not respond to circumstances similarly, people tend to

assume there is a situational difference between observers that can be overcome by

additional information.[116] When attempts to overcome situational differences do not

create agreement, people often assume dispositional differences, such as the other person

being selfish, stupid, or hopelessly biased, to explain the disagreement.[117] People may

engage in this behavior because they are motived to maintain a positive self-concept and

because they take cues from their social groups as to which perceptions are correct.

Naïve realism played a significant role in the continuation of the Gamergate

conflict. Gamergate supporters frequently assumed that women like Quinn and

Sarkeesian were too stupid to understand why video games were not sexist or were just

whining about small slights.[118] The idea that reasonable, intelligent people could support

---

[114] Emily Pronin, Thomas Gilovich & Lee Ross, *Objectivity in the Eye of the Beholder: Divergent Perceptions of Bias in Self Versus Others*, 111 PSYCHOL. REV. 781, 783 (2004).

[115] *Id.* at 789.

[116] *Id.* at 783.

[117] *Id.*

[118] *See* Jef Rouner, *8 "Criticisms" of Anita Sarkeesian That Are Utter Bullshit*, HOUSTON PRESS (Oct. 21, 2014, 7:00 AM), https://www.houstonpress.com/arts/8-criticisms-of-anita-sarkeesian-that-are-utter-

diversity and feminism and also enjoy video games was incompatible with the worldview

of Gamergate supporters. This made it difficult for members of the movement to engage

substantively with the arguments of Gamergate critics because they assumed critics were

speaking credulously or disingenuously. This worldview made it easier to disregard the

other side by using language that attacked critics' competence or character, rather than

the substance of their arguments.

Naïve realism also created the perception that the one side's arguments were

totally incompatible with the other's. Gamergate supporters' narrative was pro "ethics in

games journalism" and, because they were fighting against Gamergate critics, they

framed their critics as anti "ethics in games journalism." When prominent games

journalist Leigh Alexander wrote a strongly worded article entitled *"Gamers" Are Over*,

criticizing the use of a consumerist hobby as a social identity, Gamergate supporters

rallied to get Intel to withdraw advertisements on Gamasutra, where the article was

published.[119] The movement took this action based on the belief that Alexander's article

was insulting and degrading to gamers, not on the basis of any unethical practices related

to the article. Gamasutra had published articles related to ethics in games journalism in

the past, but that was irrelevant to Gamergate supporters working to take away an

important source of its funding. Anti-Gamergate agreed that games journalists should be

ethical,[120] but the perception that Gamergate critics were the enemy motivated Gamergate

---

bullshit-6382966; Gabenus Trollucus, *Zoe and Anita in Dumb and Dumber*, KNOW YOUR MEME (Aug. 12, 2015, 9:11 AM), https://knowyourmeme.com/photos/1004399-zoe-quinn.

[119] Wagner, *supra* note 2.

[120] RPS Hivemind, *Hello: Videogames Are For Everybody*, ROCK PAPER SHOTGUN (Sept. 8, 2014, 1:00 PM), https://www.rockpapershotgun.com/2014/09/08/videogames-are-for-everybody/.

supporters to see everything the other side did as wrong and incompatible with their own

beliefs.

As the two sides struggled for status, viewing the opposing side as fundamentally

unreasonable, they entered an escalating conflict spiral.[121] In a conflict spiral, participants

reciprocate each other's aggressive actions.[122] As participants experience aggressive

behavior from each other, they negatively readjust their perceptions of the other.[123] On an

intergroup level, the conflict spiral predicts that as social groups respond with aggressive

behavior they become increasingly bitter and caustic in their communications with each

other.[124] As such, two people communicating over Twitter may never have interacted

before, but their aggressive communication may be a reflection of the vitriol that they

have experienced from other members of the participants social group. The conflict spiral

is self-continuing because participants predisposed to perceiving each other negatively

are more likely to blame each other, perceive ambiguous actions as more threatening,

have reduced inhibitions against retaliation, and avoid those who they perceive

negatively.[125] These negative attitudes reduce empathy.[126] Pre-existing social bonds

encourage participants to give a little in their position and think of conflict like problem

---

[121] Raymond A. Friedman & Steven C. Currall, *Conflict Escalation: Dispute Exacerbating Elements of E-mail Communication*, 56 HUM. REL. 1325, 1330 (2003).

[122] *Id.*

[123] *Id.*

[124] Pettigrew et al., *supra* note 93, at 273.

[125] Friedman & Currall, *supra* note 121, at 1330.

[126] *Id.*

solving.[127] Communication on Twitter, on the other hand, very frequently occurs between people who have never communicated with each other before.

Within the conflict spiral, the groups on either side of Gamergate began to regard each other as enemies and engaged in the breakdown of empathy through delegitimization. Delegitimization is the term for a psychological process in which ingroups deny outgroups moral concern.[128] It encompasses the concepts of dehumanization and infrahumanization.[129] Dehumanization is the theory that ingroup members develop enemy images of certain outgroups and label them as inhuman and not worthy of moral consideration.[130] For example, during World War II, the Nazis referred to Jews as diseases or disease-carrying vermin in order to relegate the Jewish people to subhuman status;[131] while during Rwandan genocide Hutus called the Tutsis cockroaches.[132] Members of dehumanized outgroups are not seen as deserving of basic human rights, leading to the justification of violence toward or extermination of the outgroups. Infrahumanization is a milder phenomenon that occurs when people treat outgroup members as less human than ingroup members.[133]

---

[127] *Id.*

[128] Neta Oren & Daniel Bar-Tal, *The Detrimental Dynamics of Delegitimization in Intractable Conflicts: The Israeli–Palestinian Case*, 31 INT'L J. INTERCULTURAL REL. 111, 112 (2007).

[129] *Id.*

[130] Michelle Maiese, *Dehumanization*, BEYOND INTRACTABILITY (Jul. 2003), https://www.beyondintractability.org/essay/dehumanization.

[131] DAVID LIVINGSTONE SMITH, LESS THAN HUMAN 15 (2011).

[132] *'Less Than Human': The Psychology of Cruelty*, NPR (Mar. 29, 2011, 1:00 PM), https://www.npr.org/2011/03/29/134956180/criminals-see-their-victims-as-less-than-human.

[133] Nick Haslam & Steve Loughnan, *Dehumanization and Infrahumanization*, 65 ANN. REV. PSYCHOL. 399, 402 (2014).

The stereotype content model of dehumanization suggests that people think of group stereotypes along dimensions of warmth and competence.[134] This model predicted that groups who were seen as low on both dimensions are targets of dehumanizing language. Additionally, infrahumanization processes may be applied to women in circumstances in which women are sexualized, making them seem "lacking human nature, as well as warmth, morality, and competence."[135]

For Gamergate supporters, social identity dynamics created the sense that women in gaming culture were part of an outgroup that was engaged in the violation of traditional video game culture. The need to enforce social norms and improve group status caused Gamergate supporters to form enemy images of Quinn, Sarkeesian, and the people who supported them. As a result, Gamergate supporters frequently used sexualized epithets to denigrate both Quinn and Sarkeesian. Quinn's sexuality, in particular, was frequently used to discredit both her and her work. In the Gamergate narrative, she was someone who slept around to get good reviews for her game, and she only received critical praise because of her sex, not because of quality of her product. The blogpost about Quinn and the ensuing harassment included allegations that denied her vital characteristics and fed into stereotypes of women and femininity. Gamergaters denied Quinn's competence by asserting that she was incapable of making a video game that was successful on its own merits. They denied Quinn's morality by alleging that she slept with journalists in exchange for coverage. They denied Quinn's warmth by accepting and propagating her ex-boyfriend's claim that she emotionally abused him. Not

---

[134] *Id.* at 403–04.

[135] *Id.* at 408.

only did these allegations dehumanize Quinn, they specifically targeted qualities that masculine culture values in women, particularly kindness and chastity.

Gamergate supporters even took to referring the Quinn as "Literally Who."[136] They claimed this was an attempt to get people to stop focusing on Quinn the person and instead focus on the "real issue," ethics in games journalism, but in fact it served to deny Quinn the validity of a real name as attacks against her from the movement continued unabated. Using "Literally Who" as a signifier delegitimized Quinn as a person: the words not only covered up her identity, but also showed disdain and incredulity at her prominence and popularity.

This delegitimization of Quinn served as a motivator for mobilization among Gamergate supporters. Calling their enemies names like "Literally Who" shows other people on Twitter what side you are on, who the group dislikes, that the other side is not deserving of moral consideration, and that attacks on the other side are necessary and deserved. Delegitimization also helped provide an explanation to Gamergate supporters as to why people like Quinn were popular. Quinn was a social status threat, and sexualizing her, accusing her of infidelity and ethical violations were ways to explain why someone deemed incompetent was still successful: she cheated and manipulated her way to the top. Her harassment was a direct outcome of this rationalization. If she did not come upon her success legitimately, then it was only fair to bring her down to her rightful place. While no physical violence occurred, delegitimization created a social dynamic that justified continuous, violent harassment directed at Quinn to get her to leave the games industry.

---

[136] Singal, *supra* note 19.

The conflict spiral of Gamergate was further exacerbated by nature of communication on the Internet.[137] Internet communication lacks a key component of empathy creation called "grounding."[138] Grounding is "the process by which two parties in an interaction achieve a shared sense of understanding about a communication and a shared sense of participation in the conversation."[139] Grounding allows people in communication feel as they are being heard and understood by the other parties. In an in-person interaction, there are six methods of grounding: 1) co-presence, allowing participants to see and hear the others' environment and experience what they are experiencing; 2) visibility, allowing the participants to see each other; 3) audibility, allowing the participants to hear vocal nuances, such as timing and intonation; 4) co-temporality, allowing participants to hear the other at the time that they speak; 5) simultaneity, allowing participants to both send and receive messages at the same time; and 6) sequentiality, allowing participants to communicate in turn without messages getting out of order.[140] Communication over social media does not allow for any of these six methods.

In addition, communication over social media has four major characteristics that make it more likely to produce conflict than in-person communication[141]: diminished feedback, minimal social cues, length, and excess attention. Due to the asynchronicity of

---

[137] Friedman & Currall, *supra* note 121, at 1326.

[138] *Id.* at 1328.

[139] *Id.*

[140] *Id.*

[141] Friedman and Currall propose the Dispute-Exacerbating Model of E-mail (DEME), which, although created before the rise of social media, fits well with modern uses of social media. *Id.* at 1331. I am adapting the model to examine social media's effect on conflict.

the medium, when participants communicate with one another, they are not able to give

or receive feedback in real time, nor do they receive the nuances of feedback that occur

during in-person communication.[142] As a result, participants are not given the opportunity

for self-correction and may interpret each other's communications more negatively than

they intended, leading to conflict.[143] Online interaction lacks facial expressions and

verbal intonation, making it more difficult for participants to read social cues.[144] Without

social cues, participants do not care as much about getting members of outgroups to like

them, which, in turn, encourages conflict.[145] Relational values are reduced and

communication becomes focused on argument.[146] The lack of social cues also reduces the

influence of politeness norms in an interaction[147]: people are often more blunt or more

aggressive on social media than in person. Communication over social media also occurs

in a different environment than in-person communication. People engage in social media

while they are isolated from the other participants in the communication, reducing

empathy toward each other.[148]

　　　　The length of the messages over social media also affects conflict escalation. On

Twitter, messages are restricted to 280 characters. The limited space for messages often

results in communication that is clipped and abbreviated, which makes the intended

---

[142] *Id.* at 1333.

[143] *Id.*

[144] *Id.* at 1335

[145] *Id.*

[146] *Id.*

[147] *Id.* at 1336.

[148] *Id.* at 1329.

message more difficult to understand and creates more opportunities for misunderstanding if important words are left out to fit the given space. The shorter message length does not fit with communication norms: not enough information is provided in a single Tweet, and a series of Tweets can present information in a way that is cluttered and hard to follow. Further, participants, faced with excessive information but given limited space, are more likely to pick and choose points from another's message that feel most important and heated to them—thereby causing participants to discuss the subjects that are most likely to create conflict while ignoring points of agreement that might help alleviate it.[149]

Over the Internet, participants may give messages excess attention because they can review and revise their messages for as long as they wish. Focusing on negative comments and considering how to respond may cause participants to perceive problems as larger than they are, making them angrier than they otherwise might be and resulting in conflicts that are more difficult to resolve.[150]

Without grounding, participants must work harder to understand each other's communications.[151] Tweets, by forcing participants to expend more energy to understanding each other, reduce participants' ability to expend energy toward empathy and connection with each other.[152]

---

[149] *Id.* at 1338–39.

[150] *Id.* at 1339.

[151] Friedman & Currall, *supra* note 121 at 1328.

[152] *Id.* at 1330

In addition to the lack of grounding, the anonymity of many Twitter users makes it difficult for those using Twitter to set social norms.[153] If users are not identifiable, then they cannot be held accountable for their actions. Without accountability, anonymous users can act however they want, while public users are exposed to the pervasive risks of harassment that can leak into everyday life. Public users can do little to set the norms for anonymous users, so harassment continues unchecked.[154] One study showed that when participants felt anonymous, their emotional arousal increased their aggressive behavior.[155] Anonymity makes participants feel as though their actions will not have negative consequences, for how can one be held responsible if no one knows who acted?[156] Anonymity can be further misused by those engage in "sock puppetry," the practice where one user creates multiple accounts to make it seem like their message has greater support than it does.[157]

The anonymity of the Internet made the victims of Gamergate particularly susceptible to harassment. As discussed earlier in this section, the victims used Twitter to support their work and livelihood. They were not anonymous figures and depended on Twitter to interact with their audience and promote their work. Their attackers, on the other hand, were often anonymous and could attack these public figures with impunity. Furthermore, because of Twitter's early dedication to free speech and the hands-off nature of its moderation, anonymous users, not Twitter, created Twitter's social norms.

---

[153] James Grimmelmann, *The Virtues of Moderation*, 17 YALE J. L. & TECH. 42, 77 (2015).

[154] *Id.* at 78.

[155] *Id.*

[156] *Id.*

[157] *Id.*

These anonymous users developed their own norms for acceptable conduct that were more forgiving of abusive language than most in-person communities. Once these standards became entrenched, they became normalized and expected aspects of everyday communication. As a result, Twitter effectively ignored Gamergate's pattern of constant harassment because the community standards had been set so low.

CHAPTER V

WHAT CAN TWITTER DO?

Despite the pervasive and ongoing harassment of women on Twitter, it can do more to address abuse and make the Twitter community more inviting to vulnerable populations. The way to fix Twitter's toxic culture is to focus on real community-building. While Twitter has been attempting to bolster the effectiveness of its reporting process, it appears to think that these improvements will eventually fix the toxic culture of Twitter. They will not. While an effective reporting system is a worthy goal, it will not, by itself, change the way users interact with each other. For that, Twitter will have to carefully develop a whole new culture in conjunction with its users. Twitter cannot promote instant, widespread communication while also effectively protecting its users from harassment. The design goals of Twitter are to increase engagement with the service and the users on the service, and such goals do not recognize Twitter as a public space for community.

To set community norms, Twitter users need to experience more accountability for their actions, and the abuse reporting system needs to be more transparent. Twitter users who experience harassment are currently unable to effectively create or enforce social norms that function to keep people acting within the range of desirable behavior. Additionally, Twitter's current social norms encourage inflammatory behavior with more engagement and a higher follower count. A more transparent punishment system would let users know which actions are acceptable and which actions cross the line.

Furthermore, Twitter needs to make sure that owners of accounts banned for abuse face lasting consequences. Right now, it is trivially easy for the owner of a banned

account to simply set up new profile. Banning accounts does nothing to prevent anonymous users from continuing their abuse, while it has a huge impact on public users whose work and public life are tied to their accounts.

Twitter needs to ensure that its abuse report system does a better job of addressing both the emotional and practical needs of users who are experiencing harassment. Twitter could take a big step in ensuring that the people in charge of creating and maintaining the abuse report systems understand the realities of being a woman or a person of color on Twitter by hiring a more diverse workforce.

One way to improve the responsiveness of a system is to include the community in setting community standards and delivering punishment. Including the community in the creation of community standards would allow Twitter to use its reporting system to reinforce a new culture. Additionally, involving peers in the site's justice system creates the possibility of abusers being more open and understanding of the punishment they are receiving.[158]

An example of such a design comes from *League of Legends*, an online multiplayer video game. In *League of Legends*, players are frequently matched with a random selection of four other teammates and placed into a competitive match against another team of five. Because of the nature of the matchmaking system, a player's success in a given match is at the mercy of strangers' ability to play the game well. As a result, frustrated players often used the in-game chat to berate and harass teammates who they perceived as letting them down. In response to toxic player behavior, the developer of *League of Legends* implemented a number of changes. These changes created a system

---

[158] BAILEY POLAND, HATERS: HARASSMENT, ABUSE, AND ONLINE VIOLENCE 233–34 (2016)

in which players would be immediately notified of poor behavior through player

feedback at the end of the match.[159] Players are given a report card on their behavior

during the match, outlining any community standards they violated and the reasons for

their punishment, if needed.[160] The new system also required players to opt-in to the in-

game chat if they wanted to see or use it, resulting in a decrease in negative chat behavior

of 30 percent and a 30 percent increase in positive chat behavior compared to the week

before the changes.[161]

While these changes have shown improvements in player behavior in *League of

Legends*, there are some challenges to implementing such a system on Twitter. In *League

of Legends*, encounters with other users occur over a discrete timeframe, the length of a

game. On Twitter, it is more difficult to determine when interactions with other users

end. *League of Legends* games are also not public by default, the way Tweets are. Twitter

could implement a user behavioral rating system that allows users to judge the quality of

individual Tweets that then feeds into an overall player rating. However, the publicity

and permanence of Tweets would make such a system susceptible to "review bombing"

users who have not violated conduct policies but have upset particular groups of users.

---

[159] *New Player Reform System Heads into Testing*, LEAGUE OF LEGENDS, https://na.leagueoflegends.com/en/news/game-updates/player-behavior/new-player-reform-system-heads-testing (last visited May 8, 2019).

[160] POLAND, *supra* note 158, at 227 (citing Lauren Keating, *'League of Legends' Battles Harassment with Automated System That Could Ban Gamers*, TECH TIMES (May 26, 2015, 2:02 PM), http://www.techtimes.com/articles/55629/20150526/league-legends-battles-harassment-automated-system-ban-gamers.htm).

[161] Laura Hudson, *Curbing Online Abuse Isn't Impossible. Here's Where We Start*, WIRED (May 15, 2014, 6:30 AM), https://www.wired.com/2014/05/fighting-online-harassment/.

Another change *League of Legends* implemented was the inclusion of its players in its dispute resolution process through a jury-like system called the Tribunal.[162] In the Tribunal, player-volunteers are given the responsibility of reviewing abuse reports and determining whether reported players deserve punishment.[163] The Tribunal gives players the opportunity to engage in the setting of community norms and gives them a sense of agency and ownership over the actions that take place in their community. Player-involved punishment, combined with both an explanation of what the offending player did wrong and what they could do in the future to avoid further punishment, had a notable impact on player behavior. Players who received a three-day suspension had 13 percent fewer reports, and players who received a fourteen-day suspension had 11 percent fewer reports.[164] While these changes did not magically solve all the problems of abuse and harassment in *League of Legends*, the improvements provide evidence that communication design and an emphasis on community involvement can have a marked impact on the way people choose to communicate with one another.

In its current form, Twitter is a space for conflict, not dialogue. It is good at engendering moral conflicts, where the people in engaged in conflict "articulate the conflict in different ways because they have distinct value orientations toward what is important."[165] In moral conflicts, typical modes of rhetoric do not work. "Successfully persuading people to join the other side[] or compromising one's own values and virtues

---

[162] POLAND, *supra* note 158.

[163] *Id.*

[164] *Id.* (citing DANIELLE CITRON, HATE CRIMES IN CYBERSPACE 237 (2014)).

[165] J. Kevin Barge, *Dialogue, Conflict, and Community*, *in* THE SAGE HANDBOOK OF CONFLICT COMMUNICATION 517, 517 (John G. Oetzel & Stella Ting-Toomey eds., 2006).

is unlikely."[166] The resolution of moral conflicts requires an opportunity for people to engage in dialogue. In dialogue, participants do not try to convince the other side of their rightness, instead they "listen deeply to each other's moral orders, . . . explore the particular rationality that each uses, . . . and create new categories that allow the competing moral orders to be compared and weighed."[167]

The solution to change how intergroup contact functions on Twitter is to change how people communicate on Twitter. The design of a communication system "can inhibit and enhance certain kinds of speech over others."[168] What Twitter needs is the ability for its users to create communities like ones they find offline. In her article *Curbing Online Abuse Isn't Impossible. Here's Where We Start*, Laura Hudson writes: "Shouting racial slurs and rape threats at someone in public often has consequences, but on the Internet . . . it almost never does."[169] One of the primary reasons for this is Twitter's dedication to free speech over community norms.

> Twitter does not consider a threat against another user a violation of its terms of service unless that threat is "direct and specific." As one of the company's PR representatives, Jim Prosser, explains, "It's not just that something should happen to you; it's that something is going to happen to you. Where it will happen, from what, with what. Rather than just 'I hate you, go die in a fire.' You have something more specific there."[170]

Twitter wants to be a place where people congregate, share information, and engage in informed dialogue. But it seems unwilling to create a sense of shared

---

[166] *Id.*

[167] *Id.*

[168] SARAH JEONG, THE INTERNET OF GARBAGE 71 (1.5 ed. 2018).

[169] Hudson, *supra* note 161.

[170] *Id.*

community where this could happen. Communities "are defined by shared values rather than by the outer limits of the law."[171] Rather than having a complicated set of abusive behavior and hateful conduct policies (policies that explicitly except "offensive" speech from breaking the terms and conditions), Twitter "could take a strong and meaningful stand against harassment simply by applying the same sort of standards in their online spaces that we already apply in our public and professional lives."[172]

In order to reduce the negative effects of anonymity, Twitter could make anonymity conditional based on good behavior.[173] For example, all users could choose to be anonymous when creating their account, but if a user is flagged for harassing behavior too many times could be sanctioned with a withdrawal of their anonymity. Twitter could require accounts that violate conduct policies to provide their email address, phone number, and real name, even a physical ID, to continue using the site. This allows Twitter to maintain the benefits of anonymity that many Internet users value, while enforcing community norms and punishing users who violate harassment policies. "The knowledge that losing site access or the ability to participate anonymously may act as a deterrent for many abusers."[174]

Twitter needs to change its design, but a "solution to online harassment cannot be a top- down initiative developed by a heavily male or all-white team; such a team will never be able to fully understand the impact of abuse and will miss important factors of

---

[171] *Id.*

[172] *Id.*

[173] POLAND, *supra* note 158, at 226 (citing CITRON, *supra* note 164).

[174] *Id.*

that abuse while working on solution design."[175] In moving towards a more community-oriented design, Twitter must listen to abuse victims. "[L]earning about the intersections of abuse they face is the only way sites will be able to develop harassment policies that are proactive enough to actually prevent the abuse people are dealing with . . . ."[176]

Twitter must adjust its values away from being the "free speech wing of the free speech party." Its "focus on freeing up communication placed no emphasis on what types of communication users might want to avoid[.]"[177] As a result, it is "unprepared to deal with the fallout of the abuse and harassment that develop."[178] While Twitter's attempts to improve its reporting process and respond to abuse and harassment is a start, ultimately it must create a community that is more focused on protecting its users than on the instant sharing of ideas with no barriers. Truly free speech must be free not only from the silencing of censorship, but from the silencing of harassment as well. Twitter has prioritized the former at the cost of the latter.

In order for the building of community on Twitter to be effective, it must "proceed from the top down and the bottom up simultaneously."[179] The grassroots need demand for Twitter to protect its most vulnerable users exists; Twitter must now rethink what it means to design a worldwide system of communication and create real change for users from the top down.

---

[175] *Id.* at 234.

[176] *Id.*

[177] *Id.* at 233.

[178] *Id.*

[179] Daniel Bar-Tal, *Peace Building: Processes and Methods*, in INTRACTABLE CONFLICTS: SOCIO-PSYCHOLOGICAL FOUNDATIONS AND DYNAMICS 400, 400 (2013).

CHAPTER VI

CONCLUSION

Twitter has a long way to go before it will create a community that is truly open to diverse users. Research shows that "women, African-American, and Hispanic users report disproportionate levels of harassment."[180] The nature of electronic communication exacerbates the already conflictual and polarized character of public discourse. Harassment on Twitter is inevitable because Twitter was not designed to build community or address the natural tendencies of human communication in its absence.

As Twitter moves into the future, it needs to consider how the mechanics people use to communicate on a website affects the ways in which people communicate. After thirteen years of Twitter, we can see how instant communication with everyone leads to unchecked, oppressive systems targeting society's most vulnerable populations.

---

[180] Eva Galperin & Dia Kayyali, *Abuse and Harassment: What Could Twitter Do?*, ELECTRONIC FRONTIER FOUND. (Feb. 20, 2015), https://www.eff.org/deeplinks/2015/02/twitter-harassment-what-can-do.

REFERENCES CITED

1. Brianna Wu (@BriannaWu), TWITTER (Oct. 10, 2014, 5:57 PM),
    https://twitter.com/BriannaWu/status/520739878993420290.

2. Kyle Wagner, *The Future of the Culture Wars Is Here, And It's Gamergate*,
    DEADSPIN (Oct. 12, 2014, 3:46 PM), https://deadspin.com/the-future-of-the-culture-
    wars-is-here-and-its-gamerga-1646145844.

3. Anita Sarkeesian, *Tropes vs. Women in Video Games*, KICKSTARTER,
    https://www.kickstarter.com/projects/566429325/tropes-vs-women-in-video-games
    (last visited May 20, 2019).

4. Angela Watercutter, *Feminist Take on Games Draws Crude Ridicule, Massive
    Support*, WIRED (June 14, 2012, 6:30 AM), https://www.wired.com/2012/06/anita-
    sarkeesian-feminist-games/.

5. Simon Parkin, *Zoe Quinn's Depression Quest*, THE NEW YORKER, (Sept. 9, 2014),
    https://www.newyorker.com/tech/elements/zoe-quinns-depression-quest.

6. Amanda Marcotte, *Gaming Misogyny Gets Infinite Lives: Zoe Quinn, Virtual Rape,
    and Sexism*, THE DAILY BEAST (Aug. 22, 2014, 5:45 AM),
    https://www.thedailybeast.com/gaming-misogyny-gets-infinite-lives-zoe-quinn-
    virtual-rape-and-sexism.

7. Nellie Bowles, *How 'Doxxing' Became a Mainstream Tool in the Culture Wars*, NY
    Times (Aug. 30, 2017), https://www.nytimes.com/2017/08/30/technology/doxxing-
    protests.html.

8. Adam Baldwin (@AdamBaldwin), TWITTER (Aug. 24, 2014, 6:22 PM),
    https://twitter.com/AdamBaldwin/status/504801169638567936
    [https://web.archive.org/web/20140903012212/https://twitter.com/AdamBaldwin/sta
    tus/504801169638567936].

9. Devin Faraci, *Joss Whedon On #GamerGate, Jurassic World, Adam Baldwin And
    Speaking Out*, BIRTH MOVIES DEATH (Apr. 12, 2015),
    https://birthmoviesdeath.com/2015/04/12/joss-whedon-on-gamergate-jurassic-world-
    adam-baldwin-and-speaking-out.

10. MannoSlimmins, *Depression Quest Dev Claims Harassment and Misogyny. Facts
    Come Out Showing She's Lying*, REDDIT (Aug. 17, 2014, 9:44 AM),
    https://www.reddit.com/r/TumblrInAction/comments/2dt5hu/depression_quest_dev_
    claims_harassment_and/.

11. Joe Rogan Bible, *"GAMERGATE"! Zoe Quinn the Fake Victim*, YOUTUBE (Jun. 4,
    2018), https://www.youtube.com/watch?v=KOtjR2GWQLQ.

12. Chris Tognotti, *What Is "#Gamer Gate"? It's Misogyny, Under the Banner Of "Journalistic Integrity"*, BUSTLE (Sept. 5, 2014), https://www.bustle.com/articles/38742-what-is-gamer-gate-its-misogyny-under-the-banner-of-journalistic-integrity.

13. Caitlin Dewey, *The Only Guide to Gamergate You Will Ever Need to Read*, WASH. POST (Oct. 14, 2014), https://www.washingtonpost.com/news/the-intersect/wp/2014/10/14/the-only-guide-to-gamergate-you-will-ever-need-to-read/.

14. Jesse Singal, *Gamergate Should Stop Lying to Journalists—and Itself*, N.Y. MAG.: THE CUT (Oct. 20, 2014), https://www.thecut.com/2014/10/gamergate-should-stop-lying-to-itself.html.

15. Frank A. Dubinskas, *Culture and Conflict: The Cultural Roots of Discord*, *in* HIDDEN CONFLICT IN ORGANIZATIONS: UNCOVERING THE BEHIND-THE-SCENES DISPUTES 187 (Deborah M. Kolb & Jean M. Bartunek eds., 1992).

16. Jay Hathaway, *The Angry Ex Who Ignited Gamergate Has No Regrets*, GAWKER (Oct. 16, 2014, 1:40 PM), https://gawker.com/the-angry-ex-who-ignited-gamergate-has-no-regrets-1647186033.

17. Jessica Megarry, *Online Incivility or Sexual Harassment? Conceptualising Women's Experiences in the Digital Age*, 46 WOMEN'S STUD. INT'L F. 46 (2014).

18. AMNESTY INT'L, #TOXICTWITTER: VIOLENCE AND ABUSE AGAINST WOMEN ONLINE 50 (2018), https://www.amnesty.org/download/Documents/ACT3080702018ENGLISH.pdf.

19. AMANDA LENHART, MICHELE YBARRA, KATHRYN ZICKUHR & MYESHIA PRICE-FEENEY, ONLINE HARASSMENT, DIGITAL ABUSE, AND CYBERSTALKING IN AMERICA 3 (2016).

20. Maeve Duggan, *Online Harassment*, PEW RESEARCH CENTER (Oct. 14, 2014), http://www.pewInternet.org/2014/10/22/online-harassment/.

21. Kate Klonick, *Re-Shaming the Debate: Social Norms, Shame, and Regulation in an Internet Age*, 75 MD. L. REV. 1029, 1044 (2016).

22. DANAH BOYD, IT'S COMPLICATED: THE SOCIAL LIVES OF NETWORKED TEENS 146 (2014).

23. Jessica Valenti, *Zoe Quinn: After Gamergate, Don't Cede the Internet to Whoever Screams the Loudest*, GUARDIAN (Sept. 24, 2017), https://www.theguardian.com/technology/2017/sep/24/zoe-quinn-gamergate-online-abuse

24. William Turton, *Gamergate is Never Going Away*, OUTLINE (Sept. 7, 2017), https://theoutline.com/post/2218/gamergate-will-never-die-alt-right-trump?zd=1&zi=rtxk4tdn.

25. Jesse Singal, *Why the Video-Game Culture Wars Won't Die*, N.Y. MAG. (Sept. 30, 2016), http://nymag.com/selectall/2016/09/why-the-video-game-culture-wars-wont-die-two-years-later.html.

26. Brett Molina, *Twitter Overcounted Active Users Since 2014, Shares Surge on Profit Hopes*, USA TODAY (Oct. 26, 2017, 8:39 AM), https://www.usatoday.com/story/tech/news/2017/10/26/twitter-overcounted-active-users-since-2014-shares-surge/801968001/.

27. Azmina Dhrodia, *Unsocial Media: The Real Toll of Online Abuse against Women*, MEDIUM (Nov. 20, 2017), https://medium.com/amnesty-insights/unsocial-media-the-real-toll-of-online-abuse-against-women-37134ddab3f4.

28. Betsy Schiffman, *Twitterer Takes on Twitter Harassment Policy*, WIRED (May 22, 2008, 5:16 PM), https://www.wired.com/2008/05/tweeter-takes-o/.

29. Charlie Warzel, *"A Honeypot For Assholes": Inside Twitter's 10-Year Failure To Stop Harassment*, BUZZFEED NEWS (Aug. 11, 2016, 8:43 AM), https://www.buzzfeednews.com/article/charliewarzel/a-honeypot-for-assholes-inside-twitters-10-year-failure-to-s.

30. Sarah Jeong, *The History of Twitter's Rules*, MOTHERBOARD (Jan. 14 2016, 7:00 AM), https://motherboard.vice.com/en_us/article/z43xw3/the-history-of-twitters-rules.

31. *The Twitter Rules*, TWITTER (Jan. 14, 2009), http://twitter.zendesk.com/forums/26257/entries/18311 [https://web.archive.org/web/20090118211301/http://twitter.zendesk.com/forums/26257/entries/18311].

32. Josh Halliday, *Twitter's Tony Wang: 'We Are the Free Speech Wing of the Free Speech Party'*, GUARDIAN (Mar. 22, 2012, 11:57 AM), https://www.theguardian.com/media/2012/mar/22/twitter-tony-wang-free-speech.

33. Shreyas Doshi, *Policy and Product Updates Aimed at Combating Abuse*, TWITTER (Apr. 21, 2015), https://blog.twitter.com/official/en_us/a/2015/policy-and-product-updates-aimed-at-combating-abuse.html.

34. Issie Lapowsky, *Why Twitter Is Finally Taking a Stand Against Trolls*, WIRED (Apr. 21, 2015, 2:14 PM), https://www.wired.com/2015/04/twitter-abuse/.

35. Vijaya Gadde, *Twitter Executive: Here's How We're Trying to Stop Abuse while Preserving Free Speech*, WASH. POST (Apr. 16, 2015), https://www.washingtonpost.com/posteverything/wp/2015/04/16/twitter-executive-heres-how-were-trying-to-stop-abuse-while-preserving-free-speech/.

36. Megan Cristina, *Fighting Abuse to Protect Freedom of Expression*, TWITTER (Dec. 30, 2015), https://blog.twitter.com/en_au/a/2015/fighting-abuse-to-protect-freedom-of-expression-au.html.

37. *The Twitter Rules*, TWITTER, https://help.twitter.com/en/rules-and-policies/twitter-rules (last visited May 3, 2019).

38. Ingrid Lundgren, *Twitter Updates Its Abuse Policy and Adds Muting and Reporting Tools to Combat Trolls*, TECHCRUNCH (Nov. 15, 2016), https://techcrunch.com/2016/11/15/twitter-updates-abuse-policy/.

39. Sherisse Pham, *Twitter Tries New Measures in Crackdown on Harassment*, CNN (Feb. 7, 2017, 9:57 PM), https://money.cnn.com/2017/02/07/technology/twitter-combat-harassment-features/.

40. Greg Evans, *Twitter Policy Changes Under Review: Glorification of Violence, Unwanted Sexual Advances Targeted*, DEADLINE (Oct. 18, 2017, 9:38 AM), https://deadline.com/2017/10/twitter-policy-harassment-hate-speech-jack-dorsey-rose-mcgowan-1202190479/.

41. Vijaya Gadde & Del Harvey, *Creating New Policies Together*, TWITTER (Sept. 25, 2018), https://blog.twitter.com/official/en_us/topics/company/2018/Creating-new-policies-together.html.

42. Louise Matsakis, *Twitter Releases New Policy on 'Dehumanizing Speech'*, WIRED (Sept. 25, 2018, 9:00 AM), https://www.wired.com/story/twitter-dehumanizing-speech-policy/.

43. Donald Hicks & David Gasca, *A Healthier Twitter: Progress and More to Do*, TWITTER (Apr. 16, 2019), https://blog.twitter.com/en_us/topics/company/2019/health-update.html.

44. Kurt Wagner, *Twitter Says It's Getting Better at Detecting Abusive Tweets without Your Help*, VOX (Apr. 16, 2019, 3:35 PM), https://www.vox.com/2019/4/16/18410931/twitter-abuse-update-health-technology-harassment.

45. Briane Greene, *How Twitter Shapes Global Public Conversation: Jack Dorsey Speaks at TED2019*, TEDBLOG (Apr. 16, 2019, 2:41 PM), https://blog.ted.com/how-twitter-shapes-global-public-conversation-jack-dorsey-speaks-at-ted2019/.

46. *Abusive Behavior*, TWITTER, https://help.twitter.com/en/rules-and-policies/abusive-behavior (last visited May 3, 2019).

47. *Hateful Conduct Policy*, TWITTER, https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy (last visited May 3, 2019).

48. *Report Abusive Behavior*, TWITTER, https://help.twitter.com/en/safety-and-security/report-abusive-behavior (last visited May 3, 2019).

49. *Our Range of Enforcement Options*, TWITTER, https://help.twitter.com/en/rules-and-policies/enforcement-options (last visited May 3, 2019).

50. *Twitter Rules Enforcement*, TWITTER, https://transparency.twitter.com/en/twitter-rules-enforcement.html#twitter-rules-enforcement-jan-jun-2018 (last visited May 3, 2019).

51. *Our Range of Enforcement Options*, TWITTER, https://help.twitter.com/en/rules-and-policies/enforcement-options (last visited May 3, 2019).

52. Charlie Warzel, *"It Only Adds to the Humiliation" — How Twitter Responds to Harassers*, BUZZFEED NEWS (Sept. 22, 2016, at 9:01 PM), https://www.buzzfeednews.com/article/charliewarzel/after-reporting-abuse-many-twitter-users-hear-silence-or-wor.

53. Ash Sarkar (@AyoCaesar), TWITTER (Jan. 18, 2018, 3:58 AM) https://twitter.com/AyoCaesar/status/953959860403687424.

54. Monique Judge, *Twitter Has a Serious Harassment and Abuse Problem but Doesn't Seem to Want to Cure It*, ROOT (Oct. 30, 2017, 4:59 PM), https://www.theroot.com/twitter-has-a-serious-harassment-and-abuse-problem-but-1819979725.

55. Rob Beschizza, *Popehat Suspended from Twitter for Sharing a Threat He Received*, BOINGBOING (Aug. 3, 2017, 5:26 PM), https://boingboing.net/2017/08/03/popehat-suspended-from-twitter.html.

56. Ken White, *How I Got Barred from Posting on Twitter*, POPEHAT (Aug. 3, 2017), https://www.popehat.com/2017/08/03/how-i-got-barred-from-posting-on-twitter/.

57. Chris Peterson, *Twitter Suspended Me for Tweeting Feminist Academic Research. Here's Why That's a Problem.*, CIVIC MEDIA (Sept. 29, 2018), https://civic.mit.edu/2018/09/29/twitter-suspended-me-for-tweeting-feminist-academic-research-heres-why-thats-a-problem/.

58. Cory Doctorow, *Twitter Suspends Academic Who Quoted Feminist STEM Research*, BOINGBOING (Sept. 29, 2018, 12:06 PM), https://boingboing.net/2018/09/29/platform-censorship.html.

59. Amber Athey, *Twitter Restricts People Who Tweet "Learn to Code" — Even If They Aren't Engaging in Harassment*, DAILY CALLER (Feb. 13, 2019, 5:31 PM), https://dailycaller.com/2019/02/13/twitter-restricts-learn-to-code/.

60.	Jon Levine (@LevineJonathan), TWITTER (Jan. 28, 2018, 9:38 AM), https://twitter.com/LevineJonathan/status/1089940895418466305.

61.	Shérazade Faynel, *A Case of 'Double Standards' in Twitter's Enforcement Policy?*, JOURNAL INTERNATIONAL (Mar. 13, 2018), http://www.lejournalinternational.info/en/deux-poids-deux-mesures-lapplication-de-politique-de-twitter/.

62.	Charlie Warzel, *Twitter Is Still Dismissing Harassment Reports And Frustrating Victims*, BUZZFEED NEWS (July 18, 2017, at 4:39 PM), https://www.buzzfeednews.com/article/charliewarzel/twitter-is-still-dismissing-harassment-reports-and.

63.	Jake New, *Twitter Still Isn't Doing Enough about Rape Threats against Women*, TEEN VOGUE (Apr. 20, 2017), https://www.teenvogue.com/story/twitter-petition-rape-threats.

64.	Terry L. Cooper, *Building Ethical Community*, 41 AM. REV. FOR PUB. ADMIN. 3, 5 (2011).

65.	Sarah J. Jackson, Moya Bailey & Brooke Foucault Welles, *#GirlsLikeUs: Trans Advocacy and Community Building Online*, 20 NEW MEDIA & SOC'Y 1868, 1876 (2017).

66.	Moya Bailey, *#transform(ing)DH Writing and Research: An Autoethnography of Digital Humanities and Feminist Ethics*, 9 DIGITAL HUMAN. Q. (2015).

67.	ROBERT N. BELLAH, WILLIAM M. SULLIVAN, RICHARD MADSEN, ANN SWIDLER & STEVEN M. TIPTON, HABITS OF THE HEART: INDIVIDUALISM AND COMMITMENT IN AMERICAN LIFE 72–73 (1st ed. 1985).

68.	Thomas F. Pettigrew, Linda R. Tropp, Ulrich Wagner & Oliver Christ, *Recent Advances in Intergroup Contact Theory*, 35 INT'L J. OF INTERCULTURAL REL. 271, 273 (2011).

69.	Thomas F. Pettigrew, *Intergroup Contact Theory*, 49 ANN. REV. PSYCHOL. 65, 76 (1998).

70.	*About Direct Messages*, TWITTER, https://help.twitter.com/en/using-twitter/direct-messages#basics (last visited May 15, 2019).

71.	Rupert Brown, *Social Identity Theory: Past Achievements, Current Problems and Future Challenges*, 30 EUR. J. SOC. PSYCHOL. 745, 747 (2000).

72.	David Wong, *5 Reasons It's Still Not Cool to Admit You're a Gamer*, CRACKED (May 24, 2010), https://www.cracked.com/article_18571_5-reasons-its-still-not-cool-to-admit-youre-gamer.html.

73. Ricardomz, *Why Are Videogames So 'Uncool' on the Social Media?*, GAMESPOT, https://www.gamespot.com/forums/offtopic-discussion-314159273/why-are-videogames-so-uncool-on-the-social-media-29357431/ (last visited May 14, 2019).

74. MDK12345, *Afraid of Getting Judged for Being a Gamer*, GAMESPOT, https://www.gamespot.com/forums/games-discussion-1000000/afraid-of-getting-judged-for-being-a-gamer-31167494/ (last visited May 14, 2019).

75. Jesse Fox & Wai Yen Tang, *Sexism in Online Video Games: The Role of Conformity to Masculine Norms and Social Dominance Orientation*, 33 COMPUTERS HUM. BEHAV. 315 (2014).

76. Emily Pronin, Thomas Gilovich & Lee Ross, *Objectivity in the Eye of the Beholder: Divergent Perceptions of Bias in Self Versus Others*, 111 PSYCHOL. REV. 781, 783 (2004).

77. Jef Rouner, *8 "Criticisms" of Anita Sarkeesian That Are Utter Bullshit*, HOUSTON PRESS (Oct. 21, 2014, 7:00 AM), https://www.houstonpress.com/arts/8-criticisms-of-anita-sarkeesian-that-are-utter-bullshit-6382966.

78. Gabenus Trollucus, *Zoe and Anita in Dumb and Dumber*, KNOW YOUR MEME (Aug. 12, 2015, 9:11 AM), https://knowyourmeme.com/photos/1004399-zoe-quinn.

79. RPS Hivemind, *Hello: Videogames Are For Everybody*, ROCK PAPER SHOTGUN (Sept. 8, 2014, 1:00 PM), https://www.rockpapershotgun.com/2014/09/08/videogames-are-for-everybody/.

80. Raymond A. Friedman & Steven C. Currall, *Conflict Escalation: Dispute Exacerbating Elements of E-mail Communication*, 56 HUM. REL. 1325, 1330 (2003).

81. Neta Oren & Daniel Bar-Tal, *The Detrimental Dynamics of Delegitimization in Intractable Conflicts: The Israeli–Palestinian Case*, 31 INT'L J. INTERCULTURAL REL. 111, 112 (2007).

82. Michelle Maiese, *Dehumanization*, BEYOND INTRACTABILITY (Jul. 2003), https://www.beyondintractability.org/essay/dehumanization.

83. DAVID LIVINGSTONE SMITH, LESS THAN HUMAN 15 (2011).

84. *'Less Than Human': The Psychology of Cruelty*, NPR (Mar. 29, 2011, 1:00 PM), https://www.npr.org/2011/03/29/134956180/criminals-see-their-victims-as-less-than-human.

85. Nick Haslam & Steve Loughnan, *Dehumanization and Infrahumanization*, 65 ANN. REV. PSYCHOL. 399, 402 (2014).

86. James Grimmelmann, *The Virtues of Moderation*, 17 YALE J. L. & TECH. 42, 77 (2015).

87. BAILEY POLAND, HATERS: HARASSMENT, ABUSE, AND ONLINE VIOLENCE 233–34 (2016)

88. *New Player Reform System Heads into Testing*, LEAGUE OF LEGENDS, https://na.leagueoflegends.com/en/news/game-updates/player-behavior/new-player-reform-system-heads-testing (last visited May 8, 2019).

89. Lauren Keating, *'League of Legends' Battles Harassment with Automated System That Could Ban Gamers*, TECH TIMES (May 26, 2015, 2:02 PM), http://www.techtimes.com/articles/55629/20150526/league-legends-battles-harassment-automated-system-ban-gamers.htm.

90. Laura Hudson, *Curbing Online Abuse Isn't Impossible. Here's Where We Start*, WIRED (May 15, 2014, 6:30 AM), https://www.wired.com/2014/05/fighting-online-harassment/.

91. DANIELLE CITRON, HATE CRIMES IN CYBERSPACE 237 (2014).

92. J. Kevin Barge, *Dialogue, Conflict, and Community*, *in* THE SAGE HANDBOOK OF CONFLICT COMMUNICATION 517, 517 (John G. Oetzel & Stella Ting-Toomey eds., 2006).

93. SARAH JEONG, THE INTERNET OF GARBAGE 71 (1.5 ed. 2018).

94. Daniel Bar-Tal, *Peace Building: Processes and Methods*, in INTRACTABLE CONFLICTS: SOCIO-PSYCHOLOGICAL FOUNDATIONS AND DYNAMICS 400, 400 (2013).

95. Eva Galperin & Dia Kayyali, *Abuse and Harassment: What Could Twitter Do?*, ELECTRONIC FRONTIER FOUND. (Feb. 20, 2015), https://www.eff.org/deeplinks/2015/02/twitter-harassment-what-can-do.