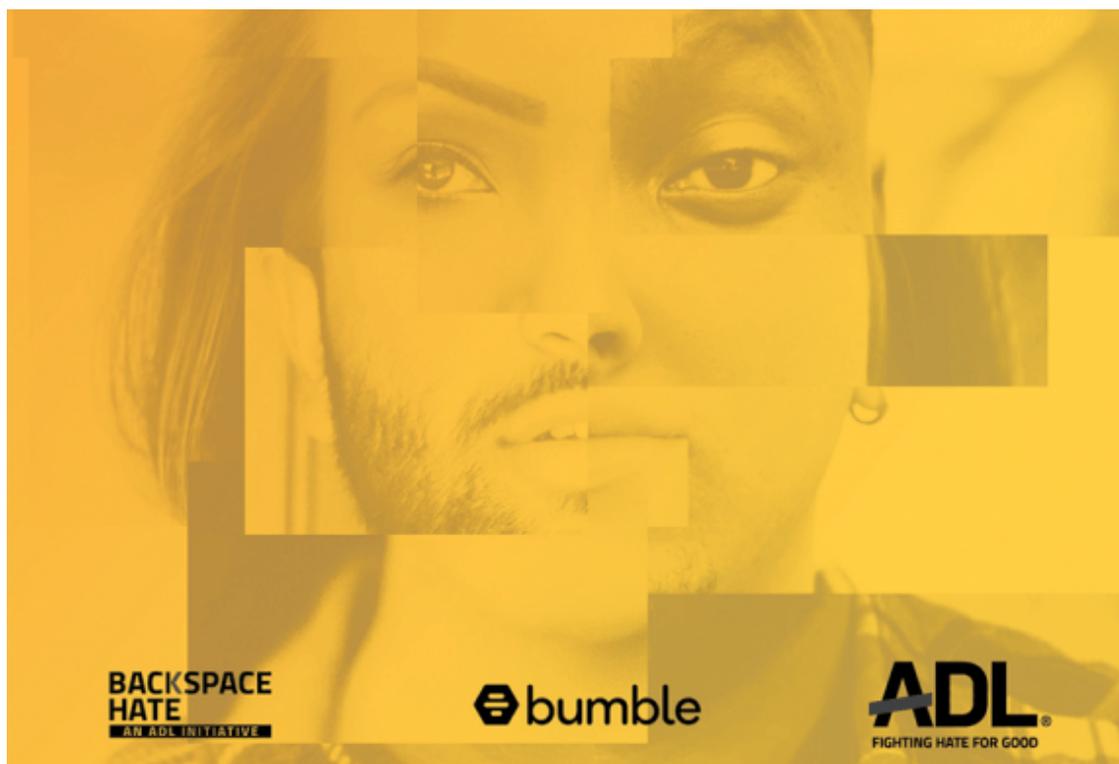


There's No Time To Lose: Donate to ADL today to take the good fight against anti-Semitism into the new year! >



The Trolls are Organized and Everyone's a Target: The Effects of Online Hate and Harassment



Chapters

- 1 Executive Summary
- 2 Introduction
- 3 Findings
- 4 Recommendations
- 5 Moving Forward: Protecting Targets of Hate Through Better Government Regulation
- 6 Conclusion
- 7 Appendix: Research Methods
- 8 Endnotes

Sign the Backspace Hate Petition

Tell Lawmakers to Support Victims and Targets of Online Hate and Harassment

EXECUTIVE SUMMARY

Online hate and harassment¹ have increasingly become a common part of the online experience. Public attention has usually focused on harassment of celebrities and public figures. However, our recent work has shown that a substantial swath of the American public has experienced online harassment, with 37 percent of adults having experienced severe online harassment², defined by the Pew Research Center as including physical threats, sexual harassment, stalking and sustained harassment³. For this study, we wanted to examine the effects of online hate and harassment on private individuals—the type of people whose experiences represent the bulk of that statistic. We engaged in an extensive literature review and also conducted 15 in-depth qualitative interviews to better understand and chronicle the full experience of being a target of online harassment. We explore the personal stories of targets of online hate in an attempt to paint a more complete picture of the ways in which harassment can envelop multiple facets of a person's life. We hope this report sheds light on the statistics.

Five findings stand out from the literature review and interviews:

1. **Online hate incidents are frequently connected to the target's identity**

Whether it was simply being a Jewish business owner or authoring a blog post on feminism, the online hate incidents experienced by our interviewees were frequently centered around issues of identity.

2. **Harassers use platform design to their advantage**

Coordinated attacks often caused harm to a target by leveraging key features of social media platforms. This included the ability to be anonymous online,

to create multiple accounts by one person, the fact that there is no limit to the number of messages one user can send to another, and the use of personal networks as weaponized audiences.

3. Online hate can cause significant emotional and economic damage

Targets of harassment reported deep and prolonged emotional difficulties. Additionally, harassers often targeted individuals' economic wellbeing by trying to tarnish their reputation or by contacting their employers.

4. Harassers attack and impact others in the target's community

Interviewees revealed experiences of harassment where perpetrators would also attack their relatives, friends and employers. Targets were highly disturbed by the spillover of hate into their offline lives and felt that the increase in radius of attack was meant to cause further harm to them.

5. Social media platforms are not adequately designed to remove or efficiently review hateful content

Respondents were universally unhappy with the processes and functions of the reporting systems on major social media platforms. Interviewees expressed frustration in having to wait weeks for the content moderation teams to respond to their reports of harassment. They also felt that the ability to only report one piece of content at a time created a bottleneck in content flagging. The lack of options when reporting harassment and an attitude described by targets as a general disinterest towards their plight caused further unease for interviewees.

The interviews and the review of the literature also point to ways to prevent or mitigate the impact of hate and harassment on victims. These include:

1. Increase users' control over their online spaces

Interviewees felt like they had no control over their profiles, pages or accounts to prevent attackers from targeting them relentlessly. Platforms

could provide more sophisticated blocking features like blocking a user's ability to stalk someone across a platform. Platforms should also allow users to designate friends' accounts as co-moderators, with specific permissions to assist with harassment management and moderation.

2. Improve the harassment reporting process

Companies should redesign their reporting procedures to improve the user experience. Platforms should provide step-by-step tracking portals, so users can see where their abuse report sits in the queue of pending reports.

Platforms should also allow bulk reporting of content, consider harassment occurring to the target on other platforms, and respond to targets quickly. Platforms could set up hotlines for people under attack who need immediate assistance and assign case managers to help targets of hate through the process.

3. Build anti-hate principles into the hiring and design process

Safety, anti-bias and anti-hate principles should be built into the design, operation and management of social media platforms. Platforms should prioritize diversity in hiring designers, including individuals who have been targets of online harassment. Platforms should create user personas and use cases that address the needs of vulnerable populations. Platforms should also weigh tool functionality against increased opportunities for harassment before implementing new features. We also recommend that input from a diverse set of community representatives and outside experts should be solicited before additions to or changes in platforms features are made.

While interviewees did not directly comment on the Government's role in passing legislation that holds perpetrators accountable for their actions, it's important that federal and state governments strengthen laws that protect targets of online hate and harassment. Many forms of severe online misconduct are not consistently covered by cybercrime, harassment, stalking and hate crime law. Legislators have

an opportunity, consistent with the First Amendment, to create laws that hold perpetrators of severe online hate and harassment more accountable for their offenses.

INTRODUCTION



According to ADL's 2019 Online Hate and Harassment survey,⁴ 53 percent of American adults have experienced online harassment and 37 percent have experienced severe harassment.⁵ The prevalence of online toxicity is even higher in video games, a series of platforms which have recently emerged as a leading social space. In a first-of-its-kind survey, ADL found that 74 percent of adults who

play online multiplayer games in the US experience some form of harassment while playing games online, while 65 percent of players experience some form of severe harassment.

Relying on both a literature review of online harassment and in-depth interviews with targets of hate, this report describes experiences of the people behind these stark numbers. The ethnographic study that informs this report has been designed using an extensive literature review on the topic of cyber harassment. Through that exercise, we explored the definitions of hate and online harassment and studied the history and evolution of these concepts. We also provided a comprehensive taxonomy of the different forms through which online hate manifests itself. We pored over studies that explored potential causes of online hate, including behavioral reasons and the features on platforms that are ripe for abuse.

A few of the more prevalent themes in cyber hate literature included:

Gender and Harassment

Some of the most widely reported incidents of campaign harassment (the ability of harassers to use online networks to organize campaigns of hate) and networked harassment (the weaponization of a target's online network) have been waged against women and the LGBTQ+ community. Such incidents demonstrate a new way to police and attempt to control these targets' ability to speak and participate in public life.

Racism, Islamophobia and Anti-Semitism

Racism, Islamophobia and anti-Semitism are also common motivating factors in networked harassment episodes. In fact, these three kinds of hate often appear together.

Targeting Intersectional Identities

Hate campaigns often target individuals with complex identities or affiliations

with special vigor. Scholars and activists use the term “intersectionality,” the idea that bias, discrimination and oppression manifest differently when identifiers like race and gender overlap.⁶

Harassment of Professionals

People are also targeted because of their professions. For example, journalists and academics are frequently targeted due to the public nature of their careers. This places them in a double bind since it’s imperative to their careers to publish and promote their work online.

For this report, ADL conducted 15 in-depth interviews with individuals about their experiences with online hate and harassment. For the safety of the interviewees, this report does not publish their names or personally identifiable details. ADL selected subjects with diverse backgrounds, identity characteristics and professional experiences.

We deliberately selected and interviewed subjects who are not famous or public figures, as are well-known targets Carlos Maza (a video producer for Vox who was subjected to homophobic slurs by an alt-right commentator on a popular YouTube channel⁸) and Zoë Quinn (a non-binary video game developer who suffered sustained harassment including threats of rape and death⁹). Such cases have been extensively covered in news articles and research papers. Social media platforms are also more reactive to cases of harassment in which public personalities are targeted,¹⁰ and these individuals might have more financial and social capital to respond or recover.

Our interview subjects represent a collection of experiences that have been described as distressing but are also displays of powerful resilience against a barrage of hate. Embedded in their stories are tales of setback, courage and resistance. But beyond compelling narratives, they also serve a more practical function—these interviews help us more fully understand the dynamics of online harassment at a depth that would be very challenging to extract from survey

results. Moreover, these interviews shine a light on how harassers exploit the design of social media platforms.

The Trolls are Organized and Everyone's a Target
The Effects of Online Hate and Harassment

"Being trans and successful is all it takes to attract harassment"

adl.org Anti-Defamation League @ADL @adl_national



FINDINGS

The Trolls are Organized and Everyone's a Target
The Effects of Online Hate and Harassment

A family business run by a Jewish couple found its reputation in shambles when angry reviews claimed that the business was run by a "Jewy Jew" and the type of person "who gave rise to Hitler."

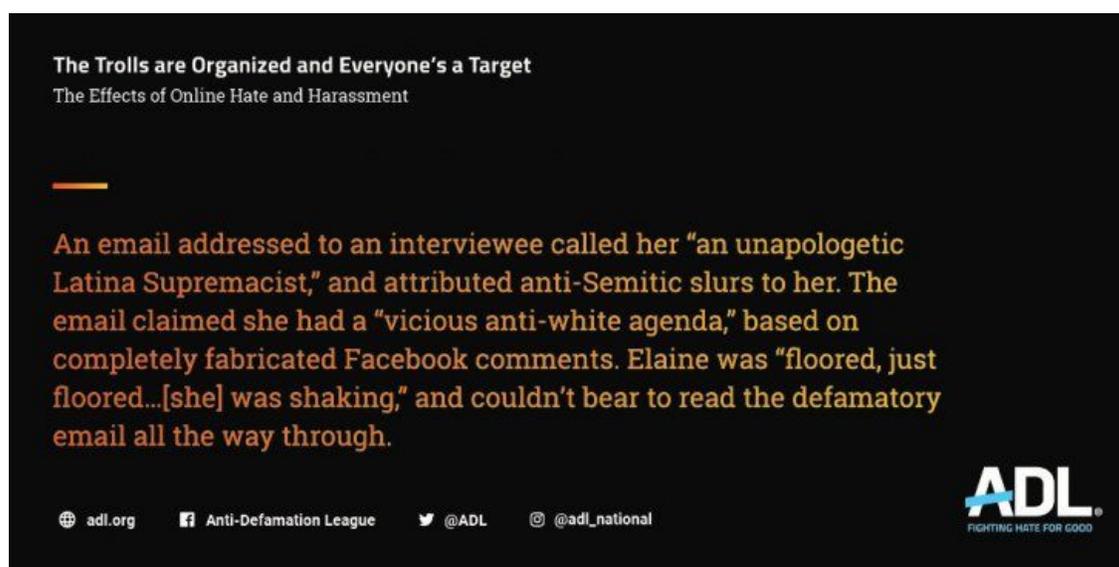
adl.org Anti-Defamation League @ADL @adl_national



Our in-depth interviews show that online harassment and hate come in a variety of forms, ranging from single but intense episodes of hate to months-long sustained harassment campaigns. They cross from online-only events to offline incidents. They can target one person or seek to disrupt entire personal and professional networks.

While the breadth of strategies available to attackers can be an overwhelming topic to explore, the impetus for harassment appears to be especially myopic. More often than not, targets felt they were attacked because of an identity-related attribute. Equally troubling is the fact that targets felt they did not have any legitimate recourse for action. In their attempts to remove hateful content, they felt stymied by the content reporting mechanisms across major social media platforms.

Finally, interviewees and other targets experience harassment on a wide variety of platforms, including mainstream social media (especially Twitter, Facebook and YouTube), anonymous or pseudonymous web boards (such as 4chan, 8chan and Reddit), gaming and related sites (Twitch and Discord), online review sites (Yelp and Google), publishing platforms (Wordpress, Medium and other online publications), and alt-right websites (Breitbart, Quillette and Stormfront). In addition, email was often used.



Key Finding 1: Online hate incidents are frequently connected to the target's identity

While the breadth of strategies available to attackers can be an overwhelming topic to explore, the impetus for harassment appears to be especially myopic. More often than not, interviewees felt they were attacked because of an identity-related attribute. The specifics of how identity resulted in harassment varies, but

the common theme in hateful incidents across interviews was identity. Being a member of a marginalized community makes an individual particularly vulnerable to hate and harassment online. People who spoke out against sexist, racist and anti-Semitic injustice also attracted vicious attacks, typically in response to their condemnation of hate and bias. Overall, we found that it required very little to be targeted online—it could be one comment or a single publication, but the overwhelming focus was related to identity, as opposed to political belief.¹¹

Recent research has found that individuals from marginalized communities experience online harassment at much higher rates. ADL's 2019 Online Hate and Harassment survey report¹² revealed that identity-based harassment was most common against LGBTQ+ individuals, with 63 percent of LGBTQ+ respondents experiencing harassment because of their sexual orientation or gender identity. In fact, around one-third (32 percent) of Americans who had been harassed reported that they thought the harassment was a result of their sexual orientation, religion, race or ethnicity, gender identity or disability. Similarly, in the ADL games survey report,¹³ 53 percent of online multiplayer gamers who experienced harassment attributed their targeting to their race/ethnicity, religion, ability, gender or sexual orientation.

Our interviews found that harassers often weaponized aspects of their target's characteristics by using slurs, offensive imagery and threats of violence based on their target's real or perceived identities. For example, themes of misogyny were prevalent when our female interviewees were attacked on social media. Perpetrators' comments to women were often premised on a woman's appearance or their "proper place" being in the home, doing housework and bearing children.

Some people received harassment solely because of their identity. For example, both trans women reported harassment just because they were trans and publicly visible online. In one instance, a trans woman who is a gaming professional reported harassment on Twitch, including receiving invasive comments about her body and gender identity. In some cases, the transgender interviewees shared that

their gender identity was ignored by harassers who would purposefully misgender them.

In other cases, people were targeted because of an identity-related article or statement they published or posted online. For example, an academic who researches gender in the United States was harassed following the publication of her scholarly essay about masculinity and consumer culture. On Twitter and Instagram, harassers followed her, insulted her appearance and attacked her with threats of rape and death. The far-right news site Breitbart wrote critically about her article. The Breitbart coverage led to her receiving hundreds of harassing messages on social media. The harassment stemming from this one article lasted for nine months.

Similarly, one interviewee, a researcher, also faced public threats after arguing with a prominent white feminist and activist on Twitter about the portrayal of a woman of color in popular culture. The prominent feminist shared the researcher's tweets in a mocking way, triggering other (mostly) white women to begin harassing her for being "too sensitive." Far-right harassers picked up on the argument among white feminists and a piece was written about the poster on Breitbart, sending a second wave of harassment her way. Fortunately, her employer helped keep her safe in her workplace during this period of sustained harassment.

Women of color faced similar harassment tactics and vulnerabilities. In certain instances, when our interviewees commented on social media about topics related to disparate impact and injustice, they became targets of racist, xenophobic and misogynistic comments that triggered persistent harassment episodes.

"It's just infuriating...someone sits behind their screen and tries to destroy everything I've worked so hard for. They don't have any consequences and I have to defend everything I do."

First generation college grad and non-profit professional

Key Finding 2: Harassers use platform design to their advantage

Through their design, platforms encourage (or discourage) certain practices and behaviors.¹⁴ Many of the platforms' designs make harassment easy, and harassers are effective at taking advantage of platforms' features. For example, the video game streaming site, Twitch, makes it easy to ban users from the chat feature of a streamer's individual channel, but it is quite difficult to ban users from viewing a stream on an individual channel and it is especially difficult to ban users across the platform from different streams and chats. An interviewee who live streams explained that this allowed their stalkers to continue tracking them by viewing their streams even when users were banned from the chat feature on an individual's channel. It also allowed users to continue engaging in harassing activity by attacking the intended target in chat rooms on different channels.

Multiple accounts controlled by a unique user—often for the purpose of dominating chat forums, harassing a target or spreading information—are known informally as “sock puppet” accounts. Interviewees cited the ability to create multiple accounts operated by the same person on Twitter as an example of a social media feature that can be easily abused by harassers. This feature can be useful when someone, for example, wants to have a professional and personal account; however, it can also be used as a means to continue cyber stalking or harassment even after the target blocks the initial account.

Another common feature most platforms have is the ability to send unlimited messages to another user. In the cyber harassment context this feature can prove to be incredibly problematic. If a user does not block or mute a harasser on a platform, the harasser can send them hundreds of messages at a time.

Additionally, most interviewees reported being targeted across multiple online platforms in campaigns that weaponized the target's networks. For the purposes of this report, we will refer to this phenomenon as "networked harassment."¹⁵

Networked harassment uses the target's online networks and the openness and features of social media platforms against the target. This includes inserting embarrassing and fallacious news into a target's professional networks, engineering the disruption of important identity-based alliances by turning groups against each other, and resharing posts to incite further harassment. Features such as interlinked user profiles and user-created shareable content increased the ease with which a perpetrator could harass someone across multiple platforms.

In some instances of networked harassment, perpetrators discussed their motivations. In one instance, there was explicit conversation about making the target, a woman of color, angry with the intention of making her unemployable. In another example, there was explicit conversation about increasing distrust between white women and feminists of color. In a third example, an article picked up by a far-right website was paired with a call to action to harass the author. This call-to-action resulted in an individual target receiving thousands of hateful messages for months.

In another instance, a graduate student living in the United Kingdom tweeted a question on Twitter inviting followers to describe a time they were "mansplained." This question quickly went viral and resulted in a harassment campaign against her on both Twitter and Facebook. As a dedicated researcher, she set out to find where harassers were coordinating their efforts and discovered a lengthy thread

on 4chan that tracked women of color and coordinated harassment campaigns against them. There, she found her name, as well as the names of other female journalists of color.

The Trolls are Organized and Everyone's a Target
The Effects of Online Hate and Harassment

—

"Don't want to give troll power ... you feel annoyed but does not want to show those feelings. revealing them give harassers gives them power."

What is it like to hide those feelings?

"It's really hard! It produces stress after the show is over. The emotional stress is exhausting, often end steam earlier, leads to lethargy. As you become used it just wears you down."

adl.org Anti-Defamation League @ADL @adl_national

ADL
FIGHTING HATE FOR GOOD

Key Finding 3: Online hate can cause significant emotional and economic damage

Our interviewees detailed the emotional effects and economic burdens of being the target of Internet abuse. Targets described their experiences as stressful and demoralizing, often isolating and traumatizing, and sometimes fear-inducing.

Some individuals interviewed actively worked to compartmentalize their feelings when trying to document and report their harassment, so as not to internalize the hate spewed at them across the screen. Many interviewees acknowledged that desensitization was necessary to get through a comprehensive review of the harassing and hateful messages they received in order to address them. This process forced the subjects into states of numbness in order to avoid experiencing the full depth and breadth of the attacks. Relatedly, in at least one case, being subject to this abuse led to extreme social isolation. Due to intense fear, these subjects kept their experiences private, and felt detached and alone as a result.

Targets also reported feeling exhausted and frustrated by the process of "cleaning up the hate" while also trying to maintain their distance from it. Removing an onslaught of hateful comments involved deep engagement with abusive images

and messages, since the targeted individuals were solely responsible for screen-capturing, saving, deleting, documenting and reporting cross-platform abuse. Unfortunately, this process did not help these interviewees feel safer and took several hours to complete. The inadequacy of the platform reporting mechanism served as a prevalent source of frustration for the subjects we interviewed. Many of our interviewees responded by limiting their online presence or visibility, or by posting less frequently.

Uncertainty was cited as a common driver for targets' emotional and psychological response to the abuse. Many interviewees explained that not knowing how long an attack would last, the ultimate effects on their personal relationships and professional standing, or in stalking cases, why they were targeted and by whom, were major components of their distress. Interviewees who experienced uncertainty due to cyberstalking¹⁶ were often the most stressed and upset and the most likely to take additional security measures or contact law enforcement. One interviewee felt exceedingly frustrated that Facebook wouldn't provide any information about her stalkers, despite the target having access to their IP addresses.

The impact on targets went beyond emotional suffering, affecting their economic sustainability and job prospects. Subjects noted that their harassers often targeted their colleagues and supervisors in order to tarnish their reputation and with the goal or feared outcome of making them unemployable.

One subject discussed how her attacker publicly accused her of being anti-white. A streaming professional we interviewed explained that, as a result of a harassment incident, viewers unsubscribed from her channel, a portion of whose subscription fee had been income for the interviewee. A non-profit employee who was a first-generation college graduate was falsely accused of being anti-Semitic in a defamatory email and spent weeks working with a university ombudsperson to craft a response explaining the incident to colleagues. The interviewee said that her colleagues did not understand the nature of the harassment and assumed

the defamatory email was true. Instead of supporting her, they said, “See, that’s why we don’t use social media,” even though social media was crucial to the young woman’s career. Another woman said that she lost her job because her company was unable to support her in handling the grueling and unrelenting harassment.

One interviewee discussed how her harassment experience centered around the perpetrator attacking her family business. Rather than focusing on the individual’s professional online presence, the attacker targeted the company. It was common knowledge in her community that her husband, who was also her business partner, was active on Jewish Facebook groups. One day, the couple discovered vitriolic reviews of their business on Yelp, Google Reviews and Glassdoor posted by a fake Facebook account. These reviews plummeted the business’s online ratings and slowed growth. The reviews themselves attacked the business at large but made specific references to the husband being a “Jewy Jew” and the type of person “who gave rise to Hitler.” This episode resulted in months of lost potential income.

There are also costs associated with cyber harassment protection. One interviewee describes installing security cameras in her home. Another interviewee installed an expensive security system and hired private security guards for her home, fearing that her online stalker was a former client who had the potential to be dangerous.

"I have an obligation being a public facing scholar because we have a mission to educate... the museum social media asks me to be front and center on the [museum] pages and those moments I end up attracting more attention than I seek otherwise."

Museum researcher and educator

Key Finding 4: Harassers also attack and impact others in the target's community

Networked and campaign harassment often involved not just the targeted victims but also their friends, family, colleagues and broader communities. In some interviews, targets explained the depths to which their networks experienced harassment. One professor was shocked and horrified when alt-right harassers targeted his wife and daughter. His distress continued when, in response to the defamatory statements spewed by the perpetrators, a local elected official called for an investigation of the professor and not the harassers. These actions left the victim's family and employer—a public university—tasked with managing the actual harassment as well as the fallacious investigation.

A game designer and researcher described the gut-wrenching feeling of finding out that her attackers attempted to swat her mother. Swatting is the act of falsely reporting a serious crime with the aim of drawing a massive police response to the home of an unsuspecting target (often a SWAT team with extreme weaponry)¹⁷. Swatting has resulted in at least one fatality in the United States.¹⁸

Beyond family and friends, harassment episodes have permeated entire companies. When a law firm owner was the target of online harassment, his entire firm lost business, spent money on security, and sacrificed productivity. The owner and his employees spent hours tracking, analyzing and reporting hate. They spent additional time meeting together to discuss whether the stalker was

someone they knew and if firm employees were in physical danger. In another instance, a woman who was defamed to her employer found that her colleagues began to doubt her decisions and actions—affecting productivity and team-oriented tasks.

Finally, this harassment often occurs in ways that other members of the public witness, which can have a negative impact of silencing or bystander effect on witnesses. For example, on Twitch, viewers witnessed certain instances of harassment as it unfolded. As mentioned earlier, the streamer interviewees reported losing followers and donations, and hypothesized that this change in behavior was due to their followers wanting to remove themselves from hostile environments or avoid inadvertently subjecting themselves to harassment.



Key Finding 5: Social media platforms are not adequately designed to remove or efficiently review hateful content

Every person we interviewed used their respective platform's reporting tools when attempting to mitigate their online harassment. Most interviewees said that they did not receive meaningful or helpful responses to a majority of their abuse reports for weeks or months—some reports went completely unanswered. The subjects noted that when platforms did send responses, the messages lacked empathy for or a real understanding of the broader difficulties created by

networked harassment.

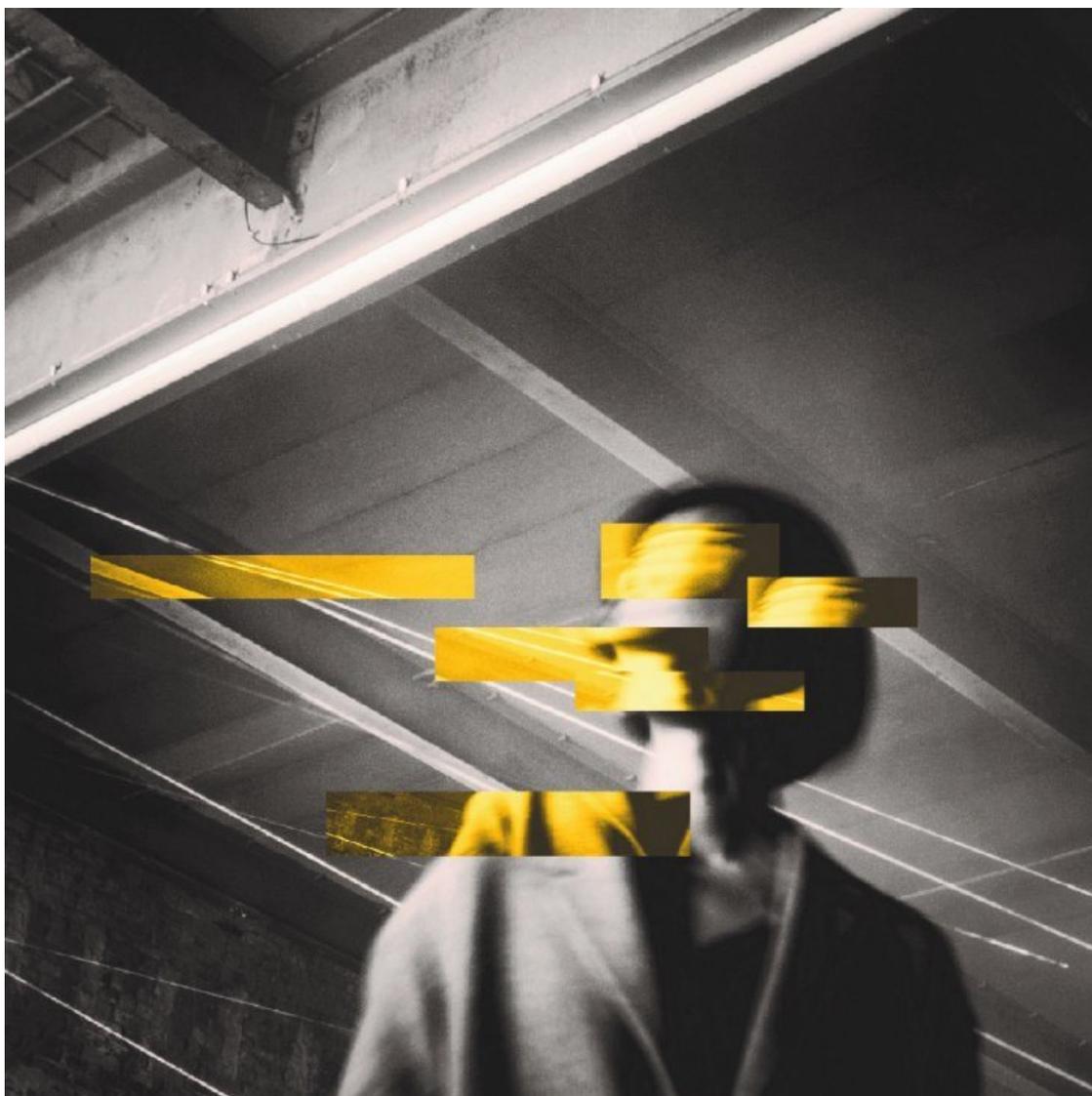
Most platforms have reporting systems and structures for users to contact them about harassment, spam, hate speech and other violations of their Terms of Service (TOS). Notably, a majority of our study interviewees criticized the current systems, deeming them largely ineffective.¹⁹ One of the deepest sources of frustration our subjects highlighted was that platforms make it extremely difficult to accurately, thoroughly and quickly report massive attacks; most platform reporting systems are designed so users can only report one hateful post or account at a time. Twitter does allow users to aggregate five posts in one complaint, but this was inadequate for our interviewees who faced hundreds or thousands of harassing messages at a time.

Interviewees expressed the belief that reporting harassment to the platforms had limited efficacy. They explained that reporting mechanisms do not provide easy ways to explain and substantiate the depth and breadth of a harassment incident. In fact, many of the harassing comments interviewees reported were deemed “admissible” upon platform review, meaning that an initial evaluation by human moderators—employees or contractors working for the individual platform—determined the harassing comments did not violate the platform’s terms of service.

Interviewees noted that some technology platforms had more effective reporting tools. They felt most secure when platform responses validated and supported them, when they had access to appropriate and effective resources specifically related to the type of harassment that occurred, and when they had a better idea of who was targeting them and through what means. For example, some interviewees said that platforms where an individual could see a perpetrator’s referring website or track harassers’ IP addresses were far more helpful in the target’s quest to manage harassment than sites that merely allowed a target to report harassing activity.

Based on our interviews, the platforms that offered the most effective reporting tools were Discord, Medium and Wordpress. For example, Wordpress's webmaster tools give users more autonomy and control in spam filtering through a mechanism called Akismet. Additionally, Medium and Wordpress display referring sites of visitors. This information allows administrators to see, for example, if harassment being directed at an individual is by a perpetrator coming from a specific website, like 4chan. Websites like Reddit and Twitch allow for volunteer moderation in certain instances. One benefit of volunteer moderation is that some communities can set their own standards to better protect community members from harassment.

RECOMMENDATIONS



As an outgrowth of our interviews and supporting research, we include three broad recommendations that platforms can implement to decrease hate on their platforms.

Recommendation 1: Increase users' control over their online spaces

Targets of online harassment often felt like they had no control over their profiles, pages or accounts because perpetrators invaded them relentlessly and without consent.

Platforms can support targets by giving them more control over moderating their individual online spaces. Some of the interviewees reported satisfaction in managing harassment when they had the ability to block harassers at the IP level rather than blocking a single account. Interviewees also reported satisfaction in managing harassment when they could control which users could engage in their space and enforce community norms. Platforms should provide more sophisticated blocking features.

Another way platforms can support targets is to allow users to designate trusted friends as moderators to help targets control the traffic on their page. Many of our interviewees said that they shared usernames and passwords with friends and colleagues to help them regulate the influx of harassing messages they received. Instead, as a safer and more streamlined alternative to this method, platforms should allow users to designate sub-accounts as co-moderators, with specific permissions to assist with harassment management and moderation.

Recommendation 2: Improve the harassment reporting process

Companies should redesign their reporting procedures to improve the user experience and to make it more responsive to networked harassment.

Platforms' reporting features should allow users to thoroughly explain and demonstrate their abuse to the moderation team. At present, there is no

streamlined way to aggregate campaign or networked harassment in one report. Platforms should implement reporting systems that allow for detailed, nuanced and comprehensive reporting.

Users need clear, transparent means to track and report harassment—these processes should not require specialized knowledge or tech expertise. In our interviews, the only subject who was successful in getting meaningful responses from moderation teams was an individual who called in a personal favor and had social networking professionals advise her on submitting her complaint. Platforms should provide step-by-step tracking portals, so users can see where in the queue their abuse report sits, how it is being processed and who in the company is the assigned case manager.

Finally, platforms should respond to targets quickly and with compassion. Platforms could set up hotlines for people under attack who need immediate assistance, assign internal caseworkers to assist with aggregating and investigating the harassing activity, or develop other more effective response procedures.

Recommendation 3: Build anti-hate principles into the hiring and design process

Safety, anti-bias and anti-hate principles should be built into the design, operation, and management of social media platforms. To create safer and more hospitable online spaces, platforms should appropriately train designers to understand Internet abuse so they can anticipate and prevent new opportunities for harassment as they build and architect platform features. We recommend that companies solicit input from a diverse set of community representatives and outside experts before they make additions to or changes in platforms' features.

Companies can also build anti-hate principles into cross-industry initiatives. Interviewees spoke at length about how attackers would use multiple platforms to target someone. Designing content moderation processes that result in an

automatic sharing of information between companies safety teams could improve safety standards for the entire industry. Admittedly this is a complicated solution that would require different platforms reaching a common understanding of online harassment definitions. However, the benefits to users who are vulnerable to online hate and harassment would be immense.

Platforms should prioritize diversity in hiring designers, including individuals who have been targets of online harassment. These individuals have a unique perspective on how the tools, features and mechanisms being built by social media companies can be weaponized. In addition to adding diverse individuals—especially targets—on design teams, companies should educate designers on the specific details of extreme harassment cases from both their own sites and other sites. Designers can also learn through the production of user personas and journey maps detailing extreme harassment episodes.

Platforms should also weigh tool functionality against increased opportunities for harassment before implementing new features. Features that promise to increase performance metrics or reduce friction sometimes exacerbate already abusive dynamics or inadvertently create opportunities for new harassment. If a new feature has a high likelihood of facilitating harassment, it should be modified. Similarly, features that have led to abuse in the past should be adjusted to reduce abuse.

MOVING FORWARD: PROTECTING TARGETS OF HATE THROUGH BETTER GOVERNMENT REGULATION

BACKSPACE HATE™

AN ADL INITIATIVE

It's important to note that this ethnographic study does not represent the full breadth of views and experiences embodied by the victims and targets of online hate and harassment. That said, it is our hope that through this report we can encourage social media platforms, government and civil society to take ownership of their roles in addressing these issues. The best way to approach these issues is through a multi-faceted lens.

Federal and state governments should strengthen laws that protect targets of online hate and harassment. Many forms of severe online misconduct are not consistently covered by cybercrime, harassment, stalking and hate crime law. States should close the gaps that often prevent government agencies from holding individual perpetrators accountable because the laws do not appropriately capture online misconduct. Many states have intent, threat, harm or "directed at" requirements that prevent prosecution of online behavior that would otherwise easily fit the definitions of stalking or harassment statutes. Many states do not have swatting or doxing laws on the books.

Legislators have an opportunity, consistent with the First Amendment, to create laws that hold perpetrators of severe online hate and harassment more accountable for their offenses. Additionally, governments can require better data from law enforcement agencies regarding online harassment investigations and prosecutions. Finally, social media companies should be required to increase

transparency around online hate and harassment on their platforms, so we better understand the scale and scope of hate and harassment on social media.

CONCLUSION



Ethnographic interviews can expose important details about individual experiences that are difficult to capture through other methods. In the absence of an experience as visceral as sustained harassment, the underlying assumptions, hypotheses and priorities developed by academics that guide quantitative research might not mesh with the sentiments of targets of hate. For example, we didn't go into this study with a specific intention to learn about the ways in which friends, colleagues and relatives are exploited to increase harm and reputational damage against one target. Moreover, the details surrounding networked harassment exposed through these interviews helped us craft recommendations, like user journey maps, as a means to plug holes in platform design.

The semi-structured interviews conducted for this study remind us of the value in listening to an individual narrate a complete story, which can reveal new, unpredicted insights. And the findings of this study provide a sobering reminder of the emotional and economic harms that vulnerable communities shoulder when using online platforms. While much damage has already been done, and many lives upended because of online hate, it's not too late for social media

companies to learn about their users and redesign their platforms to mitigate further harassment.

APPENDIX: RESEARCH METHODS

This study was conducted using semi-structured, ethnographic interview methods. Fifteen interviewees were recruited and interviewed remotely, following a pre-established interview guide. Two professionals with extensive experience in ethnographic research established the guide and conducted the interviews.

Recruitment

Fifteen people who were targets of extensive cyber harassment were recruited and interviewed for this study. Interviewees were asked to indicate characteristics with which they identified or associated. Interviewees included thirteen women (including two trans women) and two men. Nine of the interviewees were white, one was Black, three were Latinx and two were Asian. Four were researchers or academics, five were in the gaming industry (players and designers), three were media professionals (though two were in different careers at the time of the interview), two owned or founded businesses, and one worked for a social justice non-profit.

Interviewees were recruited via snowball sampling. Initial interviewees were recruited through professional networks and open calls publicized on ADL and Implosion Labs' social media channels. Then, the variety and number of interviewees "snowballed." Each person was asked to recommend others who might be eligible and interested in participating.

Process

Two researchers developed and used an interview guide to conduct interviews via the videoconferencing platform Zoom, using audio only or video plus audio based on the interviewee's preference. Interviews covered the following topics:

- An explanatory introduction to the study, including the interviewees right to withdraw and a description of security measures in place;
- An invitation for the interviewees to describe themselves demographically, professionally and biographically;
- A walkthrough of a specific episode of cyber harassment, including events, platforms and responses;
- Reasons why the interviewees thought they were targeted;
- A description of how cyber harassment affects the individual and their networks or communities; and
- An opportunity to reflect on what they wish had gone differently and what could be changed to prevent similar cyber harassment from happening in the future.

Recruitment and interviewing began in February 2019 and was completed in April 2019.

ENDNOTES

1. Cyber harassment has been defined as “to involve the intentional infliction of substantial emotional distress accomplished by online speech that is persistent enough to amount to a ‘course of conduct’ rather than an isolated incident.” Citron, D. K. (2014). *Hate Crimes in Cyberspace*. Harvard University Press.

2. <https://www.adl.org/onlineharassment>

Duggan, M. (2018, January 3). Online Harassment 2017. Retrieved from <https://www.pewinternet.org/2017/07/11/online-harassment-2017/>

3. <https://www.adl.org/onlineharassment>

“Severe harassment” as defined by the Pew Research Center includes physical threats, sexual harassment, stalking and sustained harassment. ADL has adopted this definition for its research on online harassment. Duggan, M. (2018, January 3).

Online Harassment 2017. Retrieved from

<https://www.pewinternet.org/2017/07/11/online-harassment-2017/>

4. <https://www.adl.org/free-to-play>

5 Coleman, A. L. (n.d.). What Is Intersectionality? A Brief History of the Theory. Retrieved from <https://time.com/5560575/intersectionality-theory/>

6. Rosenberg, E. (2019, June 5). A right-wing YouTuber hurled racist, homophobic taunts at a gay reporter. The company did nothing. Retrieved from <https://www.washingtonpost.com/technology/2019/06/05/right-wing-youtuber-hurled-racist-homophobic-taunts-gay-reporter-company-did-nothing/>

7. Marcotte, A. (2019, August 28). Gamergate is back, because it never went away: Zoë Quinn faces new round of attacks. Retrieved from <https://www.salon.com/2019/08/28/gamergate-2-0-zoe-quinn-accuses-game-developer-alec-holowka-of-abuse/>

8. Spangler, T. (2019, June 6). YouTube Says It's Reviewing Harassment Policies, After It Condoned Sustained Attacks on Gay Latino Journalist. Retrieved from <https://variety.com/2019/digital/news/youtube-harassment-policies-attacks-gay-latino-carlos-maza-steven-crowder-1203234645/>

9. For the purposes of this study, “identity” refers to how interviewees self-define, often in terms of gender and sexuality, race or ethnicity, and sometimes in terms of immigration status, nationality, profession, religious affiliation or perceived religious affiliation, and political beliefs.

10. <https://www.adl.org/onlineharassment>

11 <https://www.adl.org/free-to-play>

12 Boyd, Danah. 2014. *It's Complicated*. New Haven: Yale University Press., 233–53. SAGE Publications Ltd.; Bucher, Taina, and Anne Helmond. 2019. "The Affordances of Social Media Platforms." In *The SAGE Handbook of Social Media*, 233–53. SAGE Publications Ltd.; Dijck, José Van, and Thomas Poell. 2013. "Understanding Social Media Logic." *Media and Communication*, 1 (1): 2–14; Majchrzak, Ann, Samer Faraj, Gerald C Kane, and Bijan Azad. 2013. "The Contradictory Influence of Social Media Affordances on Online Communal Knowledge Sharing." *Journal of Computer-Mediated Communication* 19 (1): 38–55; Costa, Elisabetta. 2018. "Affordances-in-Practice: an Ethnographic Critique of Social Media Logic and Context Collapse." *New Media and Society* 20 (10): 3641–56.

13. The term has been adapted from Caroline Sinderson's concept of "campaign harassment." Campaign harassment is a phenomenon in which harassers organize using online platforms to target an individual for cyber harassment en masse. Caroline Sinderson, 2018, An Incomplete (but growing) History of Harassment Campaigns since 2003. Medium. Accessed Dec. 20, 2018. <https://medium.com/digitalhks/an-incomplete-but-growing-history-of-harassment-campaigns-since-2003-db0649522fa8>

14 Defined as "an online "course of conduct" that either causes a person to fear for his or her safety or would cause a reasonable person to fear for his or her safety."

15. Citron, D. K. (2014). *Hate Crimes in Cyberspace*. Harvard University Press.

16. Madani, D. (2019, March 30). Serial 'swatter' Tyler Barriss sentenced to 20 years for death of Kansas man shot by police. Retrieved from <https://www.nbcnews.com/news/us-news/serial-swatter-tyler-barriss-sentenced-20-years-death-kansas-man-n978291>

17. Miller, R. W. (2019, March 30). His 'swatting' prank call caused a man's death. Now he'll serve 20 years in prison. Retrieved from

<https://www.usatoday.com/story/news/nation/2019/03/30/swatting-tyler-barriss-get-20-years-hoax-911-call-andrew-finch-death/3320061002/>

18. See also Charlie Warzel, "A Honeypot For Assholes": Inside Twitter's 10-Year Failure To Stop Harassment, *Buzzfeed News*, Aug. 2016,

19. <https://www.buzzfeednews.com/article/charliewarzel/a-honeypot-for-assholes-inside-twitters-10-year-failure-to-s>.