

Sara Morrell
12 September 2026

This content is early research output and has not been peer-reviewed.

Take No BS: The Role of Emotional Labor in Countering Extremism and Nurturing Community on Twitch

Abstract

Discrimination and extremism in gaming spaces are well documented, as is the emotional labor marginalized content creators perform to exist in these spaces. This study explores how streamers and moderators on Twitch counter harmful behaviors and the subsequent impacts on their communities. I conducted 21 interviews with content creators from February to May 2026 to understand why they selected particular moderation strategies and how these strategies shaped their communities. Qualitative content analysis of interview data revealed that emotional labor (e.g. taking no BS) can nurture inclusive and caring communities, in addition to countering immediate harms. Based on these results, I conclude that moderation grounded in content creators' lived experiences not only makes gaming spaces less hospitable to extremism, but further enables marginalized community members to express themselves freely without fear of harassment. This research points to bottom-up strategies for countering online extremism by supporting creators.

Keywords: video games, women in gaming, gender inequality, racism, platform governance, emotional labor, online harassment, extremism, Twitch

Introduction

Discrimination and extremism in gaming spaces are well documented (Schlegel and Kowert, 2024), as is the emotional labor marginalized content creators perform to exist in these spaces (Tran, 2024a; Guajardo, 2022). This research examines how streamers nurture inclusive communities within hostile gaming-adjacent spaces (Harris et al., 2026) and the broader political implications of this labor for online safety and equity. The two questions guiding this research are *why do streamers choose specific measures to combat harassment?* and *how do these measures impact their communities?* I interviewed 21 Twitch streamers and moderators, most of whom identify as women or femme-presenting or with other marginalized identities, on why they choose specific measures to combat harmful behaviors, how those measures impact their communities, and what recommendations they have for platform moderation. These interviews provide direct insight into the decision making processes that underlie how moderation is performed in practice. This research contributes to the literature on virtual communities and content moderation by revealing that emotional labor (e.g. taking no BS) is not only expended to

counter harms, but can also nurture caring communities that are less vulnerable to extremism. It also provides actionable policy and technical recommendations for platform governance and online safety of interest to practitioners.

The next section reviews the literature on virtual communities on Twitch, moderation, and extremism, and bridges the gap between the literature on emotional labor and on counterspeech. The methods section reviews how interviews were conducted and research ethics. The findings and discussion section presents the results of qualitative content analysis of interview themes and patterns in theme co-occurrence, how these results relate to the research questions, and policy recommendations. These themes reveal ongoing processes of trial and error shaping moderation strategies, within the context of interviewees' experiences and adversarial shifts by perpetrators. I conclude with a summary of the findings, limitations, and directions for future research.

Literature Review

Twitch is a platform where individuals live-stream themselves engaging in activities, such as cooking, making art, or, most commonly, playing video games, while engaging with viewers through a chatbox adjacent to the stream (Taylor, 2018). The platform affordances for interactivity enable “real-feeling relationships between creators and audiences” (Tran, 2022: 513). As of 2023, Twitch had over “2.5 million concurrent viewers per month (on average)” (Navarro and Tapiador, 2023). Gaming is intertwined with politics (Hope et al., 2025; Ruiz-Bravo et al., 2022) and online visibility carries risks for marginalized individuals (Hewa and Tran, 2024; Tran, 2024b). Community is central to Twitch (Hamilton et al., 2014) and, according to Starks et al. (2026), community forms the basis of safety for marginalized users. Some streamers choose to include their community of viewers in moderation processes (Thach et al., 2024) and moderators who monitor the chat are often community members (Cai and Wohn, 2023).

Moderation is defined as “the governance mechanisms that structure participation in a community to facilitate cooperation and prevent abuse” (Grimmelmann, 2015: 52, found in Bulat et al., 2025). Common moderation strategies include working with human moderators (Seering et al., 2019), filtering particular words (Jhaver et al., 2022), sharing block lists of suspicious users, limiting which accounts can post in the chat (Meisner, 2023), and detecting potential ban evaders (Thach et al., 2024). Moderator activity has been associated with “more individual participation and financial commitment for all streamers...” (Wolff and Shen, 2024) and streamers may use moderation practices to shape their community and teach chatters community norms. Moderation approaches are customizable, resulting in a wide variety of moderation strategies across channels (Seering et al., 2017). In addition to providing flexibility, this community-led moderation approach also places a significant burden on streamers and moderators to respond to harmful behaviors and content.

Despite near gender parity among gamers in the United States (Entertainment Software

Association, 2021) and Europe (Kuss et al., 2022), gaming is often considered masculine (Deng, 2026) and discrimination is rampant (Koung et al., 2025; Kasimov, 2025; Deng, 2026; Zorlu and Özkan, 2020). For example, streamers with marginalized identities have faced hate raids, where the chats adjacent to their Twitch streams are overrun with hateful comments (Cai et al., 2023; Powell and Williams-Johnson, 2025). In addition to fostering toxic online environments, misogyny has been identified as a gateway to engaging with extremist content (Weiss et al., 2025), and video game spaces and aesthetics have been exploited by extremists (Kingdon, 2024; Schlegel and Kowert, 2024). For example, attackers have mimicked video game livestreams and made gaming cultural references while carrying out mass violence (Lankford et al., 2024; Macklin, 2019). In some instances, radicalization towards violence has been furthered by the community dynamics associated with online video games as users form close social bonds (Koehler et al., 2023). Systemic failures to counter this hostility in gaming spaces lead to a toxic environment (Kasimov, 2025) in need of more effective moderation.

Marginalized individuals in gaming spaces may be further disadvantaged by the moderation systems meant to protect against abuse (Delmonaco et al., 2024; Mayworm et al., 2024). Despite its many policies focused on countering abuse (Hallinan et al., 2026), Twitch has been criticized for practices sustaining white supremacy (Powell and Williams-Johnson, 2025) and policies reenacting discrimination (Thach et al., 2024; Zolides, 2021; Harris et al., 2026). According to Harris et al. (2026: 1564), Twitch's "governance framework overwhelmingly shifts responsibility onto streamers, requiring them to navigate ambiguities and contradictions in Twitch policies while bearing risks of enforcement failures." Based on interviews with women and LGBTQ streamers, Uttarapong et al. (2021: 17) find that "marginalized streamers rely heavily on human labor with insufficient technical support from the platform to handle harassment..." While Twitch has made more technical tools available to streamers in recent years (Twitch, n.d. a), much of the moderation burden still falls to streamers and moderators.

In addition to physical and intellectual labor, streamers perform emotional and affective labor as part of moderation, (Guajardo, 2022; Hu and Ge, 2025; Harris et al., 2026; Carradore and Pirola, 2025; Woodcock and Johnson, 2019). Emotional labor "requires one to induce or suppress feeling in order to sustain the outward countenance that produces the proper state of mind in others" (Hochschild, 1983: 21, found in Guajardo, 2022: 4). This emotional labor may involve maintaining a positive demeanor in the face of threats and harassment while simultaneously taking or coordinating real-time moderation decisions (Bulat et al., 2025). Guajardo (2022: 5), summarizes the emotional labor performed by women streamers:

...[W]omen streamers are entering a space which actively seeks to gatekeep them out, does not take their skill seriously, polices their bodies, and often directly harasses and threatens their very existence on the platform. And they are expected to do it with a full face of makeup and a smile on their face (Guajardo, 2022: 5).

Examples of the emotional labor that streamers perform to counter harassment include confronting the perpetrator(s), cognitively developing “thick skin”, and regulating their feelings while on stream (Uttarapong et al., 2021: 16). While emotional labor focuses on regulating personal emotions, affective labor “influences the emotional states of others” (He and Agur, 2026: 2761). Affective labor in streaming includes “being compelling to watch and friendly to viewers, soliciting donations, building parasocial intimacy with spectators, and engaging audiences through humor” (Woodcock and Johnson, 2019: 814). Both of these concepts describe the labor performed by streamers as they respond to harmful behaviors, teach viewers community norms (Thach, 2023), keep their communities “‘healthy’, ‘clean’, ‘safe’ and ‘peaceful’” (Carradore and Pirola, 2025: 118), and develop a “sense of spending time with a friend” (Woodcock and Johnson, 2019: 817). Prior work has found that performing this emotional and affective labor has significant costs for streamers (Woodcock and Johnson, 2019), in some cases leading to burnout (Carradore and Pirola, 2025).

While the literature on affective and emotional labor often presents them as another burden placed on marginalized streamers, the literature on counterspeech potentially calls this into question by pointing to how this labor may have positive impacts beyond responding to a seemingly endless barrage of hate. Counterspeech can be defined as direct responses to “hateful and dangerous speech online” in order to “refute or undermine it” (Buerger, 2021: 1). One form of emotional labor is directly responding to harmful content by confronting the perpetrator (Uttarapong et al., 2021). This confrontation may serve as a form of counterspeech, through which streamers may shape their communities, as well as broader norms in gaming spaces. According to Meisner (2023: 6), “[d]uring and after hate raids, some streamers felt that their continued presence on Twitch was an important show of strength for their community...” This show of strength could be considered a form of counterspeech and affective labor, as streamers may seek to impart feelings of strength onto their community. Prior literature has found that prosocial community behavior on Twitch can be encouraged through imitation effects (Seering et al., 2017: 120). Emerging research has also identified the potential of grassroots counterspeech to address extremism online (Czerwinsky, 2026; see also O’Connor, 2021). This indicates that the emotional and affective labor of speaking back against harmful behaviors could shift norms within individual communities, and potentially within gaming spaces overall, to be less hostile to marginalized streamers.

This study explores two research questions. The first research question is further broken down into two hypotheses:

RQ 1: Why do streamers choose specific measures to combat harassment?

H 1: Community values shape decisions about which measures to use to combat harassment

H 2: Tool accessibility shapes decisions about which measures to use to combat harassment

RQ 2: How do these measures impact their communities?

The first question contributes to the literature on moderation in gaming spaces and the role of community. Given the extent to which streamers are expected to learn moderation tools (Harris et al., 2026), it also addresses the practicalities of tool accessibility. The second question contributes to the literature on counterspeech and whether emotional and affective labor can further prosocial community behaviors.

Methodology

I conducted semi-structured interviews with 21 Twitch streamers and moderators. I then analyze themes in interviews using qualitative content analysis and co-occurrence network analysis. This mixed-method approach affords both qualitative ‘thick’ description and systematically mapping out decision making processes. These processes shape how moderation decisions counter immediate harms and adversarial shifts, as well as develop online community norms.

Data Collection

Most interviewees are associated with the Twitch Unity Guilds and have been on Twitch for over three years. The Twitch Unity Guild program, started in 2023, provides streamers from underrepresented identities an opportunity to connect with one another. As of this writing, there are four Unity Guilds: the Women’s Guild, the Black Guild, the Latin Guild, and the Pride Guild. The Unity Guilds’ mission statement reads:

We believe that our community should make a space where all can share their authentic voice and experiences with peers within a safe space where privacy is respected, concerns are acknowledged, and differences are celebrated (Twitch, 2026)

The statements of purpose for the Women’s Guild and the Black Guild also refer to creating safe spaces (Twitch, n.d. b; Twitch, n.d. c). These statements, in addition to patterns of harassment against streamers with marginalized identities (Cai et al., 2023; Powell and Williams-Johnson, 2025), indicate that streamers in these guilds likely have experience creating and maintaining safe spaces.

Most participants were recruited from the Unity Guilds’ public members lists, others were recruited through snowball sampling. Given the focus of this study on women and femme experiences, participants were first recruited from the Women’s Guild, resulting in approximately the first 13 interviews. Participants identifying as women or femme were then recruited from the three other guilds. For the purpose of this research, I considered streamers who were part of the Women’s Guild, used she/her pronouns, used ‘femme’, ‘fem-’ (including femboy), Ms., or Mrs

in any of their self descriptions, or used nouns associated with women to describe themselves, to be women/femme streamers. Streamers recruited through snowball sampling could be of any gender. Many participants also have overlapping guild memberships. Participants were contacted over email. I collected emails at various points throughout the week and weekend to obtain a wide variety of streamers (Johnson, 2025), as there may be systematic differences in levels of chat interaction for streamers who stream during the workday, as opposed to those who stream during the evenings or weekends, necessitating different levels of moderation. I also selected streamers with a range of viewership counts.

Interviews

Interviews were conducted over Zoom (participants were not required to turn their video on) and typically lasted between 15 minutes and an hour. Questions focused on understanding the streamer's community, experiences with harassment or other harmful behaviors, how they addressed those harmful behaviors, the associated impacts on their communities, and platform level moderation. The interview protocol is available in Appendix A. Interview notes were coded inductively for primary themes. The analysis conducted here is similar to that performed by Kounig et al. (2025) in their study on gendered toxicity in gaming and by Thach et al. (2024) in their study of moderation practices (see also Obreja, 2023; Tran, 2024b; Kojah et al., 2026; Uttarapong et al., 2021). Interview notes were first reviewed to identify primary themes. More specific codes were then developed and each interview was coded for the presence or absence of each code. The initial set of codes was created after conducting 13 interviews, when persistent themes became apparent. It was then iteratively revised with subsequent interviews and interview notes were re-coded as needed. In total, 21 interviews were conducted and coded for the presence or absence of 243 thematic codes, sorted into 37 categories. A full codebook detailing categories, codes, and code frequency is available in Appendix B.

Ethics

Privacy, confidentiality, and participant well-being were prioritized throughout this research. It was approved by the Northeastern IRB and designated Exempt, Category #2(ii) (IRB #: 25-11-48). The IRB approval letter is available in Appendix C. Given the sensitivity of the subject matter and the risk of harassment, interviews were not recorded. Instead, notes were taken during and after the interviews. This study focuses on general themes and recommendations to minimize re-identification risks. It does not share individual-level results. It also does not systematically collect participant attributes beyond gender expression, guild membership, time on Twitch, and whether they describe playing single or multiplayer games, or neither, as reporting additional characteristics may enable re-identification. Interviews were conducted using trauma-informed practices, such as reminding participants they can choose not to answer any question and can take a break or stop the interview at any time.

Positionality

While I did not discuss my identity with participants, I am white and femme-presenting, which participants could observe from my email photo and over Zoom. These attributes may have influenced aspects of participants' experiences they shared and how I understand those experiences (Kojah et al., 2026; Archer, 2002). Feminist research has explored how interviewer and interviewee identities may influence the interview process (Thwaites, 2017). While it is possible that my woman/femme identity enabled me to develop rapport with femme-presenting interviewees, leading to more open discussions of gender-based harms, my identity as a white person may have led some participants to avoid discussing issues of racism, leading to underreporting. My positionality may also shape how I interpret participants' stories. I do not have personal experience facing racism, limiting my ability to understand the experiences of people of color with the same nuance that another person of color might be able to understand them with (Archer, 2002; Collins, 2009). This may have impacted my ability to follow up on particular statements in participants' answers and identify patterns related to race or ethnicity in the data. I am also heterosexual and cis-gender. This aspect of my identity could similarly impede my ability to recognize patterns related to harassment targeting LGBTQIA+ participants. These limitations highlight the importance of increasing researcher diversity and engaging in reflexive research approaches (Morales et al., 2026), as I do not know what I do not know.

Findings and Discussion

This section first presents a summary of common themes in the interviews and co-occurrences among them. It then discusses themes and processes relevant to each of the research questions. Last, it presents recommendations from participants.

Qualitative Content Analysis

Codes that appeared in 8 or more interviews and their categories, excluding time on Twitch, are listed in Table 1. The full codebook, with frequencies for all codes, is available in Appendix B.

Table 1

Frequent Categories and Codes

Category	Code
Community Sentiment	Like community: Expresses that they like their community

Category	Code
Harassment Experience	Harassment decreasing: Describes harassment, in general, decreasing over the years Harassed less than average: Describes themselves as being harassed less than average or describes others as having experienced more severe harassment
Harassment Responses	Maintaining composure: Describes streamers learning how to maintain composure and not let negative behaviors ruin their experience
Moderation Community Impact	Positive impact: Describes how moderation measures have a general positive impact on a community Safe environment: Describes a streamer providing a safe or comfortable environment
Moderation Development	Trial and error: Describes how streamers or moderators develop moderation strategies through a process of trial and error that often involves experiencing harm until effective strategies are developed Personal preferences: States that the right moderation approaches or strategies for the streamer depends on the streamer's personal preferences
Sexism	Appearance comments: Describes perpetrators making comments about appearance or saying a streamer should smile more
Social Management Strategies	Directly addressing harms: Describes directly addressing harmful behaviors
Social Network Targeting	Friend harassment: Describes how friends in video game communities were also harassed
Streamer Identity Targeting	Identity-based harassment: Describes harassment as being based on attributes of the targeted person's identity
Technical Improvements	Moderation tools improving: Describes how technical moderation tools in general on Twitch have improved over time or that moderation on Twitch is improving in terms of technical tools available
Technical Moderation Strategies	Banning and/or blocking: Describes banning and/or

Category	Code
	blocking perpetrators on Twitch and/or on other contact methods
	Verified chatters: Describes requiring that a viewer have a verified username or phone, or meet a criteria for a certain length of time, before that viewer is able to chat
	Text monitoring: Describes using a blocked words list or text monitoring
Technology-Enabled Harms	Hate raids: Describes streamers receiving hate raids
	General botting: Describes streamers receiving general botting
Threats	Sexual advances: Describes perpetrators sharing unsolicited sexual content and/or making advances
Twitch General	Twitch improving: States that Twitch is improving

Note. This table contains codes that appear in eight or more interviews and their categories.

Figure 1 displays the frequency of each of these codes, colored by category.

Figure 1

Code Frequency

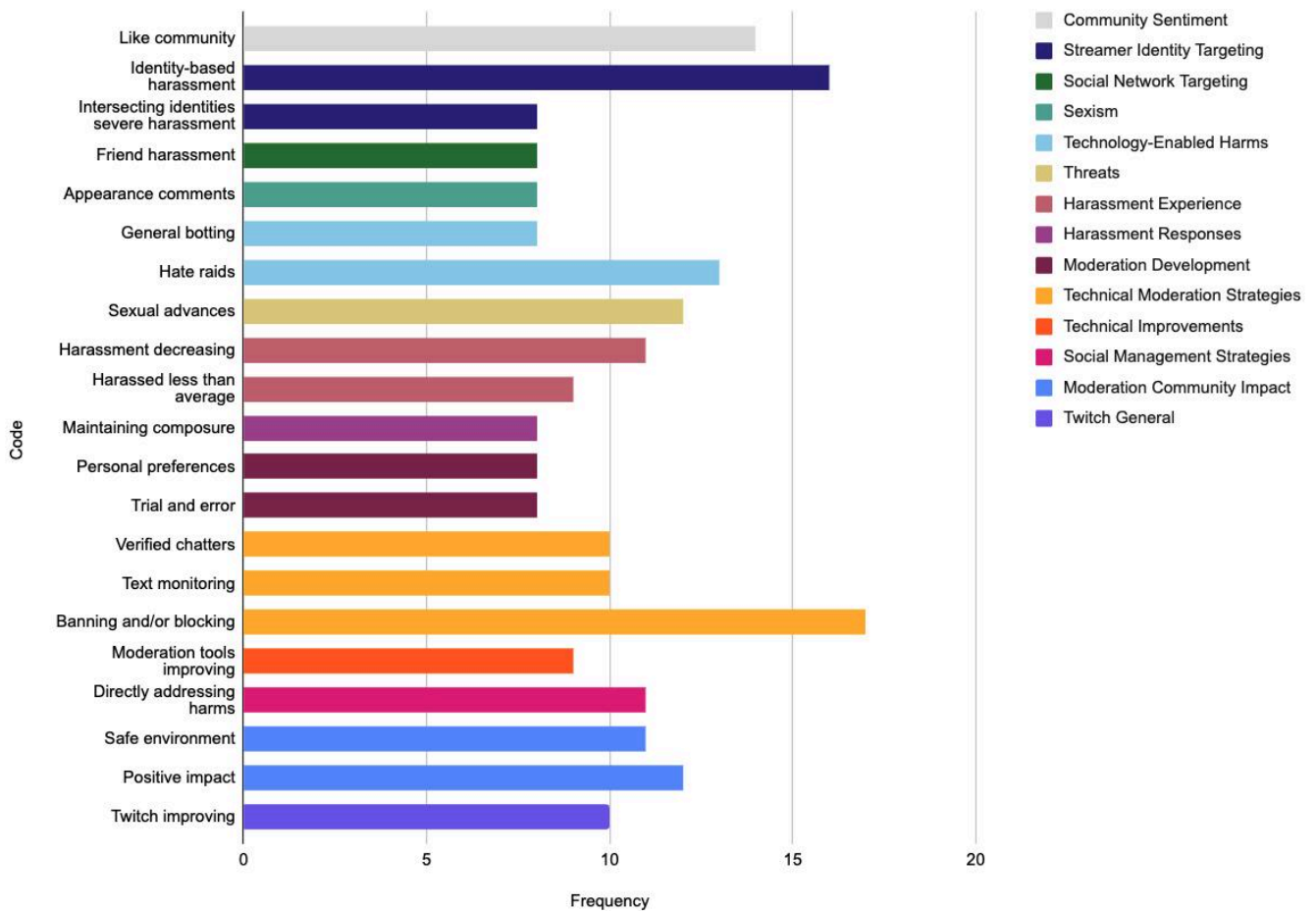
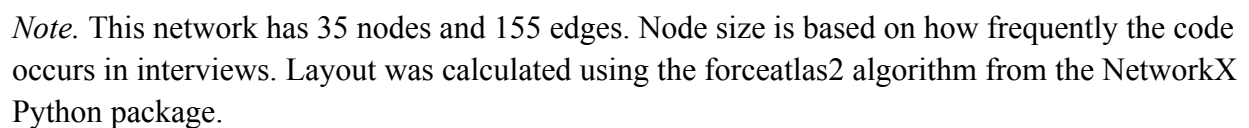


Figure 2 displays the co-occurrence of frequent codes across the interviews. For two codes to be connected, they must co-occur in six or more interviews.

Network of Codes with Six or More Co-occurrences



11

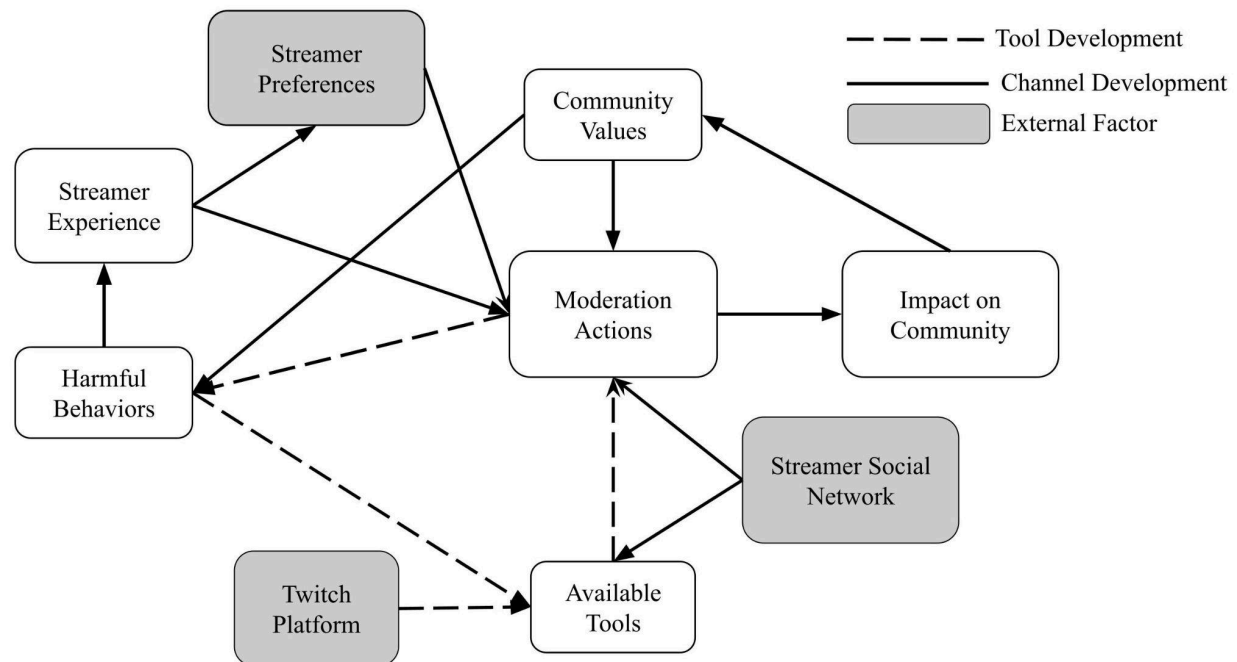
such as sexist content. Community 2 may represent the manifestation of identity-based harassment in hate raids and common responses to hate raids such as maintaining composure and leveraging tools such as text monitoring and only allowing verified accounts to post in chat. Community 3 may represent the positive aspects of improvements in moderation tools and resultant decrease in harassment and benefits to streamers' communities. Last, community 4 may represent the role of banning and/or blocking perpetrators in protecting a diverse community, as well as perpetrators' negative reactions and attempts to evade the ban. Python code for developing this network and identifying communities is available in Appendix D.

RQ 1: Why do streamers choose specific measures to combat harassment?

Results provide some support for H1 and H2. However, while tool accessibility and community values are two of the factors shaping decisions, moderation decisions are also informed by the streamer's experience, preferences, and social network; and are made through ongoing processes of trial and error contextualized within the streamer's broader ambitions to build the type of community they want to interact with. Figure 3 outlines these overlapping processes.

Figure 3

Moderation Processes



These processes fall into two primary categories: channel development (indicated with a solid line) and tool development (indicated with a dashed line). Factors external to these processes are indicated by a gray square; these are ‘Streamer Preferences’, ‘Streamer Social Network’, and ‘Twitch Platform’.

Several participants stated that communities change over time ($N=6$). In the channel development processes, some interviewees described how newer streamers may be unsure of how to articulate their preferences ($N=2$) or stop harmful behaviors ($N=3$). Participants also described how some streamers may be afraid to ban people or set boundaries because they do not want to be perceived as annoying or lose viewers ($N=5$), feel like they cannot say they are uncomfortable ($N=2$), or do not want to hurt a viewer’s feelings ($N=1$). Participants also described how economic precarity, or reliance on Twitch for income, and small size can lead to riskier moderation decisions to avoid losing potential donors ($N=2$). Without sufficient moderation, participants described how harassment will worsen ($N=3$). This process can be characterized as a negative feedback loop, where a lack of streamer experience and uncertainty about preferences leads to insufficient moderation actions which can then allow further harmful behaviors.

Participants described how streamers or moderators develop moderation strategies through a process of trial and error that often involves experiencing harm until effective

strategies are developed ($N=8$), and one participant stated that if new streamers cannot learn moderation skills quickly they may leave streaming ($N=1$). Participants described learning how to present themselves online ($N=3$) and where their boundaries are ($N=4$). Several described how they learned to troll back the trolls ($N=5$). Many participants emphasized that the right moderation approaches or strategies for the streamer depends on the streamer's personal preferences ($N=8$) and moderation approaches can vary widely across streams ($N=5$). While one participant described their moderation capabilities as self-taught ($N=1$), others described how streamers share tools and tips to help educate one another ($N=5$), give advice to new streamers ($N=4$), and mentor others in the video game space ($N=5$). Moderation tools are also often shared through word of mouth ($N=2$). Two participants pointed out that streamers sharing tools and tips can lead to conflicting advice ($N=2$) and several described needing more comprehensive education and guidance on Twitch capabilities ($N=3$).

In addition to learning personal preferences and boundaries, harmful behaviors also shape streamers' experiences, which then influence the types of moderation they prioritize. The most common form of online harm participants described was hate raids ($N=13$). This type of harmful behavior can often be countered through restrictions on who can chat ($N=10$) and text filtering ($N=10$). In contrast, some streamers described more interpersonal and persistent threats such as stalking ($N=4$) or harmful parasocial behaviors ($N=6$). These harms diverge from more common chat-based harms, as perpetrators may be willing to spend time circumventing common countermeasures, and streamers sometimes take more costly measures, such as moving ($N=1$) or avoiding in-person events ($N=2$). Many participants also stated that they were harassed less than average ($N=9$). This is striking given the prominence of harmful behaviors described by many participants. Interviewees also emphasized the importance of not sharing location information, for example through in real life (IRL) streams, and the importance of not having children on stream.

Tools may be developed by third parties to counter specific problems ($N=2$) and then integrated more formally into the platform by Twitch ($N=1$). Several participants stated that third party bots are useful ($N=5$). Tool availability shapes what moderation actions can be taken. In some cases, streamers may find out about a tool through word of mouth ($N=2$), making streamer social networks an important mediating variable between available tools and moderation actions. Overall, many participants emphasized that the technical tools provided by Twitch have improved over time ($N=9$). Moderation actions (or lack thereof) then shape harmful behaviors, as described above.

Some streamers saw chat-based harassment as a moment for teaching proper behaviors to the offender ($N=4$) and teaching their viewers and community members how to counter hatred ($N=3$). Some described how they give offenders a second or third chance ($N=5$) (see also Utarapong et al., 2021). Participants also emphasized the importance of directly addressing harmful behaviors ($N=11$), taking no BS ($N=5$), and repairing the harm done to other viewers who were exposed to problematic behavior by addressing the problematic behavior and not exclusively punishing the perpetrator ($N=1$). One participant noted that if harassment is not

directly addressed it can persist over time, even if the original perpetrator was banned ($N=1$). Participants noted that harassment decreased over time as they put up with it less ($N=3$). Once rules are well established in a community, some participants described how the community can become self moderating ($N=4$) and some viewers can act as ‘knights in shining armor’ that help keep the community safe ($N=5$). Many participants described how streamers deliberately build the type of community they want to interact with, or when streamers start they do not have a good community but they shape it to be better over time ($N=7$). These learning processes on behalf of the streamer and the community can be characterized as a positive feedback process where the streamer’s known preferences inform their moderation actions, which then shape their community over time, leading to less harmful behaviors.

Another participant noted, however, that white cisgender femme presenting streamers usually have safer communities and are more likely to have protective community members ($N=1$). They are also able to be more direct in setting boundaries, while femme presenting creators of color might feel pressure to say things more softly to avoid being perceived as harsh ($N=1$). This participant’s observations point to ways in which race shapes not only the harmful behaviors streamers are targeted with, but also how the social tools available for countering harms may be limited by societal stereotypes. Jones and Norwood (2017: 2049) describe negative stereotypes applied to Black women as including being “disagreeable, overly aggressive, physically threatening, loud (even when she speaks softly), and to be feared.” In her discussion of controlling images, Collins (2009: 84) writes that “... labeling Black women unfeminine and too strong works to undercut U.S. Black women’s assertiveness.” These labels and stereotypes may still be active in online contexts, leading femme creators of color to moderate their assertiveness in setting boundaries more than their white counterparts, who may be considered in more need of ‘protection’ according to societal biases (Phipps, 2021).

RQ 2: How do these measures impact their communities?

Interviewees consistently emphasized the positive impacts of moderation on their communities ($N=12$), although some did note that it limited their growth ($N=2$). These positive impacts included providing a safe or comfortable environment ($N=11$), protecting community members ($N=6$), and increased freedom of expression ($N=5$). Harassment also decreased over time as streamers put up with it less ($N=3$) and some communities became more self-moderating ($N=4$), as described above. Some participants also stated that viewers appreciate that they will not be harassed for being themselves in a community ($N=3$). These impacts are not only the result of technical moderation strategies, but also the result of the emotional and affective labor performed by streamers and moderators to build healthy communities by directly addressing harmful behaviors ($N=11$). This aligns with prior work on affective labor, counterspeech, and prosocial community behaviors (Mathew et al., 2019; Carradore and Pirola, 2025; Seering et al., 2017). The importance of this labor also implies that exclusively punishment focused approaches may not be effective for fostering healthy online communities.

Recommendations

At the conclusion of each interview, I asked: *If you could have any moderation tool from the Twitch platform, what would that look like?* Participants provided recommendations for technical improvements, policy refinements, and additional communication strategies. Technical improvements included the below recommendations:

- Accessibility: Participants stated that moderation tools should be more accessible and/or not buried or hard to find ($N=3$)
- Text: One participant shared they would like to have an automated message displayed to a chatter who breaks the channel rules ($N=1$). Others stated they would like improved text filtering for slurs ($N=1$) and the ability to ban phrases ($N=2$).
- Platform-Level Shared Ban List: Participants wanted easier access to shared ban lists ($N=2$), potentially including a platform-level shared ban list ($N=2$).
- Automatic Reports: A participant recommended an option to automatically send a report to Twitch when a user is banned, with the ability to specify why the user was banned ($N=1$).
- Restrictions: Many participants discussed the ease with which perpetrators create multiple accounts to continue harmful behaviors ($N=7$). To help address account creation issues and prevent stream access while logged out, some proposed that bans include an IP ban ($N=3$). Participants also proposed that banning one account would also ban all accounts with that same verified phone number ($N=1$), and recommended that accounts not be able to access profiles from which they are banned ($N=2$).
- Moderators: A participant described wanting an easier way to communicate with moderators on the Twitch platform ($N=1$). Another participant recommended allowing moderators to continue viewing the channel during ads ($N=1$) and to invoke commands without displaying their name in the chat ($N=1$).
- Age Verification: One participant identified age verification as a way to protect children ($N=1$).
- Bots: Participants noted Twitch should avoid disrupting useful third party tools ($N=1$) and flagged harmful botting as a common issue ($N=8$), with follow botting being the most common specific form discussed ($N=5$).

Policy refinements included the below:

- Consistent and Unbiased Enforcement: Some participants described wishing that Twitch would not tolerate streamers who engaged in harmful behaviors ($N=3$), being frustrated with Twitch for not doing more to punish streamers who engage in harmful behaviors toward other streamers ($N=2$), and wanting Twitch staff to be unbiased in interacting with all streamers ($N=3$). Another participant recommended that if someone says they want to rape or assault someone, they should be banned from the platform ($N=1$).
- Reports and Investigations: Participants would like to have reports to companies taken seriously and acted on ($N=3$). One participant also described wanting an actual human they can talk to at Twitch when problems arise ($N=1$).

- Event Moderation: Participants expressed the need for additional moderation support and guidance for events highlighting creators with marginalized identities (e.g. heritage months), given the foreseeable harassment these events draw ($N=2$).
- Proactive Solutions: One participant described having to go through harmful experiences to then get tools to address those harms and wishing they did not have to go through those experiences ($N=1$).
- Additional Options: One participant recommended developing options for addressing harmful behaviors beyond exclusionary punishments ($N=1$).
- Mental Health: Participants recommended improved moderation around mental health ($N=2$).
- Inclusive Policy Development: One participant stated that Twitch needs to have members of marginalized communities on staff to understand potential harms to underrepresented communities that policies could cause ($N=1$). Another recommended additional training on diversity, equity, and inclusion ($N=1$).

Lastly, participants proposed improvements in how Twitch communicates with streamers:

- Education: Participants expressed the need for additional education about moderation tools. One described wanting regular communications on how to use moderation tools (e.g. emails or streams) ($N=1$). Another described wanting asynchronous educational onboarding for moderators on potential harms ($N=1$).
- Responsiveness: Some participants felt that Twitch failed to help with a particular problem or was slow in responding ($N=3$), while others stated that Twitch does try to help ($N=4$). One described wanting Twitch to communicate when they are working on a problem ($N=1$).
- Resources: Participants also wanted better resources for individuals experiencing harassment ($N=2$).

Some of these recommendations may already be in place, but participants may not know about them, highlighting the need for additional education and communication. Many participants also stated that Twitch is improving ($N=10$), and some stated that Twitch does a good job of providing streamers with moderation tools ($N=6$). Participants also stated that more harassment happens on other platforms than on Twitch, or that Twitch is better than other platforms in terms of harassment ($N=7$). Lastly, two described wishing that other platforms would adopt some of the tools Twitch has ($N=2$).

Conclusion

This study examines the research questions: *why do streamers choose specific measures to combat harassment?* and *how do these measures impact their communities?* through interviews with 21 streamers and moderators. It finds that moderation decisions are shaped by multiple processes involving not only community values and tool accessibility, but also streamer preferences, experiences, and social network as well as trial and error. A major component of effectively countering harassment was directly calling out harmful behaviors (e.g. taking no BS).

This emotional and affective labor, in conjunction with other moderation measures, positively impacted streamers' communities, and, as harassment decreased, community members were able to express themselves more freely.

This work contributes to the literature on emotional and affective labor and digital communities by showing the critical role that these forms of labor play in developing healthy inclusive communities. The emphasis placed on directly countering issues and calling out problematic behaviors by participants ($N=11$) partly contradicts prior work showing that female gamers often hesitate to confront sexism (Deng, 2026). It is possible that this difference is due to a streamer's role as a community leader. It could also be a result of self-selection into this study, as those who are willing to confront harassment may be more willing to participate in a study on harassment. This work also shows that experiences vary significantly based on identity (Powell and Williams-Johnson, 2025), as identity shapes the harms streamers are targeted with and the social tools they have available to counter these behaviors may be influenced by societal stereotypes (Jones and Norwood, 2017; Collins, 2009; Phipps, 2021). Lastly, the results show that moderation can support engagement and free speech in online communities. This work points to bottom-up strategies for countering extremism online (see also Czerwinsky, 2026) by providing digital leaders the resources, tools, and support they need to keep their communities safe.

The primary limitations of this study are the small sample size and potential sampling biases. Sampling biases include survivorship, self-selection based on experiences and interests, availability to volunteer time, and language. By interviewing Twitch streamers and moderators with publicly available emails, streamers who stopped streaming due to harassment are excluded. Therefore, some of the worst harms may not be represented. Many participants also had a preexisting interest in, and experience with, working against harassment. This could lead to the over-representation of participants with an above-average understanding of moderation tools and strategies. This study did not offer monetary compensation to participants and was therefore limited to participants who could afford to volunteer an hour of their time. More economically precarious streamers may be under-represented in this research. Recruitment emails were also sent, and interviews were conducted, in English. This excludes streamers whose primary language may not be English. It is possible that those facing linguistic marginalization have different harassment experiences and moderation strategies. This is also a small- n study and the experiences of the 21 streamers and moderators are not representative of all streamers. Lastly, interview notes were coded by the sole author, therefore, inter-rater reliability was not calculated, potentially limiting the reliability of the coding results. Future work could test the theoretical insights and proposed relationships identified by this research. This work could be carried out in the form of surveys that can reach a larger number of streamers, potentially with monetary compensation and in multiple languages.

AI Use Statement

During the preparation of this work the author used Elicit and Litmaps to assist in identifying related academic work and Google Gemini to assist in Python code generation and revision for analysis. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Funding Statement

The author received no financial support for the research, authorship, and/or publication of this article.

Acknowledgements

I would like to thank the Twitch streamers and moderators who generously contributed their time and experience to this study. I would also like to thank Dr. Kirsten Rodine-Hardy, Dr. Denise Garcia, and Dr. Nicholas Beauchamp for their valuable feedback and insights on the methods and topics addressed in this paper, as well as the members of the Extremism and Gaming Research Network for their thoughtful feedback on this work.

References

- Archer L (2002) 'It's Easier that you're a Girl and that you're Asian': Interactions of 'Race' and Gender between Researchers and Participants. *Feminist Review* 72(1): 108–132.
- Buerger C (2021) Counterspeech: A Literature Review. Report, Dangerous Speech Project. University of Connecticut, USA, June.
- Bulat B, Wang H, Fujimoto S and Frey S (2025) The psychology of volunteer moderators: Tradeoffs between participation, belonging, and norms in online community governance. *New Media & Society* 27(11): 5962–5985.
- Cai J, Lin YF, Zhang H and Carroll JM (2024) Third-party developers and tool development for community management on live streaming platform twitch. In: *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (eds FF Mueller and P Kyburz), Honolulu, HI, USA, 11–16 May 2024, pp. 1–18. New York: Association for Computing Machinery.
- Cai J and Wohn DY (2023) Understanding Moderators' Conflict and Conflict Management Strategies with Streamers in Live Streaming Communities. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (eds A Schmidt et al.), Hamburg, Germany, 23–28 April 2023. New York: Association for Computing Machinery.
- Carradore R and Pirola T (2025) Streaming out of the spiral of silence. Women's visibility in

- gaming community and on twitch. *Internet Histories* 9(1-2): 107–127.
- Collins PH (2009) *Black Feminist Thought*. London: Routledge.
- Czerwinsky A (2026) From the Ground Up: Citizen-Led Counterspeech Against Male Supremacism on Reddit. Report, Global Network on Extremism & Technology. Available at: <https://gnet-research.org/2026/08/07/from-the-ground-up-citizen-led-counterspeech-against-male-supremacism-on-reddit/> (accessed 10 September 2026).
- Delmonaco D, Mayworm S, Thach H, Guberman J, Augusta A, and Haimson, OL (2024) "What are you doing, TikTok?": How Marginalized Social Media Users Perceive, Theorize, and "Prove" Shadowbanning. In: *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCWI) (ed J Nichols), April 2024, pp. 1-39. New York: Association for Computing Machinery.
- Deng Z (2026) "They don't mean to hurt": Female gamers' reluctance in recognizing and confronting sexism in gaming as an online-offline juxtaposition. *New Media & Society* 28(1): 270–288.
- Entertainment Software Association [ESA] (2021) Essential Facts about the Video Game Industry. Available at: <https://www.theesa.com/resource/2021-essential-facts-about-the-video-game-industry/> (accessed 10 September 2026).
- Grimmelmann J (2015) The virtues of moderation. *Yale Journal of Law & Technology* 17: 42–109.
- Guajardo AM (2022) "It sucks for me, and it sucks for them": The emotional labor of women Twitch streamers. In: *Proceedings of DiGRA 2022 conference: Bringing worlds together*, 1 January 2022. Digital Games Research Association.
- Hallinan B, Reynolds C, Scharlach R, Theiler D, Niv N, Rothenstein O, Knief I and Kuperberg Y (2026) Priorities and exclusions within Trust and Safety industry standards. *New Media & Society* 28(8): 3986–4010.
- Hamilton WA, Garretson O and Kerne A (2014) Streaming on twitch: fostering participatory communities of play within live mixed media. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, Toronto, CA, April 26–May 1, 2014, pp. 1315–1324 New York: Association for Computing Machinery.
- Harris BC, Tran CH and Persaud CJ (2026) Contextual governance & androcentric hegemony on Twitch. tv. *New Media & Society* 28(4): 1550–1570.
- He T and Agur C (2026) The platformization of emotions: Managing affective labor in platform-mediated game work. *New Media & Society* 28(6): 2760–2778.
- Hewa N and Tran CH (2024) Verified play, precarious work: GamerGate and platformed authenticity in the cultural industries. *New Media & Society* 26(11): 6695–6714.
- Hope R, Tran C and Taylor N (2025) Containing games. *Critical Studies in Media Communication* 42(5): 345–352. DOI: 10.1080/15295036.2025.2599286.
- Hochschild AR (1983) *The Managed Heart: Commercialization of Human Feeling*. Berkley:

University of California Press.

- Hu T and Ge L (2025) Affective liquidity, synthetic bonds: VTubers as posthuman mediators of digital intimacies. *New Media & Society*. Epub ahead of print 24 December 2025. DOI: 10.1177/14614448251403091.
- Jhaver S, Chen QZ, Knauss D and Zhang AX (2022) Designing word filter tools for creator-led comment moderation. In: *Proceedings of the 2022 CHI conference on human factors in computing systems* (eds S Barbosa et al.), New Orleans, USA, April 29–May 5 2022, pp. 1–21. New York: Association for Computing Machinery.
- Johnson MR (2025) Time zones, time slots and Twitch: When do game streamers go live, and why? *New Media & Society* 28(9): 4925–4941.
- Jones T and Norwood KJ (2017) Aggressive encounters & white fragility: Deconstructing the trope of the angry black woman. *Iowa Law Review* 102: 2017–2069.
- Kasimov A (2025) Act like a woman, play like a man: Manhood acts and the gendered and racialized organization of online professional streamers. *Men and Masculinities* 28(3): 247–266.
- Kingdon A (2024) Beauty is Power: The Use of Gaming References and Gaming Aesthetics in Extremist Propaganda. In: L Schlegel and Kowert R (eds) *Gaming and Extremism: The radicalization of digital playgrounds*. England: Routledge, pp. 130–147.
- Koehler D, Fiebig V and Jugl I (2023) From gaming to hating: extreme-right ideological indoctrination and mobilization for violence of children on online gaming platforms. *Political Psychology* 44(2): 419–434.
- Kojah SA, Thach H, Haimson OL and Thomas M (2026). The Burden of Bearing Witness: Digital Practices of Marginalized Social Media Users in High-Stakes Contexts. In: *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems* (eds N Oliver et al.), Barcelona, Spain, 13–17 April 2026, pp.1–16. New York: Association for Computing Machinery.
- Koung E, Zhang Z, Gui X and Kou Y (2025) Gendered Toxicity in Competitive Gaming: Women’s Perceptions and Responses. *Proceedings of the ACM on Human-Computer Interaction* 9(6): 415–446.
- Kuss DJ, Kristensen AM, Williams AJ and Lopez-Fernandez O (2022) To Be or Not to Be a Female Gamer: A Qualitative Exploration of Female Gamer Identity. *International Journal of Environmental Research and Public Health* 19(3): 1–18.
- Lankford A, Allely CS and McLaren SA (2024) The Gamification of Mass Violence: Social Factors, Video Game Influence, and Attack Presentation in the Christchurch Mass Shooting and Its Copycats. *Studies in Conflict & Terrorism*. Epub ahead of print 13 October 2024. DOI: 10.1080/1057610X.2024.2413184.
- Macklin G (2019) The Christchurch attacks: Livestream terror in the viral video age. *CtC Sentinel* 12(6): 18–29.
- Mathew B, Saha P, Tharad H, Rajgaria S, Singhania P, Maity SK, Goya P and Mukherjee A (2019) Thou Shalt Not Hate: Countering Online Hate Speech. In: *Thirteenth*

- International AAAI Conference on Web and Social Media*, Munich, Germany, 11–14 June 2019. PKP Publishing Services.
- Mayworm S, Li S, Thach H, Delmonaco D, Paneda C, Wegner A and Haimson OL (2024) The Online Identity Help Center: Designing and developing a content moderation policy resource for marginalized social media users. *Proceedings of the ACM on Human-Computer Interaction* 8(CSCW1): 1–30.
- Meisner C (2023) Networked responses to networked harassment? Creators’ coordinated management of “hate raids” on Twitch. *Social Media+ Society* 9(2): 1–12.
- Morales E, Divon T and Lundqvist M (2026) Context matters: Understanding the platformization of violence. *New Media & Society* 28(4): 1395–1411.
- Navarro A and Tapiador FJ (2023) Twitch as a privileged locus to analyze young people’s attitudes in the climate change debate: a quantitative analysis. *Humanities and Social Sciences Communications* 10(1): 1–13.
- Obreja DM (2023) Toward a multidimensional streaming: A thematic case study of two Twitch channels. *New Media & Society* 25(6): 1354–1373.
- O’Connor C (2021) The Extreme Right on Twitch. Report, Institute for Strategic Dialogue, September.
- Phipps A (2021) White tears, white rage: Victimhood and (as) violence in mainstream feminism. *European Journal of Cultural Studies* 24(1): 81–93.
- Powell A and Williams-Johnson D (2025) “You dumb cracker b* tch”: The legitimizing of White supremacy during a Twitch ban of HasanAbi. *New Media & Society* 27(3): 1318–1335.
- Ruiz-Bravo N, Selander L and Roshan M (2022) The Political Turn of Twitch—Understanding Live Chat as an Emergent Political Space. In: *Proceedings of the 55th Hawaii International Conference on System Sciences*, Maui, USA, 4–7 January 2022, pp. 3170–3179. Honolulu: University of Hawai’i at Mānoa Library.
- Schlegel L and Kowert R (eds) (2024) *Gaming and extremism: The radicalization of digital playgrounds*. UK: Taylor & Francis.
- Seering J, Kraut R and Dabbish L (2017) Shaping pro and anti-social behavior on twitch through moderation and example-setting. In: *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, Portland, USA, February 25–March 1 2017, pp.111–125. New York: Association for Computing Machinery.
- Seering J, Wang T, Yoon J and Kaufman G (2019) Moderator engagement and community development in the age of algorithms. *New Media & Society* 21(7): 1417–1443.
- Starks DL, Thach H, DeGuia A, Dillahunt TR, and Haimson OL (2026). Safety With Community: Technologies of Care, Connection, Collective Safety, and Mutual Aid for Transgender Black, Indigenous, and People of Color (TBIPOC). In: *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems* (eds N Oliver et al.), Barcelona, Spain, 13–17 April 2026, pp.1–16. New York: Association for Computing Machinery.
- Taylor TL (2018) *Watch me play: Twitch and the rise of game live streaming*. Princeton:

- Princeton University Press.
- Thach H (2023) *"Should We Unban, Chat?": Communal Content Moderation on Twitch*. Masters thesis, University of Illinois at Chicago, US.
- Thach H, Mayworm S, Delmonaco D and Haimson O (2024) (In)visible moderation: A digital ethnography of marginalized users and content moderation on Twitch and Reddit. *New Media & Society* 26(7): 4034–4055.
- Thwaites R (2017) (Re)examining the feminist interview: Rapport, gender “matching,” and emotional labour. *Frontiers in Sociology* 2(18): 1–9.
- Tran CH (2022) “Never Battle Alone”: Egirls and the Gender(ed) War on Video Game Live Streaming as “Real” Work. *Television & New Media* 23(5): 509–520.
- Tran CH (2024a) Twitch spouse: Livestreaming and the legacy of spousal labour in the video game industry. *Global Media and China* 9(2): 173–187.
- Tran CH (2024b) Room with a viewership: visibility work & Twitch. tv in the domestic context. *Information, Communication & Society* 27(12): 2281–2296.
- Twitch (n.d. a) Setting Up Your Chat. Available at: https://safety.twitch.tv/s/article/Chat-Tools?language=en_US (accessed 11 September 2026).
- Twitch (n.d. b) Twitch Women’s Guild. Available at: <https://www.twitch.tv/team/womensguild> (accessed 11 September 2026).
- Twitch (n.d. c) Twitch Black Guild. Available at: <https://www.twitch.tv/team/blackguild> (accessed 11 September 2026).
- Twitch (2026) Twitch Community MeetUps @PAX East 2026. Available at: <https://meetups.twitch.tv/events/details/twitch-twitch-community-meetups-events-present-s-twitch-community-meetups-pax-east-2026/cohost-boston/> (accessed 11 September 2026).
- Uttarapong J, Cai J and Wohn DY (2021) Harassment experiences of women and LGBTQ live streamers and how they handled negativity. In: *Proceedings of the 2021 ACM international conference on interactive media experiences*, Virtual Event, USA, 21–23 June 2021, pp. 7–19. New York: Association for Computing Machinery.
- Weiss P, Koban K and Matthes J (2025) A narrow gateway from misogyny to the far right: empirical evidence for social media exposure effects. *Information, Communication & Society* 28(13): 2377–2395.
- Wolff GH and Shen C (2024) Audience size, moderator activity, gender, and content diversity: Exploring user participation and financial commitment on Twitch.tv. *New Media & Society* 26(2): 859–881.
- Woodcock J and Johnson MR (2019) The affective labor and performance of live streaming on Twitch. tv. *Television & New Media* 20(8): 813–823.
- Zolides A (2021) Gender moderation and moderating gender: Sexual content policies in Twitch’s community guidelines. *New Media & Society* 23(10): 2999–3015.
- Zorlu D and Özkan N (2020) Women on Twitch Turkey: Affective communities and female

solidarity under patriarchy and postfeminism. *Gender Forum: An Internet Journal for Gender Studies* 19(3): 105–127.