7-3-2020

Inclusion at Scale: Deploying a Community-Driven Moderation Intervention on Twitch

Johanna Brewer
Massachusetts Institute of Technology, jbrewer@smith.edu

Morgan Romine
AnyKey

T. L. Taylor
Massachusetts Institute of Technology

Recommended Citation

Inclusion at Scale: Deploying a Community-Driven Moderation Intervention on Twitch

Johanna Brewer
MIT CMS/W
Cambridge, MA 02139, USA
johannab@mit.edu

Morgan Romine
AnyKey
Burbank, CA 91605, USA
morgan@anykey.org

T.L. Taylor
MIT CMS/W
Cambridge, MA 02139, USA
tltaylor@mit.edu

ABSTRACT

Harassment, especially of marginalized individuals, on networked gaming and social media platforms has been identified as a significant issue, yet few HCI practitioners have attempted to create interventions tackling toxicity online. Aligning ourselves with the growing cohort of design activists, we present a case study of the GLHF pledge, an interactive public awareness campaign promoting positivity in video game live streaming. We discuss the design and deployment of a community-driven moderation intervention for GLHF, intended to empower the inclusive communities emerging on Twitch. After offering a preliminary report on the effects we have observed based on the more than 370,000 gamers who have participated to date, the paper concludes with a reflection on the challenges and opportunities of using design activism to positively intervene in large-scale media platforms.

Author Keywords

Live streaming; platform moderation; video games; design activism; inclusion and equity; online harassment.

CCS Concepts

•**Human-centered computing** → **Empirical studies in HCI;** *Empirical studies in collaborative and social computing; Social media;*

INTRODUCTION

Live streaming – where people use networked broadcasting platforms such as Twitch or YouTube to share gameplay in real time with an audience – has been become an important site for research regarding online communication, community moderation, and cultural change. Much of the foundation for inquiry into this topic comes from anthropologists, sociologists, and media studies scholars who have traced both the rise of esports as a broadcast product, as well as the popularity of "variety streamers" who, as entertainment-focused users, share their gameplay with the public [9, 42, 47, 48]. Importantly, this research has explored the techniques streamers have developed to manage the online communities growing up around them

and the challenges marginalized broadcasters in particular face when going live.

HCI scholars have built upon this work, generating findings that resonate with prior studies in the field of computer-mediated communication. Researchers have observed how streams can function like online "third places" where participatory communities form around their shared experiences, noting the important influence the streamer has over the emergent culture of the group [22, 36, 46]. While some investigations into large-scale live stream chats draw parallels to well-worn concepts like information overload and communication breakdown, there is increasing evidence to indicate that the massive, fast-moving chats that characterize the most popular streams actually represent a new, legible form of communication called "crowdspeak" [14, 21, 24]. Recent inquiries have also begun to examine the economic and cultural implications of other live streaming-related practices like sending gifts and donations or volunteering to moderate for a streamer [31, 52].

As is customary, the results of HCI studies exploring the experiences of streamers, chatters, viewers, supporters, and moderators are typically framed in terms of implications for design. The recommendations are often insightful, but given that researchers have no direct control over the features and practices implemented by industry media platforms, they tend to have little influence on the real world landscape of technology. Promisingly though, the HCI community has created several design interventions that were able to be evaluated within the live streaming ecosystem.

Live Streaming Design Interventions

There have been a number of technically robust interventions developed by HCI researchers seeking to enhance the engagement of streamers and their audiences. Some have sought to advance the state of the art within the live streaming ecosystem by developing tools that support emergent forms of community gameplay, such as the Audience Participation Games designed by Seering et al. [44]. Others, like Helpstone, offer an additional layer of communication between gamer and viewer by allowing spectators to provide tips to a streamer playing Hearthstone [28]. More subtly, stream add-ons like All the Feels use wearable sensors and webcam data to create an enhanced overlay displaying real-time biometric data about the broadcaster [40].

Preliminary studies conducted with streamers on Twitch indicated that interventions like these can improve communi-

cation and perceived connection between the streamer and their spectators. However, these design probes were typically only able to be briefly tested (on the order of minutes) with a handful of individuals (typically in the low tens) in relatively controlled situations. Nonetheless, they demonstrate the enormous potential that HCI research has to directly engage with the development of a new medium of communication. Other design concepts, while intriguing, are less successful in their ability to integrate with live streaming platforms. In the case of TwitchViz, an information visualization tool for stream chats, the authors were unable to overcome the technical challenges of implementing a real-time responsive interface, and instead conducted a feasibility study with just four participants [35].

Given that there are currently thousands of tools used by an ever-growing market of millions of streamers, it is clear that it is not for want of participants that research prototypes remain comparatively underdeployed on live streaming platforms. While it is neither desirable, nor reasonable, for HCI researchers to directly compete with industry inventions, we also submit that our community has a critical, reflective role to play in shaping the landscape of communication technologies adopted by a mass audience, and therefore we ought to explore methods for increasing our direct engagement with the broader public. Specifically, we see both an opportunity and an urgency to offer a critical response to the ethical questions posed by the operational practices of live streaming platforms. Chief amongst those issues are the ways in which user behavior is, or is is not, moderated by such large social media sites.

Platform Moderation
Harassment online has been extensively studied and identified as a significant issue especially for marginalized individuals like women, people of color, members of the LGBTQ+ community, and disabled folks [4, 8, 15, 19, 48]. Generally, online platforms tend to employ two related moderation mechanisms. First, they ask users to report offensive or abusive content created by other users, which moderators then review and may remove. Secondly, most platforms employ technical solutions that automatically filter content before it is published. Apart from basic tools like rate-limiting users' posts, commercial platform moderation generally centers on content rather than behavior. Accordingly, systemic campaigns of harassment, especially when involving multiple actors or targets, often go unchecked or even, chillingly, appear to be rewarded. Through a large-scale analysis of abusive accounts on Twitter, researchers found that users who engaged in harassment related to the Gamergate controversy were disproportionally *not* suspended from the platform in comparison to other random users who had been reported [6].

Not only are these moderation systems failing to adequately protect the people most vulnerable to online harassment, they often have a compounding effect. The unseen human moderators who screen the flagged content are often marginalized themselves, and thus as these "ghost workers" absorb the harassment on the behalf of the platform's users, the cycle of trauma is not broken, merely hidden away from public scrutiny [17, 20, 33, 39]. Though dominant social media sites like Facebook, Twitter, and YouTube continue to utilize this centrally controlled content filtering process, as a new wave of social platforms emerges, we are seeing a shift away from this commercial approach to moderation [45]. In the case of Twitch, currently the largest live streaming platform, streamers are subject to centralized oversight via the site administrators, but the behavior of viewers is by and large moderated by other members of the community, with virtually no supervision from Twitch employees.

Community-Based Movements
Given the systemic failure of platforms to effectively curb online harassment targeted at marginalized users, a patchwork of community-driven approaches has developed over the past decade. Women of color began collectively organizing to resist their mistreatment on Xbox Live [18]. On Twitter, a variety of groups adopted the practice of creating community-generated blocklist tools to fight harassers targeting multiple individuals [16, 23, 26]. Organizations like Black Girl Gamers and Transmission Gaming have created cross-platform networks of support that act as distributed safe spaces for women of color and trans gamers [30, 38]. Most recently, some gaming communities (e.g. Super Smash Brothers) have formed grassroots coalitions to improve the inclusivity of their competitive events by taking steps such as creating an official code of conduct that is enforced by an independent panel [34, 53].

Though there have been comparatively fewer attempts by HCI researchers to deploy interventions related to online harassment, there are several notable examples. As a tool for friend-sourced email moderation, SquadBox offers a platform empowering individuals to create an ad hoc response group to handle an influx of harassing messages. While SquadBox has been open sourced to the public, citing concerns of perpetuating further harm, the creators limited their study of the system to a four-day field engagement with five pairs of friends [32]. This caution is prudent, yet, as expressed earlier, we believe there is an imperative to more directly engage with the public.

In this belief we are joined by the growing cohort of interaction design activists whose work highlights the power of taking an intentional stance on social change. While the concept of persuasive technology has been around for two decades [13], more recently, researchers from sustainable technology have begun to argue that persuasion in HCI focuses too narrowly on individuals, overlooking the broader sociocultural systems that surround them [5, 37]. In an effort to explore how we might engage with grassroots movements through design research, Bisker et al. demonstrate how technology prototypes might be adopted by communities interested in shaping their urban environments through distributed action [2]. This community-centered perspective has also been embraced by platforms like Hollaback and Heartmob which provide social support to the targets of harassment through collective storytelling and visible classification of abusive behaviors [3, 12]. Other tools, like Botivist, offer a technical platform that assists activists in calling volunteers to action by means of Twitter bots [41]. And most significantly, interventions like Turkopticon have sought to spur change by working to increase accountability and transparency for the ghost workers powering massive platforms like Amazon's Mechanical Turk [27].

DESIGN ACTIVISM FOR LIVE STREAMING

As the core members of AnyKey – an advocacy organization originally sponsored by Intel and ESL that is focused on diversity, inclusion, and equity in esports and live streaming – we align ourselves with other design activists who value an explicit orientation to social justice goals, place marginalized people at the center of design, and take a decided stance on the pressing issues of our day [10, 25]. Through our organization, we seek to undertake initiatives that directly address important topics in game culture. This paper presents a detailed case study of our efforts to mount one such initiative motivated by the question: *how can we design a positive intervention to increase inclusion in an online platform at a significant scale?*

Our work has been inspired by recent research in live streaming which suggests the use of proactive moderation tools that amplify the actions of upstanding individuals in a stream community could have a significant positive impact on the behavior of other members in chat [43]. Such insights point to the need not only for "banhammers" but for practices and tools that help cultivate and support desirable behaviors [29]. Furthermore, there is strong evidence that individuals who engage in "public promising" (declaring a mutual commitment with others) even through decentralized, mediated communication channels like chats, can increase the level of social cooperation within their communities [1]. Seeing an opportunity for a proactive design intervention in live streaming, we set out to grapple with the design challenges of attempting to positively influence behavior on a massive social media site, namely, Twitch.

A CASE STUDY: THE GLHF PLEDGE

This paper presents a detailed case study of an initiative designed to intervene on a live streaming platform at a significant scale. We describe our effort to spur the adoption of the GLHF pledge, a code of conduct that players and spectators can sign to earn the global AnyKey chat badge on Twitch which signifies their commitment to be welcoming and positive. In the coming pages, we recount our effort to empower the grassroots groups who create safe live streaming spaces through the design and deployment of a community-based moderation system for the GLHF pledge, and offer a preliminary report on the effects we have observed based on the more than 370,000 gamers who have participated to date.

In order to adequately contextualize the community-driven moderation intervention we developed, it is necessary to first expound upon the genesis of this project. Accordingly, in this section we will describe the origin of the GLHF pledge initiative, recount our experiences from development to first launch, and reflect on the results of our initial experiment. Later, we will turn to the second iteration of the pledge and present a broader series of reflections.

Building on the Keystone Code

Driven to foster change and break down barriers in gaming, AnyKey focuses its efforts on developing innovative, research-based programs that meaningfully improve the world of esports for marginalized gamers. From early work engaging the community through participatory workshops, the organization identified a variety of basic resources, guides, and tools that

---

GLHF is a promise to...

1. Be a good sport whether I win or lose

2. Know that people online are real people and my words have real impact

3. Set a positive example with my behavior

4. Speak up against discrimination, hate speech, harassment, and abuse

5. Show integrity by honoring the rules, my opponents, and my teammates

6. Stop, listen, and reassess if I'm told that my words or actions are harmful

7. Respect others, even if their sincere opinions are different from my own

Figure 1. Seven promises of the GLHF pledge.

---

could better support the needs of the growing movement to create more welcoming gaming spaces [50, 51].

The Keystone Code, a freely downloadable code of conduct that esports groups were encouraged to adapt, was the first such resource we shared in 2016[1]. Adopted by leading collegiate esports programs at UC Irvine and Robert Morris University, as well as industry events like the SKYLLA tournament series and TwitchCon, the Keystone Code demonstrated that we were capable of helping these established organizations improve their individual communities.

Our initial success disseminating a standardized code of conduct relied on direct collaboration with operational leaders from these large gaming organizations. After making progress with this top-down approach, we began to ask ourselves: *would it be possible to spur wide-spread adoption of an inclusive code of conduct for online gaming from the bottom up?*

A Public Awareness Campaign

As a welcoming gesture in competitive gaming, players often type GLHF ("good luck, have fun") before they start a match. Building on this practice, we created the GLHF pledge, a simple code of conduct that anyone online can make a promise to uphold (see Figure 1 for details). Similar to the interface of an online petition, the GLHF website outlines the seven promises a player commits to by taking the pledge and provides a simple form that requests the name and email address of each signer. After successfully registering a pledge, the GLHF system prompts the user to share a pre-composed post on Twitter or Facebook, and also offers the opportunity for the signer to unlock the AnyKey global chat badge on Twitch by authenticating their account.

---

[1]Publicly available at: **https://anykey.org/keystone-code**.

Badges are small icons which appear next to a user's name in a live stream chat (see Figure 2). Similar to flair on popular message board sites like Reddit and Stack Overflow, badges on Twitch must be earned before they can be equipped. Additionally, most badges, such as the icons symbolizing channel moderators, VIPs, and subscribers, are stream-specific. Apart from those indicating a user's status as staff, partner, or conference attendee, there are only a handful of badges that Twitch has made available on a platform-wide level. Most crucially, though all badges on the platform reveal a title (e.g "GLHF Pledge") when hovered, the AnyKey badge has the distinction of being one of the few that opens a website when clicked.

Each pledger who equips the AnyKey global chat badge brings an actionable, visual cue into every live stream where they are active, providing the means for anyone in the space, even lurkers, to learn about and join the inclusive gaming movement. In this way, the GLHF pledge was originally conceived of as an interactive public awareness campaign that would leverage the allure of an exclusive badge to promote discussion about the social value of being a good digital citizen, and offer an alternative practice to rewarding those who take on the role of funny troll, "shitposter," or "edgelord"[2]. The goal of this campaign was to provide members of the gaming community on Twitch with an easy way to declare their support for inclusion and positivity, spark conversations in the online spaces they frequent, and identify the other good actors they encountered.

Technical Specifications of v1.0
AnyKey's original website was built using a customized WordPress installation running on Dreamhost. The first version of GLHF pledge was powered by a PHP template that integrated a mix of native and third-party form widgets to collect the names and email addresses of pledgers. That information was stored in the managed SQL database, and could be accessed through the WordPress administrative control panel. Additionally, a series of PHP templates were developed that facilitated the Twitch OAuth authentication flow and fired the API call to a custom endpoint they created specifically to allow AnyKey to unlock the badge for users who successfully completed the authorization. Records of those authentications were also stored in the the SQL database, but were not accessible through the WordPress admin interface.

Launch & Promotion of v1.0
The GLHF pledge launched on October 20, 2017 for the opening day of TwitchCon, a public convention organized by the platform and attended by tens of thousands of users annually. AnyKey was provided with space on the main showroom floor to promote the pledge. Informational material about GLHF was on display and laptops were available for attendees to take the pledge. Signers were encouraged to share a pre-composed post on Twitter or Facebook, and they also received a free enamel pin they could wear to show their support IRL ("in real life"). Approximately 1,500 visitors took the pledge at the AnyKey booth over the course of the three day convention.

[2]Digital citizenship lacks a singular definition, but we include all senses of the term as outlined by Choi (i.e., ethics, media & information literacy, participation/engagement, and critical resistance) [7].
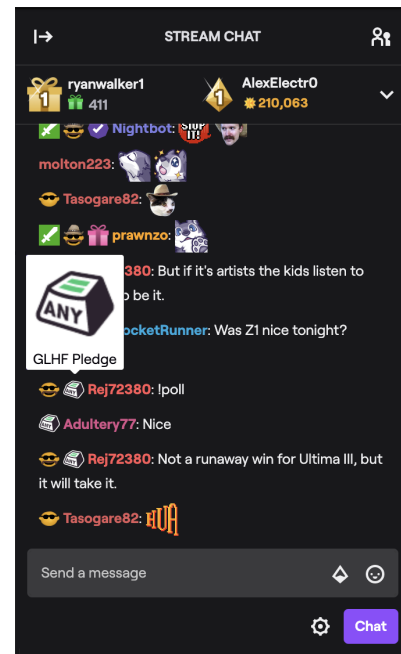

Figure 2. AnyKey badges in a Twitch stream chat.

Apart from the onsite activation and these basic seeds of a viral campaign, we did very little direct promotion of the pledge. Though AnyKey, along with sponsors Intel and Oculus, tweeted in support of the pledge shortly after it was launched, unlike other public awareness campaigns, we intentionally chose not to mount a concerted marketing effort for GLHF. The motivation for this choice was threefold. First, and most simply, as an advocacy organization which relies on sponsorships to operate, AnyKey has a very limited budget and creating even a small-scale traditional campaign would have been well out financial reach. Second, and most saliently, our goal with this work was to explore the feasibility of spurring bottom up change by creating lightweight support for spreading inclusion on the grassroots level, and therefore a large-scale campaign would have undermined the experiment. Finally, and most significantly, our team had the very real concern that the pledge system would attract trolls, or be subjected to a coordinated attack, neither of which were were adequately prepared to handle, and consequently we opted not to direct mainstream media attention towards the project.

Results of v1.0
Because of concerns about trolling, it was decided that there would be a fixed expiration date placed on AnyKey's global Twitch chat badge. While the pledge system would remain active indefinitely, the badges were set to disappear after December 31, 2017. This predetermined lifespan for the campaign ensured that if the GLHF pledge was overrun with trolls, we would have a clear recourse to quash the toxicity. However, we are pleased to report that over the course of this 73 day campaign, our fears were never realized.

On the contrary, we found ourselves surprised by both the positive response to, and extraordinary uptake of, GLHF. More

GLHF Phases	v1.0	Dormant	Leak	Closed	v2.0
Dates	*Oct 20 - Dec 31, 2017*	*Jan 1, 2018 - July 14, 2019*	*July 15 - 18, 2019*	*July 19 - Sept 16, 2019*	*Sept 17 - Dec 31, 2019*
Duration	73 days	560 days	4 days	60 days	106 days
Site Visits	565,975	40,709	54,611	142,556	411,370
Pledges	204,778	17,082	25,737	11	126,441
Twitch Authentications	191,400	14,257	21,611	1	116,532
Visit Rate	7,753/day	73/day	13,653/day	2,376/day	3,881/day
Pledge Rate	2,805/day	30/day	6,434/day	0.2/day	1,193/day
Authentication Rate	94%	84%	84%	9%	92%
Email Duplication Rate	29%	50%	33%	82%	0%

Table 1. Key usage statistics for the five distinct phases of the GLHF pledge's lifetime.

than half a million people visited the campaign website between October 20th and December 31st in 2017. During that time, 204,778 unique individuals took the pledge, and 94% of those signers completed the authentication procedure to unlock the AnyKey global chat badge on Twitch (see Table 1).

During the time the campaign was active, we received both public and private messages through social media expressing appreciation and enthusiasm for the pledge. Though most inbound communication was supportive, occasionally we received negative comments of a fatalistic nature (e.g. "you can't stop harassment online"). While our organization was not trolled, we did receive several reports from signers regarding other users who had clearly equipped the badge in bad faith, and were seen violating the promises of the pledge. In the approximately 15 cases in which we received clear evidence that someone was flagrantly abusing the badge, our partners at Twitch were able to manually revoke the AnyKey badge for those users at our request.

Building on the community's established pattern of speaking out about their thoughts on the campaign, the day after the badges were deactivated, we published a tweet explicitly seeking more feedback[3]. From the responses, we learned that the appearance of a global chat badge which symbolized positive values actually contributed to making chat streams feel more welcoming. Furthermore, several respondents expressed sadness and even mild grief over "losing" the AnyKey badge, and others directly asked for it to be reinstated.

Reflections on v1.0
Candidly, no member of our team anticipated the level of response we observed. We were worried that the campaign would go unnoticed or have only a superficial impact. On the contrary, we realized that the extent of public adoption had greatly exceeded our aspirations. Most notably, we were impressed by the apparent power of the visual symbol which we had designed to represent the inclusive gaming movement.

Our organization's name is a playful reference to the phrase "press any key to continue." We chose it because we believe that anybody should be welcome in gaming. In keeping with this theme, our logo is a mechanical keycap with an equality

---

[3]The results of this informal survey can be viewed on Twitter at: https://twitter.com/anykeyorg/status/947914991012737024.

sign on top that is intended to be a literal representation of that elusive key. We used the AnyKey logo as a global chat badge on Twitch with the hope of creating a universal symbol of inclusion for welcoming gaming communities. Given that the AnyKey badge was unlocked by more than 1% of Twitch's daily active users, the GLHF pledge campaign had substantial reach. One pledger tweeted:

> it felt great going into communities and their chat and see so many people with the badge. helped get rid of the anxiety of talking to strangers knowing ahead of time i was safe. then when i saw our community almost unanimously get the badge too... made my heart soar

Though there were many positive outcomes indeed, this initial experiment was far from flawless. The majority of those failings were largely of our own making. Given both the limited budget of our advocacy organization, and the speculative nature of this project, we did not substantially invest in the technical infrastructure of the GLHF pledge. The PHP template which gathered the names and email addresses of signers was developed by a small web design firm who graciously offered their services at a discounted rate. However, it was beyond the capability of this firm to implement the integration with Twitch we had planned. To develop this authentication flow, our sponsor, Intel, contracted the services of a third-party agency. The result of this patchwork approach was a brittle system which was rife with bugs.

The form which gathered email addresses did not have any type of validation which resulted in a 29% duplication rate and caused the public counter of pledges to consistently display an inflated tally. Additionally, because the fragmented system architecture stored the email addresses separately from the records of Twitch authentications, there was no feedback presented to users having difficult activating their badges. We received several hundred requests for support, and while in most cases the issue was resolved by an explanatory GIF demonstrating the procedure for equipping a badge on Twitch, we were unable to diagnose or offer assistance in cases of real technical failure. And though we were thankful that the documented instances of abuse were relatively low, if it were not for the like-minded and generous partners who supported this initiative at Twitch, we would have not been able to quash the cases of flagrant violation that community did report.

In short, though the GLHF campaign was largely a success, the fragile system we had created was straining under the pressure. The first incarnation of the GLHF pledge was created with an exploratory mindset and the results far exceeded our own expectations. As it became clear that it was indeed possible to spur wide-spread adoption of an inclusive code of conduct from the bottom up, we began asking ourselves: *how might we create sustained support for the growing movement of inclusive gaming communities online?*

REACTIVATING GLHF

Because of the significant uptake and strong support we received from the public, no sooner were the badges deactivated than we began exploring the possibility of a permanent revival of the campaign with Twitch. In this section we will describe the evolution of the GLHF pledge, recount our experiences from the sunset of the initial experiment to the relaunch, and reflect on the results of this ongoing experiment.

Supporting the Grassroots Movement

AnyKey, as an advocacy organization, found itself in a unique position by the end of the first campaign. Both our partners at Twitch and more than 190,000 members of their user community had grown to trust in the integrity of the badge. In order to maintain that trust in the long term, we recognized that preventing bad actors from abusing the badge would be crucial. Indeed, several pledgers expressed their desire for more moderation, with one tweeting:

> This was a great initiative, however it was sad when some people obviously pledged just for the badge and continued toxic behavior, even if reminded of their pledge. Maybe it would have been nicer to have a report system for offenders who didn't live by the pledge?

Given the small size and limited resources of our organization, we were at once daunted and emboldened by the prospect of intervening in platform moderation on such a broad scale. Twitch has often been, and continues to be, criticized for its approach to moderation, especially when it comes to protecting marginalized folks from harassment [11]. Though the platform offers a reporting mechanism, users often express that it appears no one at Twitch is actually listening or taking action on the reports they file. As any systemic failure of the platform to protect vulnerable users would likely be a contributing factor in the uptake of the AnyKey badge, we concluded that permanently reviving the GLHF campaign would require a moderation system capable of providing meaningful support to the community.

A Community-Driven Moderation Intervention

In order to rise to this sociotechnical scalability challenge, we completely rebuilt the GLHF pledge system to offer a clear procedure for reporting badge abuse as well as a moderation tool robust enough to address the community's concerns. Most fundamentally, we redesigned the pledge page itself to prominently feature an informative call to action that urges pledgers to report instances of badge abuse. Additionally, the new system sends each pledger an email which reviews the promises they have made and reminds them of the link they can use

to file a report. Finally, the footer of our website now also features a direct link to the reporting system. Taken together, these messages are intended to convey a clear signal that our organization is both listening to the community and relying on each individual to maintain the integrity of the movement.

The badge abuse report form is designed to gather the necessary details a human moderator needs to investigate the issue, make a judgement about the incident, and potentially intervene. In order to be accepted, the report must include at a minimum the Twitch username of the person being reported, the stream where the pledge violation occurred, and a brief description of the incident. Additionally, we encourage people to add a date and a screenshot with as much context from the stream chat as possible. Users are required to provide a valid email address and also asked to share their Twitch username before submitting, and so accordingly, anonymous reports are not permitted. Finally, we solicit the reporter's opinion about how to handle the situation by posing the optional question: *what do you think we should do?*

Complimenting AnyKey's public user interface for flagging violations of the GLHF pledge, we also developed an internal administrator system that allows moderators (mods) to review and react to the reports the community files. Mods receive an email notification every time a report is submitted which prompts them to access a secure dashboard where they can see and sort reports according to their status. In addition to the basic information submitted by the reporter, the interface presents mods with further contextual details about whether the reported user exists on Twitch and if they took the pledge. After reviewing a report, the system prompts mods to dismiss it or take action on it by either issuing a warning or revoking the badge (see Figure 3 for flow details).

Moderators are directed to offer a clear and considerate explanation to the pledger about why their actions violated the promises they made, and after resolving a report, that human-generated response is sent to the reported user via email. Warnings are given in cases where the mod sees the potential for a "teachable moment" that will educate the pledger about inclusion. If, however, a mod observes a flagrant, bad faith violation and chooses to revoke a pledger's badge, we also provide a channel for appeal. Finally, in order to demonstrate to the community that their concerns are being addressed, the system automatically sends an email to the individual who reported an incident whenever a mod issues a warning or revokes a badge.

Tech Specs of v2.0

AnyKey's new site was built as a custom Ruby on Rails web application using the latest stable versions at the time of development (Ruby 2.6 and Rails 6.0). The web app was designed to be fully mobile responsive, relying on HTML5, CSS3, and JavaScript to create a bespoke user interface. The new GLHF system is backed by a streamlined and semantically enhanced SQL database which stores relational data for each pledge and report. Overviews of this data can be accessed through our custom web-based administrator system and details are easily accessible through remote connections to the database or console. Additionally, the application integrates a Redis cache coupled with an Amazon S3 store for fast loading of
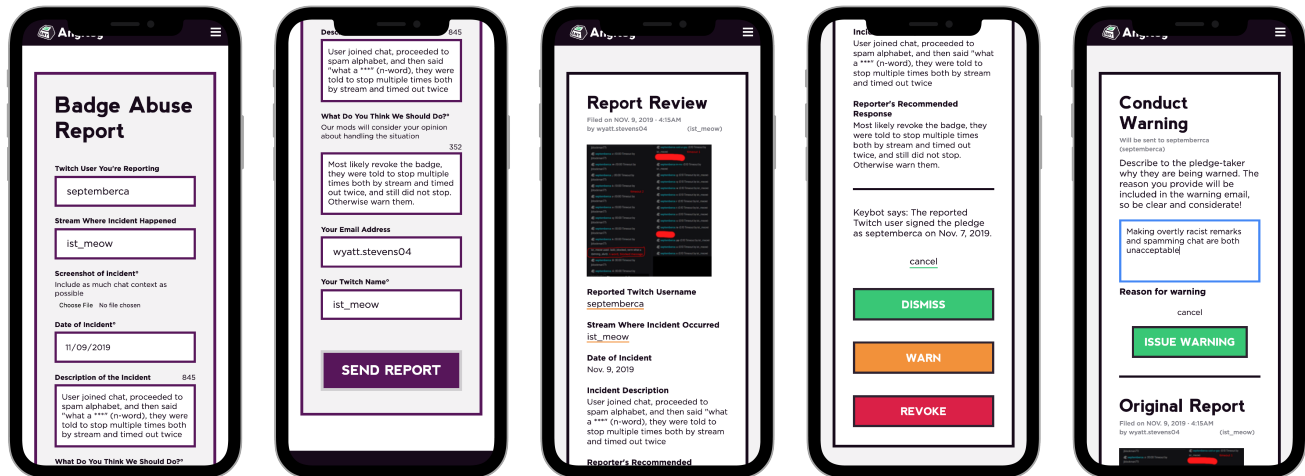
Figure 3. Moderation system flow on mobile which highlights the public badge abuse form and internal review interface.

user-submitted screenshots. Currently, the application is running on Heroku, making use of Cloudflare for SSL, utilizing SendGrid for transactional email delivery, and tracking user behavior with Google Analytics tools. Integration with the Twitch API has been fully updated to comply with current standards of authentication and utilize new custom endpoints for granting and revoking the AnyKey badge.

Launch & Promotion of v2.0
Though the campaign had gone dormant at the end of 2017, and all AnyKey badges had been deactivated on Twitch, visitors steadily returned to the site to take the pledge. In the summer of 2019, we began working with Twitch in preparation to relaunch GLHF, and in the process of refreshing and testing our integration with the new system, suspended endpoints were reactivated. Unintentionally, our old system began granting badges to any first-time pledgers who authenticated their accounts. This leak resulted in a completely unexpected viral snap. During the 560 days the campaign had been dormant, 17,000 users had taken the pledge. Yet in the four days from July 15th to 18th when the badges were temporarily obtainable, more than 25,000 individuals joined the movement. While we had originally intended to keep the return of the badges a secret until the fall, the accidental reveal of our plans served to demonstrate the persistent demand from the community. Indeed, after closing the floodgate by deactivating the old system, users continued to return to our site and social channels to ask when the badges would be available for all again.

Following this unintentional teaser, the GLHF pledge officially relaunched on September 17, 2019 as the flagship feature of a new interactive site reflecting a refreshed identity for the organization. Coinciding with the launch, an updated badge design was introduced to reflect the "glow up" the organization had undergone in order to better embrace the aesthetic of the community. Additionally, the AnyKey badges of all pledgers who had previously authenticated their accounts with Twitch were reinstated. Clicking on the global Twitch badges now

brings visitors to the new media-rich pledge page highlighting the community involvement in the GLHF movement. Inspired by an analysis of the spread of the summer leak (which traced the initial spike to members of the Fortnite community in Russia), the pledge page was internationalized to support the seven most frequently represented languages of our past users[4].

Promotion for the relaunch began on September 27th, following a similar strategy to the first version of the campaign. AnyKey was provided with space outside of the Unity Lounge at TwitchCon 2019 to advertise the pledge to passersby with laptops, informational materials, enamel pins, and vinyl stickers, all sporting the refreshed look of GLHF. Additionally, the initiative was highlighted during broadcast coverage of the ESL One New York tournament through both onscreen graphics as well as a 90 second video trailer featuring prominent gamers and streamers who've taken the GLHF pledge.

Apart from this initial push to announce the return of the pledge, we once again did little in the way of direct marketing. Instead, we focused our efforts on enhancing the grassroots spread of the movement. Most simply, we improved the pre-composed social posts that pledgers are encouraged to share by adopting a tone more consistent with the the live streaming community's vernacular, and adding a rich visual representation of the campaign which highlights the new design and badge (see Figure 4). Finally, we introduced a new referral system and corresponding leaderboard with a call to action that encourages pledgers to share a unique link with their community in order to earn the bragging rights for being the most inclusive gamer and help us in our quest to reach one million pledges by the end of 2020.

Results of v2.0
Because the campaign is currently ongoing at time of the writing of this paper, we will limit our discussion of results to the 106 day period from the relaunch of GLHF on through

---

[4]English, French, German, Italian, Portuguese, Russian, and Spanish.

Figure 4. Pre-composed tweet shared by GLHF pledger.

December 31, 2019. During that time more than 400,000 people visited the campaign website. 126,441 of those individuals took the pledge, bringing the total number of signers to 374,049 by the end of the year. To date, 92% of all pledgers unlocked their AnyKey global chat badge on Twitch (refer back to Table 1 for comparison details).

The new pledge system was designed to detect returning plegders by their email address or Twitch account, and provide them with contextual feedback such as offering the ability to activate their badge or instructions on how to equip it. 19% of users who submitted a valid email address for the pledge had already done so before, and those users were redirected to the appropriate follow up action. Accordingly, the email duplication rate of pledgers dropped to 0%. Furthermore, our updated system was able to detect and prevent the 5% of cases in which a pledger used a new email address to reactivate the badge on a previously authenticated Twitch account.

The more robust management of the pledge data proved essential to provide the context necessary to address the 309 badge abuse reports that were filed following the relaunch. 25% of those reports generated a warning and 13% resulted in a revocation. More precisely, 40 users have had their badges revoked, and the new system was able to support that moderation decision by effectively blocking the 26 cases in which a Twitch user was attempting to obtain the AnyKey badge again by concealing their identity. However, at the close of 2019, 50% of reports remained unresolved, having not yet been dismissed or acted upon by our mod team.

The community also responded well to the newly introduced leaderboard, with 1,337 signers joining the competition by successfully referring at least one other pledger. Overall, 3% of pledges registered after the relaunch came through a referral. On average an active referrer converted three of their contacts, with the most prolific user generating 186 referrals to earn the top spot on the board (see Table 2 for a summary of results).

Reflections on v2.0
Though some members of our team had concerns about the inclusion of the optional field in the report form which so-

licited a recommended response to the badge abuse that was being flagged, it proved to be one of the most frequently, and positively, used features of the moderation system. While 50% of reports included a screenshot, a surprising 90% offered opinions on how we ought to react. These recommendations tended to be brief calls for revocation based on what the reporter thought was a flagrant violation of the pledge. However, a number of reports proffered more nuanced requests like:

> I believe the individual should be made aware of this actions and how it may be taken as offensive and harmful to the viewer that they were interacting with. Responding immediately with toxicity and aggressive behavior is inappropriate and disrespectful.

Our moderation strategy was to adopt an empathetic and educational stance towards even the most serious offenders. In a handful of cases, that yielded unexpectedly positive results. After having received an email notification that their AnyKey badge had been revoked, several pledgers replied to apologize for their behavior. Others reached out through the appeals process asking what they could do to have their badges reinstated. Mods took these opportunities to further clarify what it means to be part of an inclusive community, and offer the offenders a chance for redemption by explaining GLHF to their friends and using their referral link to recruit a set number of new folks. Our team developed this ad hoc process in response to a perceived opportunity for constructive dialogue, and it appears to be a promising method for cultivating meaningful reform.

That potential for growth was also observed in several interactions our mod team had with pledgers who appeared to be gaming the leaderboard. Five cheaters were flagged by our moderators for their rapid advancement up the ranking, accomplished by generating pledges using fake, but valid, email addresses. In the instances where it was unclear if the pledger themself was behind the falsification, they received a warning. We learned, from their responses, that two of these pledgers were streamers who had been promoting GLHF on their channels, one of them going so far as to recite the seven promises at the beginning of each broadcast. Together we concluded that an over-eager viewer had likely been trying to pad their scores, and both of these streamers took the opportunity to use the incidents as teachable moments for their communities.

As with the last iteration of the campaign, we thankfully experienced few instances of trolling apart from one pledger who attempted to get their edgy username on the leaderboard. However, we did receive a handful of false reports which featured insults about our team, fatalistic comments about the mission, and, in one case, a threat to "take down" our system. While none of these harassing reports included explicitly offensive screenshots, instead opting to upload memes, we nonetheless became more acutely aware of our inability to shield our moderation team from the risk of trauma in screening toxicity.

FUTURE WORK FOR GLHF
Through our experience with the second iteration of the GLHF pledge it has become clear that sustained support of the grassroots inclusive gaming movement requires significant labor to achieve effective community moderation (for more on this

GLHF Reactivation	v2.0
Dates	*Sept 17 - Dec 31, 2019*
Duration	106 days
Total Pledges	374,049
New Pledges	126,441
Referrers	1,337
Badge Abuse Reports	309
Referral Rate	3%
Avg. Referrals	3/referrer
Highest Referrals	186/referrer
Warning Rate	25%
Revocation Rate	13%
Unresolved Rate	50%

Table 2. Summary of results from the relaunch of GLHF.

topic see [49]). As evidenced by the high rate of unresolved reports, our mod team was woefully understaffed, with only one member of AnyKey able to devote limited time to review. Though our intention for the relaunch of GLHF was to hire a dedicated moderator to work with the community, as an advocacy organization we continue to face challenges raising adequate funding to cover the cost of operating our initiatives.

Beyond reckoning with our pressing need for additional staff, we also identified several key improvements we believe will help make the administrative dashboard more effective and efficient for our existing moderators. We plan to expand the contextual feedback offered about reports to include information about related incidents and to provide mods the opportunity to email the reporter or add notes to each case, enhancing our ability to support meaningful reform. Beyond these essential investments in the GLHF moderation infrastructure, we are also driven to better understand both the depth and breadth of the influence the campaign has had.

Though we have presented compelling evidence to suggest that it is possible to design a positive intervention to increase inclusion in an online platform at a significant scale, we have yet to thoroughly examine the effects of the GLHF pledge. Accordingly, we intend to conduct a large-scale qualitative study of the pledgers to assess the impact of our intervention on the Twitch community. Finally, we are eager to integrate the AnyKey badge with other services popular amongst gamers (e.g. Discord, Mixer, etc.). Mounting a cross-platform campaign that relies on a robust moderation intervention presents further social and technical challenges, but in so doing, our intention will be to explore what a more holistic approach to quelling systemic harassment might look like in practice.

IMPLICATIONS FOR DESIGN ACTIVISM
Our ongoing experiment demonstrates that interventions like the GLHF pledge have the potential to address pressing issues facing the communities we study at a meaningful scale. In the remainder of this paper, we will offer a series of critical reflections which, though prompted by our own experiences, apply more broadly to design activists seeking to positively influence online platforms.

The Power of Distributed Microdisplays
As stated earlier, the widepsread uptake of the AnyKey badge greatly exceeded our expectations. The newest wave of social platforms (e.g. Twitch, TikTok, Snapchat, and Discord) are distinguished from their predecessors not least by the ways in which they support the development of a visually-rich media landscape characterized by an emote-laden aesthetic language. Though the symbol we introduced to Twitch users was a mere 18x18 pixels – just one tiny image amongst the millions of emotes and badges that scroll through chats every day – it proved to be capable of conveying real meaning for users.

While visual designers always hope their work will resonate with the public, the meaningfulness of the AnyKey symbol is more rightfully attributable to the way it was collectively adopted by the community. Many of the tweets we received in response to the informal survey following the first iteration of the campaign highlighted the power of repeatedly encountering the AnyKey badge in chats across Twitch, with two pledgers commenting:

> It felt amazing to see the badge right by so many other names as I went from stream to stream. It gave me a bit of hope and courage to try streaming on my own.

> It wasn't just good to know I was rocking an emblem for respect and kindness in the community, but seeing it in droves in a Twitch chat has always been a good sign that I can expect good behaviour there!

In the same way public harassment acts not only as an individual harm for the target, but also as a social one by infusing the space with toxic values that wash over everyone present, microdisplays like the AnyKey badge become countervailing actors. They not only represent a personal commitment but, when taken collectively in a space like a Twitch channel, can offer a powerful symbolic and affective social intervention.

With the introduction of the community-driven moderation system we saw further evidence of the power of this symbol had gained. Several pledgers who filed reports offered suggested courses of action that subtly alluded to the significance of the badge by referring to it with the definite article:

> I respect the AnyKey and the GLHF Pledge a lot. I'll leave the decision to the mods, but I do think the user should not have an AnyKey.

> Maybe have a chat with the person about what the AnyKey stands for? Your call.

The bottom-up nature of our intervention campaign served to facilitate the spread of a visual cue signaling safe spaces, and though the community is largely responsible for that emergent effect, it is clear that design activists can play an important role in supporting the growth of a grassroots movement through material contributions to its aesthetic language.

The Need for Cross-Platform Interventions
One of the most striking observations from our ongoing work with the community is their continued pleas for more comprehensive intervention. These requests typically manifest themselves through reports which trace an abuser's behavior

across multiple platforms. For instance, one pledger flagged an incident in which another user sporting an AnyKey badge:

> Entered streamers Twitch chat and commanded for Discord link. Joined Discord and proceeded to abuse the voice channel and use inappropriate language while the content creator was broadcasting.

Though in this case the reporter was able to ban the harasser from the Discord server they managed, other reports have aired frustrations from the community that their moderation requests for stricter enforcement on serial abusers are going unanswered on Twitch and Twitter alike.

Online harassment transcends the boundaries of a single platform, and we believe that independent groups, both research and advocacy, are uniquely positioned to tackle societal issues that social media corporations have failed to address. We see an opportunity for design activists to create interventions which center the needs of the members of marginalized social groups rather than users of a specific platform. The work of building welcoming communities does not take place within the confines of a single service, and as researchers we are well-positioned to envision new forms of intervention designed for integration across the complex modern media ecosystem.

The Potential for Meaningful Reform

Restorative practices are often promoted by the inclusive community on Twitch. While bans and timeouts remain useful tools, streamers, mods, and regular chatters often adopt a more nuanced approach, attempting to educate offenders about the values of the group by explaining the sorts of behaviors that are expected or unacceptable. Though the moderation system of the GLHF pledge was designed to encourage this type of community-based intercession, the instances of apparently genuine reform observed by our moderation team were among the most unexpected results of this entire experiment.

While significant attention, both from academia and industry, has been paid to developing algorithmically-driven moderation systems that automatically filter content, far less emphasis has been placed on designing interventions to proactively cultivate meaningful change in user behavior on social media. Our work highlights the potential for reform represented by community-driven moderation systems that foster human dialogue.

Though we have repeatedly emphasized the significance of deploying interventions at scale, it is important to stress that a solely quantitative analysis would be an insufficient measure of our success. Accordingly, though few in number, we take the cases of meaningful reform facilitated by our mods to reflect the one of the most promising avenues for further exploration. As design activists we have an opportunity to propose such alternative metrics for success and to pursue paths unlikely to be taken by the developers of online platforms. By exploring how we might take a restorative approach to the design of community moderation systems, we can actively contribute to, rather than solely critique, the current state of social media.

The Impetus for Collective Activism

As an advocacy organization grounded in research and sponsored by corporations, AnyKey sits on the boundary of two realms, giving it both unique advantages as well as challenges. Redesigning, reengineering, and relaunching the GLHF pledge system required platform development expertise and connections to industry that are not typically available to most research groups. Our ability to effect change at scale would not have been possible without the relationships one of our core members had previously established with key stakeholders at Twitch. Likewise, our internal capacity to redesign the campaign and rebuild its infrastructure did not develop until a researcher with a decade of startup experience joined us.

While AnyKey may appear to occupy an enviable position through these close ties to industry, this proximity to corporate influence also renders it vulnerable. On occasions where the goals of our design activist research diverge from the priorities of our sponsors and partners, the precarious financial and material supports for our endeavors are often placed even further into jeopardy. Like the creators of Turkopticon, we find that we must constantly attend to the intricacies of navigating the adversarial nature of advocacy work. These efforts are only made more challenging in the cases where the continued existence of our initiatives are dependent on the support of the very platforms we are attempting to positively influence.

To succeed in having a significant and sustained impact on the pressing issues faced by marginalized communities on social media platforms then, clearly we cannot afford to work alone. Our hope in presenting this detailed case study is to inspire other activists in HCI, and across the academy, to join us in a growing collective effort to support grassroots movements for social change through critically reflective design.

CONCLUSION

How can we positively intervene in online platforms through design activism? First and foremost, as advocates we must continually strive to learn from and amplify the work of the underrepresented individuals at the vanguard of the inclusive movement. Our best intentions as designers will fail to effect social change if they are not grounded in the hard-won experiences of marginalized streamers, independent content creators, underappreciated community managers, and countless volunteer moderators. In addition to seeking out and centering those on the front lines of change in the process of design, we would also do well to continue that direct engagement throughout the lifespan of our interventions. By fostering an ongoing dialogue with users of our systems, and encouraging the community to offer critical feedback, we can conduct our research in service to the greater public. Most importantly, for those of us within the university, we ought to use our privileged positions to push large-scale media platforms to take meaningful steps towards reforming their moderation practices. For motivation to engage in such collective action, we need look no further than the combined power of 370,000 gamers brave enough to stand up and speak out against toxicity.

ACKNOWLEDGEMENTS

REFERENCES

[1] C. Bicchieri and A. Lev-On. 2007. Computer-mediated communication and cooperation in social dilemmas: an experimental analysis. *Politics, Philosophy & Economics* 6, 2 (2007), 139–168. DOI: http://dx.doi.org/10.1177/1470594X07077267

[2] S. Bisker, M. Gross, D. Carter, E. Paulos, and S. Kuznetsov. 2010. Personal, Public: Using DIY to Explore Citizen-Led Efforts in Urban Computing. In *Extended Abstracts of Proceedings of Conference on Human Factors in Computing Systems (CHI EA '10)*. ACM, 3547–3552. DOI: http://dx.doi.org/10.1145/1753846.1754016

[3] L. Blackwell, J. Dimond, S. Schoenebeck, and C. Lampe. 2017. Classification and Its Consequences for Online Harassment: Design Insights from HeartMob. *Proceedings on Human-Computer Interaction* 1, CSCW (2017). DOI:http://dx.doi.org/10.1145/3134659

[4] A. Braithwaite. 2016. It's About Ethics in Games Journalism? Gamergaters and Geek Masculinity. *Social Media + Society* 2, 4 (2016). DOI: http://dx.doi.org/10.1177/2056305116672484

[5] H. Brynjarsdóttir, M. Håkansson, J. Pierce, E. Baumer, C. DiSalvo, and P. Sengers. 2012. Sustainably Unpersuaded: How Persuasion Narrows Our Vision of Sustainability. In *Proceedings of Conference on Human Factors in Computing Systems (CHI '12)*. ACM, 947–956. DOI: http://dx.doi.org/10.1145/2207676.2208539

[6] D. Chatzakou, N. Kourtellis, J. Blackburn, E. De Cristofaro, G. Stringhini, and A. Vakali. 2017. Hate is Not Binary: Studying Abusive Behavior of GamerGate on Twitter. In *Proceedings of Conference on Hypertext and Social Media (HT '17)*. ACM, 65–74. DOI:http://dx.doi.org/10.1145/3078714.3078721

[7] M. Choi. 2016. A Concept Analysis of Digital Citizenship for Democratic Citizenship Education in the Internet Age. *Theory & Research in Social Education* 44, 4 (2016), 565–607. DOI: http://dx.doi.org/10.1080/00933104.2016.1210549

[8] M. Consalvo. 2012. Confronting Toxic Gamer Culture: A Challenge for Feminist Game Studies Scholars. *Ada: Journal of Gender, New Media, and Technology* 1 (2012). DOI:http://dx.doi.org/10.7264/N33X84KH

[9] M. Consalvo, K. Mitgutsch, and A. Stein. 2013. *Sports Videogames*. Routledge. DOI: http://dx.doi.org/10.4324/9780203084496

[10] S. Costanza-Chock. 2020. *Design Justice: Community-Led Practices to Build the Worlds We Need*. The MIT Press.

[11] L. Dale. 2017. How Those Most Affected Feel About Twitch's Lax Moderation. (2017). https://www.kotaku.co.uk/2017/08/15/how-those-most-affected-feel-about-twitchs-lax-moderation.

[12] J. Dimond, M. Dye, D. Larose, and A. Bruckman. 2013. Hollaback! The Role of Storytelling Online in a Social Movement Organization. In *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW '13)*. ACM. DOI: http://dx.doi.org/10.1145/2441776.2441831

[13] B. J. Fogg, J. Grudin, J. Nielsen, and S. Card. 2002. *Persuasive Technology: Using Computers to Change What We Think and Do*. Morgan Kaufmann.

[14] C. Ford, D. Gardner, L. Horgan, C. Liu, Nardi B. tsaasan, a., and J. Rickman. 2017. Chat Speed OP PogChamp: Practices of Coherence in Massive Twitch Chat. In *Extended Abstracts of Conference on Human Factors in Computing Systems (CHI EA '17)*. ACM, 858–871. DOI: http://dx.doi.org/10.1145/3027063.3052765

[15] J. Fox and W. Tang. 2014. Sexism in online video games: The role of conformity to masculine norms and social dominance orientation. *Computers in Human Behavior* 33 (2014), 314–320. DOI: http://dx.doi.org/10.1016/j.chb.2013.07.014

[16] R.S. Geiger. 2016. Bot-based collective blocklists in Twitter: the counterpublic moderation of harassment in a networked public space. *Information, Communication & Society* 19, 6 (2016), 787–803. DOI: http://dx.doi.org/10.1080/1369118X.2016.1153700

[17] T. Gillespie. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press.

[18] K. Gray. 2013. Collective Organizing, Individual Resistance, or Asshole Griefers? An Ethnographic Analysis of Women of Color in Xbox Live. *Ada: Journal of Gender, New Media, and Technology* 2 (2013). DOI:http://dx.doi.org/10.7264/N3KK98PS

[19] K. Gray. 2016. "They're just too urban": Black gamers streaming on Twitch. *Digital Sociologies* (2016). DOI: http://dx.doi.org/10.2307/j.ctt1t89cfr.28

[20] M. Gray and S. Suri. 2019. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Houghton Mifflin Harcourt.

[21] O. Haimson and J. Tang. 2017. What Makes Live Events Engaging on Facebook Live, Periscope, and Snapchat. In *Proceedings of Conference on Human Factors in Computing Systems (CHI '17)*. ACM, 48–60. DOI: http://dx.doi.org/10.1145/3025453.3025642

[22] W. Hamilton, O. Garretson, and A. Kerne. 2014. Streaming on Twitch: Fostering Participatory Communities of Play within Live Mixed Media. In *Proceedings of Conference on Human Factors in Computing Systems (CHI '14)*. ACM, 1315–1324. DOI: http://dx.doi.org/10.1145/2556288.2557048

[23] R.L. Harper. 2014. Good Game Auto Blocker. (2014). https://github.com/freebsdgirl/ggautoblocker.

[24] D. Harry. 2012. *Designing Complementary Communication Systems*. Ph.D. Dissertation.

[25] T. Hirsch. 2009. Learning from Activists: Lessons for Designers. *Interactions* 16, 3 (2009), 31–33. DOI: `http://dx.doi.org/10.1145/1516016.1516024`

[26] J. Hoffman-Andrews. 2017. BlockTogether. (2017). `https://blocktogether.org`.

[27] L. Irani and M. S. Silberman. 2013. Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk. In *Proceedings of Conference on Human Factors in Computing Systems (CHI '13)*. ACM, 611–620. DOI: `http://dx.doi.org/10.1145/2470654.2470742`

[28] P. Lessel, A. Vielhauer, and A. Krüger. 2017. Expanding Video Game Live-Streams with Enhanced Communication Channels: A Case Study. In *Proceedings of Conference on Human Factors in Computing Systems (CHI '17)*. ACM, 1571–1576. DOI: `http://dx.doi.org/10.1145/3025453.3025708`

[29] C. Lo. 2018. *When all you have is a banhammer : the social and communicative work of Volunteer moderators*. Master's thesis.

[30] J. Lopez. 2015. Black Girl Gamers. (2015). `http://theblackgirlgamers.com`.

[31] Z. Lu, H. Xia, S. Heo, and D. Wigdor. 2018. You Watch, You Give, and You Engage: A Study of Live Streaming Practices in China. In *Proceedings of Conference on Human Factors in Computing Systems (CHI '18)*. ACM. DOI:`http://dx.doi.org/10.1145/3173574.3174040`

[32] K. Mahar, A. Zhang, and D. Karger. 2018. Squadbox: A Tool To Combat Online Harassment Using Friendsourced Moderation. In *Extended Abstracts of Conference on Human Factors in Computing Systems (CHI EA '18)*. ACM. DOI: `http://dx.doi.org/10.1145/3170427.3186498`

[33] C. Newton. 2019. The Trauma Floor: The secret lives of Facebook moderators in America. (2019). `https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona`.

[34] E. O'Neal. 2018. Harassment Task Force Releases Community Code of Conduct. (2018). `https://smashboards.com/threads/harassment-task-force-releases-community-code-of-conduct.458377`.

[35] R. Pan, L. Bartram, and C. Neustaedter. 2016. TwitchViz: A Visualization Tool for Twitch Chatrooms. In *Extended Abstracts of Conference on Human Factors in Computing Systems (CHI EA '16)*. ACM, 1959–1965. DOI:`http://dx.doi.org/10.1145/2851581.2892427`

[36] A. Pellicone and J. Ahn. 2017. The Game of Performing Play: Understanding Streaming as Cultural Production. In *Proceedings of Conference on Human Factors in Computing Systems (CHI '17)*. ACM, 4863–4874. DOI: `http://dx.doi.org/10.1145/3025453.3025854`

[37] S. Prost, J. Schrammel, and M. Tscheligi. 2014. "Sometimes It's the Weather's Fault": Sustainable HCI & Political Activism. In *Extended Abstracts of Conference on Human Factors in Computing Systems (CHI EA '14)*. ACM, 2005–2010. DOI: `http://dx.doi.org/10.1145/2559206.2581358`

[38] V. Ripley. 2016. Transmission Gaming. (2016). `https://www.twitch.tv/team/transmission`.

[39] S. Roberts. 2019. *Behind the Screen: Content Moderation in the Shadows of Social Media*. Yale University Press.

[40] R. Robinson, Z. Rubin, E. M. Segura, and K. Isbister. 2017. All the Feels: Designing a Tool That Reveals Streamers' Biometrics to Spectators. In *Proceedings of Conference on the Foundations of Digital Games (FDG '17)*. ACM. DOI: `http://dx.doi.org/10.1145/3102071.3102103`

[41] S. Savage, A. Monroy-Hernandez, and T. Höllerer. 2016. Botivist: Calling Volunteers to Action Using Online Bots. In *Proceedings of Conference on Computer-Supported Cooperative Work and Social Computing (CSCW '16)*. ACM, 813–822. DOI: `http://dx.doi.org/10.1145/2818048.2819985`

[42] R. Scully-Blaker, J. Begy, M. Consalvo, and S. Ganzon. 2017. Playing along and Playing for on Twitch: Livestreaming from Tandem Play to Performance. In *Proceedings of Conference on System Sciences (HICSS '17)*. ScholarSpace. DOI: `http://dx.doi.org/10.24251/HICSS.2017.246`

[43] J. Seering, R. Kraut, and L. Dabbish. 2017a. Shaping Pro and Anti-Social Behavior on Twitch Through Moderation and Example-Setting. In *Proceedings of Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, 111–125. DOI:`http://dx.doi.org/10.1145/2998181.2998277`

[44] J. Seering, S. Savage, M. Eagle, J. Churchin, R. Moeller, J. Bigham, and J. Hammer. 2017b. Audience Participation Games: Blurring the Line Between Player and Spectator. In *Proceedings of Conference on Designing Interactive Systems (DIS '17)*. ACM, 429–440. DOI: `http://dx.doi.org/10.1145/3064663.3064732`

[45] J. Seering, T. Wang, J. Yoon, and G. Kaufman. 2019. Moderator engagement and community development in the age of algorithms. *New Media & Society* 21, 7 (2019), 1417–1443. DOI: `http://dx.doi.org/10.1177/1461444818821316`

[46] M. Sjblom and J. Hamari. 2017. Why Do People Watch Others Play Video Games? An Empirical Study on the Motivations of Twitch Users. *Computers in Human Behavior* 75 (2017), 985–996. DOI: `http://dx.doi.org/10.1016/j.chb.2016.10.019`

[47] T.L. Taylor. 2012. *Raising the Stakes: E-Sports and the Professionalization of Computer Gaming*. The MIT Press.

[48] T.L. Taylor. 2018. *Watch Me Play: Twitch and the Rise of Game Live Streaming*. Princeton University Press.

[49] T.L. Taylor, C. Lo, and M. Romine. 2018. Live Streaming Moderation Best Practices for Event Organizers Whitepaper. (2018).

[50] T.L. Taylor and M. Romine. 2015. Women in Esports Whitepaper. (2015).

[51] T.L. Taylor and M. Romine. 2016. Diversity & Inclusion in Collegiate Esports Whitepaper. (2016).

[52] D. Wohn. 2019. Volunteer Moderators in Twitch Micro Communities: How They Get Involved, the Roles They Play, and the Emotional Labor They Experience. In *Proceedings of Conference on Human Factors in Computing Systems (CHI '19)*. ACM. DOI: `http://dx.doi.org/10.1145/3290605.3300390`

[53] S. Wright. 2019. How Fighting Game Tournaments are Dealing With Toxic and Unruly Players. (2019). `https://www.theverge.com/2019/8/16/20807484/fighting-game-smash-bros-tournaments-esports-ceo-bad-behavior-toxic-players`.