

Ignoring As a Moderation Strategy for Volunteer Moderators on Twitch

Na Li

Pennsylvania State University
University Park, USA
nzl5264@psu.edu

Jie Cai

Pennsylvania State University
University Park, USA
jie.cai@psu.edu

Donghee Yvette Wohn

New Jersey Institute of Technology
Newark, USA
yvettewohn@gmail.com

ABSTRACT

Content moderation is a crucial aspect of online platforms, and it requires human moderators (mods) to repeatedly review and remove harmful content. However, this moderation process can lead to cognitive overload and emotional labor for the mods. As new platforms and designs emerge, such as live streaming space, new challenges arise due to the real-time nature of the interactions. In this study, we examined the use of ignoring as a moderation strategy by interviewing 19 Twitch mods. Our findings indicated that ignoring involves complex cognitive processes and significant invisible labor in the decision-making process. Additionally, we found that ignoring is an essential component of real-time moderation. These preliminary findings suggest that ignoring has the potential to be a valuable moderation strategy in future interactive systems, which highlights the need to design better support for ignoring in interactive live-streaming systems.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**.

KEYWORDS

volunteer moderator, content moderation, ignoring, moderation strategy, cognitive load, context moderation, live streaming, real-time moderation

ACM Reference Format:

Na Li, Jie Cai, and Donghee Yvette Wohn. 2023. Ignoring As a Moderation Strategy for Volunteer Moderators on Twitch. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)*, April 23–28, 2023, Hamburg, Germany. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3544549.3585704>

1 INTRODUCTION

Online harassment has become a significant social issue as user-managed communities thrive rapidly. In order to effectively identify and remove harassment content, different online communities may adopt various strategies to sustain a productive and civil discourse (e.g., [3, 6, 18, 21, 28, 29]). While the role of moderation tools are likely to grow over time and curb content at scale, at present, the human moderator is an irreplaceable element during the moderation process [37]. Live streaming is popular with high-fidelity of

video streaming and low-fidelity of text chat [15], spurring much research in different contexts, ranging from gaming [31] and education [7] to e-commerce [34] and virtual avatar [25]. When the streamer is broadcasting with viewers interacting in the chatroom, some viewers might break the channel-specific rules and norms through harassing the streamer/viewers by sending toxic messages. To reduce toxicity, streamers often appoint some active viewers as volunteer moderators (mods) to manage the chatroom by removing toxic content or banning violators [38]. However, in some cases, mods might choose to ignore certain messages without taking any moderating action [3, 6].

The moderation task in live streaming communities is challenging due to the interactivity and ephemerality of the live text-based communication in the chat [6]. Mods had to make a decision with the time constraint, which forced them to experience information overload and emotional toll [32, 38]. Though previous studies have discussed a variety of strategies mods used to handle toxic messages [2, 6], none of the work examined mods' ignoring behaviors, even though they are common occurrences in fast-paced synchronous live-streaming space. To fill the research gap, we interviewed 19 volunteer mods on Twitch to understand how ignoring works as a moderation strategy, and the reasoning behind this strategy, with potential design interventions to facilitate moderation and reduce mods' cognitive overload.

2 RELATED WORK

2.1 Content Moderation Strategies

Grimmelmann [14] defined content moderation as the "governance mechanism that structures participation in online communities to facilitate collaboration and prevent abuse." According to Grimmelmann, content moderation combines algorithm and human content mods, including the administrators and mods who own the power to remove inappropriate content and the system design that decides the users' engagement in online communities [13, 14]. Although algorithmic moderation plays a vital role in curating content and determining what should be presented to online users, the design of the algorithmic system is an opaque and secretive black box [11], receiving critics from different stakeholders [9]. For example, a widespread concern is the incapability of predictive classifiers in making contextual and accurate decisions [13] even though it can get rid of inappropriate content faster and effectively, which, however, on the other hand leads to perceived injustice and distrust among users [12, 18, 37]. However, this weakness could be compensated by human mods who perform much better in differentiating between various contexts and making situated decisions to improve moderation fairness and justice [4, 19].

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI EA '23, April 23–28, 2023, Hamburg, Germany

© 2023 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-9422-2/23/04.

<https://doi.org/10.1145/3544549.3585704>

Many scholars have explored various moderation strategies from both proactive and reactive perspectives. Proactive strategies (e.g., rule echoing, word blocking) are always employed before users engage in harmful behaviors, including norm-setting [29], communicating and educating [6], flagging and filtering [21]. In contrast, reactive strategies are employed after users behave negatively, such as content removal [18], ban and block [20], contestability [36]. All of these strategies involve visible actions to the stakeholders. However, moderation by human mods might also involve invisible activities. Some studies mentioned message dismissal (ignoring) worked as a way for mods to avoid attention seekers [3, 6]. On the other hand, deliberate ignoring can reduce cognitive load, boosting people’s mental sustainability when immersed in seamless digital information [22]. Therefore, mods should choose "what to ignore and where to invest limited attentional capacities" [23]. Although previous studies (e.g., [6]) lightly mentioned ignoring strategy in moderation, there needs to be more understanding about the reasons behind such ignoring behaviors. In this study, we extend the line of research by identifying the reasons behind mods’ ignoring behaviors in live streaming communities, and we chose to focus on Twitch live streaming platform as our research locale.

2.2 Moderation in Live Streaming Communities

Moderation strategies on community-based sites have been explored in previous studies [30]. For instance, compared with top-down moderation dependent on company policy, user-governed communities (e.g., Reddit, Twitch, Facebook Groups) rely on their users to handle the most moderation work from the bottom up. As user-governed communities, Twitch, Reddit, and Facebook Group have become commonly-used online spaces where numerous users can communicate and share interests with others. However, unlike Reddit and Facebook, where people engage in asynchronous text-based discussions, Twitch is a live-streaming space where people interact with others synchronously [15]. While synchronous live chatting is more engaging and interactive than forum-based communities, such synchronicity needs more immediate attention from mods on a large volume of messages that flow dynamically and disappear fast, resulting in information overload and emotional toll [38]. Thus, understanding moderation strategies is pivotal to identifying mods’ challenges and providing possible design interventions to reduce mods’ cognitive load and improve their psychological well-being.

In live-streaming communities, mods applied combined social and technical strategies to combat harassment; The live chatting on Twitch provided immediate feedback to the viewers about mods’ moderation, and vice versa [6, 35]. Mods had to spend time explaining and educating the violators; Sometimes, if the viewers were unsatisfied with the decision, they had to spend more time combating the problematic viewer [6]. However, sometimes, mods intentionally ignore toxic messages to avoid attention seekers who might otherwise continue producing harmful content [3, 6]. We extend their studies by exploring the factors that affect mods’ ignoring activities.

Table 1: Demographic and Experience of Participants

ID	Viewership	Category	Experience (yrs)	Age	Race	Gender
P1	18-20	Gaming	4	21	Hispanic	F
P2	10-15	Gaming	4	19	African American	M
P3	70-100	Art, body painting	2.5	23	Hispanic	M
P4	—	Gaming	1.5	18	White	M
P5	—	Gaming	3	27	African American	F
P6	—	Gaming	3.5	34	White	F
P7	30-35	Rhythm & music game	0.5	18	White	M
P8	15-20	Gaming, video editing	4	18	White	F
P9	—	Gaming	1	19	White	M
P10	130-150	Gaming	2	18	Asian	M
P11	—	Gaming	3	19	White	F
P12	650-1400	Gaming, IRL	2	21	Asian	M
P13	—	Gaming, IRL, Drama	3	29	White	M
P14	—	Gaming, IRL	8	28	White	F
P15	800-1000	Gaming, IRL, eSports	6	31	White	M
P16	—	Gaming, IRL	3	24	Pacific Islander	M
P17	9000-11000	Gaming	1.5	21	White	M
P18	3000-4000	Gaming	1	20	White	F
P19	—	Gaming	5	26	Asian	M

3 METHOD

3.1 Participant Recruitment

This study was approved by Institute Review Board (IRB). We recruited 19 volunteer mods on Twitch from three different sources to diversify the samples. First, we recruited six mods via emails to the streamers and mods that participated in one of our previous projects about live streaming. Second, we recruited an additional five mods through snowball sampling, utilizing the network of a research assistant who is also a moderator on Twitch. Finally, we recruited eight additional mods through random sampling. Four researchers used their personal Twitch accounts to browse randomly recommended channels on their homepages. They first joined the channel to observe the chat, and if they saw active mods in the chat, they used the Twitch Whisper function to reach out to them with recruitment messages. Out of the nineteen mods, twelve were male and seven were female. Most moderated gaming channels with diverse viewership, ranging from tens to thousands. The average time spent moderating was three years (with a range of 0.5-8 years). Most of the mods were young, with an average age of 23. In terms of race, the majority were White (11), followed by Asian (3), African American (2), Hispanic (2), and Pacific Islander (1). More detailed information can be found in Table 1.

3.2 Interview and Data Analysis

We conducted all the interviews on Discord, using questions designed to elicit information about the mods’ experiences with moderation. The interviews began by asking questions to understand the mods, including the platforms they moderate and the length of time they have been working as mods. We then asked questions about the mods’ ignoring behaviors during the moderation process, such as whether they ever ignored toxic content, in what situations, and why. We also asked about their thoughts on ignoring behaviors. Based on their responses, we asked follow-up questions to gain a deeper understanding, such as requesting specific examples of ignoring situations or asking for further explanation. We ended the interviews by gathering demographic information. After all the interviews were completed, we used speech recognition software to generate transcripts, which were also double-checked by the researchers.

We used Taguette ¹, a free and open-source qualitative data analysis tool, for collaborative coding. Initially, two researchers independently read all transcripts to develop a general understanding of the content. They then selected one informative transcript and began independent coding. They also had weekly meetings to discuss the codes and ensure clear definitions, which were all archived. Through this process, the two researchers developed an initial codebook. The rest of the transcripts were then coded by one researcher independently and reviewed by a second researcher to achieve consensus. Any new codes added by the first researcher were reviewed by the second researcher for agreement before being included in the updated codebook. After completing the coding, the researchers exported the codes to a Google Doc and organized them under the research question to yield categories and subcategories.

4 FINDINGS

Mods might unintentionally let go harmful messages if they did not notice the toxic content, or they might intentionally ignore some toxic messages in specific situations. In this section, we discuss the reasons why mods tend to turn a blind eye to some "toxic" messages and intentionally ignore them without any moderation action.

4.1 Familiar with the Streaming Contexts

We define the streaming context as a multidimensional construct that includes the streaming content, the interactions between the streamer and viewers, and the overall atmosphere created by the content and interactions. We found that the level of familiarity mods have with the streaming context greatly influences their decisions on whether to delete or ignore harmful messages. Mods may choose to ignore these messages if they are familiar with the context, indicating that they proactively choose to ignore among other possible options.

4.1.1 When It Is Not a Big Issue to Streamers. Mods' familiarity with the streamers factored into their decisions about how to deal with toxic messages. If mods watch a streamer for an extended time, they could predict mods' reactions to viewers' comments, and their attitudes toward toxic messages etc. Further, the understanding of streamers also hinges upon the communication between mods and streamers. For instance, the streamers might specifically communicate to mods that "some messages are annoying", signaling to mods that they should handle the annoying messages. Such familiarity between mods and streamers makes it easier for the mods to understand the perspectives of the streamers on which they will rely on to decide whether they need to deal with the toxic messages or not. P17 described the moderation work as "*seeing fit from the streamer's perspective, and gauge what is toxic for streamers and not what is toxic for him*". Therefore, if streamers are unable to tolerate toxic content, mods will remove the harmful messages. On the other hand, if mods are indifferent towards negative messages, they may choose to ignore the content. Like P17 said, "*The more you watch a streamer, the more you know his reaction to the comments and you know what to do in that situation. When you see something that needs to be moderated or not depending on the streamer. It's like for once*

you remember, maybe this is okay and for another it's like, well, why did you moderate that? I don't give a fuck."

4.1.2 When the Intentions of Violators are Not Toxic. Mods can develop a deep understanding of the audience after working in the same live-streaming space for an extended period of time. In some cases, mods may choose to disregard negative messages if they are familiar with the viewers and can discern their intentions. By spending a significant amount of time in a live-streaming space, mods can gain a thorough understanding of their regular audience, which allows them to quickly identify whether viewers are engaging in "*playful banter*" or sending out "*racist or violent*" messages. If the mods can confirm that the message senders are harmless people, even though the messages are considered "toxic", they may choose to ignore them. On the other hand, if the violator is a newcomer, the moderator will handle toxic messages with more caution, as they may not have the same level of understanding of the newcomer and thus cannot accurately assess the intent of their messages. For example, P6 said, "*Other things, you know, like regulars in our streams, it's like I know what they're, what they really mean. Cause you know, we've got, I know we've had one person that's been a regular and she'll often do the backhanded threat of I will cut you. And it's like we know she's not going to, but it's more of that feisty spirit more than anything. So it's like we know she's not really going to attack this person, we're not worried about that. But if someone randomly comes in and says, wow, I'm going to cut you, we might be like, okay, you need to back off. We don't know you very well. We don't know if that means you're seriously threatening someone or if you're pat them on the back.*"

4.1.3 When the Streaming Atmosphere is Harmful. When mods are familiar with the streaming atmosphere, they have a better understanding of the content, interactions, and activities taking place in the space. This allows them to make more informed judgments about whether the atmosphere is positive and welcoming or toxic and harmful. For example, when the atmosphere is harmful, the mods would probably ignore the toxic messages that appear much more overwhelmingly than in the space where the atmosphere is positive. Based on this assessment of the streaming atmosphere, mods can make decisions about whether to ignore or take action to remove toxic messages. P15 said, "*It's very much a fluid thing. You got to understand the current, the current climate, how people are responding to things, how the streamer is doing, and then depending on how things are going, that's when you decide, okay, I'm going to allow that even though maybe I'll say something and try to deflect them from what they're saying. Um, so it's very different. You have to understand the feel of what's going on to be able to determine when you're going to be able to ignore a message.*"

4.2 The Limited Impact of Toxicity on Viewers

4.2.1 Evaluating Potential Negative Impact on Viewership. The limited impact of toxicity on the viewers is also a reason responsible for ignoring behaviors. Some mods drew the line between what harmed people and what did not. For example, suppose the offensive messages are "*racist or sexist*" or anything that clearly "*insults and hurts*" people. In that case, the mods would take action by deleting the messages or timing out the violators. Additionally, if small-scale

¹<https://www.taguette.org/>

toxic conversations began to negatively influence the rest of the audience, the mods would intervene to address the problem rather than allowing it to resolve on its own. P19 said, *"Um, just because it doesn't lay himself bothering the rest of the people in the chat, like also has a clogging the entire conversation, like some of those who will be arguing about politics takes for no reason and it'll get an old track the rest of the chat too. And that's when we step in. But otherwise, if it's just been like a few people, it doesn't really make sense to leave it alone."* However, if the impact of toxic messages is limited without hurting people or causing larger-scale negative influence on the audience, the mods will probably ignore them.

4.2.2 Estimating the Frequency and Intensity of Toxic Content. In this category, we focus on the toxic nature of chat messages and emphasize the interactions among viewers without the streamer's direct involvement. The level of toxicity, frequency, and severity of violations also play a role in determining moderation strategies. For example, if a message is just a *"suggestion"* or *"a little jab"*, the mods may allow it depending on the streamer and the streaming environment. This type of subjective judgment on the level of toxicity is also based on an understanding of the specific streaming context. If the mods have a good understanding of the streaming context, they can easily distinguish between major and minor violations, allowing them to make informed decisions about when to let toxic messages pass and when to delete messages or impose time-outs on the senders. In addition to the level of toxicity, if violators send frequent toxic or strange messages, the mods will likely intervene to warn them a few times before eventually imposing a time-out. On the other hand, if the violator sends a toxic message sporadically and does not cause any serious repercussions, the moderator may choose to ignore the message. For example, P7 said, *"If it's like a drug-related comment that's a little out of hand and it's just one message, another mod may look at it and say 'don't say that' or they'll just ignore it because why fuel the fire more"*. In this case, *"one message"* can be ignored because further intervention might *"fuel the fire more"*. Similarly, P13 said, *"I would generally say no [to toxic messages], unless there's some understanding of the user and who they're addressing. If they're kind of bantering back and forth, but not blatant racism or violence or something. I mean you don't want that. Don't want that at all"*. According to P13, some content is so toxic and should not be allowed.

4.3 Unfamiliar with Streaming Context

When mods are not familiar with the streaming context, such as the general interactive dynamics, the streaming content, they may be unsure of how to handle the *"toxic messages,"* whether they should delete them or let them pass. In this case, they may choose to either *"sit back and let the streamers handle the situation"* or simply ignore the toxic messages. P12 said, *"If I ignore something, it's either because I'm not entirely sure or it's because, uh, well actually yeah, it's because I'm not entirely sure. So generally when that happens, I'll let it go and then if the streamer does something about it, then I'll know for the next time it's like, okay, I need to take care of that next time."* This indicates that ignoring is a way for mods to learn about the streamer's expectations and community norms for future moderation. Though they may let it go, they assumed the streamer would

take action. To some extent, mods are forced to ignore without any other options.

4.4 Limited Mod's Capability and Potential Support from Other Community Members

Mods' attitude and working style also influence their decisions. Instead of paying full attention to catch any inappropriate or toxic content in the live streaming space, mods tend to work on multiple tasks while moderating. As P14 mentioned, *"I play league game at the same time when moderating... I mean, it's not like we sit focused on chat the whole entire time. You can't sit and focus on chat the whole entire time, Steven streams, because sometimes he will stream for 12 hours. And I adore Steven, but I'm not sitting there staring at chat for 12 hours ignoring my real life or, or my social life, you know."*

Besides, mods regard the moderation as a volunteering work (without payment) rather than a job requiring great time and efforts, so that they might not take the moderation job in a very serious way to the degree that they should catch on every small mistake during moderation. For example, P8 said, *"Of, yeah, of course. I'm always worried that I'm going to miss something. It's just you have to remember you're one person, you know, and you can't catch everything. And there is a reason why I'm not the only mod and I have to remember that, you know."*

Some viewers could react to the harmful messages by reminding the viewers or asking them to stop. In this case, viewers' willingness to moderate the content to some degree saved the mods' efforts to deal with the harmful content. As P3 said, *"If somebody asks a really weird question sometimes it isn't even me that calls them out and says that's really making them uncomfortable. It's actually people in the channel to say, Hey, that's really weird to ask. You really shouldn't ask that. It's happened enough for me to at least be confident that in certain chats, if something does happen and we don't catch it immediately, somebody will say something, which is good."*

5 DISCUSSION

5.1 Ignoring as the Complex Cognitive Activities of Human Moderators

Our interview data revealed that mods were engaged in complicated invisible work before deciding to ignore toxic messages. The activities ranged from evaluating the chat atmosphere to the intensity and frequency of toxic content. In general, invisible work refers to tasks that are hidden from view or socially or economically devalued, such as informal or unpaid labor [16]. In the context of content moderation, some moderation strategies can be considered as visible activities, as they or their consequences can be seen by viewers and streamers, such as banning, removing toxic content, and flagging [6]. Opposite to the publicly visible moderation work, the mental activities of mods are invisible to the public. Our study focuses specifically on the cognitive activities that are not visible to the public or the mods themselves but play a crucial role in their decision making.

The cognitive activities are influenced by the familiarity mods have with the streaming context. For instance, when mods are familiar with the context, they must consider various contextual factors (e.g., streaming atmosphere, streamer-viewer interaction,

and streaming content) before dismissing toxic content, which we refer to as proactive ignoring. In this case, mods need to judge the nature of the streaming atmosphere (friendly vs. toxic), the streamers' attitude towards toxicity (concerned VS. indifferent), and the violators' intention (unintentional VS. intentional). However, mods are forced to ignore toxicity when they are unfamiliar with the context as they had no clue how to deal with the "toxic content". We define this ignoring as reactive ignoring. While proactive or reactive ignoring are invisible activities, they prove to be a complex cognitive labor that warrants attention.

In addition to cognitive work, previous studies have also examined another form of invisible work - emotional labor - that mods experienced due to repeatedly encountering toxic content [8, 38]. Taking both invisible work into consideration, we believe that future research should aim to improve the visibility of mods' cognitive labor. With more investigative efforts into examining the traces of moderation actions (e.g., ignoring), the visibility of mods' work could go further into the field of public vision. This is important to reduce critics from the users that mods are indifferent to some violations. Besides, the higher visibility of cognitive labor would bring mods' work to the attention of the public audience, who will become more appreciative of the mods' efforts in the community. For example, future work could investigate how to visualize the mods' cognitive work by encouraging mods' to report their working hours [24], or revealing their cognitive decision-making process by presenting a cognitive working map that represents the procedures when mods make ignoring decisions under different situations. In addition, it is important for live-streaming platforms to develop a classification system that can identify and explain mods' physical and cognitive work, thereby offering higher visibility for mods.

5.2 Ignoring as a Part of Real-time Moderation

Different from traditional asynchronous communities such as Reddit and Wikipedia, live-streaming platform like Twitch is considered as a synchronous space, with unique characteristics such as simultaneity [27] and authenticity [33]. This allows users to engage in real-time interactions with streamers and other viewers through video and chat [26]. Toxic content on asynchronous platforms remains static, easy to be caught and removed by either human mods or technical interventions. However, the simultaneity of live-streaming requires mods to be highly attentive on a large amount of flowing messages in a fast speed in real time [10]. This high attention requirement can lead to overwhelming stress among mods who are supposed to handle all toxic content in a space with many concurrent viewers [38]. In addition, the synchronicity of live-streaming means that everyone, including other mods and viewers, are online at the same time [5], potentially allowing for collective management of toxic content. As such, ignoring is part of the real-time moderation process, with the expectation that other online members will provide additional support.

In dynamic decision-making environment, subjects perform poorly as a result of lacking features needed to make informed decisions [1] in their mental models. For example, our study shows that the mods' unfamiliarity with streaming context leads to the forced (reactive) ignoring. However, as "ignoring" has been used as a common moderation strategy in synchronous spaces (e.g., Twitch),

there is a need for tools and resources to help mods transform reactive ignoring into proactive ignoring in this type of environment. One potential solution is that live-streaming platforms (e.g., Twitch) could design context-specific instructional guide that help mods to proactively ignore toxic content by drawing on different variables, which includes the streaming context, the mods' attitude, and the violators' intentions etc. Compared with filtering features used by Twitch, the instructional guide could help mods, especially those novice, quickly gain a thorough understanding of the streaming communities, the streamers' and viewers' attitudes towards toxicity etc. While designing instructional guide for proactive ignoring, preserving mods' agency is important, which means that they can choose to adopt ignoring strategy, or decline to engage with it [17] based on personal needs and preferences. In the end, we need to note that over-ignoring behaviors might negatively impact the entire streaming environment where toxic content might increase. To address the issue, potential research direction could be investigating how to strike a balance between ignoring with not affecting the whole streaming environment.

5.3 Limitations and Future Work

First, we used an interview method to understand the cognitive activities of mods. We believe that advanced technology and experiments could better capture cognitive activity data through quantitative analysis, providing more accurate results, with this preliminary study serving as a guide for future experiment settings. For example, future work could run experiments to understand how different factors (e.g., context familiarity, toxicity of content) interact with each other in influencing mods' decision making. Besides, our study interviews only include mods working on the Twitch live-streaming platform, and future work could expand on more diverse livestreaming platforms beyond Twitch to improve the generalisability of our findings. Second, there are some areas of interest that we do not have enough data to support. For example, future research could examine how human mods and viewers collaborate on moderating the chat, taking into account their differing roles and power dynamics, as well as how to better understand the subjective decision-making processes of mods in light of the complex context of the streaming environment. Besides, although mods' ignoring behaviors were interpreted in our study, the impact of such behaviors on different stakeholders such as viewers or streamers were under-explored. In the future, a potential direction is to examine the influence of ignoring behaviors not only on mods, but also on streamers/viewers and the whole streaming community.

6 CONCLUSION

This study examined mods' ignoring behaviors on Twitch. Through interviews with 19 mods, we identified the circumstances when mods ignored toxic messages, and the reasons behind the ignoring. In particular, we focus on explaining how mods' familiarity and unfamiliarity with streaming context impact ignoring behaviors, and how mods' perceived impact of toxicity on viewers as well as their reliance on other stakeholders influenced their moderation decisions. Through analyzing the reasoning behind mods' behaviors, the study contributed to understanding the invisible work of mods, that is, the cognitive level of activities unseen to the public.

In the end, we derived lessons to improve the visibility of mods' cognitive work, and the design insight to effectively support mods' proactive ignoring on synchronous platforms.

ACKNOWLEDGMENTS

This research was funded by by National Science Foundation (Award No. 1928627).

REFERENCES

- [1] Berndt Brehmer. 1990. Strategies in real-time, dynamic decision making. In *Insights in decision making: A tribute to Hillel J. Einhorn*. University of Chicago Press, Chicago, IL, US, 262–279.
- [2] Jie Cai and Donghee Y. Wohn. 2019. Categorizing Live Streaming Moderation Tools: An Analysis of Twitch. *International Journal of Interactive Communication Systems and Technologies* 9, 2 (July 2019), 36–50. <https://doi.org/10.4018/IJICST.2019070103>
- [3] Jie Cai and Donghee Y. Wohn. 2019. What are Effective Strategies of Handling Harassment on Twitch?: Users' Perspectives. In *Conference Companion Publication of the ACM on Computer Supported Cooperative Work and Social Computing*. 166–170. <https://doi.org/10.1145/3311957.3359478>
- [4] Jie Cai and Donghee Yvette Wohn. 2021. After Violation But Before Sanction: Understanding Volunteer Moderators' Profiling Processes Toward Violators in Live Streaming Communities. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 1–25. <https://doi.org/10.1145/3479554> Publisher: Association for Computing Machinery.
- [5] Jie Cai and Donghee Yvette Wohn. 2022. Coordination and Collaboration: How do Volunteer Moderators Work as a Team in Live Streaming Communities?. In *CHI Conference on Human Factors in Computing Systems*. ACM, 1–14. <https://doi.org/10.1145/3491102.3517628>
- [6] Jie Cai, Donghee Y. Wohn, and Mashaal Almoqbel. 2021. Moderation Visibility: Mapping the Strategies of Volunteer Moderators in Live Streaming Micro Communities. In *Proceedings of ACM International Conference on Interactive Media Experiences*. 61–72. <https://doi.org/10.1145/3452918.3458796>
- [7] Zhilong Chen, Hancheng Cao, Yuting Deng, Xuan Gao, Jinghua Piao, Fengli Xu, Yu Zhang, and Yong Li. 2021. Learning from Home: A Mixed-Methods Analysis of Live Streaming Based Remote Education Experience in Chinese Colleges during the COVID-19 Pandemic. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–16. <https://doi.org/10.1145/3411764.3445428>
- [8] Bryan Dosono and Bryan Semaan. 2019. Moderation Practices as Emotional Labor in Sustaining Online Communities. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. 1–13. <https://doi.org/10.1145/3290605.3300372>
- [9] Motahhare Eslami, Kristen Vaccaro, Min Kyung Lee, Amit Elazari Bar On, Eric Gilbert, and Karrie Karahalios. 2019. User Attitudes towards Algorithmic Opacity and Transparency in Online Reviewing Platforms. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 1–14. <https://doi.org/10.1145/3290605.3300724>
- [10] Colin Ford, Dan Gardner, Leah Elaine Horgan, Calvin Liu, a m Tsaasan, Bonnie Nardi, and Jordan Rickman. 2017. Chat Speed OP PogChamp: Practices of Coherence in Massive Twitch Chat. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 858–871. <https://doi.org/10.1145/3027063.3052765>
- [11] Tarleton Gillespie. 2018. *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85051469782&partnerID=40&md5=8d850b5298b7e5dc1a1fcf4c427fe3da> Publication Title: Yale University Press.
- [12] Tarleton Gillespie. 2020. Content moderation, AI, and the question of scale. *Big Data & Society* 7, 2 (July 2020), 1–5. <https://doi.org/10.1177/2053951720943234> Publisher: SAGE Publications Ltd.
- [13] Robert Gorwa, Reuben Binns, and Christian Katzenbach. 2020. Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society* 7, 1 (Jan. 2020), 1–15. <https://doi.org/10.1177/2053951719897945>
- [14] James Grimmelmann. 2015. The Virtues of Moderation. *Yale Journal of Law and Technology* 17, 1 (2015), 68. <https://digitalcommons.law.yale.edu/yjolt/vol17/iss1/2>
- [15] William A. Hamilton, Oliver Garretson, and Andruid Kerne. 2014. Streaming on twitch: Fostering participatory communities of play within live mixed media. In *Conference on Human Factors in Computing Systems - Proceedings*. ACM, 1315–1324. <https://doi.org/10.1145/2556288.2557048>
- [16] Erin Hatton. 2017. Mechanisms of invisibility: rethinking the concept of invisible work. *Work, Employment and Society* 31, 2 (April 2017), 336–351. <https://doi.org/10.1177/0950017016674894>
- [17] Ralph Hertwig and Till Grüne-Yanoff. 2017. Nudging and Boosting: Steering or Empowering Good Decisions. *Perspectives on Psychological Science* 12, 6 (Nov. 2017), 973–986. <https://doi.org/10.1177/1745691617702496>
- [18] Shagun Jhaver, Darren Scott Appling, Eric Gilbert, and Amy Bruckman. 2019. “Did you suspect the post would be removed?”: Understanding user reactions to content removals on reddit. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–33. <https://doi.org/10.1145/3359294> Publisher: Association for Computing Machinery.
- [19] Shagun Jhaver, Iris Birman, Eric Gilbert, and Amy Bruckman. 2019. Human-machine collaboration for content regulation: The case of reddit automoderator. *ACM Transactions on Computer-Human Interaction* 26, 5 (2019), 35. <https://doi.org/10.1145/3338243> ISBN: 10.1145/3338243.
- [20] Yubo Kou. 2021. Punishment and Its Discontents: An Analysis of Permanent Ban in an Online Game Community. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 1–21. <https://doi.org/10.1145/3476075> Publisher: Association for Computing Machinery.
- [21] Yubo Kou and Xinning Gui. 2021. Flag and Flaggability in Automated Moderation: The Case of Reporting Toxic Behavior in an Online Game Community. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3411764.3445279>
- [22] Anastasia Kozyreva, Stephan Lewandowsky, and Ralph Hertwig. 2020. Citizens Versus the Internet: Confronting Digital Challenges With Cognitive Tools. *Psychological Science in the Public Interest* 21, 3 (Dec. 2020), 103–156. <https://doi.org/10.1177/1529100620946707> Publisher: SAGE Publications Inc.
- [23] Anastasia Kozyreva, Sam Wineburg, Stephan Lewandowsky, and Ralph Hertwig. 2022. Critical Ignoring as a Core Competence for Digital Citizens. *Current Directions in Psychological Science* (Nov. 2022), 09637214221121570. <https://doi.org/10.1177/09637214221121570> Publisher: SAGE Publications Inc.
- [24] Hanlin Li, Brent Hecht, and Stevie Chancellor. 2022. All That's Happening behind the Scenes: Putting the Spotlight on Volunteer Moderator Labor in Reddit. *Proceedings of the International AAAI Conference on Web and Social Media* 16, 1 (2022), 584–595. <https://ojs.aaai.org/index.php/ICWSM/article/view/19317>
- [25] Zhicong Lu, Chenxinran Shen, Jiannan Li, Hong Shen, and Daniel Wigdor. 2021. More Kawaii than a Real-Person Live Streamer: Understanding How the Otaku Community Engages with and Perceives Virtual YouTubers. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3411764.3445660>
- [26] Jie Lv, Cong Cao, Qianwen Xu, Linyao Ni, Xiuyan Shao, and Yangyan Shi. 2022. How Live Streaming Interactions and Their Visual Stimuli Affect Users' Sustained Engagement Behaviour—A Comparative Experiment Using Live and Virtual Live Streaming. *Sustainability* 14, 14 (Jan. 2022), 8907. <https://doi.org/10.3390/su14148907> Number: 14 Publisher: Multidisciplinary Digital Publishing Institute.
- [27] Katrin Scheibe, Kaja J. Fietkiewicz, and Wolfgang G. Stock. 2016. Information Behavior on Social Live Streaming Services. *Journal of Information Science Theory and Practice* 4, 2 (June 2016), 6–20. <https://doi.org/10.1633/JISTAP.2016.4.2.1>
- [28] Joseph Seering, Tianmi Fang, Luca Damasco, Mianhong Cherie Chen, Likang Sun, and Geoff Kaufman. 2019. Designing User Interface Elements to Improve the Quality and Civility of Discourse in Online Commenting Behaviors. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 1–14. <https://doi.org/10.1145/3290605.3300836>
- [29] Joseph Seering, Robert Kraut, and Laura Dabbish. 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*. ACM, 111–125. <https://doi.org/10.1145/2998181.2998277>
- [30] Joseph Seering, Tony Wang, Jina Yoon, and Geoff Kaufman. 2019. Moderator engagement and community development in the age of algorithms. *New Media & Society* 21, 7 (July 2019), 1417–1443. <https://doi.org/10.1177/1461444818821316>
- [31] Jeff T. Sheng and Sanjay R. Kairam. 2020. From Virtual Strangers to IRL Friends: Relationship Development in Livestreaming Communities on Twitch. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 34. <https://doi.org/10.1145/3415165>
- [32] Miriah Steiger, Timir J. Bharucha, Sukrit Venkatagiri, Martin J. Riedl, and Matthew Lease. 2021. The Psychological Well-Being of Content Moderators: The Emotional Labor of Commercial Moderation and Avenues for Improving Support. In *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery, 1–14. <https://doi.org/10.1145/3411764.3445092>
- [33] John C. Tang, Gina Venolia, and Kori M. Inkpen. 2016. Meerkat and Periscope: I Stream, You Stream, Apps Stream for Live Streams. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 4770–4780. <https://doi.org/10.1145/2858036.2858374>
- [34] Ningjing Tang, Lei Tao, Bo Wen, and Zhicong Lu. 2022. Dare to Dream, Dare to Livestream: How E-Commerce Livestreaming Empowers Chinese Rural Women. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3491102.3517634>

- [35] Hibby Thach, Samuel Mayworm, Daniel Delmonaco, and Oliver Haimson. 2022. (In)visible moderation: A digital ethnography of marginalized users and content moderation on Twitch and Reddit. *New Media & Society* (July 2022), 1–22. <https://doi.org/10.1177/14614448221109804>
- [36] Kristen Vaccaro, Ziang Xiao, Kevin Hamilton, and Karrie Karahalios. 2021. Contestability For Content Moderation. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 1–28. <https://doi.org/10.1145/3476059> Publisher: Association for Computing Machinery.
- [37] Sarah Myers West. 2018. Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society* 20, 11 (2018), 4366–4383. <https://doi.org/10.1177/1461444818773059>
- [38] Donghee Yvette Wohn. 2019. Volunteer moderators in twitch micro communities: How they get involved, the roles they play, and the emotional labor they experience. *Conference on Human Factors in Computing Systems - Proceedings* (May 2019). <https://doi.org/10.1145/3290605.3300390> ISBN: 9781450359702 Publisher: Association for Computing Machinery.