

# Practicing Moderation: Community Moderation as Reflective Practice

AMANDA L. L. CULLEN, University of California, Irvine, USA

SANJAY R. KAIRAM, Twitch/Amazon, USA

Many types of online communities rely on volunteer moderators to manage the community and maintain behavioral standards. While prior work has shown that community moderators often develop a deep understanding of the goals of their moderation context and sophisticated processes for managing disruptions, less is known about the processes through which moderators develop this knowledge. In this paper, we leverage Donald Schön's concept of *reflective practice* as a lens for exploring how community moderators develop the 'knowledge-in-action' that they use to perform their work. Drawing on interviews with 18 Twitch moderators, we conceptualize moderators as reflective practitioners, iteratively encountering novel situations and adjusting their practices and mental models. Our findings provide detailed insight into how community moderators reflect-in-action, re-evaluating in real-time their mental models of viewer intent and community goals, and reflect-on-action, conducting post hoc assessments of individual incidents and long-term changes to adjust their practice over time. Moderators working in teams reveal specific aspects of reflection facilitated by cooperative discussion, which we call 'groupwise reflective practice'. By identifying community moderation as a form of reflective practice, we can leverage insights gained from studying practitioners in other fields, providing theoretical and practical implications for the study and support of community moderation.

CCS Concepts: • **Human-centered computing** → **Empirical studies in collaborative and social computing**; • **Information systems** → **Chat**; **Social networks**.

Additional Key Words and Phrases: community moderation, online communities, online governance, reflective practice, live streaming, Twitch

## ACM Reference Format:

Amanda L. L. Cullen and Sanjay R. Kairam. 2022. Practicing Moderation: Community Moderation as Reflective Practice. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1, Article 111 (April 2022), 32 pages. <https://doi.org/10.1145/3512958>

## 1 INTRODUCTION

Many of the largest social systems on the Internet (e.g. Twitch, Reddit, Discord, Facebook Groups) rely on some combination of centralized and community-based layers of moderation. Centralized moderation layers typically focus on evaluating whether behavior falls outside of a set of established system-wide guidelines or policies, with the assumption that moderators can operate according to a 'playbook' [14, 44]. Community moderators, in contrast, must apply their contextual understanding of a group's values and goals in order to shape and monitor communities. Communities on Reddit can differ substantially in terms of stated rules [12], implicit norms [7], and what behavior is rewarded [19]. On Twitch, for example, community-specific differences may impact how channels promote 'positive' interactions [21] or the ways in which they do or don't embrace 'negativity' [39].

---

Authors' addresses: Amanda L. L. Cullen, University of California, Irvine, Irvine, California, USA, [cullena@uci.edu](mailto:cullena@uci.edu); Sanjay R. Kairam, Twitch/Amazon, San Francisco, California, USA, [sanjay.kairam@gmail.com](mailto:sanjay.kairam@gmail.com).

---

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).

© 2022 Copyright held by the owner/author(s).

2573-0142/2022/4-ART111

<https://doi.org/10.1145/3512958>

What behavior is encouraged or discouraged within a channel may depend on the goals of the streamer [52], helping to craft a distinctive performance style [41] or supporting particular types of community interaction, such as peer feedback or mentoring [11, 42].

In prior research spanning various online communities, community moderators have typically described a process of ‘learning by doing’, where the technical and contextual knowledge and strategies required to moderate effectively are gained over time [10, 23, 32, 54, 61]. However, while many prior studies of community moderation have documented the strategies that moderators have developed to cope with challenges, there remains a substantial gap in our understanding of *how* specifically community moderators, often starting as untrained volunteers, eventually develop the knowledge and strategies required to successfully engage with and manage communities.

In this work, we argue that community moderators can be conceptualized as “reflective practitioners”, following the definition proposed by educator and learning theorist Donald Schön [48], who develop a repertoire of knowledge and actions as a result of frequent interaction with the environment, in which they act on and update knowledge and strategies for managing situations both familiar and new. According to Schön, professionals become reflective practitioners through displays of both *reflection-in-action*, adjustments made during the course of an action to account for surprises, and *reflection-on-action*, analysis of responses made after the fact in order to adjust patterns of action for the future. This conceptualization of professional work forms a sharp contrast with one rooted in ‘technical rationality’, which posits that practitioners ‘solve well-formed instrumental problems by applying theory and technique derived from systematic preferably scientific knowledge.’ ([49], p.3).

In this paper, we study community moderation through the lens of reflective practice, drawing on interviews with 18 moderators across a variety of Twitch communities. Findings from these interviews confirm moderation as a ‘reflective practice’, highlighting how moderators rely on implicit knowledge and ‘learning by doing’ to fill gaps associated with the limited structured onboarding, in order to develop a professional practice around moderation work. We observe common patterns of reflection-in-action, in which moderators re-evaluate in real time both aspects of individual situations and overall goals and practices for moderation, and identify how many arrive at a system of escalating punishments in order to accommodate this reflection. With respect to reflection-on-action, we identify that expectations for behavior within a channel and how moderators should respond evolve over time through reflection on specific incidents, long-term changes, and even streamer behavior. Among moderators who worked within larger teams, we identify some specific processes of reflection associated with collaborative discussion, which we refer to as ‘groupwise reflective practice’, in which teams alternate between operating independently and convening as a group in order to actively refine mental models and practices. We summarize this processes in Figure 1. One moderator that we interviewed characterized the primacy of reflection and refinement within their moderation practice as follows:

*“Moderation is a social science and social science has a lot of gray area and nuance to it.”*

In summary, our findings make the following contributions:

- (1) We demonstrate how community moderation in live or distributed moderation contexts can be framed as a ‘reflective practice’, in which moderators learn primarily through engagement with the community, the streamer, and other moderators, as they develop a ‘professional’ orientation towards moderation work.
- (2) We illustrate how ‘reflection-in-action’ and ‘reflection-on-action’ aid moderators in refining their mental models and strategies for engaging with the community.

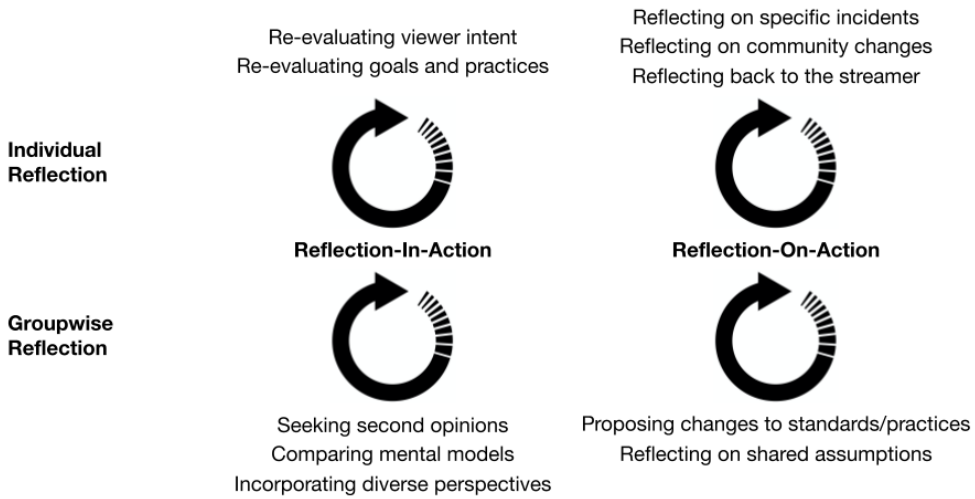


Fig. 1. An overview of reflective practice in the context of volunteer moderation on Twitch. The left side of the figure captures observations about how moderators reflect-in-action while making specific moderation decisions, and the right side captures observations about reflection-on-action, typically after a stream has ended. The top of the figure highlights our findings about individual-level reflection; the bottom highlights findings about groupwise reflection. These processes work in tandem to support moderators in refining their goals, strategies, and technical practices associated with moderation.

- (3) We explain how moderators working in teams reveal specific benefits associated with groups of professionals reflecting together in order to understand their work, which we term ‘groupwise reflective practice’.

We conclude by examining some of the broader implications of connecting community moderation and reflective practice. We outline how prior study of reflective practitioners can help inform the design of interventions to support community moderators. We also consider how studying community moderators raises theoretical implications for thinking about reflective practice in CSCW contexts more generally. Finally, we consider how some of our findings about how moderation teams approach groupwise reflective practice could inform research and design for distributed teams conducting professional work in underspecified or ambiguous environments.

## 2 BACKGROUND

This paper explores how community moderation can be conceptualized and analyzed as a form of reflective practice. We start by providing some background on reflective practice, including prior application of this framework to HCI/CSCW contexts. We review related work on community moderation, in particular on Twitch, which serves as our research context for this study.

### 2.1 Reflective Practice

Donald Schön introduced the concept of the “reflective practitioner” as a professional who engages with practical scenarios in a scientific way, reconceptualizing and testing goals and strategies for action in the face of novel situations [48]. This notion of professional practice forms a stark contrast with the notion of ‘technical rationality’ – effectively the notion that professionals can select solutions to practical problems from a repertoire of strategies. Donald Schön argues that this viewpoint ignores how practitioners ‘set’ problems in situations where the decisions to be

made or the means which are available to accomplish actions are unclear. Instead, he points to the kind of muscle-memory that practitioners employ when reacting to situations, which he refers to as ‘knowledge-in-action’. In Donald Schön’s formulation, reflection is a response to novel situations, where a practitioner’s ‘knowledge-in-action’ fails. This reflection can be categorized into: (1) reflection-in-action, the reconceptualization and adjustment to a situation as it unfolds, and (2) reflection-on-action, the analysis of situations and responses after the fact in order to adjust patterns of action in the future. Donald Schön’s model of reflective practice has been integrated and studied widely in healthcare (e.g. [26–28, 33, 34]), education (e.g. [17, 29]), and other professional contexts (e.g. [6, 9]).

Donald Schön’s theories have, to a limited degree, also appeared in CSCW and HCI work examining practices and processes of action and reflection in relation to technology design in professional contexts [35, 55], the learning and training of professionals [8, 30, 36], and online collaboration and problem-solving [22]. The principles of reflection on and in action have also appeared in larger conceptual discussions of how reflection and collaboration are framed and studied within CSCW [2, 45]. Baumer, specifically grounded reflection as a concept in HCI literature to better support the design of personal informatics technology to promote reflection [2].

In this work, we apply the theoretical framework of reflective practice to better understand a specific type of computer-supported cooperative work – community moderation. In doing so, we develop not only novel ways of thinking about and extending the work of community moderators, but also insights which generalize to related CSCW domains.

## 2.2 Community Moderation

Many of the largest online social services, including Facebook, Twitter, and YouTube, rely on a centralized pool of content moderators, or commercial content moderators [14, 15, 44]. The benefits of this moderation approach include the ability to increase the speed and scale with which human moderation of content can be completed. Typically focused on enforcing service-wide community guidelines or terms of service, these moderation teams are often equipped with complex playbooks that guide how moderators should respond to specific scenarios [40].

In many online communities, such as Discord, Reddit, and Twitch, there exists a separate layer of community moderators, appointed at the level of individual communities, rather than by the service as a whole. These volunteer moderators differ from commercial moderators in that they typically perform a variety of functions within their communities, such as socializing newcomers, managing disputes, assessing content, and making important decisions for the success and safety of their communities [25, 54, 61]. In addition, rather than focusing on enforcing sitewide community guidelines, community moderators are typically oriented towards maintaining the explicit standards and implicit norms that can vary from community to community within a service [7, 12].

Across these services, prior work has detailed how these community moderators are often elevated into the position after having been recognized as standout community members [32, 54, 61]. A cross-service study by Seering et al. found that formal education or onboarding for new moderators is typically rare and that the majority of moderators learned ‘on the job’ as they went [54]. Prior research has identified that community moderators approach their work with a number of different metaphors or analogies to various kinds of professional work, which undoubtedly shape the mental models and strategies available when making decisions in specific practical scenarios [50, 51, 61]. In turn, moderators also build mental models and metaphors for the communities that they moderate [32].

An extensive interview study of community moderators by Seering et al. cited the problem of ‘making nuanced decisions about punishments on a case-by-case basis’ as ‘one of the greatest challenges’ that moderators faced [54]. In this study, we aim to identify how moderators develop

the mental models and strategies which allow them to surmount this challenge, and the ways in which they reflect on these decisions, as they work to develop a ‘knowledge-in-action’.

*2.2.1 Community Moderation on Twitch.* Twitch<sup>1</sup> is a service that allows individuals (known as streamers) to broadcast their activities live to an audience. While user-created content on Twitch has historically been oriented towards video gaming, the site also features content ranging from creative endeavors from musicians and artists to talk shows and educational programming. Twitch provides streamers with a number of features for social interaction, in order to support them in building ‘participatory communities’ [16]. Community members interact with each other, and with the creator, through a chat functionality available to each channel, and other interactive features, such as polls, predictions, and multimedia extensions. Community members can support streamers financially by subscribing to their channels or by donating using a feature called ‘bits’. As of July 2021, Twitch reported that they had over 7 million unique content creators broadcasting each month and approximately 30 million daily visitors, on average [59].

While Twitch does use algorithmic and commercial moderation methods, communities on Twitch depend on volunteer moderators for the majority of moderation tasks. These moderators work within the channel in order to guide the community and maintain standards for behavior. Twitch moderators have a range of technical options for directly managing and interacting with community members in the chat, including deleting content, banning members from the community (temporarily or indefinitely), or making behind-the-scenes adjustments to channel infrastructure, such as editing custom commands or automated moderation tools [4]. Twitch moderators also frequently rely on the assistance of customizable chat bots to aid in their work, performing actions like responding to command prompts from community members or welcoming newcomers [53]. This technical expertise, beyond merely facilitating the work of live moderation, can also be used by moderators as tools for reflecting on viewer behavior and community norms [5]. More than just managing content, community moderators on Twitch often serve as role models; prior research has shown that the linguistic patterns of moderators are imitated at higher rates than those of other community members [52].

*2.2.2 Moderation as Professional Practice.* Several prior studies on Twitch have demonstrated how streamers establish practices of professionalization [38, 46, 57]. Prior research has also shown that moderators on Twitch often approach their roles and responsibilities as jobs in which they develop particular areas of expertise [4, 50, 51, 54, 61]. While Twitch moderators are often volunteers, they are also arguably members of social media-enabled communities of practice [1], which provide opportunities for sharing knowledge, reflecting on norms and standards, and developing skills and identities as professionals. Such communities of practice provide opportunities to learn through observation and imitation, and raise awareness of unique social conventions [36].

Prior research on Twitch moderators has identified that many of these moderators approach their work as a professional practice, that they typically learn on the job, and that they develop a repertoire of actions over time. Drawing on these insights, we believe that this is the first investigation into the work of community moderation specifically from the perspective of reflective practice. In doing so, we provide valuable context for the processes through which the outcomes described above for community moderators unfold, and develop an analogy to a body of work on reflective practice that can inform future research and design to support community moderators.

---

<sup>1</sup><https://twitch.tv>

### 3 METHODS

To address these research questions, we conducted in-depth, semi-structured interviews with 18 Twitch channel moderators. We outline participant recruitment, the interview protocol, and descriptive statistics about our interview participants in more detail below.

#### 3.1 Participant Recruitment

We recruited participants by reaching out to members of a panel of Twitch users who had previously opted in to consideration for survey and interview studies. A screener survey was distributed in July 2020 to 100 members of this panel who had engaged in moderation actions on Twitch within the preceding 28 days. Using a combination of demographic data provided during panel recruitment (e.g. age, gender, race/ethnicity) and information about the channels moderated (e.g. channel content, viewership), we selected 100 potential participants with the goal of representing a diverse set of backgrounds and perspectives with respect to channel moderation on Twitch.

The screener survey (measures provided in Appendix A) asked moderators to identify how many channels they regularly moderated, their moderation styles (adapted from [61]), the channel that they spend the most time moderating on Twitch, and additional demographic questions. We received 38 completed responses and selected 20 to participate in an interview. This strategy allowed us to identify interview participants who reflect a diverse set of moderator characteristics, perspectives, channel types, and moderation styles. Two participants did not complete their interviews, leading to the final set of 18 participants discussed below.

Details about the interview participants, including self-reported demographics, are summarized in Table 1. Participants ranged in age from 18 to 47 ( $M = 28.7$ ,  $SD = 8.1$ ). Of our 18 participants, 9 (50%) identified as men, 8 (44%) as women, and 1 as non-binary. 78% identified with at least one ethnic group other than 'White/Caucasian'. Twenty-eight percent identified as LGBTQ and 16% chose not to disclose their sexual orientation. Participants "primary" channels were evenly divided (33% each) across three high-level categories: gaming, creative, talk/other. We note that the demographics of our participants differ substantially from those observed in prior research on services like Twitch, which has often relied on convenience sampling methods that skew towards the most widely represented groups in these spaces. We specifically recruited participants to capture a diverse set of perspectives on how moderator identity, streamer identity, and channel content could contribute to the experience of moderating on Twitch.

#### 3.2 Interview Procedure and Analysis

The interviews were each approximately 60 minutes long and conducted over a one-week period in July 2020 using Google Meet. For the majority of the interview, participants were asked to focus on the channel that they had identified as their "primary" channel in the screener survey. Interviews began by asking participants to describe how they became moderators, how they learned to moderate, and how they worked with other moderators. The interviews then asked a series of questions on community standards, communication, and enforcement. These questions specifically covered topics such as how first-party and third-party tools were used for moderation and communication, how standards for community interaction were established, maintained, and modified, and how they responded as moderators to specific categories of user behavior. The interview script (provided in Appendix B) offered launching points for discussion, but we adopted a semi-structured approach throughout, to allow participants to dive deep into those topics particularly relevant to their experience on Twitch.

The audio for each interview was transcribed and anonymized. The interview transcripts were first coded by the authors using a deductive approach. While designing both the screener survey

| ID | Primary Channel |                          |            | Self-Reported Demographics |     |                         |
|----|-----------------|--------------------------|------------|----------------------------|-----|-------------------------|
|    | Category        | Content                  | Viewership | Gender                     | Age | Race/Ethnicity          |
| A  | Creative        | Music / Performance      | 10-100     | W                          | 24  | Asian-American          |
| B  | Gaming          | <i>League of Legends</i> | 10-100     | M                          | 23  | White                   |
| C  | Gaming          | <i>Call of Duty</i>      | 100+       | W                          | 23  | Native American         |
| D  | Creative        | Music / Performance      | 10-100     | M                          | 34  | White                   |
| E  | Talk / Other    | Outdoors                 | 100+       | M                          | 40  | White                   |
| F  | Talk / Other    | Just Chatting            | 100+       | W                          | 25  | Black                   |
| G  | Creative        | Music / Performance      | 100+       | M                          | 37  | White                   |
| H  | Gaming          | <i>Borderlands</i>       | 100+       | M                          | 32  | Native American / White |
| I  | Gaming          | <i>Animal Crossing</i>   | 100+       | M                          | 47  | Black                   |
| J  | Talk/Other      | Just Chatting            | 100+       | W                          | 25  | Asian-American          |
| K  | Creative        | Music / Performance      | 10-100     | NB                         | 20  | White                   |
| L  | Talk / Other    | ASMR                     | 10-100     | M                          | 27  | Undisclosed             |
| M  | Gaming          | <i>Fortnite</i>          | 10-100     | W                          | 23  | Black                   |
| N  | Talk            | Just Chatting            | 100+       | W                          | 25  | Native American / White |
| O  | Gaming          | <i>Apex Legends</i>      | 10 - 100   | W                          | 41  | Native American         |
| P  | Creative        | Food & Drink             | 10 - 100   | W                          | 19  | Undisclosed             |
| Q  | Creative        | Music / Performance      | 100+       | M                          | 30  | Asian-American          |
| R  | Talk            | Just Chatting            | 100+       | M                          | 21  | Asian-American          |

Table 1. Descriptive information for interview participants. Content, category, and viewership are provided for the primary channel that each participant moderated (the channel discussed in our interview), along with participants' self-reported demographic information. To protect participant privacy, we refer to each using an assigned letter (ID). We refer to channels with 10-100 viewers throughout as 'medium'; 100+ as 'large'.

and interview protocol, the co-authors developed an initial codebook of expected themes and topics, drawing on our review of prior work relevant to moderation on Twitch (e.g. [54, 61]). Building on themes raised by moderators regarding the role of reflection in their work, the co-authors approached the transcripts a second time, inductively developing a set of themes related specifically to learning, reflection, and professionalism. Through these, we developed the sub-categories relevant to reflection-in-action, reflection-on-action, and groupwise reflective practice which are discussed in detail in our results below.

#### 4 COMMUNITY MODERATORS AS REFLECTIVE PRACTITIONERS

In contrast with many of the professions typically studied in the context of reflective practice, such as educators (e.g. [17, 29]) and healthcare workers (e.g. [26–28, 33, 34]), community moderators are often thrust into the role with especially little training and sometimes little notice. Because few channels provide access to structured onboarding, moderators typically start out relying on the implicit knowledge they have gained through participating in the channel or on Twitch. They learn the practice of moderation largely by performing it, gaining autonomy and increasing responsibility over time. Despite the informal way that many moderators begin their work, those who persist in the role often adopt a professional orientation towards their practice, nonetheless. In these ways, we argue that community moderators display many of the characteristics of reflective practitioners.

##### 4.1 Drawing on implicit learning through participation

Many community moderators were initially chosen because they had been recognizable to the streamer through regular participation in the community, contributing by answering questions or being helpful in other ways. These have been identified previously by Wohn as 'glorified

viewers' [61] or by Seering et al. as 'standout members of the community' [54]. L, a moderator for a medium-sized ASMR (autonomous sensory meridian response) channel, described being recruited into the moderator role as a result of having played such a role within the channel as a viewer:

*Well, I would go on her stream pretty much every day, get to know her, talk and get to know her more. She would ask me questions and eventually she needed some people to help her out when her channel grew. She asked me if I wanted to help her out by becoming a mod and I was happy to help her. (Moderator L)*

Long-term participation in the channel serves not only to demonstrate an interest in the community, but also affords potential moderators with the opportunity to observe the channel and how the community operates, as described by F, who moderates a large 'Just Chatting' channel:

*Just watching the chat itself. Seeing how things can be misinterpreted if they didn't read the prior message, or just seeing how everyone communicates with each other. Little stuff like that. Just really being observant and seeing how everyone else maneuvers with each other. (Moderator F)*

Even before being placed in the role, potential moderators are evaluating and reflecting on chat, studying how it operates and the boundaries which shape behavior within the channel. Cases where moderators had not previously been regular participants in the channel are less common, but these moderators still leveraged insight from participation elsewhere on Twitch. O, a moderator for a medium-sized *Apex Legends* channel explained learning how to moderate, as follows:

*From hanging out in Twitch. Just hanging out in a bunch of streams and seeing what the streamer expects from their moderators and asking moderators questions. (Moderator O)*

D, a moderator for a medium-sized Music channel, similarly explained how watching a wider category of Creative channels allowed him to develop an understanding of what moderation strategies are more or less effective for crafting a desired atmosphere within the channel:

*So, I saw certain channels that I enjoyed, and the moderators in there had the vision that I saw like that would be perfect if every Twitch channel had that type of relationship. The mods were good friends with the streamer and then vice versa, but they're also friendly to the chat without being jerks. So, that gave me an idea of, it's like, sure you can be a jerk, because I've seen it with certain mods in certain streams, but then it maybe turns people off from watching. So I thought, if I want somebody's channel to grow, I would rather do my best to not be a jerk, and then also do my best to change the atmosphere, to be influence, be the change, you know that quote, "Be the change you want to see in the world," that kind of thing. (Moderator D)*

In this case, D's prior experience on Twitch gave him insight into not only how the atmosphere can vary from channel to channel, but also how the active participation of moderators could help to influence the atmosphere. Sometimes, moderators are able to leverage their knowledge of Twitch culture and practices in order to provide guidance for streamers. K, a moderator for a medium-sized Creative channel, described working as a moderator for a newer streamer:

*The streamer was super, super new. It was actually a lot of us helping him with a lot of things with a lot of Twitch culture and what are raids and things like that. (Moderator K)*

Finally, some of the moderators that we interviewed indicated that they were also streamers. It is not uncommon for streamers to engage in reciprocal moderation relationships in order to support each others' channels. J, who moderates a large Just Chatting channel, described how her particular understanding of how to moderate combined knowledge from prior participation as a viewer in the channel and her own experience as a streamer:

*I think for me, being a good moderator is more about understanding the community. So, I think I started off very active in the community. Then you start trying to be a helpful person, and then just try to be generally a go-to person, or active person in this community. I think I draw on my own experiences from being a streamer on Twitch. I think it is up to streamer at the end of the day, because everyone has different boundaries for what they think is and isn't okay for their chat or their community. (Moderator J)*

J's comment echoes the findings that each streamer creates different boundaries for their respective community, and describes how she used implicit learning – both from participation and from streaming on their own channel – as a mechanism for sensemaking about these differences. These findings illustrate the various ways implicit learning through participation informs the practice of moderation. For community members who are quickly thrust into the role of moderator, this implicit learning represents the main source of knowledge on which they draw when starting out.

#### 4.2 Learning by doing or 'trial and error'

Learning to moderate requires learning across two domains: the technical aspects of how to use moderation tools and the behavioral knowledge of the community and its goals which dictate how and when moderation tools should be applied. Typically, streamers are not deeply involved in the onboarding process for moderators. While some streamers do require a trial period for moderators, in which a new moderator's autonomy or abilities may be limited, this is typically overseen by other moderators. Instead, most moderators describe the process of learning how to moderate as 'trial and error' or a 'trial by fire', with roughly half of the moderators that we interviewed using one of those phrases exactly. When asked how she learned to moderate in a medium-sized gaming stream, Moderator M described the process as follow:

*It really it was just trial and error. I tried something, it worked, and then that's what I kept doing. Or I tried something, it wasn't as effective as I wanted it to be, so I sought some other avenue for getting that done. (Moderator M)*

In this case, M explains how trial and error plays an important role in learning effective strategies for accomplishing moderation goals within the channel. However, much of the learning required to moderate successfully also involves learning what the specific moderation goals for the channel actually are. I, a moderator for a large gaming channel, described how he used trial and error specifically to understand better the standards and tone that he was supposed to enforce within the channel:

*It seemed a lot of it is trial by fire, a lot of it is learning on the job. And then every now and then there's a check in with somebody who maybe made a decision that didn't make sense and the rest of us will figure out what or why. It doesn't ever feel like any of us are acting of our own accord unless we're just the only one around, but it's mostly watching and learning and getting the tone of the channel. (Moderator I)*

The notion of 'learning on the job' described by Moderator I aligns closely with observations from prior research about how users learn to moderate on Twitch and related services [32, 54]. Q, who moderated a large music stream, identified the process of interpreting a streamer's goals as the primary difficulty in learning how to moderate.

*The challenge for me was more so what kind of expectations the streamer has, because every streamer has a different view of what they want their community to be. The way I see it, my role as a moderator is to capture that vision and enforce it. Because every streamer has their own specific vision, one particular action that I take may be against that. A particular streamer might be more tolerant of certain behavior, but I won't be, so I*

*adapt my own style accordingly. I think that was the only real challenge that I had on Twitch. There wasn't really much of a technical side of things. It was just more different streamers have different expectations. (Moderator Q)*

Here, Q has emphasized the role that reflection plays in helping moderators to develop a mental model of the specific vision that a streamer has for their channel. Through this overall process of iteratively trying things and reflecting on outcomes, moderation evolved from the initial common sense gained from participation in the channel to an 'instinctive' practice developed over time, as M continued:

*I think we've all been in the channel for long enough that we all have a sense of what is and isn't okay. At the beginning, we'd message or mod chat like, "Do you think we should do this? Should a person be timed out? Do you want to change this? Do you want to up the moderation settings or anything like that?" But I think now it's just instinct at that point. We've all been doing it for long enough. (Moderator M)*

This 'instinct' that Moderator M describes having developed over time is the 'knowledge-in-action' described by Donald Schön [48]. Through this iterative process of building mental models, encountering novel situations, and refining one's practice, moderators eventually gain autonomy and increase their levels of responsibility within the channel.

### 4.3 Adopting a professional stance

Despite their typically volunteer status and the unstructured manner in which they enter the role, many moderators develop a professional stance towards their work, often referring to the practice of moderation as a 'job'. They feel that they are in a position of trust, essential to the community, and therefore should act in a manner that reads as professional, and be recognized as such. They frame professionalism against a ground of what 'unprofessional' moderation behavior looks like.

Moderators and creators who don't take the moderator role seriously are a substantial source of concern. Moderators in our interviews specifically referenced 'friend mods' or 'unprofessional mods', describing a class of actors similar to the 'token mods' described by Wohn [61] or the 'badge hunters' referred to by Lo [32]. Friend mods were often described as someone with an offline/personal relationship with the creator who has been modded in the channel to reward them for that relationship and/or to protect them from actions taken by other moderators. While friend mods aren't typically thought of as malicious, they are seen as not meaningfully contributing to the work of upholding channel standards and goals. Friend mods who didn't also start out as channel regulars are also far less likely to enter the role with an understanding of these things.

A larger concern was unprofessional moderators. These were described as mods who abused their moderation powers. This abuse described by moderators in the interviews ranged from "timing someone out because they felt like it" to "undercover trolls" who engage in behavior that actively and deliberately harms the channel once they are modded. Most moderators attributed problems caused by unprofessional mods to creators not being careful about who they choose as moderators, and moderators lacking 'training' or guidance about how to approach moderation situations, as expressed by H, who moderated for a large *Borderlands* channel:

*They might mean well, they just don't know how to express that because they're not trained. They are an all volunteer moderator force. They're not trained in how to properly handle these situations (Moderator H)*

This combination of a professional stance and recognition that most learning happened 'on the job' creates a common sentiment that more standardization in the onboarding process would be beneficial for moderators. In ideating about how to improve moderation, J described a highly-professionalized image of moderation practice:

*I think the way I think of it is, I play Magic the Gathering. There's an association of judges, where they're licensed, and there's a training course on how to use the rule book, and the rulings. It's something that you have to have some training for, but then, at the same time, you are compensated when you go to an event or something like that, or a certification that you can get. I think that would be a good way for people to take moderating a little more seriously. (Moderator J)*

This conceptualization of moderation work combines notions of accreditation, standardization, and compensation which invite comparisons to typical forms of paid professional work. This common stance towards 'unprofessional' moderators underscores the professionalism with which community moderators approach their practice, despite starting as volunteers with little training. This focus on professionalism fuels the reflective processes which allow moderators to build and refine their practice over time.

## 5 HOW COMMUNITY MODERATORS REFLECT-IN-ACTION

During a stream, moderators assess behavior in real-time, developing situational awareness of the participants currently active and monitoring when action may be required. During typical moderation scenarios, two types of reflection-in-action are common. The first type is re-evaluation of a viewer's intent through continued interaction or after receiving additional information. The second type is in-the-moment re-evaluation of the expectations for acceptable behavior in the channel and strategies for maintaining these expectations. In order to create space for evaluation and response, moderators typically adopt a process of escalating responses, allowing them to account for ambiguity about viewer understanding and intent and collect additional information.

### 5.1 Re-evaluating viewer intent

In learning how to respond to situations, one of the principal considerations for moderators as they evaluate and plan strategies for intervention is viewers' intent; understanding the context for viewer actions and ascertaining their intentions was described as an essential part of moderation. Is this individual engaging in disruptive behavior because they intended to break the rules, or do they simply lack knowledge about the expectations for behavior within the channel? Many of the moderators that we interviewed believed that experienced mods could identify at a glance, based on the content and tone of a message, whether someone had made a mistake or had intentionally acted to cause a problem. As Moderator M described, their in-the-moment assessment of a specific viewer's intent would lead them to generate a different moderation response for the same behavior:

*I think it's more so based on the intent of the trolling because some people are just, they're bored and they just want to have a wild time in some random chat where they won't be held accountable. That's understandable, that's okay. But I think when we can tell that you're purposely trying to offend somebody or hurt somebody, that's when we definitely draw a hard line. (Moderator M)*

However, because viewer interaction is largely limited to expression in chat, intent is not always easy to identify. Sometimes moderators will make an assessment, and then receive additional information about the viewer that requires them to re-assess and modify their response. Moderator I described a situation in which they updated their response to a viewer after finding out that viewer had been engaging in serial harassment behavior across channels:

*Basically, the harassment started. I looked at what they said as – maybe it was taken out of context or misunderstood by the streamer. And I was talking to mods about that, I think that maybe they just said something and it was misunderstood. And the other mods didn't agree and that was fine. We found out that I was way wrong because the person continued*

*to go from channel to channel. Yeah, I've definitely had that change because I'm like, okay, I read that situation wrong. And I feel like I'm pretty good at reading those situations, but I still screw up. Yeah. (Moderator I)*

For regular participants in the channel, moderators can develop long-term mental models around intent and knowledge of appropriate behavior. For this reason, channel regulars are typically given more leeway when misbehaving. Moderator L described how decision-making might vary for the same infraction, depending on the amount of accumulated knowledge of the user responsible:

*If somebody's first message is a rule break, we'll usually time them out, depending on kind of a sliding scale if it's an accidental thing. But if somebody comes in and says some lewd content or things like that, usually it's an immediate kind of thing. But then if it's a person who's been chatting for a long time and they say something that totally breaks a rule, it's a little bit tough. That one is super, super tough in terms of making a call. Usually that's another time where we'll do an emote that's just like, "Oh, you really said that?" But then other times, it's like somebody that the streamer knows who's one of his streamer friends who says something that's pretty not family friendly. The streamer will interact with that. That's kind of like a, "All right. I guess ..." (Moderator L)*

These findings highlight an important dimension of reflection-in-action in the context of community moderation – the fact that effective practice involves reflection about the mental states of others. In the case of healthcare, the ability to adopt an ‘interpretative’ or ‘realistic’ orientation towards patients, in which situational information is used to develop minute-by-minute updates of a patient’s state, compared with a ‘static’ or ‘objective’ orientation towards the patient, can help practitioners cope with unique or uncertain situations [27, 28]. Similarly, the ability to assess viewer’s psychological states and build models of these over time, seem to be a critical component of reflection-in-action for community moderators.

## 5.2 Re-evaluating moderation goals and practices

In some cases, specific interactions may lead moderators to re-evaluate, on the fly, either the goals of the channel or the strategies used to accomplish these goals. Often, in-the-moment adjustments in how moderators approach a particular type of behavior are prompted by a surprising community response to that behavior. Moderator B described a rule-change in his medium-sized *League of Legends* channel, which came about in this way:

*So, this came up semi-recently. She had somebody donate 200,000 bits, which is huge. Great. Congratulations. But then, people in chat were 'at'-ing the person that donated and being like, "Wow, that's amazing. I wish somebody would do that for me." Not directly asking for money but being kind of that way. So, that rule got added in, so that way, we can delete stuff. So, we deleted it then, and then we messaged them saying like, "Hey, that's not okay." But then, we added an actual rule to it. If you harass or beg for money from anyone in the channel, you will be immediately banned. (Moderator B)*

This uniquely large donation led to a novel community response, which was counter to the standards for the channel; this led to an on-the-fly modification of both the rules and the community-facing mechanisms for surfacing those rules. Sometimes a community response may instead lead moderators to reconsider how appropriate a certain type of behavior is for the channel, prompting a change to the commonly-understood rules, as described by J:

*We're going through this phase where there's a lot of people who I would say aren't as well versed in finance and stock trading, and are more curious about it, or came from the Just Chatting category in general, and are just looking for entertainment, and think of it*

*as more of an entertainment stream. So, there's times where someone posts an edgy meme, or something that could be offensive to some people, but then is generally accepted. Then you have this debate about whether or not it's okay or not okay. (Moderator J)*

In this case, J had assumed that a certain type of content would be considered offensive by the community, but the community's response differed from this expectation, leading to a reflection about whether this content should be allowed by moderators. Even when reflection doesn't involve reconsidering standards for behavior, how the audience responds may lead to adjustments in how a moderator addresses that behavior. H described how he adjusted the ways in which he approached infractions, in order to avoid an undesired audience response.

*Somebody is live on Twitch right now receiving basically a scolding from some sort of authority figure and that [crosstalk 00:17:57] can feel bad depending on how you handle that particular action... If I'm live on the channel and issuing a timeout or a ban, I'm going to say it. I'm going to say why but I also don't want add to the pile-on so you have to do it in a nuanced way where chat doesn't feel like they need to defend you because that's how toxicity also increase (Moderator H)*

Thus, reflection-in-action may lead to adjustments to a moderator's mental models of both the expected goals for the channel and the means of achieving these. While moments of reflection-in-action are hard to capture as they happen, it's clear that these moments happen frequently, due to the liveness, changing audience, and ambiguity around intent and channel goals which are common to the community moderation environment.

### 5.3 Escalation as accommodation

In the interviews, we asked moderators to walk through common moderation scenarios and the actions that they were likely to take, along with their reasoning. Our expectation had been that there would be a clean mapping between infractions of increasing severity (e.g. spam, harassment, hate speech) and punishments of increasing severity (i.e. deletion, timeout, ban). Instead, moderators typically described a 'three-strike' system of escalation, often in the following form: content deletion with a warning, a temporary timeout, then a ban from the channel. This three-strike system appears to be a common strategy among moderators to accommodate for ambiguity in both viewer intent and channel standards, providing additional time to process the situations while making these considerations. This process is similar to the strategy of 'graduated sanctions' described by Cai et al [5] and escalation processes observed by Seering et al [54] and Jiang et al [20].

Responses to behavior seen as disruptive to the stream, but not explicitly malicious or harmful (e.g. backseat gaming), typically started with a warning. Content deletion is typically paired with an explanation, although some moderators explained that deleting a message in chat is itself a warning to viewers about what language would be deemed acceptable. In the three-strike system, warnings largely serve two functions. First, warnings provide a mechanism for alerting viewers of expectations for behavior which are specific to the community, rather than enforced across the entire service. G, who moderates a large Music channel explained:

*So for things like violating the family friendly, like swear or whatever, or spelling out a swear word, that usually will get a warning because it's not every channel that is family friendly. (Moderator G)*

Here, G recognizes that viewers may be carrying over behavior that is acceptable in other channels because they believe it is acceptable in this channel. Many moderators recognize that differences in channel standards can be difficult for viewers, so a "three-strike" system gives viewers time to adjust to channel expectations without discouraging participation. The second function of warning is providing a mechanism for assessing the viewer's intent. By warning a viewer and

observing their subsequent behavior, moderators can assess whether a viewer has made a mistake or whether they intend to keep disrupting the stream, as in this example provided by O:

*There's one kid that I've encountered in a few chats and he comes in and just types meow. I mean, he does it a lot. Meow, meow, meow. You got to be like, "Dude, chill out with the meow. Then he'll be like, "But I'm a cat." Then you go 'timeout'. (Moderator O)*

In this case, even though the infraction in question was relatively benign, the escalation process allowed O to assess that the viewer's intended to persist in the behavior and then respond accordingly. There were, of course, some behaviors that many moderators described as meriting a timeout without a warning being necessary, like trash talking the streamer, because the behavior is more egregiously disruptive, offensive, or even harmful. Moderators skip to timeouts in cases where they consider it reasonable to assume that viewers would already know what they did was wrong.

Bans are used as the final corrective measure in the three-strike system for users who did not correct their behavior after a warning or timeouts and have created major disruption in the stream. Moderator D explained how a ban was used as a 'last resort' when a situation couldn't be resolved:

*We're trying to make sure that we can resolve things outside of chat until it gets to a point where we have to say something, we have to time someone out, we have to ban someone as a last resort. (Moderator D)*

The other rationale for enacting a channel ban was that there are some behaviors that can never be tolerated (racist, sexist, homophobic language being the most common examples), regardless of viewer intent; these warrant an immediate ban in the channel, as Moderator D continued:

*So, we do have like a...three step system to weed that out. If it's instantly something awful, like telling somebody to kill themselves, that's instant...that's a straight up ban because that's just, they jumped off the deep end. There's really no, 'Oh, I misunderstood,' kind of thing. (Moderator D)*

While moderators described using a "three-strike" system as internal guidelines, this system was never articulated to viewers to avoid giving potential bad actors insight that could be used to disrupt the channel. The three-strike system also provides time and space for moderators working in a live and interactive context to reflect in the moment on viewer intent and behavior and to adjust their responses accordingly without discouraging viewer participation, a common concern among Twitch moderators.

## 6 HOW COMMUNITY MODERATORS REFLECT-ON-ACTION

After a stream has ended, moderators often engage in reflection about events which occurred. Through an iterative process of confronting and reflecting on moderation incidents, moderators change their mental models for the standards within a channel and the practices best suited to maintaining those standards. In our interviews, moderators typically described a process in which streamers started out with an overall 'vibe' for the channel, and then adjusted their expectations and standards over time, as they gained experience streaming and interacted more with the audience. Sometimes, these changes were responses to specific incidents, while in other cases, these changes were responses to gradual changes within the community or outside. Though streamers define the standards for a channel, reflection among moderators sometimes leads to feedback about how streamers could modify their own behavior to meet the channel's goals and norms.

### 6.1 Reflecting on Specific Incidents

In many cases, changes to moderation practices or explicitly-stated rules are prompted by reflection on specific incidents that happen within the channel. Moderator E (large, Outdoors channel), for

example, described a situation that prompted the channel to make technical changes, in the form of updating their list of banned words, in response to the harassing behavior of a single viewer:

*She's a mother. So, she had her kids out with her. Here lately, one guy came in, and he took this whole mom fantasy thing too far. Then that kind of turned, and now those are banned terms. So, it evolves. It's a constantly evolving constitution, if you will. (Moderator E)*

While these refinements sometimes happen out-of-view of community members, they also sometimes prompt changes to the explicitly-stated rules for a channel. Changes to rules may also occur as a result of reflection on a new behavior that has developed over time, rather than one which occurred within a single incident. Moderator Q, for instance, described how a new rule around song requests was developed in response to a specific category of incidents that were being observed more frequently:

*Typically, rules come up based on necessity. Say for instance, a particular music streamer is getting a lot of demands to play a specific song or a specific set of songs, that will probably get added to the list saying that, "No, you're not allowed to make random demands." You have to say, "A lot of these musicians have monetized song requests, so you have to pay up to make a request kind of thing. You can't shout the song in chat and expect it to get played, for instance. Those kinds of rules, they get added over time. (Moderator Q)*

In some cases, the reflection which occurs after an incident involves assessment of responses with the streamer, the ultimate arbiter of what would or would not be allowed within the channel. This form of active reflection using the streamer is used to calibrate and modify moderators' individual practices, as in this example provided by Moderator K about how they refined their understanding of what is acceptable in the chat for their medium-sized Music channel:

*Usually what'll happen is, we'll ask the streamer afterwards. It will be like, "Hey, this person said this at this time, what are you feeling about that?" (Moderator K)*

In other cases, this reflective dialogue with the streamer will lead a moderator to learn that their existing mental model of the desired standards for the channel has differed from the streamer's. Moderator N described an incident where a guest streamer swearing in the channel led them to a realization about the streamer's preference for swearing:

*Just prior to that any swears we would just delete, because she doesn't swear. As moderators, we just assumed since she doesn't swear, we don't want swears in the chat. But then when this came up and she spoke up and said, "Yeah, I don't actually mind if they swear. I just don't swear. But I don't want them to get crazy with it either." (Moderator N)*

Prior work has demonstrated that this ability to adapt governance strategies to rapid changes can play a critical role in sustaining communities over time [24, 31]. On Twitch, where shifts in social norms and the culture occur rapidly, reflection on specific incidents and why they were so surprising can provide moderators with opportunities to examine the impact of those cultural shifts on their own communities and make decisions about future changes to the channel.

## 6.2 Reflecting on Community Changes Over Time

In other situations, changes to moderators' models of channel standards and how they should be enforced happened gradually, in response to changes in the audience composition and size, as well as the experience level of the streamer and moderators. In some cases, changes to moderation practices were a natural consequence of a channel successfully growing, attracting new audiences and challenges as it grew:

*When there are smaller streams, I feel like it's just like trying to keep people engaged but I think at least as the channel has gotten bigger there of course more trolls, more people*

*in chats and more opportunities for different conversations to happen in chat whether positive or negative. (Moderator A)*

In this case, the steady growth of the channel led to a recognition that moderation focus would have to shift from community engagement to stronger enforcement. In other cases, increased familiarity with streaming means that moderation practices could relax over time. For some, like the streamer for whom L moderated, this meant initially establishing standards that were very strict in anticipation of harassment being directed at a woman streamer, but then changing their approach to one that was more nuanced and adaptable to specific situations, as the streamer and the moderators become more comfortable on Twitch:

*“I think it’s loosened a little bit over time more than anything. She’s a female streamer, so with that comes a lot of guys being disgusting, things like that. I think at first, we all took that very seriously, we’re like, ‘No, get out now.’ But as for now we’re just like, ‘Okay, it happens. It’s not that offensive.’ We’ll let it slide and obviously we’ll correct people, that was not just an instant ban or anything. It’s like, ‘Hey, that’s not cool. Cut it out.’ Then if they continue, then we’ll time them out or ban them, but if they pull back, so then it’s like, ‘Okay, you get it. This is fine.’” (Moderator L)*

The factors which can lead to changes in moderation practices over time can also be external; the COVID-19 pandemic or the Black Lives Matter movement in 2020, represent the types of extreme exogenous events that might prompt changes to the underlying goals of many Twitch channels. Some moderators we interviewed indicated that their channels responded by trying to promote a ‘stress-free’ environment where viewers can escape current events. This resulted in some channels choosing to adopt stricter rules around what topics could be discussed in the channel, as in the case of the stream for which L moderated:

*We did add a few rules recently as well due to COVID. We had to maintain a stress free environment in the stream. As soon as those things started happening, we made a rule that we’re not going to talk about politics or world views in the channel. People come in the stream to forget about everything else. That would be a good place to let them know that we know this is going on but everybody is trying to forget it. It’s better if we don’t talk about it. (Moderator L)*

In contrast, other channels sought to create spaces for individuals to discuss and process these events. H described how their channel, a gaming channel attached to a corporate sponsor, relaxed its rules around political discussion to accommodate conversations around Black Lives Matter:

*We felt that it was something that was okay to talk about because we had a backing from [corporate sponsor] and they were also being involved in this. To be fair, when you look at it, Black Lives Matter movement goes beyond the political. It does kind of have a bit of a realm into the gaming sphere as well. That’s okay. You know what, the fact that they backed us and allowed us to have that kind of conversation, that feels good. It feels real good. (Moderator H)*

Here, the channel for which H moderated chose to introduce a standing rule around non-gaming conversations to help support the members of the channel in what they believed to be an important topic. Twitch channels may have the leeway to make their own rules and moderators largely work to reflect on the context of a specific channel, but these examples further emphasize how Twitch channels are not isolated islands and how channel moderators are often called to act and reflect on events, beliefs, and values that originate from beyond the channel and even beyond Twitch.

### 6.3 Reflecting back to the Streamer

A predominant theme across our interviews is that – while moderators are primarily performing the practice of moderation – the streamer is still considered the ultimate authority in a Twitch channel. Many of our moderators described themselves as reflections of the streamer and their values. As a result, adjustments made by the streamer will lead to shifts in how moderators operate within the channel, as described by Moderator I:

*They definitely have changed, a lot of it changes just in relation to how the streamer is feeling about the way of things. She may add new rules. She may take rules away. That is something that she consults with us on. But again, we can only offer and voice our opinion, the final say is hers. (Moderator I)*

However, there are also cases in which moderators are reflecting on streamer and community behavior in order to provide feedback back to the streamer. Once standards for behavior have been established, many moderators feel that a single ‘slip-up’ from the streamer could lead to viewers patterning off of a negative behavior, interpreting the streamer’s behavior as a license to engage in similar bad behavior. As Moderator M described:

*I think whatever the streamer or the moderators do, because those are the people who you’d obviously think are the power and who make the rules of the chat. What they do definitely has an effect on how chat behaves. If they see this person’s okay with saying this, then I can say that too. (Moderator M)*

While prior work has demonstrated that viewers imitate other viewers and moderators [52], these observations highlight the powerful effect that streamers can have in guiding behavior. As a result, several moderators described examples where they provided feedback to streamers about their behavior, particularly in cases where that behavior might be in conflict with the goals of the channel or Twitch’s policies. In the example presented by Moderator O below, where a guest streamer used language typically disallowed in chat, this led to a conversation with the streamer and re-adjustment of what was acceptable:

*I will enforce that if that’s what you want us to enforce. But if you would like to maybe have a conversation with [guest creator] about his language, and we can bring that down a little bit and adjust our chat expectation up a little bit and meet in the middle, I would suggest that. (Moderator O)*

Through reflection on a streamer’s actions and how a community reacts, moderators can also provide an importance source of guidance for streamers and how they can better practice their craft. In this way, they can reverse the flow of how behavioral expectations are operationalized, in order to optimize the channel for managing community behavior. So while it is true that moderators, through their selection and practices as moderators, act as reflections of the streamer, it is also true that moderators reflect on the actions of the streamer and how these actions impact the work of moderation as well as the well-being of the community.

## 7 GROUPWISE REFLECTIVE PRACTICE IN MODERATION

Many moderators that we interviewed were part of larger moderation teams within a channel. While moderators working in teams typically still operate independently when making in-the-moment moderation decisions, there are times when they might work in tandem, relying on second opinions for moderation decisions. Some channels develop formalized structures of support, often featuring a hierarchy with ‘top’ or ‘lead’ mods, who supervise and mentor other mods, who themselves may fall into specialized roles. The lead mod is typically someone who has been a longtime member of a channel, has a wide range of moderation knowledge, and who may make

final decisions without consultation with other moderators [5]. Groupwise processes occur during both reflection-in-action and reflection-on-action, but deliberation is typically hidden from the community so that moderators can present themselves as a unified team.

### 7.1 Group processes in the moment

When moderators are working on a team, in-the-moment reflection often involves bringing in other moderators for second opinions or support. This behavior, also observed by Cai, et al. in their interviews with Twitch moderators [5], allows moderators to test and refine their mental models against those of other professionals working in their context. Moderator N recounted a process described frequently in our interviews, in which moderators use a secondary moderator chat (often in Discord) as a way to get real-time feedback about in-the-moment moderation decisions:

*If there's a message that somebody is unsure about where other people might not have caught it but somebody saw it and they were like, "I don't know how to feel about this, this is borderline a little creepy or whatever." So sometimes I'll take a screenshot and put it in the mod chat and be like, "Hey, I don't know if you saw this message, should I delete this, is this okay?" And then we'll just respond about it in there because it's easier to do that off stream than put it in chat and be like, "Hey, I'm going to delete this." (Moderator N)*

All but one of the 18 moderators we interviewed described Discord as essential to their moderation work for these kind of discussions, which they felt should not be visible to everyone in Twitch chat. Because Twitch currently lacks features that support group conversation between moderators, dedicated and private "mod only" channels in a Discord server allow moderators to coordinate with each other in real-time when responding to situations on Twitch.

In some cases, specific incidents provided a mechanism for moderators to compare and contrast their mental models regarding what is acceptable within the channel. In this way, a specific message can play a role which is similar to a *boundary object* [56], in that it provides a common reference point for aligning two different view points, but different, in that here the message serves the function of helping bring parties into a consensus view of how the channel should be moderated. Moderator F described how these discussions would turn individual reflection into a group process of reflection:

*I think it's an individual thing. What might be acceptable to me might not be acceptable to another mod. It's kind of like, "Okay, well, if this is going on here, and this person doesn't like it, all right, maybe we should all implement this into our way of thinking." (Mod F)*

These points of contention can also allow moderators to reflect on their individual identity and how that may inform a different interpretation of how the team should respond to situations. Moderator J described the experience of being the only woman moderator in a finance-related stream, and how that led her to push the team to reflect on the currently-existing boundaries:

*Because this is a finance related channel, I'm probably one of maybe four females on the entire Discord server. So, sometimes there are things that would make me uncomfortable, or other women have said, "This makes me uncomfortable," that everyone else thinks is fine. But I think it's important. I'm glad that [streamer] moderated me, because then I can say, "No, this is not acceptable." If I were a newcomer to this stream, just because you boys think it's okay doesn't mean that it is. I think that's where boundaries are always being redrawn or shifted. It's always a gray area. (Moderator J)*

Twitch moderators may be recruited to the channel for a variety of reasons, often independently and without input from current moderators, but being able to rely in realtime on the expertise of a group of professionals with shared knowledge strengthens the capabilities of moderators while

developing common points of reference that can be used to align informal standards and metrics for their communities of practice.

## 7.2 Group processes after the fact

The moderators we interviewed also talked about using secondary discussion spaces like Discord for hosting “staff meetings” after the stream, in which they can engage in reflection on actions taken in the channel and propose changes to standards and practices within the channel, as described by Moderator Q:

*[Discord] is more for coordination purposes. We interact primarily off stream. We discuss anything. We discuss improvements to the stream, for instance. We discuss things that [streamer] could do that could be beneficial for our channel, and we bounce ideas off of each other.*

Discord is a commonly-valued tool in part because it saves a log of the conversation for inactive moderators to review, reflect, and comment on later. Spaces like a mod-only Discord channel provide an opportunity for moderators to discuss not only individual cases, but also larger categories of viewer behaviors, in order to coordinate and adjust their understanding of the bounds of acceptable behavior. Moderator O described a situation in which a guest streamer’s use of profanity on the channel prompted a discussion among the moderators about the channel’s rules:

*I went into the mod lounge in Discord and I was like, "Okay, I think we need to establish some ground rules on language and chat because I'm deleting messages and people are getting pissed at me for deleting the messages because they're hearing [guest streamer] drop an F bomb every three words while you're playing together." I can't justify to them that they can't use those words when they're hearing them in the chat without some conversation about it.*

In this scenario, a specific incident prompted the group to reflect on the long-standing assumption that the streamer wanted all profanity moderated out of the chat. This incident prompted them to share their assumptions with the group and identify that the boundary for acceptable behavior needed to be moved and defined, in order to align more closely with the streamer’s true preferences. After reflecting on this scenario as a group, the moderation team recognized that the streamer was accepting of some level of profanity, and modified their practices accordingly. These examples illustrate how group conversations about events in the stream can lead to long-term changes in how individuals then approach moderation work in the future.

## 7.3 Presenting a united front

As a rule, moderators aim to keep both of these types of groupwise reflection in secondary channels, outside of the public spaces on Twitch or Discord, where viewers could observe it. Moderators wish to hide these discussions for viewers for two primary reasons. First, when reflecting in-the-moment, it’s easier from a practical perspective for moderators to coordinate in a separate space, so that active conversations about moderation aren’t lost in the behavior that’s being discussed.

*We actually have a voice chat in our Discord that is literally called, "The War Room," where we go whenever there's a crisis. Whenever we're being spammed by bots, or something happens to the streamer, that's where we'd go to enmesh and make sure that everything is okay. (Moderator J)*

In addition to practical reasons for keeping these conversations separate, many moderator teams also felt that keeping these conversations ‘behind the curtain’ reduced the potential for problems to

escalate and avoided cases where the streamer or another moderator might accidentally (or intentionally) undermine a moderator's decision in public. D explained how secondary communication channels are used to ensure that moderators were enforcing a common standard:

*Obviously most streamers have Discord. So, we do have a separate channel for moderators that only mods can see, and of course the streamer, where we can address things privately about certain viewers or certain members that, "Hey, something's going on. We might need to ban this person. We might need to reach out." Or, "So and so said this, but so and so said that", but we don't want to look biased publicly. (Moderator D)*

Here, the secondary communication space is allowing a moderation team to compare responses to two viewers in the channel to avoid concerns about bias. The language used by D demonstrates an awareness of how the moderation team represents itself to the community, with an eye toward appearing objective and professional. Finally, these secondary communication channels provided an avenue for coordinating critical enforcement decisions in real-time with the streamer:

*Usually, it goes to the streamer. So, if it's over a ban or something like that, we usually try and keep a united face on the stream, but in Discord, we'll be talking about it. And if it gets too much, we will message the streamer. It sucks to stop the flow of the stream or whatever, but she'll read it and she will make the decision. (Moderator B)*

The comment about a "united face" from B is illustrative of a sentiment common across the moderators we interviewed – that their primary role is to support the streamer and that moderators in a channel need to be on the same page in terms of standards and their enforcement. Previous research has described how subreddit moderators similarly find it convenient to move discussions to private communication channels that are 'invisible' to the community [10]. Accordingly, in cases where moderators might disagree with each other or even the streamer, they will all defer to the streamer. A critical part of consistently enforcing the standards was representing the moderation team as in lock step with the streamer's wishes.

## 8 DISCUSSION

In this work, we were motivated by the question of how community moderators develop the intuitive knowledge that allows them to operate independently, in real-time, in their roles managing communities on Twitch. Through interviews with 18 moderators, we explore community moderation as a reflective practice, in which moderators develop this 'knowledge-in-action' through their frequent interactions with community members, other moderators, and the streamer. We observe a common pattern in which moderators enter the role with limited structured onboarding, instead relying frequently on implicit learning gained through participation, and develop over time an increasingly instinctive practice and professional stance regarding their work. When approaching moderation scenarios in the moment, moderators often 'reflect-in-action', re-evaluating both the specifics of the situation and the overall goals of the community, and employ a series of escalating responses as a strategy for accommodating the time and cognition required for problem-setting. After the stream, it is common for moderators to 'reflect-on-action', evaluating specific incidents or long-term changes to the community in order to re-assess moderation strategies or goals; these sometimes include re-assessments of how the streamer should behave in order to achieve the goals of the community. Finally, we observe that moderators working in teams often engage in reflection as a group, both during and after the stream, though they typically attempt to hide this 'groupwise reflection' from the community.

Taken together, this work demonstrates how community moderators, who typically start out as untrained volunteers, develop the sophisticated knowledge and strategies required to manage communities in the context of Twitch. We observe that the reflective practice of community

moderation contrasts starkly with the work of commercial content moderators, who typically receive structured training about how to moderate against the goals of the service, which may be spelled out in both community guidelines and internal moderator playbooks [14, 44]. In this way, commercial content moderation appears to be emblematic of the ‘technical rational’ approach against which Donald Schön contrasts his formulation of reflective practice [48]. In the sections below, we consider first how we can further support moderators in incorporating reflection into their practice, and then how we can potentially build scaffolding for ‘learning by doing’ by documenting some of the common strategies and mental models observed across community moderation contexts.

## 8.1 Supporting Reflection in Community Moderation

The role of reflective practice and how it can be supported has received considerable attention over the past few decades in a range of fields, particularly in healthcare and education. By connecting the work of community moderation to this larger field of inquiry, we have an opportunity to leverage insights gained in these fields by applying them to the design of interventions to support community moderators. A number of prior studies have demonstrated that structured interventions designed to support reflection-on-action can lead to measurable changes in practitioners’ levels of reflection (e.g. [6, 17]). Interventions in these professions to support reflection-in-action, however, are much more complicated; one can’t interrupt in a clinical setting to encourage practitioners to engage in structured reflection. The computer-mediated environment of community moderation, in contrast, provides some opportunities for supporting both reflection-in-action and reflection-on-action. To aid this discussion, we leverage Baumer’s paradigm of *breakdown*, *inquiry*, and *transformation* [2] as a structured way of developing ideas for technical interventions.

**8.1.1 Supporting reflection-in-action.** A critical component of problem-setting in moderation scenarios is determining whether a breakdown is taking place. One side of this is being clear about the rules of the community; as is evident from our interviews, streamers’ expectations for behavior in their channels are not always clear and may change over time. This provides opportunities for surfacing and specifying the existing knowledge about community rules and expectations to support real-time decision-making among mods, such as providing spaces within first- and third-party moderation tools for documenting and communicating rules, along with examples of infractions and potential consequences. The other challenge is determining the intent of the community members involved in the scenario: is this well-intentioned conversation that went too far or was the intent to disrupt or harm others? A clear opportunity arises for making available more information about viewers and their prior behavior based on the “three strike” model, in order to support assessment of what kind of breakdown is taking place. These could either come in the form of signals that are mined automatically from the viewer’s behavior in this and other channels (e.g. previous timeouts and bans), or as a log generated manually by moderators within the channel that documents the history of interactions with individual community members. One such example of how companies can respond to this opportunity is the ‘Suspicious User Detection’ tooling introduced by Twitch in late 2021, which highlights accounts that may be trying to evade a previous channel ban so that moderators can take action [58]. Making both types of information easily searchable or browsable can support patterns of inquiry, allowing moderators to test their hypotheses about the situation and viewer intentions. Transformation may be more difficult to achieve in the moment, during the fast-paced work of real-time moderation, but tools which help moderators notate incidents or information that they wish to revisit later, could facilitate later reflection which leads to change.

**8.1.2 Supporting reflection-on-action.** In designing for reflection-on-action, there are more opportunities to leverage and extend interventions tested in other professional practices. Prior studies

have established that reflecting on videos of work (one's own, or others) can prompt valuable insight and reflection (e.g. [6, 17]). In this case, we have the opportunity to summarize and surface breakdowns which happened within the stream using views such as moderator action logs, which can prompt reflection after the stream has ended. Prior work has also identified that the use of standardized questions to evaluate breakdowns can support reflection [29]; either through informal processes or built into an interface for reflection, perhaps there is a way to structure post-stream inquiry into moderation scenarios and breakdowns. Through aggregation of data about viewer activity or summarization of content in chat, it may also be possible to support reflection about longer-term changes in the community, beyond a single specific incident, such as surfacing to streamers and moderators when certain types of incidents are becoming more common. In either case, these systems should support moderators and streamers in recording what they have learned about the goals of the community and expectations for behavior, both internally to support moderation processes and externally to surface these to the community.

## 8.2 Building scaffolding for learning by doing

Though communities on Twitch may have disparate goals, our interviews demonstrated that moderators across a diverse set of communities often arrive at common strategies for managing moderation work. As noted by Kinsella, Donald Schön's criticism of the 'technical rational' viewpoint reflects the idea that not all practical knowledge can be documented, rather than a complete repudiation of the idea of documenting practical knowledge [26]. Can we encode parts of the 'muscle memory' demonstrated by the moderators in our interviews into documentation that could support new or existing moderators in their work and provide scaffolding to accelerate future learning? Here, we consider some potential challenges associated with relying on reflection to develop knowledge-in-action and how these might be addressed.

*8.2.1 Challenges associated with learning by doing.* Our findings show that moderators typically rely on implicit learning through participation in order to compensate for a lack of structured onboarding to the channel-specific expectations for moderation. One potential consequence of this practice is a tendency towards homogenization and concerns with 'culture fit' within Twitch channels. This could lead to concerns around 'groupthink' in critical moderation scenarios, and the exclusion of perspectives which differ from the dominant viewpoints represented within the audience of the channel. As exemplified by Moderator J, who viewed her status as a woman in a finance-related stream as outside the norm, adding non-normative viewpoints to discussions can push teams to reconsider how the channel wishes to approach moderation. As in other workplace contexts, channels could benefit from recognizing the value of diversifying their moderator teams and shared practices [18]. Onboarding materials that could support potential moderators in learning the norms of the channel without the requirement of long-term participation could enable streamers to select more diverse moderation teams.

The 'three-strike' system of escalation appeared to be another form of compensation caused by two forms of ambiguity: viewers' intent and creators' expectations for behavior. Instead of taking immediate, decisive action, moderators engage in an extended interaction with viewers, sometimes consisting of multiple turns, in order to determine whether a rule was broken and whether it was done so intentionally. For moderators working in larger channels, it's possible that they could be juggling more than one of these situations at a time, leading to situations that are cognitively taxing. When these situations involve long-term community members, potential outcomes such as timeouts or bans may be charged emotionally or may lead to additional future complications. Previous research has identified that community moderation work is frequently cognitively and emotionally taxing [3, 54, 61]. One potential source of ambiguity – those concerning the rules of

the channel – could be reduced substantially through the creation of onboarding materials that support moderators in understanding the expected rules of the channel and how infractions should be handled, potentially reducing the cognitive and emotional burdens of moderation

**8.2.2 Opportunities for documenting practices.** As noted above, moderators across a variety of communities arrived at many commonly-shared practices for handling moderation scenarios. This indicates that much of the ‘knowledge-in-action’ that moderators are currently developing through their reflective practice could be documented in the form of moderation onboarding materials available to moderators across the service. Rather than discussing channel-specific practices, technical systems, or rules, these could provide general best practices or strategies for addressing common scenarios, which could help new moderators act more effectively and efficiently. One major community-driven platform which has achieved this successfully is Discord’s ‘Moderator Academy’<sup>2</sup>. These materials provide structured rubrics for thinking through common moderation scenarios and questions to support moderators in reflecting on how they can facilitate positive environments; structured, standardized question-asking has previously been identified as an effective strategy for encouraging reflection [29]. Developing materials such as these for Twitch could accelerate moderators in learning the common knowledge required to moderate, enabling them to devote more time to reflecting on how to adapt strategies to their specific community.

Whether due to lack of time or lack of understanding of how these tools might help, few streamers explicitly articulate their community-specific vision or the standards for the channel to moderators. Instead, these expectations are often expressed implicitly, leading moderators to dedicate time and energy to uncovering these expectations, through a dialogue facilitated by incidents within the channel. These sitewide materials could be supplemented by the creation of community-specific onboarding materials, which help to address some of the challenges raised above. These materials would have to be both easy to customize, in order to account for the community-specific rules and moderation practices encouraged by each streamer, but also easy to update, to adapt for the common changes to these rules and practices over time captured in our findings.

### 8.3 Groupwise Reflective Practice in CSCW

One of the observations from this study is that it is common for moderator teams to reflect on both strategies for moderation and goals for the channel as a group. Sometimes this involves testing hypotheses in-the-moment against the intuitions of others, and other times this involves sensemaking about an incident after it has happened. Previous studies in education [17] and health-care [43, 60] have evaluated interventions to encourage reflection in groups among practitioners who would typically work separately. In some cases, positive results similar to isolated reflection were observed [17, 43]; in a study by von Klitzing [60], however, group-based reflection led nurses to shift their focus away from their practice and towards their patients. Clearly, certain types of group-based interventions will be more successful than others in helping practitioners develop the necessary responses to scenarios.

In this study, we observe that many community moderators are already engaging in active groupwise reflection as part of their work within larger moderation teams, and with the streamer. Moderators are actively testing hypotheses about new scenarios against the mental models of others in their team, and using the information received to update their own mental models of how to manage the channel. Furthermore, having these practices for discussion and reflection in place seemed to reinforce moderator perceptions of themselves as professionals, allowing them to act with confidence. Future research in the context of Twitch specifically might examine more closely the role of streamers in establishing a workplace environment for moderators, and how they may

<sup>2</sup><https://discord.com/moderation>

act as co-workers, team leads, or supervisors for moderators. Echoing the similar recommendation of Cai et al. [5], we also recommend that services like Twitch that do not offer first-party dedicated channels for professional discussion and reflection to community managers incorporate such spaces into future designs and products. Moderators in this study considered having a place to discuss issues in private to be crucial to the work of moderation and community management, with another desirable feature being records of these discussions and any resulting changes. It seems reasonable that in any context where professionals are working online, live and in public, that these practitioners need private and manageable places that facilitate important background supporting tasks and communication.

Drawing on our observations about the groupwise reflective practice exhibited by community moderator teams, it would be interesting to identify insights that could be adapted to other CSCW contexts, as areas for future research. Research and design studies targeting users engaged in cooperative sensemaking, knowledge co-production, and the use of collaborative tools for design could probe the specific benefits afforded by encouraging new forms of group reflection on the assumptions, tools, practices, and goals of these activities.

## 9 LIMITATIONS AND ETHICAL CONSIDERATIONS

Before concluding, we briefly consider some limitations of the present study which provide context for our conclusions, and some ethical considerations which guided our research. Regarding the community moderators included in our study, we made a concerted effort to represent a broad range of perspectives by recruiting a set of interview participants who were diverse with respect to both their own identities and the channels which they moderated. However, all of our participants were English-speaking, moderating for English-language channels, and were located in the United States. Future research with community moderators living outside the United States, or working with streamers located outside the United States, could explore whether the patterns we observed are consistent cross-culturally. Second, the size of our study did not provide the opportunity to understand how practices around reflection might vary according to moderator demographics; future work with larger groups of participants, such as those employing surveys, could provide the right data for making these comparisons and building on our findings here.

Our access to participants was facilitated by a relationship with Twitch; participants in this study were aware of this connection when being interviewed for this research, introducing a form of bias into our results. Previous studies of Twitch using convenience sampling have often over-represented the viewpoints of younger men; we felt that the benefits of gaining access to a more diverse moderator pool, including moderators in larger channels, make up for some of these concerns. Similarly, we recognize that academic-industry collaborations can also raise concerns about the objectivity of the findings. For this reason, we have not focused this study on capturing moderators' opinions about Twitch policies or products or on comparing community moderation practices on Twitch to those on other services. Instead, we focus on providing a detailed glimpse into moderation processes which have been documented at a higher level in prior work [5, 51, 54, 61].

We note that we interviewed these 18 moderators a few months after the start of the COVID-19 pandemic in 2020. COVID-19 and associated work-from-home mandates brought a number of changes to Twitch and its viewership, some of which were reflected on by moderators in our interviews. It's possible that responses to these changes may have led moderators to approach or conceptualize their work in ways that differ from how they might before or after the pandemic, leading to concerns about generalizability. Many of our findings delved more deeply into phenomena already observed in prior work, such as the need to 'learn by doing' [54] or developing systems of escalating responses [5, 20, 54]. For this reason, we feel that many of the findings presented here are likely to continue to generalize beyond the pandemic.

Our interviews focused primarily on how channels develop and maintain standards of behavior among their communities, with conversations often leading towards practices associated with enforcement. Previous work has indicated that enforcement composes only a part of the overall work that community moderators perform, along with engaging newcomers, facilitating discussions, assisting the streamer, and other actions to help manage the channel and community (e.g. [37, 54, 61]). Future work on how community moderators hone their practical knowledge could explore more fully the role that reflection plays in facilitating these processes outside of enforcement.

Finally, we followed the ethical guidelines for online research recommended by the Association of Internet Researchers [13], and we considered participant privacy and autonomy in a variety of ways. We stored all participant data within a securely locked network, and in our data collection and analysis procedures, we used identifiers that obscured personal names and publicly-facing screen names of the moderators themselves, as well as those of any moderator or streamers they discussed during the interviews. Both the screener survey and the interview protocol used in this research were approved by the Internal Review Board (IRB) at a university.

## 10 CONCLUSION

We have demonstrated how Twitch moderators act as reflective practitioners and professionals, but we wish to conclude by further advocating for a reframing and recognition of moderation as stewardship, due to the important role moderators play as individuals sharing themselves and nurturing a community [3, 51] and because moderation is an emotionally intensive and laborious role that is often unrecognized and largely a volunteer effort. Many moderators describe themselves as motivated by a love of community, a sense of responsibility, or a desire to “give back.” Through their growth as reflective practitioners, Twitch moderators are not only compelled to develop methods for accomplishing their work, they are further prompted to consider the needs of the community and how best to care for and nurture its members. Moderation is not only about regulating the community, it’s about the holistic sense of how we create the community we want to create. Community moderators are an integral part of the sense of community on Twitch, reflecting and reflecting on the values and viewers of Twitch. Despite our calls for more tools, more resources, and more standardization to support Twitch moderators, we are not advocating for a logic of choice where there are “one size fits all” solutions and technology and moderators themselves are simply a means to an end. This would deny the special understanding and care that moderators offer and even what is unique about communities on Twitch. Rather, we strive for an understanding of moderation as structured by a logic of care [47]. For us, what is important is recognizing how we can cultivate the logic of care that already exists on Twitch, as evidenced by the reflective practices of Twitch moderators, and offer more sources of contextualized support to those moderators who care deeply about the communities they serve.

## REFERENCES

- [1] Warren Allen. 2013. Rewarding participation in social media enabled communities of practice. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 7.
- [2] Eric PS Baumer. 2015. Reflective informatics: conceptual dimensions for designing technologies of reflection. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 585–594.
- [3] Johanna Brewer, Morgan Romine, and T. L. Taylor. 2020. *Inclusion at Scale: Deploying a Community-Driven Moderation Intervention on Twitch*. Association for Computing Machinery, New York, NY, USA, 757–769.
- [4] Jie Cai and Donghee Yvette Wohn. 2019. Categorizing Live Streaming Moderation Tools: An Analysis of Twitch. *International Journal of Interactive Communication Systems and Technologies (IJICST)* 9, 2 (2019), 36–50.
- [5] Jie Cai, Donghee Yvette Wohn, and Mashael Almoqbel. 2021. Moderation Visibility: Mapping the Strategies of Volunteer Moderators in Live Streaming Micro Communities. (2021).
- [6] Alberto AP Cattaneo and Elisa Motta. 2020. “I Reflect, Therefore I Am... a Good Professional”. On the Relationship between Reflection-on-Action, Reflection-in-Action and Professional Performance in Vocational Education. *Vocations*

and Learning (2020), 1–20.

- [7] Eshwar Chandrasekharan, Mattia Samory, Shagun Jhaver, Hunter Charvat, Amy Bruckman, Cliff Lampe, Jacob Eisenstein, and Eric Gilbert. 2018. The Internet’s Hidden Rules: An Empirical Study of Reddit Norm Violations at Micro, Meso, and Macro Scales. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 32 (Nov. 2018), 25 pages. <https://doi.org/10.1145/3274301>
- [8] Derrick Coetzee, Seongtaek Lim, Armando Fox, Bjorn Hartmann, and Marti A Hearst. 2015. Structuring interactions for large-scale synchronous peer learning. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 1139–1152.
- [9] Lynn Crawford, Peter Morris, Janice Thomas, and Mark Winter. 2006. Practitioner development: From trained technicians to reflective practitioners. *International Journal of Project Management* 24, 8 (2006), 722–733.
- [10] Bryan Dosono and Bryan Semaan. 2019. Moderation Practices as Emotional Labor in Sustaining Online Communities: The Case of AAPI Identity Work on Reddit. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) (CHI ’19). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3290605.3300372>
- [11] Travis Faas, Lynn Dombrowski, Alyson Young, and Andrew D. Miller. 2018. Watch Me Code: Programming Mentorship Communities on Twitch.Tv. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 50 (Nov. 2018), 18 pages. <https://doi.org/10.1145/3274319>
- [12] Casey Fiesler, Joshua McCann, Kyle Frye, Jed R Brubaker, et al. 2018. Reddit rules! characterizing an ecosystem of governance. In *Twelfth International AAAI Conference on Web and Social Media*.
- [13] Aline Shakti Franzke, Anja Bechmann, Charles Melvin Ess, and Michael Zimmer. 2020. Internet research: ethical guidelines 3.0. (2020).
- [14] Tarleton Gillespie. 2018. *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.
- [15] Mary L Gray and Siddharth Suri. 2019. *Ghost work: how to stop Silicon Valley from building a new global underclass*. Eamon Dolan Books.
- [16] William A. Hamilton, Oliver Garretson, and Andruid Kerne. 2014. Streaming on Twitch: Fostering Participatory Communities of Play within Live Mixed Media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) (CHI ’14). Association for Computing Machinery, New York, NY, USA, 1315–1324. <https://doi.org/10.1145/2556288.2557048>
- [17] Judith Harford and Gerry MacRuaric. 2008. Engaging student teachers in meaningful reflective practice. *Teaching and teacher education* 24, 7 (2008), 1884–1892.
- [18] Tanja Hentschel, Meir Shemla, Jürgen Wegge, and Eric Kearney. 2013. Perceived diversity and team functioning: The role of diversity beliefs and affect. *Small Group Research* 44, 1 (2013), 33–61.
- [19] Benjamin D Horne, Sibel Adali, and Sujoy Sikdar. 2017. Identifying the social signals that drive online discussions: A case study of reddit communities. In *2017 26th International Conference on Computer Communication and Networks (ICCCN)*. IEEE, 1–9.
- [20] Jialun Aaron Jiang, Charles Kiene, Skyler Middler, Jed R. Brubaker, and Casey Fiesler. 2019. Moderation Challenges in Voice-Based Online Communities on Discord. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 55 (Nov. 2019), 23 pages. <https://doi.org/10.1145/3359157>
- [21] Mark R Johnson and Jamie Woodcock. 2018. Conversation, Discourse and Play: Interaction and Moderation in Twitch. tv Live Streaming. (2018).
- [22] Young-Wook Jung, Youn-kyung Lim, and Myung-suk Kim. 2017. Possibilities and limitations of online document tools for design collaboration: The case of Google Docs. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 1096–1108.
- [23] Charles Kiene, Jialun Aaron Jiang, and Benjamin Mako Hill. 2019. Technological Frames and User Innovation: Exploring Technological Change in Community Moderation Teams. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 44 (Nov. 2019), 23 pages. <https://doi.org/10.1145/3359146>
- [24] Charles Kiene, Andrés Monroy-Hernández, and Benjamin Mako Hill. 2016. *Surviving an "Eternal September": How an Online Community Managed a Surge of Newcomers*. Association for Computing Machinery, New York, NY, USA, 1152–1156. <https://doi.org/10.1145/2858036.2858356>
- [25] Charles Kiene, Kenny Shores, Eshwar Chandrasekharan, Shagun Jhaver, Jialun "Aaron" Jiang, Brianna Dym, Joseph Seering, Sarah Gilbert, Kat Lo, Donghee Yvette Wohn, and Bryan Dosono. 2019. Volunteer Work: Mapping the Future of Moderation Research. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing* (Austin, TX, USA) (CSCW ’19). Association for Computing Machinery, New York, NY, USA, 492–497. <https://doi.org/10.1145/3311957.3359443>
- [26] Elizabeth Anne Kinsella. 2007. Technical rationality in Schön’s reflective practice: dichotomous or non-dualistic epistemological position. *Nursing Philosophy* 8, 2 (2007), 102–113.

- [27] U-M Klemola and Leena Norros. 1997. Analysis of the clinical behaviour of anaesthetists: recognition of uncertainty as a basis for practice. *Medical education* 31, 6 (1997), 449–456.
- [28] Ulla-Maija Klemola and Leena Norros. 2001. Practice-based criteria for assessing anaesthetists' habits of action: outline for a reflexive turn in practice. *Medical Education* 35, 5 (2001), 455–464.
- [29] Fred AJ Korthagen and Theo Wubbels. 1991. Characteristics of Reflective Practitioners: Towards an Operationalization of the Concept of Reflection. (1991).
- [30] Yong Ming Kow and Timothy Young. 2013. Media technologies and learning in the starcraft esports community. In *Proceedings of the 2013 conference on Computer supported cooperative work*. 387–398.
- [31] Zhiyuan Lin, Niloufar Salehi, Bowen Yao, Yiqi Chen, and Michael Bernstein. 2017. Better when it was smaller? community content and behavior after massive growth. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 11.
- [32] Claudia Claudia Wai Yu Lo. 2018. *When all you have is a banhammer: the social and communicative work of Volunteer moderators*. Ph.D. Dissertation. Massachusetts Institute of Technology.
- [33] Silvia Mamede and Henk G Schmidt. 2004. The structure of reflective practice in medicine. *Medical education* 38, 12 (2004), 1302–1308.
- [34] Karen Mann, Jill Gordon, and Anna MacLeod. 2009. Reflection and reflective practice in health professions education: a systematic review. *Advances in health sciences education* 14, 4 (2009), 595–621.
- [35] Gabriela Marcu, Anind K Dey, Sara Kiesler, and Madhu Reddy. 2016. Time to reflect: Supporting health services over time by focusing on collaborative reflection. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. 954–964.
- [36] Jennifer Marlow and Laura Dabbish. 2014. From rookie to all-star: professional development in a graphic design social networking site. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. 922–933.
- [37] J Nathan Matias. 2019. The civic labor of volunteer moderators online. *Social Media+ Society* 5, 2 (2019), 2056305119836778.
- [38] Colten Meisner and Andrew M Ledbetter. 2020. Participatory branding on social media: The affordances of live streaming for creative labor. *New Media & Society* (2020), 1461444820972392.
- [39] Teodora Mihailova. 2020. Navigating ambiguous negativity: A case study of Twitch. tv live chats. *New Media & Society* (2020), 1461444820978999.
- [40] Casey Newton. 2019. *The Trauma Floor: The secret lives of Facebook moderators in America*. Retrieved July 5, 2021 from <https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona>
- [41] Anthony J. Pellicone and June Ahn. 2017. The Game of Performing Play: Understanding Streaming as Cultural Production. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (CHI '17). Association for Computing Machinery, New York, NY, USA, 4863–4874. <https://doi.org/10.1145/3025453.3025854>
- [42] Andrew Phelps and Mia Consalvo. 2020. Laboring Artists: Art Streaming on the Videogame Platform Twitch. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*.
- [43] Hazel Platzer, David Blake, and Dorothy Ashford. 2000. Barriers to learning from reflection: a study of the use of groupwork with post-registration nurses. *Journal of advanced nursing* 31, 5 (2000), 1001–1008.
- [44] Sarah T Roberts. 2018. Digital detritus: 'Error' and the logic of opacity in social media content moderation. *First Monday* (2018).
- [45] Daniela K Rosner. 2012. The material practices of collaboration. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*. 1155–1164.
- [46] Bonnie Ruberg and Amanda L.L. Cullen. 2020. Feeling for an audience: the gendered emotional labor of video game live streaming. *Digital Culture and Society* 5, 2 (2020), 133–148.
- [47] Minna Ruckenstein and Linda Lisa Maria Turunen. 2020. Re-humanizing the platform: Content moderators and the logic of care. *new media & society* 22, 6 (2020), 1026–1042.
- [48] Donald A Schon. 1983. *Reflective practitioner*. Vol. 5126. Basic books.
- [49] Donald A Schön. 1987. *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. Jossey-Bass.
- [50] Joseph Seering. 2019. Building More Positive Online Communities through Improving Moderation and Strengthening Social Identity. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing* (Austin, TX, USA) (CSCW '19). Association for Computing Machinery, New York, NY, USA, 89–93. <https://doi.org/10.1145/3311957.3361855>
- [51] Joseph Seering, Geoff Kaufman, and Stevie Chancellor. 2020. Metaphors in moderation. *New Media & Society* (2020), 1461444820964968.

- [52] 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* (Portland, Oregon, USA) (CSCW '17). Association for Computing Machinery, New York, NY, USA, 111–125. <https://doi.org/10.1145/2998181.2998277>
- [53] Joseph Seering, Michal Luria, Connie Ye, Geoff Kaufman, and Jessica Hammer. 2020. *It Takes a Village: Integrating an Adaptive Chatbot into an Online Gaming Community*. Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376708>
- [54] 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 (2019), 1417–1443. <https://doi.org/10.1177/1461444818821316>
- [55] Petr Slovák, Kael Rowan, Christopher Frauenberger, Ran Gilad-Bachrach, Mia Doces, Brian Smith, Rachel Kamb, and Geraldine Fitzpatrick. 2016. Scaffolding the scaffolding: Supporting children’s social-emotional learning at home. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. 1751–1765.
- [56] Susan Leigh Star. 1989. The structure of ill-structured solutions: Boundary objects and heterogeneous distributed problem solving. In *Distributed artificial intelligence*. Elsevier, 37–54.
- [57] TL Taylor. 2018. *Watch me play: Twitch and the rise of game live streaming*. Princeton University Press.
- [58] Twitch.tv. 2021. *Battling Ban Evasion with Machine Learning*. Retrieved January 15, 2022 from [https://safety.twitch.tv/s/safety-news/detail?language=en\\_US&post=Using-Machine-Learning-to-Curb-Ban-Evasion](https://safety.twitch.tv/s/safety-news/detail?language=en_US&post=Using-Machine-Learning-to-Curb-Ban-Evasion)
- [59] Twitch.tv. 2021. *Twitch Press Center: Facts and Figures*. Retrieved January 15, 2021 from <https://www.twitch.tv/p/press-center/>
- [60] Waltraut Von Klitzing. 1999. Evaluation of reflective learning in a psychodynamic group of nurses caring for terminally ill patients. *Journal of Advanced Nursing* 30, 5 (1999), 1213–1221.
- [61] Donghee Yvette 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 the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3290605.3300390>

## A STUDY MATERIALS

### A.1 Screener Survey

Below, we provide all measures included in the survey which either informed our selection of participants or provided background information used in discussions with interview participants.

#### A.1.1 Survey Landing Page.

We want to understand more about your viewing experience on Twitch and would love to connect you with one of our researchers for a 1-hour chat. We hope you will accept our invitation.

Sound interesting? To see if you qualify, fill out some additional information about your Twitch usage, demographics, and contact information. It should take you about 3 minutes to complete. If you’re selected, we’ll reach out to schedule the chat.

***Additional text on this page has been removed for submission***

You must be at least 18 years old and have a PC or Mac with a webcam and microphone to participate.

#### A.1.2 Screening Measures.

- What is your current age?
  - 13 - 17 [*Respondents choosing this option were screened out.*]
  - 18 - 25
  - 26 - 34
  - 35 - 44
  - 45 - 54
  - 55 and over
- How long have you been using Twitch?

- Less than 3 months
- 3 - 6 months
- 6 months to 1 year
- 1 -3 years
- More than 3 years
- Which of the following activities, if any, have you ever performed on Twitch? [Please select all that apply]
  - Chatted with other users on Twitch while viewing a stream
  - Broadcasted yourself on Twitch
  - Played Twitch the video game
  - Paid to subscribe to a channel on Twitch, or paid to renew a subscription
  - Used the bits / cheering feature
  - Chatted in a Twitch streamer’s Discord server
  - Moderated a channel [*Respondents who did not choose this option were screened out.*]
- On how many different channels have you acted as a moderator within the past 3 months?
  - 0 (I have not moderated any channels in the past 3 months) [*Respondents choosing this option were screened out.*]
  - 1
  - 2-5
  - 6-10
- Please select all of the channel types that you moderated within the past 3 months (please select any option that applies to one or more channel that you moderate):
  - Gaming (e.g. video games, tabletop, card games)
  - Creative (e.g. painting, cooking, drawing)
  - Performing Arts (e.g. music, dancing, singing)
  - Educational (e.g. science & technology, coding, travel)
  - Talk (e.g. ‘just chatting’, IRL, talk shows)
  - Other [open-text]

### A.1.3 Primary Measures.

- Please provide the name of the channel on which you have spent the most time as a moderator over the past 3 months (please provide the lowercase name of the channel, or copy the URL for the channel page).
- When did you first become a moderator on this channel? Please choose the answer that best applies:
  - Within the past month
  - Within the past 3 months
  - Within the past year
  - More than one year ago
- How would you characterize the content of this channel? (Please choose the option that best applies, or provide your own):
  - Gaming (e.g. video games, tabletop, card games)
  - Creative (e.g. painting, cooking, drawing)
  - Performing Arts (e.g. music, dancing, singing)
  - Educational (e.g. science & technology, coding, travel)
  - Talk (e.g. ‘just chatting’, IRL, talk shows)
  - Other [open-text]

- How would you best describe your role as a moderator within this channel? Please choose the answer that best applies (or provide your own, if none are a good match):
  - The ‘Helping Hand’: My role is primarily to help the streamer and the community accomplish its goals.
  - The ‘Justice Enforcer’: My role is primarily to maintain a safe environment and punish bad behavior.
  - The ‘Neutral Observer’: My role is primarily to observe, and only participate or step in when absolutely necessary.
  - The ‘Conversationalist’: My role is to actively engage and socialize with viewers to build a community.
  - Other (please provide your own):
- How many Twitch users (including yourself) have been actively engaged in moderating this channel over the past 3 months? Please choose the answer that best applies:
  - 1 (Only Me)
  - 2
  - 3 - 5
  - 6 - 10
  - 11 or more

#### A.1.4 Demographic Measures.

- What is your gender? (You may select multiple responses):
  - Woman
  - Man
  - Non-Binary
  - Prefer not to disclose
  - Other [open-text]
- What is your race/ethnicity? (You may select multiple responses):
  - Hispanic or Latinx
  - Native American, Alaskan Native, First Peoples, or Aboriginal
  - Asian or Asian-American
  - Native Hawaiian or Pacific Islander
  - Black or African-American
  - White or Caucasian
  - Prefer not to disclose
  - Other [open-text]
- What is your sexual orientation? (You may select multiple responses)
  - Asexual
  - Bisexual
  - Lesbian
  - Gay
  - Pansexual
  - Queer
  - Heterosexual
  - Prefer not to disclose
  - Other [open-text]
- Which of the following best describes you?
  - Non-Gamer: I am not interested in playing video games
  - Casual Gamer: I dabble in video games, but in short sessions or infrequently

- Core Gamer: I regularly play video games, but I am not super-serious or competitive
- Hardcore Gamer: I play video games frequently, and I play seriously or competitively

## A.2 Semi-Structured Interview Protocol

Below, we provide a list of questions included in the interview protocol which were relevant to the present study. Interviews were approximately 60 minutes long, and the questions covered in these sections typically covered approximately 30 minutes of the interview. This list was intended to be exhaustive, including questions and follow-up questions which were not asked in every interview, designed to provide a set of branching possibilities on which the interviewer could draw, depending on the direction that the participant led the conversation.

### A.2.1 Overview [Target: 10 Minutes].

- When did you first become a part of <channel>?
- When did you first become a moderator for <channel>? How did it happen?
  - What motivated you to become a moderator in this channel?
- How would you describe your work as a moderator in this channel?
  - What are your goals as a moderator?
  - Have these changed over time?
- What does the larger moderation team for this channel look like?
  - How do you divide up tasks across the team?
  - How frequently do you interact with other mods and for what reasons?
  - How do you communicate with other mods? What platforms do you use?
- How did you learn how to be a moderator?
  - Did the streamer give you any instructions?
  - Did another moderator train you or offer advice?
  - Did you learn anything from Twitch resources? Third-party resources?
  - Do you belong to any groups/communities for moderators?

### A.2.2 Standards, Communication, & Enforcement [Target: 20 Minutes].

- How would you describe the chat experience on <channel>?
  - How would you describe the standards for acceptable behavior?
  - Have these changed over time? In what ways?
  - Have any particular situations prompted changes to these standards?
  - In what ways do these standards differ from the overall Twitch Community Guidelines? For example, what would you prohibit that Community Guidelines might allow?
- How do you (and the other moderators) decide what is acceptable behavior within this channel?
  - Describe a situation where you’ve needed to discuss this “on the fly”? What happened and how did you resolve it?
  - How and when is the streamer involved in decisions about what is acceptable in the channel?
- I’m now going to ask you to briefly explain when you implement certain mod actions:
  - What kinds of viewer actions would lead you to talk to them or warn them in chat?
  - What kinds of content would you typically delete or remove from chat?
  - What are some reasons why you might put a user on timeout?
  - What actions merit a ban in your channel? Temporary vs. Permanent?
  - What would prompt you to submit a report to Twitch?

- How do you intervene when negative interactions happen on other platforms that support the community, such as Discord?
- What information about viewers is helpful when deciding how to react to a negative interaction?
  - What information about viewers would be helpful that's not currently accessible to you?
- How do you work with other moderators to decide how a situation should be handled?
  - How would you resolve a disagreement about how a situation should be handled?
  - How and when is the streamer involved in these decisions?