"I Have Abused Someone Who Abused Me": Understanding People Who Have Experienced Both Sides of Harassment Accusations in Social VR

GUO FREEMAN, Clemson University, USA
LINGYUAN LI, The University of Texas at Austin, USA
KELSEA SCHULENBERG, Clemson University, USA

Social VR's focus on embodied and immersive experiences has led to intensified and more physicalized forms of harassment than other online contexts. Therefore, a growing body of HCI and CSCW work has explored multiple strategies and mechanisms to prevent and mitigate harassment risks in social VR. However, existing works have also highlighted a fundamental challenge in mitigating harassment in social VR – the apparent lack of consensus among social VR users on how to explicitly define harassment and what behaviors should be considered harassing in social VR. In this work, we aim to offer new knowledge on the uncertainty about how harassment is defined and perceived in social VR, particularly by learning from social VR users who have experienced *both sides of harassment accusations*. Based on interviews with 12 participants with diverse identities who have both been harassed by others and been accused of harassing others in social VR, we unpack how people justify and reflect on their behavior given their prior experiences of both being victims of harassment and being called a harasser. We thus offer unique insights into the complexity of harassment in social VR by highlighting cases of "gray areas" and critical ethical implications in such harassment accusations, which are understudied in the existing literature. We also propose two high-level design principles for new strategies and approaches to foster safe social VR spaces based on people's unique experiences of both sides of harassment accusations in social VR.

CCS Concepts: • **Human-centered computing → Empirical studies in collaborative and social computing**.

Additional Key Words and Phrases: online harassment, harassment accusations, social virtual reality

ACM Reference Format:
Guo Freeman, Lingyuan Li, and Kelsea Schulenberg. 2025. "I Have Abused Someone Who Abused Me": Understanding People Who Have Experienced Both Sides of Harassment Accusations in Social VR. *Proc. ACM Hum.-Comput. Interact.* , CSCW (2025), 26 pages.

1 INTRODUCTION

More recently, the growing popularity of social virtual reality (VR) has given rise to concerns about how social VR spaces may amplify and extend online harm and how researchers and developers can work to mitigate said harm. In social VR, multiple users can interact with one another through VR head-mounted displays in 3D virtual spaces [29, 53] while also leveraging other technological features (e.g., partial-to-fully body-tracked avatars) to simulate offline-like social interactions (e.g., touching and grabbing others) [63–65]. These unique features thus allow users to socialize in more **embodied** (i.e., experiencing a virtual body representation as one's own [68]) and **immersive** (i.e.,

Authors' addresses: Guo Freeman, Clemson University, Clemson, South Carolina, USA, 29634, guof@clemson.edu; Lingyuan Li, lingyuan.li@ischool.utexas.edu, The University of Texas at Austin, USA; Kelsea Schulenberg, Clemson University, South Carolina, USA, 29634, kelseas@clemson.edu.

2573-0142/2025/-ART
https://doi.org/

being enveloped by, included in, and interacting with the virtual environment [78]) ways than in other, screen-mediated online social spaces (e.g., social media and gaming). This focus on embodied and immersive experiences has also led to intensified and more physicalized forms of harassment in social VR compared to other online contexts, ranging from trash-talking women, drawing penises, and virtual "groping" to the most recent "rape" in the metaverse [6, 7, 23, 24, 32, 55, 60, 63–65, 67, 69, 70, 80].

Realizing the urgent need to address such harm, an emerging body of HCI and CSCW works has begun to identify various strategies and mechanisms to prevent and mitigate harassment risks in social VR, especially from the perspectives of the diverse parties being impacted by such incidents, including *victims, moderators,* and *bystanders* [6, 7, 23, 24, 32, 60, 64, 67, 80]. However, these works have also collectively highlighted that a fundamental challenge in addressing harassment risks in social VR is an apparent lack of consensus among social VR users on what social activities and behaviors should be considered harassing. This thus creates barriers to effectively defining and identifying harassment in social VR, as diverse individuals and communities may have different understandings and definitions of harassment.

Indeed, a common understanding is that harassment in the offline world is closely related to unwelcome conduct targeting one or more characteristics of an individual's identity, including but not limited to race, gender, religion, sexuality, national origin, age, disability, or genetic information [20]. Yet, harassment is often considered a culturally contextualized construct and can be defined differently across various cultures [81]. As a result, what might feel harassing to one person might not feel harassing to another when it happens to them in social VR [6, 32, 60, 64, 65]. For example, Maloney et al.'s work has noted that while children social VR users consider their curiosity-driven behaviors adventurous and playful, adult users view such behaviors as annoying, disturbing, or even harassing [50, 51]. Schulenberg et al.'s research further reveals the complex power dynamics in social VR, which affect how harassment is defined and perceived in different contexts [63]. One particularly significant example they reported is that some women social VR users would harass their harassers back. These women do not perceive their harassing behavior as harassment *per se* but a well-justified "an eye for an eye" approach. In doing so, they not only protect themselves in a cisnormative, male-dominated social VR space but also take revenge and punish their harassers by making them experience the suffering of their women victims [63].

In recognition of these uncertainties and challenges surrounding how harassment is defined and perceived in social VR, we argue for the critical need to delve into more nuanced cases and experiences of harassment accusations in social VR, such as those that do not represent easily defined bad behaviors or straightforward instances of misconduct. Therefore, in this work, we focus on how social VR users who have experienced **both sides of harassment accusations** justify or reflect on their behavior given their prior experiences. We believe that this new and understudied lens is valuable for two reasons. First, the social stigma surrounding the act of acknowledging and reflecting upon harassing behaviors often significantly hinders researchers' abilities to study why an individual is accused of harassing others. This stigma also affects understanding their own perspectives and reactions after being called a harasser [40, 42, 44, 52]. As such, learning from social VR users who have experienced both sides of harassment accusations may help overcome this social stigma because these individuals have not only been called a harasser but also been a *victim of harassment* in social VR. This dual role, in turn, may motivate them to share their unique experiences of and reactions to harassment accusations in social VR. Second, this lens would provide new and nuanced insights into cases of "gray areas" in harassment in social VR, which is understudied in existing literature. For example, it could reveal how accusations of harassment can be weaponized against marginalized social VR users rather than protecting them. This may lead to

critical ethical reflections and precautions regarding whether labeling and punishing harassers is the most effective way to address emerging harassment risks in social VR.

Based on interviews with 12 participants with diverse identities who have been both harassed by others and accused of harassment (e.g., have been called a harasser by others) in social VR, we specifically address the following two research questions:

- **RQ1: How do social VR users who have experienced both sides of harassment accusations explain such accusations against them?**
- **RQ2: How do these users react after they are accused of harassment in social VR?**

It is important to note that the main goal of our study is not to help people justify their harassing behaviors but to unpack the intricacies behind their behaviors, especially given their prior experiences of both being a victim of harassment and being called a harasser. As such, this work should be approached as an empirical investigation to promote more in-depth and nuanced conversations and discussions regarding the uncertainty and complexity of harassment in social VR rather than oversimplifying this issue.

In doing so, our work contributes to existing HCI and CSCW research on online harassment and social VR in three ways. **First**, to the best of our knowledge, we offer the first empirical investigation into the urgent issue of social VR harassment by learning from social VR users who have experienced both sides of harassment accusations. By addressing RQ1, we explore what these users did to make others accuse them of harassment in social VR and, more critically, what motivated them to perform such behaviors. As these individuals have also been victims of harassment in social VR, how they explain their behavior and motivations provides new and unique insights into why harassment happens in social VR, which may go beyond the straightforward intention to upset people and cause harm, aggravation, anxiety, and instability as reported in prior work [32]. Through investigating RQ2, we uncover these individuals' reactions to such accusations and their own reflections after being called a harasser, especially given their prior experiences as victims. This, in turn, helps us further investigate the complex and nuanced nature of harassment in social VR by focusing on people's dual role in experiencing harassment. **Second**, a critical highlight of our study is that all participants are considered marginalized and are often targeted for harassment in social VR due to their varied genders, sexualities, and ethnicities. In fact, there are no cisgender straight white men or women involved in this study. While anyone can be a harasser, how and why these marginalized users have been accused of harassment provides unique and valuable insights into the critical ethical implications and dilemmas surrounding social VR harassment. **Third**, built upon our findings and critical reflections, we also propose two high-level design principles for new strategies and approaches to foster safe social VR spaces based on marginalized users' unique experiences of both sides of harassment accusations in social VR, which go beyond just detecting and punishing easily defined bad behaviors or straightforward instances of misconduct.

2 RELATED WORKS

2.1 Existing Efforts to Understand Various Motivations and Contexts Behind Harassment in Online Social Spaces

Harassment is considered a severe issue in online social spaces that seriously damages victims' mental, emotional, and physical well-being [5, 33, 76]. Therefore, in hopes of designing and developing actionable harassment prevention and mitigation strategies, extensive HCI and CSCW works have explored how victims or targets of online harassment experience, manage and cope with the harassment incidents they encounter in diverse online contexts. For example, prior works have detailed women's experiences of marginalization and varying forms of harassment against them in online gaming and virtual worlds [3, 9, 25, 34, 47, 48, 56, 72, 73], women and LGBTQ live

streamers' strategies to handle online negativity [12, 31, 74], various methods for online users to classify and combat online harassment in social media and text-based online forums [16, 17, 21, 37–39, 57, 75, 76], and content moderation as a community effort to mitigate harassment in various online social spaces [13, 14, 19, 41, 59, 66, 71, 79].

However, what is often lacking in the existing research discourse on online harassment is an in-depth understanding of why people harass others online and how they explain or justify their behaviors after they are accused of being online harassers. Such an understanding is crucial for providing a more comprehensive image of said harassment by taking various reasons, motivations, and contexts behind harassment into account, which will help develop more nuanced harassment mitigation solutions. Yet, a major challenge to recruit and study why an individual is accused of harassing others and their self-reflections is the apparent unwillingness of people to honestly admit being accused of harassment and the social stigma surrounding harassment [40, 42, 44, 52]. Acknowledging this challenge, a growing body of HCI and CSCW works has begun to investigate "why everyday internet users participate in abusive behaviors online" [8, 40, 42–44, 46, 52, 62]. Indeed, these works have further emphasized the complexity of online harassment, which goes beyond simply bad and problematic behaviors or straightforward instances of misconduct.

Among them, Jhaver et al. point out the challenge to identify what constitutes harassment on social media platforms, as many people who were blocked as harassers on Twitter felt that they were blocked unnecessarily and unfairly [40]. Kim et al.'s interview study about "calling out" (i.e., publicly broadcasting online criticism of someone) on Twitter echoes this understanding and highlights the blurring boundary between online criticism and online harassment and the high interchangeability between callers and callees [42]. Focusing on unpacking why people engage in online harassment beyond simple bias or dislike [44, 52, 62], Lee et al.'s survey study reveals several psychological factors, rather than demographic factors, that are more associated with those who harass online than those who do not [44]. According to these works, while harassment is often deemed to be negative, damaging, and disruptive, there exist cases of "gray areas" in online harassment where either the harassment accusation is unfair, or the accused behavior may not be perceived as harassing in a specific context (e.g., where online criticism is the norm). Marwick thus proposes an explanatory model (i.e., Morally Motivated Networked Harassment) to highlight these various reasons, motivations, and contexts to explain better why people participate in networked harassment, which specifically focuses on the complicated interplay of norm violations in an online community, networked audience' shared ideological or moral framework, and justification for harassment [52]. Within this framework, certain harassing behavior can be explained and justified because it aligns with the shared community norms and moralities or is motivated by moral and ethical considerations to protect the online community [52].

Additionally, some researchers warn that harassment accusations can even be **weaponized against marginalized online users** rather than creating safe online spaces for them [35, 36]. For example, Guynn's report reveals the alarming fact that Black online users are often punished when they post about the racism they have experienced [36]. In these cases, their posts are censored as hate speech and thus removed, and their accounts are suspended or banned [36]. Gray and Stein share similar concerns, suggesting that safety features and anti-harassment policies in online spaces often disproportionately punish minority online users, such as black women [35]. Black women are targeted, punished, and labeled as harassers for violating terms of service when they speak out about racist and sexist incidents both online and offline [35]. Likewise, Are investigates malicious flagging on Instagram and TikTok as a new online abuse technique, because people who are flagged often feel targeted by both the platforms and their audiences' retaliation [1]. Others also point out that although the global #Metoo movement has empowered women all over the world to share and discuss how they experience sexual harassment on social media platforms, minority

women (e.g., Asian women, trans women, and women of lower class) are often doubted about their credibility and intentions and even are accused of being a harasser themselves, which not only silence but also further marginalize them [18, 54].

Taken together, this small body of prior HCI and CSCW works highlights the potential risk of oversimplifying online harassment and calls for more research on explicating the complicated power dynamics and nuanced experiences of online harassment and harassment accusations. Grounded in these understandings, the following section details existing efforts and challenges to define and mitigate harassment in social VR, especially highlighting a lack of nuanced insights into why a social VR user might be accused of harassment for various reasons and how they react to such accusations.

2.2 Existing Efforts and Challenges to Define and Mitigate Harassment in Social VR Spaces

There is an emerging research agenda in HCI and CSCW that focuses on understanding and mitigating new forms of harassment in novel social VR spaces, especially through (1) *victims' efforts*; (2) *moderators' efforts*; and (3) *bystanders' efforts* [6, 7, 23, 24, 32, 60, 64, 67, 80].

Mitigating Harassment in Social VR Through Victims' Efforts. The majority of prior work on social VR harassment focuses on how different groups of social VR users, especially those who belong to marginalized technology user groups (e.g., women, LGBTQ individuals, and racial minorities), experience and cope with emerging harassment risks in social VR. These works have warned that from victims' perspectives, social VR's focus on embodiment, sense of presence, body tracking, and synchronous voice conversation may allow people to verbally assault and virtually "touch" (e.g., grabbing and groping) others without their permission [6, 7, 32], the latter of which seems to simulate types of physical harassment and assault that often happen in the offline world [6, 32]. As a result, for victims, harassment in social VR may be felt as more realistic and disruptive compared to harassment in traditional online gaming and virtual worlds [32].

These works have thus collectively highlighted that social VR users often depend on two types of approaches to mitigate potential harassment. First, they have to actively use platform-specific safety tools such as muting (i.e., muting a harasser's voice), blocking (i.e., completely removing a harasser from the victim's surroundings), and a personal space bubble (i.e., preventing a harasser from getting too close to the victim physically) to protect themselves when they encounter harassment [32]. Second, they may even have to develop their own personal techniques to deal with harassment in social VR, including avoiding sharing personal information, leaving a world, room, or situation where they might encounter harassment, crafting avatar design and voice to conceal their identities; and performing personal resilience, among others [27, 29, 30, 32, 63]. Employing these existing harassment mitigation tools and strategies, unfortunately, requires a massive effort from the potential victims to develop sufficient knowledge of what safety features or personal strategies they can use, how and when to use them appropriately, and how to prove that they are indeed harassed in social VR [6, 7, 32].

Mitigating Harassment in Social VR Through Moderators' Efforts. Prior work has also explored how to leverage content moderation, especially human-based moderation, as a main approach to monitor and mitigate harassment in social VR [22–24, 60, 64]. In this moderation model, a behavior or content creation in social VR first becomes flagged or noticed as inappropriate/harassing based on the platform's pre-defined community guidelines, and then the flag or notice is reviewed by a party (i.e., a professionally hired or contracted human moderator; or a voluntary, user-driven human moderator), and finally, punishment is executed by a party [64]. In particular, Sabri et al.'s interview study highlights that to address the ephemeral nature of harassment in social VR through limited platform design choices, moderators often need to rely on several personal

moderation strategies, such as informing users of the rules, strategically positioning their avatar to increase social VR users' awareness of their presence, actively identifying and helping those struggling, and especially paying attention to specific avatars [60]. Additionally, Schulenberg et al. and Fiani et al. investigated the perceived opportunities and limitations for traditional human-based moderation to address emergent harassment in social VR, and explored potential directions for designing AI-based moderation to enhance such opportunities and address limitations [22–24, 64]. Taken together, these works represent existing efforts to achieve more nuanced and effective moderation practices to better protect social VR users from harassment, such as through helping human moderators monitor and deal with harassment in the moment [32, 60, 64] or developing new AI-based moderation or AI moderators to relieve human moderators' emotional labor and/or subjective bias [22–24, 64].

Mitigating Harassment in Social VR Through Bystanders' Efforts. Lastly, some recent research (e.g., [80]) provides insights into potential harassment mitigation approaches from bystanders' perspectives in social VR. By analyzing user-generated YouTube videos of their social VR experiences, Zheng et al. explicated how bystanders (i.e., users who are present when harassers harass their victims) and spectators (i.e., viewers who watch the YouTube videos that document such incidents and post comments) reacted to harassment incidents in social VR. Specifically, their findings reveal that the presence of bystanders alone can place pressure on the harasser, which may indirectly prevent the harasser from causing more damage and thus reduce burdens on victims and moderators to deal with such harassment on their own [80]. However, they also found that bystanders tend to accept that harassment incidents are inevitable and prevalent in social VR. Therefore, bystanders often either observe and ignore such incidents and even laugh or leave the area where harassment happened without any intervention [80]. In this sense, although bystander intervention might be a potential new approach to mitigate harassment in social VR (e.g., scaring away a harasser) without creating an excessive burden on individual victims or moderators to react to such a harassment incident after it happens, how to implement this approach has not been extensively studied.

Overall, these existing works have made significant strides in exploring multiple strategies and mechanisms to mitigate the various types, nature, and harm of harassment in social VR from the perspectives of diverse parties directly impacted by such incidents. Nevertheless, these existing works also highlight that a fundamental challenge in mitigating harassment in social VR is a distinct lack of consensus among users on what constitutes socially appropriate behaviors versus harassing behaviors in social VR [6, 32, 60, 64, 65]. For instance, Blackwell et al. describe harassment in social VR as subjective and extremely personal, ranging from verbal attacks to violations of physical and personal space [6]. In Freeman et al.'s work, social VR users seem to view "any interaction or experience that intentionally upset them and cause harm, aggravation, anxiety, and instability" as harassment [32].

Indeed, social VR users have shown differing opinions on how they distinguish harassment from behavior that is just inappropriate or "fun/play" [6, 32]. In different contexts, what might feel harassing to one person might not feel harassing to another when it happens to them [6]. One example is the mismatch between children's and adults' expectations for appropriate social behaviors in social VR [50, 51]. As reported in Maloney et al.'s work, when children consider their curiosity-driven behaviors in social VR (e.g., running around and shouting) adventurous and playful, adult users view such behaviors as annoying or even harassing, which even drive them to leave the social VR platform [50, 51]. Additionally, the complex power dynamics embedded in social VR, which is still often considered a cisnormative and male-dominated online social space, creates situations where behavior that might seem inappropriate to a group of people can be interpreted as appropriate and well-justified by a different group of people due to unequal power dynamics

between these groups. Such inconsistency in perceptions of harassment due to power dynamics is evident in Schulenberg et al.'s research on women social VR users, who are already marginalized in various tech spaces [63]. Their work especially highlights that one main strategy women use to protect themselves from harassment in social VR is to harass their harasser right back [63]. For these women, this strategy demonstrates their personal resilience to the existing power dynamics in social VR that often marginalize and disempower women. In doing so, they both effectively "scare away" and enact revenge upon their harassers by making harassers experience the suffering of their women victims [63].

However, it still remains unclear why people harass others in social VR, and if there are more intricate motivations and contexts behind such behavior beyond simply straightforward bad behaviors or misconduct. Therefore, we believe that our work, which focuses on more nuanced cases and experiences of harassment accusations in social VR by learning from social VR users who have experienced *both sides of harassment accusations*, can offer a more comprehensive image of online harassment to build up new strategies and approaches to better foster safe social VR spaces for everyone in the future.

3 METHODS

3.1 Data Collection

Recruitment and Participants. This study was a part of a multi-year research project on social experiences in social VR. Since this broader project focuses on how diverse users experience social VR, especially those who are often considered marginalized in technology spaces, such as women and LGBTQ individuals, we posted recruitment messages on various popular online forums for social VR users and queer gamers (e.g., r/VRchat, r/Recroom, and r/gaymers in Reddit) to recruit participants engaged in various social VR platforms who had personally been harassed, witnessed someone else being harassed, or had been accused of harassing others in social VR and were willing to be interviewed. To include diverse voices and perspectives, we also made it clear in our recruitment messages that harassment could happen to anyone, so anyone who had had such experiences and wished to share would be eligible to participate. We then interviewed all willing participants in March and April of 2022 (N=39) as part of the broader project via text/voice chat over Discord or video chat over Zoom, depending on participants' modality preferences.

For this study, we used the interview data from all participants who admitted that they had been accused of harassing others (e.g., being called a harasser) in social VR (N=12) out of the 39 participants. Among the 12 participants, 5 were primarily VRChat users, 5 were AltspaceVR users, one mainly used Mozilla Hubs, and one did not mention a specific main platform. Table 1 summarizes participants' self-reported demographic information and social VR experiences. It should be noted that all 12 participants in this study are considered marginalized in social VR due to their varied genders, sexualities, and ethnicities. In fact, there are no cisgender straight white men or women involved in this study. As these participants are often targeted for harassment in social VR, how they experience both sides of harassment provides unique insights into the complex nature of harassment accusations in social VR.

Interviews. Before the interviews, we provided all participants with an informed consent document based on their written communication preferences. We did not collect names or identifiable information. Interview questions were crafted using dialogic techniques to encourage participants to engage deeply with their responses [77]. These questions, detailed further below, drew inspiration from prior literature on harassment in social VR [6, 7, 32] and social interaction dynamics in social VR [26, 28, 29, 49] as well as from our own prior experiences with social VR as both researchers and users.

ID	Gender	Age	Ethnicity	Sexuality	Location	Social VR Platforms Mainly Used	Social VR Experience
P1	Woman	25	Black	Lesbian	USA	VRChat	8 years
P2	Woman	24	Black	Straight	USA	VRChat	3 years
P3	Trans Woman	26	White	N/A	USA	AltspaceVR	3 years
P4	Genderqueer Feminine Presenting	N/A	Biracial White and Black	N/A	USA	VRChat	3 years
P5	Man	25	Black	Straight	USA	N/A	3 years
P6	Man	27	White	Gay	USA	AltspaceVR	3 years
P7	Man	29	Mixed Race	N/A	USA	AltspaceVR	2 years
P8	Woman	25	White	Pansexual	Germany	VRChat	4 years
P9	Woman	24	Black	Bisexual	N/A	VRChat	3 years
P10	Man	25	Black	Straight	N/A	Mozilla Hubs	3 years
P11	Man	30	Mixed Race	Queer	USA	AltspaceVR	2 years
P12	Man	30	Hispanic	Bisexual	USA	AltspaceVR	3 years

Table 1. Self-Reported Demographic Information and Social VR Experiences of Participants. **Note:** N/A – participants did not provide information.

Interviews began with introductions, basic demographic questions, and questions regarding their level of experience in social VR. Then, participants were asked questions about their experiences with harassment in social VR. It is important to note that we understand that harassment is a culturally sensitive construct and acknowledge the significant challenge to explicitly define harassment in social VR as highlighted in prior work [6, 32, 60, 64, 65]. Therefore, in the interviews, we did not offer a specific definition of harassment but encouraged our participants to freely recount and share as much detail as they felt comfortable and appropriate regarding how they personally perceive and understand harassment in social VR. For example, participants were asked to describe how they defined and identified harassment (e.g., *"Please explain how you define harassment in social VR?"*) and describe a time they were harassed.

Interview questions specifically related to this particular study (see Appendix) focused on how participants explained their behaviors when they were accused of harassment in social VR (e.g., *"Can you describe a time when someone called you a harasser in social VR? In your opinion, why did other people accuse you?"*), their responses and self-reflections after they were accused of harassment in social VR (e.g., *"How did that experience make you feel about yourself? About social VR? Did you change your behavior because of this accusation? Why or why not?"*), others' reactions to their behaviors when they were accused of harassment (e.g., *"How did others nearby react to the situation, if at all?"*), and potential interventions to prevent their harassing behaviors from their own perspectives (e.g., *"If possible, what would you have done so others would not accuse you as a harasser?"*). Interviews lasted 102 minutes on average, and participants received a $50 Amazon digital gift card after they completed the interviews.

3.2 Data Analysis

After interviews were complete, recordings were transcribed for data analysis. We then adopted the thematic analysis approach [10, 11] to conduct an in-depth inductive qualitative analysis of the collected data. Following Braun and Clarke's [11] detailed guidelines for thematic analysis, we analyzed all collected interview data as described below.

(1) *Familiarizing ourselves with the data*: Two of the authors closely read through the participants' transcribed narratives line by line to identify information relevant to the research questions in this particular study by highlighting them and taking notes to gain a full picture of how participants

perceived their behaviors and reflections on their behaviors when they were accused of harassment in social VR [11].

(2) *Generating initial codes*: The same authors began an iterative coding process. They independently assigned preliminary codes to identified information. Then, the two authors combined the identified codes, eliminated redundant codes, and ensured that highlighted content only aligned with a single code. For example, the quote *"I can recall that vividly because I was trying to stand up for somebody, a minority as well. [...] At that point, most people there accused me of harassment, but I didn't see that as harassment. I just saw it as standing up for the minority"* was coded as "stand up for minority," "receiving criticism" and "accusations of harassment," which were then combined into "stand up for others."

(3) *Searching for themes*: These authors categorized codes into thematic topics related to our research questions and developed sub-themes based on how participants explained their perspectives and behaviors when accused of harassment in social VR. For example, codes on participants' reflections that they should have put themselves in others' shoes to understand better the consequences of their behaviors in social VR were categorized as revisiting, recollecting, and redefining.

(4) *Reviewing themes*: All authors continued to discuss, integrate, and refine themes and sub-themes to streamline participants' perceptions and reflections of their behaviors in social VR to best capture and represent our findings in relation to the research questions.

(5) *Defining and naming themes*: All authors collaborated to refine these themes further and name the final set of themes. At this stage, all authors considered themes across the entire data set and identified the "essence" of what each theme is about [11].

(6) *Producing the report*: All authors selected the most compelling quotes as examples and logically drafted the structure of the findings. This phase aimed to create a narrative structure where all findings flowed naturally and coherently [11].

3.3 Positionality Statement and Ethical Considerations

Due to the sensitive nature of our research (i.e., understanding more nuanced cases and experiences of harassment accusations in social VR), disclosing our positionality is crucial for acknowledging how our identities may influence this research and the analysis and interpretation of our data [2, 45, 61]. Our team includes three straight, cisgender women, two of whom are women of color. We all belong to marginalized communities in social VR (e.g., as women and minorities) and have extensive experience in social VR both as actual users and as researchers. Our own identities thus help us understand our marginalized participants' experiences of both sides of harassment accusations in social VR.

It is also critical to reflect on this kind of research's ethical and moral risks. We acknowledge that studying people who have been accused of harassment should be approached with extraordinary caution, as there are risks stemming from the amplification of people who have harassed others for various reasons. This is especially salient in a qualitative study that could include direct quotes that may privilege their voices and perspectives. On the one hand, since all our participants are marginalized social VR users, this work can help them have a voice and tell their stories, e.g., how harassment accusations can be weaponized to marginalize them in social VR further. In doing so, our work can seek to understand how marginalized users may act differently because of those stories [58]. However, on the other hand, our focus on these cases of "gray areas" in harassment accusations in social VR is not to justify or normalize toxic behavior in social VR but to encourage further ethical reflections on our everyday actions – "the need to consider exactly what we're putting into our networks - even when we're trying to help" [58].

With these understandings, we took additional precautions to protect the participants' identities and the collected interview data. The University's Institutional Review Board (IRB) approved this

study for research ethics before recruiting participants. All collected interview data (textual or audio) and transcriptions were stored in a password-protected main computer to which only the research team could access. All data were anonymized, and any identifiable information (e.g., usernames) was removed before data analysis.

4 FINDINGS

Similar to prior work that has highlighted the subjective and ambiguous nature of harassment in social VR from *victims'*, *moderators'*, and even *bystanders'* perspectives [6, 32, 60, 64, 65, 80], our participants who have experienced both sides of harassment accusations also exhibited diverse perceptions and understandings of what they personally defined as harassment in social VR environments. Overall, they tended to describe any behaviors or interactions that intentionally discomfort them (e.g., "*Basically for me, harassment in social VR would be some experience of ill-comforting situations in this VR room.*" - P1), go against their will (e.g., "*How I actually define harassment in social VR is when someone gets into my space, especially when I do not approve of it. Some things are being imposed on me when I do not want it.*" - P9), and cause interpersonal harm (e.g., "*anything that seeks to harm you or would cause any form of mental dissonance.*" - P7) as harassment in social VR.

To them, harassing behaviors in social VR broadly involve the following types of actions, including (1) actions that provoke aggravation, such as sexual innuendos (*probably using sexual innuendos or anything sexual, anything brutal, anything inappropriate*" - P12) and mockery ("*whenever people come up to me and start asking random questions but they're not interested in forming a friendship but make fun of you.*" - P4); (2) actions that create a sense of insecurity, such as cybersecurity concerns (e.g., "*Some people sent out some viruses that can make you not access something or can make you lose something. [...] It's misuse of social VR.*" - P2); and (3) actions that negatively impact people's mental health, such as cyberbullying ("*name calling, using abusive words, rejecting a participant from participating just for no good reasons*" - P10) and verbal abuse ("*When being vocally abused in the VR, you tend to feel it more because it seems very real because of how the 3D effects and the rest of it makes it looks realistic and other stuff. "* - P5).

These diverse interpretations and perceptions of harassment further underscore the challenge of reaching a unanimous definition of harassment in social VR spaces. This complexity and uncertainty thus make our findings particularly valuable, especially in understanding how people who have experienced both sides of harassment accusations explain the motivations behind their behaviors toward others (**section 4.1**) and their reactions and reflections (**section 4.2**) after being called a harasser in social VR. Table 2 summarizes our main findings.

4.1 How Social VR Users Who Have Experienced Both Sides of Harassment Accusations Explain the Accusations Against Them

Based on their own definitions of harassment detailed above, all participants admitted that they had been accused of harassing others in social VR. However, none of them described their behaviors as being intended to harm others. Rather, our participants highlighted three main reasons why they were accused of harassment in social VR from their own perspectives (**RQ1**), including: (1) because they stood up for others to protect others from being harassed (**Section 4.1.1**); (2) because they defended themselves and fought back against their own harassers (**Section 4.1.2**); (3) because they disagreed with others or took jokes too far (**Section 4.1.3**). For (1) and (2), participants noted the importance and necessity of their behavior as a response to harassers who targeted and attacked marginalized social VR users (either themselves or others).

Research Questions	Key Findings	Example Quotes
RQ1: How do social VR users who have experienced both sides of harassment accusations explain such accusation against them?	• Standing up to protect other marginalized social VR users from being harassed	*"I was trying to stand up for somebody, a minority as well. He was black, so he got criticized because of being black. His avatar was Black as well. So people made some racist comments against him. So I had to stand up for him. I didn't see that as harassment. I just saw it as standing up for the minority."* (P6)
	• Defending themselves and fighting back against their harassers.	*"I just do it as self defense, the way to piss people off. So at that point I will start rude and I was verbally harassing people with my use of words. But I only do that when I'm pissed off."* (P5)
	• Disagreeing with others or taking jokes too far.	*"We had a bone of contention about something that we were actually talking about, and I used a couple of abusive words. And I actually did apologize, but it did not take away the fact that I actually harassed someone on VR."* (P1)
RQ2: How do these users react after they are accused of harassment in social VR?	• Reacting to harassment accusations the same way as how they would react to harassment incidents against them	*"I just stop the game immediately because I do not want to hurt myself, because sometimes it makes me cry."* (P9)
	• Feeling hurt and disappointed but still adhering to their own agenda	*"I thought of changing my behavior, but at some point I had to encourage myself that I don't have to change because of other people's opinion about me. I felt I did the right thing."* (P7)
	• Revisiting their own understandings of harassment in social VR	*"I didn't see it as anything like harassment until when something similar was done to me. So, then I could recollect what I said, what I did, and it made me feel bad about myself."* (P11)

Table 2. Summary of Key Findings

4.1.1 Standing Up to Protect Other Marginalized Social VR Users from Being Harassed. Given their own prior experience of being a victim of harassment in social VR, all of our participants acknowledged the seemingly pervasive and severe nature of harassing incidents in social VR, especially those targeting marginalized users such as women and ethnic minorities. Therefore, several participants explained that they were accused of harassment in social VR not because they were hurting others but for the opposite reason: they were just trying to protect others, especially other marginalized social VR users, from being harassed. For example, both P3 and P6 shared their experiences,

> "It was when I stood up for someone, I tried defending someone who was literally sexually harassed in social VR. I had to stand up for her and tell the harasser this shouldn't be done. It shouldn't be like this. And that was how it all went. He called me a harasser but I just couldn't stand him harassing her that way." (P3)

> "I can recall that [being called a harasser] vividly because I was trying to stand up for somebody, a minority as well. He was black, so he got criticized because of being black. His avatar was Black as well. So people made some racist comments against him. So I had to stand up for him. At that point, most people there accused me of harassment, but I didn't see that as harassment. I just saw it as standing up for the minority." (P6)

Despite being accused of harassing other social VR users, it is clear that neither of these participants viewed themselves as harassers. Rather, they considered themselves "protectors" who stood up for actual victims of harassment and who actively intervened in incidents specifically targeting marginalized social VR users (e.g., sexual harassment towards women users and racist attacks towards people who use black avatars). According to P3 and P6, while their behavior could be viewed as harassment towards a given user (e.g., a social VR user who was sexually harassing a woman

user), they believed that it served as an intervention (i.e., stopping ongoing sexual harassment) by protecting another user.

These participants' experiences thus highlight the complexity involved in understanding harassers in the social VR context. First, these participants raise questions about whether specific motivations behind harassing behaviors and consequences of such behaviors can be used to justify harassment. In P3's example, her behavior was for *"defending someone who was literally sexually harassed in social VR."* According to P6, he *"just saw it as standing up for the minority."* Second, they also raised concerns about whether it is appropriate to use a "consensus" among social VR users to identify and define harassment in social VR. Indeed, as mentioned in section 2.2, prior work on harassment in social VR has called for creating a consensus among users on what social norms/behaviors are harassing in these spaces [6, 7, 32, 60, 64, 65]. Yet, P6's story reveals that even though most people accused him of harassment, he still did not consider his behavior harassing but rather a proactive action to protect marginalized communities in specific social VR contexts.

4.1.2 Defending Themselves and Fighting Back against Their Harassers. Our participants also felt that they were accused of harassment in social VR because other users made biased assumptions about them based on how they behaved or how their avatars looked. This spurred them to defend themselves and fight back against these biases. Several participants explained such bias in social VR:

> "It's linked to my heart. I always want to help others with my kindness. And also it's because of my expertise in social VR and my knowledge of how social VR works. So people think that I want to take advantage of them or I want to use the thing that I know much to cheat them because they know less. So they call me a harasser." (P2)
>
> "I think it [the accusation of harassment] was aiming to particulars of my avatar. Kinda like 'you look like a harasser.' " (P3)
>
> "Whenever that [accusing me of harassment] happened, I was in a different avatar. It was a bit more lewd. It didn't have necessarily regular pants on, but the face and stuff wasn't lewd. It was a pretty tame-like avatar. I feel like that people accuse me of harassment based on how I look and sound a certain type of way in social VR." (P4)

As social VR is still a relatively novel online social space, P2 understood that many people might have been unaware or still learning how it works. In contrast, she was already an expert with *"knowledge of how social VR works."* However, her intention to help others with her *"kindness"* and her *"expertise in social VR"* seemed to be viewed as trying to take advantage of or cheat others. In her opinion, this mismatch between how she viewed her behavior and how others interpreted her intentions made others accuse her of being harassing. Likewise, P3 and P4 both highlighted potential biases towards them based on their avatars. Their accounts are interesting because they seem to argue that they were accused of harassment mainly because of how they *"look and sound a certain type of way in social VR."* In P4's case, they had not been accused of harassment until they used *"a different avatar"* that *"was a bit more lewd."* P3 echoed this view and mentioned that others accused her of harassment because certain *"particulars"* of her avatar made her *"look like a harasser."* According to them, there seem to be established stereotypes or expectations about the appearance and voice of a harasser in social VR that make other users weary of interaction with them.

Therefore, participants acknowledged that they would indeed conduct harassing behaviors in response to such bias or unfair treatment from others, but only to defend themselves or as a way to punish other people who attacked them first. In this sense, participants considered their behavior a form of personal resilience in response to their harassers rather than intentional harm to others. P5 explained this process,

> "When I heard comments from people, I don't take it light with them. So I tend to use so much verbal words with some people. But I just do it as self defense, the way to piss people off. So at the point I will start rude and I was verbally harassing people with my use of words. But I only do that when I'm pissed off and I'm trying to make sure the third party or the people involved tend to feel the same way I'm feeling. When I do such, I feel I'm only embarrassing someone who actually did it at first, someone who attacked me at first. So I see it as means of defense. I don't attack people online. I only attack those who tend to bring up comments and then I blow out and use it as a way of defense. So I usually feel like I'm not doing anything bad. I'm not trying to play like I'm the weak one. I'm just confronting people and fighting back."

P5 admitted that his behavior was *"rude"* and he *"was verbally harassing people."* However, he viewed his reaction as a necessary self-defense mechanism because the other party attacked him **first**. P5 thus endeavored to portray himself as not *"the weak one"* to discourage others from committing more attacks in the future, which made him feel that he was *"not doing anything bad."* P9 echoed a similar sentiment,

> "I know I've actually abused someone who abused me. So I gave it in return. The moment when somebody actually insulted me, I actually gave it back to the person. I insulted the person too, though I actually felt bad for my actions. I only did it because I was angry and I felt harassed, so that was actually why I revenged."

P9 indeed *"abused"* and *"insulted"* other social VR users, which is typically viewed as online harassment. Yet, she considered it her revenge for how she was treated first (i.e., others abused or insulted her first). Compared to P5, who felt nothing was wrong with his behaviors, P9 *"actually felt bad"* for her actions. However, like P5, P9 also viewed her actions as inevitable steps to protect herself after being harassed by others (e.g., *"I only did it because I was angry and I felt harassed"*).

4.1.3 Disagreeing with Others or Taking Jokes Too Far. Instead of describing their behaviors as protection/intervention or self-defense mechanisms, some participants did admit that they harassed others in social VR because they disagreed with others, were not in a good mood, or took jokes too far without considering others' feelings. For example, P1 and P10 highlighted,

> "It's actually hard for me to say that [...] but that's the truth. I harassed someone when we were actually in a VR event. And we had a bone of contention about something that we were actually talking about, and I used a couple of abusive words. And I actually did apologize, but it did not take away the fact that I actually harassed someone on VR." (P1)

> "At the time we were playing a competitive game in social VR. I and my team actually won. But someone close to me accused me of harassing another player. He came back to me and was like, 'Hey, that game wasn't a fair one.' He was accusing me of name-calling the other player to make him feel soft. He was accusing me of pushing the other player to do things he wasn't supposed to do. He was accusing me of confusing the other player to take advantage of the other player so I could win." (P10)

According to these participants, accusations of harassment often emerge when people are arguing with each other in social VR. For P1, such arguments sometimes lead to *"a bone of contention"* while attending a social VR event. In her case, this triggered her use of abusive language towards others. She also admitted to being well aware that her behaviors should be considered harassment, regardless of the context. P10's case adds that competitive online contexts (e.g., VR gaming) may further escalate disagreements and tensions between social VR users when winning and losing are involved. As P10 explained, he was accused of harassing other players in social VR mainly

due to the dispute about whether he won the game fairly and ethically. While P10 considered his gameplay a fair competition, others viewed his strategies (e.g., *"make him feel soft," "pushing the other player," and "confusing the other player"*) as intentional harassment against the other player for the sole purpose of winning the game.

Additionally, our participants admitted that they might harass other social VR users to relieve negative emotions, especially when they are not in a good mood or are more agitated. P8 described,

> "I did unintentionally harass someone in public in social VR. I can vaguely remember not being in the greatest moods that time. So I commented fairly negatively towards an avatar someone was using, calling it poorly made or unoptimised. It does bother me a lot since because of everyone being able to make any kind of avatar they want, this also means it can be an avatar that would take a powerful computer to a crawl. Usually I let it go but that day I was not in a good mood and I got a bit annoyed about that."

The avatar P8 encountered in social VR is the so-called "crashers." Crashers are social VR users who use various technological capabilities, typically through the manipulation of customized avatars, to disturb other users through visual (e.g., flashing lights), auditory (e.g., blaring music), and/or technical (e.g., flying particles to slow down computers) ways. While P8 overall tended to be fairly tolerant of crashers (e.g., *"usually I let it go"*), in that specific moment, she was annoyed and became confrontational and even harassing. However, although P8 acknowledged the harassing nature of her behavior, she considered it *"unintentional"* as she did not intend to hurt others. Still, she was *"not being in the greatest moods."*

P5 also further emphasized this "unintentional" aspect of why people can sometimes be accused of harassment in social VR,

> "I feel like sometimes I take jokes too far. A few times when my friends and I were hanging out someplace. We meet some people and we were all joking, laughing, and having fun. And then maybe if I keep going too far with a joke or something like that, I've been called out for that before."

Participants like P8 and P5 agreed that their behaviors should be counted as harassment, and both took responsibility for their actions. In contrast to social VR users who justified their harassment as a defense or intervention mechanism to protect themselves or others, these participants openly acknowledged the negative consequences and damages of their harassing behaviors to others and even to the overall social VR environment. Yet, they also highlighted that it can be difficult for them to realize when and to what degree their behavior would be considered harassment towards others when they themselves do not intend to be harassing. As P5 revealed, he viewed himself as just *"joking, laughing, and having fun"* until other people started to accuse him of harassment because he took jokes too far.

4.2 How Social VR Users Who Have Experienced Both Sides of Harassment Accusations React to Such Accusation Against Them

In addition to uncovering our participants' own explanations for why they were accused of harassment in social VR, we also endeavor to reveal their reactions after being accused (**RQ2**). Our participants demonstrated three main reactions after being accused of harassing others, especially based on their own prior experience with being harassed in social VR.

First, as all participants have also been harassed by others in social VR before, they immediately used the same personal techniques that they used to deal with harassment incidents against them (i.e., when they are harassed by others) to react to harassment accusations against them (i.e., when they are called a harasser), such as quickly withdrawing from the situation where they were accused of harassment [32, 63] (**Section 4.2.1**). Second, after they withdrew from such situations,

participants who viewed their actions as defense or intervention mechanisms to protect themselves or other marginalized social VR users, often felt hurt, disappointed, and even betrayed. However, some still chose to adhere to their own agenda as they believed their behavior was justified and served as a form of personal resilience to combat harassment targeting marginalized communities in social VR (**Section 4.2.2**). Third, sometimes it was difficult for participants to realize why they were called a harasser based on their own understanding of harassment in social VR. As such, after they withdrew from such situations, they tended to revisit how they perceived the nature of their behavior and unpack when and why it became harassing and damaging to others (**Section 4.2.3**).

4.2.1 Reacting to Harassment Accusations the Same Way as How They Would React to Harassment Incidents against Them. Having experienced harassment in social VR themselves, participants have developed their own personal techniques to deal with harassment against them. These personal techniques include immediately leaving a world, room, or situation where they might encounter harassment and avoiding head-on confrontations, which have been reported in prior work on mitigating harassment in social VR as well [32, 63]. Given this, an important highlight in our findings is that participants used the same personal techniques to react to harassment accusations against them.

For example, our findings have shown that some participants viewed these accusations against them as deriving from personal disagreements (e.g., how people define winning strategies in competitive gaming) or bias towards them (e.g., automatically seeing them as harassers based on their avatar design). As a result, when they were accused of harassment in social VR, they withdrew themselves immediately from the situation to avoid potential confrontation and escalation of the conflicts. P9 described her experience,

> "If it [accusing me of harassing others] happened in a game, I just stop the game immediately because I do not want to hurt myself, because sometimes it makes me cry. I'm actually a very emotional person. So most times I just stay quiet. Sometimes I just try to cover my emotions and laugh about it."

In P9's opinion, the very reason why she was accused of harassing others lies in people's different approaches and perceptions of how a game should be played in social VR. For example, should enthusiastically or strategically trying to win a game be considered harassing in social VR? Therefore, instead of arguing with others to escalate the tension, she chose to *"stop the game immediately"* or *"stay quiet."* For her, this seems to be an appropriate approach to handle such accusations if they are based on personal disagreements – she would be able to leave the confrontational environment immediately, avoid more damage to herself and others, and have time to calm herself down (e.g., *"cover my emotions and laugh about it"*).

P10 also agreed, explaining that he would avoid direct confrontations to prevent the situation from becoming more harmful to himself and others,

> "We had different opinions, we argued, and we went our separate ways. It was a heated argument and I was accused of harassing others. I think that was just about a day or two and we came back together. It wasn't a big deal. I'm not going to lie, I don't think I would've done anything differently. Because I was actually being very competitive and I would do anything to win. I would use my full strength to win. So I don't think I would've done anything differently but I would've done more of what I did even do."

As previously discussed in section 4.1.3, P10 thought he was accused of harassing others because he played games competitively and *"would do anything to win."* Therefore, while he did not consider such accusations *"a big deal"* since they were just disagreements regarding gameplay, he avoided heated arguments with others and then came back later to smooth out these disagreements. It should

be noted that being accused of harassment did not motivate P10 to change his behaviorecause he disagreed with how others judged the nature of his behavior (i.e., being harassing vs. being competitive in gameplay), he not only decided not to do anything differently in the future but also *"would've done more of what [he] did even do."* However, in that moment when he was called a harasser, he still decided to back out of the situation to avoid immediate tensions and potential risks, regardless of whether or not he would continue his behavioral pattern in the future.

4.2.2 Feeling Hurt and Disappointed But Still Adhering to Their Own Agenda. After this first reaction (i.e., immediately withdrawing from the situation), some participants felt hurt, disappointed, and even betrayed in response to their accusations. This is especially true for participants who perceived their actions as self-defense or intervention mechanisms to stop and fight back against an ongoing harassment incident targeting themselves or other marginalized social VR users, or to prevent future harassment from happening, e.g., establishing themselves not as the "weak" ones. Most importantly, while these participants did acknowledge the harassing nature of their behaviors, they believed that they *helped* rather than harmed people in social VR by demonstrating personal resilience and by taking proactive actions to protect marginalized communities. P2 and P6 explained,

> "I felt so bad for myself. I felt so bad to be doubted. I was not trying to harass that person. And for the person, I also felt pity for him because while it's okay to be insecure, he didn't want to be helped. So I pitied the person for his lack of knowledge, especially in social VR." (P2)

> "I felt bad because what I do was right. My conscience were very clear that I did not have any bad intention. I was just standing up for the minority. Then, I felt bad because there ain't no good people in social VR. What we have are just a bunch of racists, which was bad." (P6)

P2's and P6's reactions to being called a harasser in social VR involve three levels. First, at the self-level, both P2 and P6 felt personally hurt because others seemed to doubt their intentions while they believed that what they did *"was right."* Second, at the interpersonal level, they expressed disappointment at how other social VR users misinterpreted their good intentions. For example, P2 *"pitied"* the social VR user who called her a harasser and rejected her help due to their *"lack of knowledge"* about social VR, and P6 emphasized his role as a protector of other marginalized social VR users rather than a harasser. Third, they even expressed concerns about the overall social VR atmosphere. As P6 mentioned, since he was accused of harassment while trying to protect marginalized social VR users from their harassers, he was worried that these accusations could be used to further silence and harm marginalized social VR users. In this sense, such accusations protect racists in social VR, not the victims (e.g., *"What we have are just a bunch of racists"*).

Built upon these feelings of hurt and disappointment, some participants decided to change their behaviors to avoid being accused of harassment again in social VR. In contrast, some others chose to stick to their personal agenda as they believed their behavior was justified. P2 and P7 continued to explain their different approaches,

> "I did [change my behavior]. I learned to mind my own business, do my things and just help people [...] You know, there's some people that it's difficult even to help them. So I will just help someone who understands that I am helping. I'm not trying to cheat you." (P2)

> "I didn't change my behavior though. I thought of changing my behavior, but at some point I had to encourage myself that I don't have to change because of other people's opinion about me. I felt I did the right thing." (P7)

As P2 described, after being accused of harassing others, she decided only to continue the same actions for people who would be on the same page as her (e.g., *"I will just help someone who understands that I am helping"*). She seemed to realize that what constitutes harassment can be interpreted in various ways by different social VR users (e.g., what she viewed as *"help"* was considered *"cheat"* by others). Therefore, she decided to take a more cautious approach and *"learned to mind my own business"* to avoid future accusations. In contrast, P7 decided to stick to his own agenda. Despite going through a similar experience and struggle as P2, P7 emphasized the justified nature of his behavior to protect marginalized social VR users (*"I felt I did the right thing"*) and encouraged himself not to change his behavior in the face of what he perceived to be peer pressure.

4.2.3 Revisiting Their Own Understandings of Harassment in Social VR. After participants withdrew from the situation where they were called a harasser, some also started to revisit the nature of their behavior and reflect upon why it was difficult to realize that their behavior could become harassing and damaging to others. For example, P4 and P10 shared that they felt *"bad and guilty"* after they were accused of harassing others - *"I just stop what I'm doing. I'm just like, 'Oh, I'm sorry. I didn't realize.' I don't mean to harass people if I kept going, but I always apologize"* (P4) and *"I felt very bad. Not as bad as being harassed myself though, but I felt bad and guilty"* (P10). Both participants also immediately apologized as a remedy.

Indeed, many participants highlighted that the very challenge lies in how they could arrive at the point where they could realize the damage of their behaviors to others. They explained that very often, only after others also harassed them did they start to reflect upon their own behaviors, put themselves in others' shoes, and actively revisit their own understanding of harassment in social VR given their experiences of being harassed by others. For instance,

> "Although someone called me a harasser in social VR, I really didn't see it as a big deal then until when I myself got harassed. So, when I was harassed it's like, I had a taste of my own medicine and I felt bad. I was able to put myself in the person's shoes and experience what the person did and trust me, it wasn't good. So, it made me feel bad that I could do such and say such kind of words to the person." (P7)

> "At that point, I felt I was probably just free, doing whatever I like, and that was what I did. Then I didn't see it as anything like harassment until when something similar was done to me. So, then I could recollect what I said, what I did, and it made me feel bad about myself." (P11)

These participants did not view their behaviors as *"anything like harassment"* or as *"a big deal"* when they were initially accused of harassing others, possibly due to how they understood harassment and social norms in social VR in their own ways. For P7 and P11, there seems to be an expectation that people should be able to conduct activities as they wish in a virtual world like social VR (e.g., *"I was probably just free, doing whatever I like"*). Therefore, they did not realize the mismatch between their expectations and others' expectations of how people should behave in social VR until they also suffered from the same experience themselves. As P7 confessed, *"I had a taste of my own medicine and I felt bad."* P11 also added that such experience was invaluable for reflecting upon his behavior and understanding others' perspectives. In this sense, regardless of whether they agreed with the accusation against them or not, as a response to such an allegation, they started to recollect what they did and acknowledge the importance of understanding the appropriate social dynamics in social VR not only from their own but also from others' views.

P12's description well summarizes this process of revisiting, recollecting, and redefining,

> "At first when I was accused, I really didn't feel anything. I didn't feel bad until it happened to me. Then I was able to feel something and I felt really bad. First of all,

> I felt anger. I felt shocked. Yeah, harassing someone is not really a good thing. But even when I later found out that they didn't take it that way, even when I found out that they weren't cool with it, it didn't really matter to me until when it happened to me. Yeah, until when I was treated the same way I treated others. That was when I could get a full impact of what I did and in my mind I regretted what I did. It's just automatically put this restriction in my head that probably don't do to others what you don't want on them to do to you. So I have like a mental restriction on what I say and how I say it, and also to whom I say it to."

According to P12, there seem to be three common stages associated with this revisiting process. First, they still felt confused as to why their behavior could be interpreted as harassing by others. Second, they fully experienced the damage of their behaviors to others and thus started to reflect upon what they have done (e.g., *"I could get a full impact of what I did and in my mind I regretted what I did"*), but only after they went through the same incident themselves as a harassment victim. Lastly, having experienced both sides of harassment , they began to better understand the diversity of perspectives in defining harassment in social VR. Undoubtedly, it is still challenging to bring everyone to the same page to establish a consensus on what should and should not be defined as harassment in social VR. However, as a result of this process of revisiting, recollecting, and redefining, participants like P12 can achieve a heightened self-awareness of what they say, how they say it, and to whom they would say it to in social VR (e.g., *"automatically put this restriction in my head"*).

5 DISCUSSION

Our findings have revealed how social VR users who have experienced both sides of harassment accusations explain the accusations against them (**RQ1**) as well as their reactions to such accusations (**RQ2**). In this section, we first discuss how these findings help us re-visit existing knowledge of emerging harassment risks in social VR through a unique and understudied lens, which offers new critical insights to unpack the complexity and uncertainty of understanding and mitigating harassment in emerging online social spaces. Grounded in these insights, we then explain two high-level implications for designing future interventions that take more complex and nuanced cases and experiences of harassment accusations in social VR into consideration, which go beyond just detecting and punishing easily defined bad behaviors or straightforward instances of misconduct.

5.1 Re-visiting Social VR Harassment by Learning from Both Sides of Harassment Accusations

As highlighted earlier in this paper, existing HCI literature has explored how diverse parties who are directly impacted by harassment in social VR (i.e.,*victims* [6, 7, 32], *moderators* [32, 60, 64] and *bystanders* [80]) understand and use different approaches to manage and mitigate such incidents. Taking these diverse parties' perspectives in sum, they all highlight a significant challenge in preventing and mitigating emergent harassment in social VR, namely, the absence of consensus among social VR users on what social norms/behaviors are considered harassing [6, 32, 60, 64, 65]. Participants in our study also expressed similar frustrations, which further confirms this absence of consensus. Indeed, our participants demonstrated diverse perceptions and understandings of how they defined harassment in social VR given their prior experiences of both sides of harassment accusations in social VR.

Learning from these participants' unique experiences, our work offers new in-depth insights to unpack the complexity of harassment in social VR by highlighting the understudied cases of "gray

areas" in such harassment accusations. In particular, we propose two critical reflections on this complexity.

First, **while our goal is never to help people justify their actions, it is still important to unpack specific motivations behind people's behaviors in social VR.** These understandings would be essential to explicate the multifaceted nature of harassment in the social VR context, including the roles of bystanders, the influence of stereotypes that lead to harassment, and the reciprocal harassment experiences. For example, many of our participants explained that their behaviors were unintentional and for reasonable purposes. They were especially motivated to take actions given their own prior experiences as a victim of harassment in social VR, and these specific motivations and the positive consequences of their actions should outweigh the damage of the potentially problematic execution (e.g., using abusive language or violence to scare away the harasser). As such, it is challenging for people to realize when and to what degree their behavior is considered harassment towards others, as they do not define their behavior as harassment but as protection for themselves or other marginalized individuals in social VR. As many participants reveal, they view their behavior as *"means of defense"*, and they will only abuse someone who abused them first.

Indeed, prior work has explored why everyday internet users participate in online harassment for various reasons, including the claim of harassing to get justice [4, 8, 40, 42, 44, 52, 62] and how some women social VR users would harass their harassers back as an effective harassment mitigate strategy [63]. Our research further highlights the complexity and ethical dilemma behind these more nuanced cases of harassment accusations in social VR. These findings are especially important considering the already subjective and personalized nature of harassment in social VR [6, 7, 32] and how these participants, who are also marginalized social VR users, may situate their interpretations of harassment accusations in their own prior experience of being harassed by others. For instance, should all interaction or experience that intentionally upsets any social VR user [32], regardless of the specific motivation or context, be labeled as harassment in social VR? Should behaviors that seem to be benign or positive still be folded under the general umbrella of "harassment" in social VR, a concept deemed to be negative? What is the fine line between well-intended interventions or acts of fragility to protect oneself and harassing behaviors, and how do we know certain behaviors "cross the line" in social VR? Although our study does not offer immediate answers to these questions, we point to the urgent need for unpacking specific motivations behind people's behaviors in social VR, given their own prior experiences, rather than oversimplifying and labeling them.

Second, **accusations of harassment can be weaponized against marginalized social VR users rather than protecting them.** Some of our participants considered themselves *bystanders* who stood up to protect others from being harassed, especially other marginalized users in social VR, such as racial minorities and women. However, they felt that they were called harassers in return, mainly due to the controversial methods they used to intervene (e.g., harassing the harasser). According to them, accusations of harassment can be intentionally weaponized to silence a marginalized individual for standing up for themselves or others. Similar to prior works that have warned about the use of harassment accusations to harm rather than protect victims in online social spaces [1, 18, 35, 36, 54], our participants also complain that such accusations protect and benefit actual harassers rather than victims in social VR – because victims and bystanders who would like to intervene will likely be punished (e.g., being called a harasser instead) for standing up to defend themselves or others.

Additionally, some participants felt that they were called a harasser only because others demonstrated particular discrimination or biases toward their avatar design (e.g., wearing specific outfits or displaying certain voice characteristics). For them, accusations of harassment may potentially become a new form of harassment, such as calling someone with certain stereotypes or identity traits

a harasser regardless of their actual behaviors. This could become even more risky for marginalized social VR users due to how their avatars may reveal certain gender, racial, or sexual identities. As our findings have shown, after being called a harasser, many participants seemed to naturally resort to the same personal techniques that they have used to deal with harassment against them in social VR. Some even expressed their fear of direct confrontations because that reaction may further harm them and/or others involved in the situation. In this sense, how they react to harassment accusations is somehow similar to how they react to harassment as a victim.

These observations lead to important questions regarding how, if at all, participants may view these harassment accusations as another form of potential harassment against them, given their prior experiences with harassment in social VR. Indeed, some participants admitted that their behaviors may unquestionably cross the line and should be considered harassment. However, it is essential to explore further how accusations of harassment can be abused and misused to harm certain populations, especially marginalized social VR users, and how to improve and refine future harassment mitigation systems in social VR to minimize severe and damaging risks of abusing such technologies.

5.2 Designing Safe Social VR Spaces Through Learning from Both Sides of Harassment Accusations

In light of the aforementioned findings and critical reflections stemming from our investigation into both sides of harassment accusations in social VR, two important design implications for new strategies and approaches to better foster safe social VR spaces for everyone emerge. Instead of focusing on proposing generic new design features, these implications should be viewed as higher-level design principles that can be applied to existing and future harassment mitigation approaches by taking the complexity and ethical dilemmas surrounding harassment accusations in social VR into consideration. It is important to note that these potential new directions do not aim to replace or contradict HCI and CSCW researchers' current efforts to design new safety features for harassment mitigation in social VR. Rather, we hope these implications may complement these efforts by highlighting a more comprehensive image of online harassment, which involves those nuanced cases and even "gray areas" of harassment in online social spaces.

Principle 1: New safety mechanisms beyond punishment should be designed to account for more nuanced cases or "gray areas" of harassment accusations in social VR. This principle is especially important based on the first critical reflection explained above: should marginalized social VR users who intend to stand up for themselves or others be punished if they harass their harasser back? Prior work has already shown that punishment itself, such as suspending or banning people's accounts, does not help mitigate harassment in online environments [43, 46]. In the context of our work, simply relying on existing traditional harassment mitigation methods, such as punishment, would also fail to reflect the complex nature of harassment in social VR. On the one hand, punishment may demotivate some of the seeming "harassing" behaviors reported in this study, such as taking proactive actions to protect oneself or others, which will further marginalize and de-power people already vulnerable in social VR spaces. On the other hand, it is equally dangerous to normalize toxic behavior in social VR. As we have mentioned in our **Ethical Considerations** about this study, each of our everyday actions has profound ethical implications and consequences, "even when we're trying to help" [58].

Therefore, there seems to be a crucial need for both sustaining social VR users' self-agency and personal resilience to actively foster a safe social VR environment for themselves and others while also ensuring that certain behaviors and reactions are not "crossing the line." Achieving this goal would require future design efforts to go beyond just designing and implementing punishment as

a primary harassment mitigation mechanism and investigate further these nuanced and unique experiences of harassment accusations in social VR. For example, it is essential to provide well-defined guidelines and community and bystander training on who should intervene (e.g., as a friend or as a stranger), when to intervene (e.g., during or after the incident), and how to intervene without escalating the intervention itself into another harassment incident (e.g., do not physically attack a harasser to stop their ongoing harassing behavior). In this sense, it might be beneficial to provide social VR users with an onboarding safety training process, which could focus on how to identify potential harassment incidents as a community and what appropriate intervention actions are in a given context could also be incorporated.

Principle 2: Designing to prevent personal abuses/misuses of harassment reporting in social VR. This principle directly stems from the second critical reflection: What if harassment accusations are abused to punish the victims who stand up for themselves or others, not the actual harassers? Reporting, or accusing someone of harassment, is still one of the main harassment mitigation mechanisms in most social VR spaces that is designed to give users personal agency [6, 7, 32, 63]. However, what has not been widely discussed in prior works is how reporting, if not used appropriately, can be weaponized to further marginalize and harm vulnerable social VR users by preventing them from taking actions to defend themselves. In this sense, harassment reporting, if not designed with careful consideration of unintended use, can become a new tool for harassing and silencing marginalized social VR users. Given these potential risks, the design of future harassment mitigation mechanisms in social VR must involve careful consideration of all how an individual might abuse and misuse such systems if given the opportunity. This is not to say that we should make future harassment reporting systems even more complicated and difficult to use by adding additional steps to verify accusations of harassment against social VR users. Rather, future designs of harassment reporting mechanisms should allow for further elaborating details about the instance for future investigation and linking with other harassment mitigation strategies that serve the broader social VR community (e.g., AI-based moderation systems [22, 23, 64] or social VR consent mechanics [65, 82]) to help sustain social VR users' proactivity and personal agency in reporting harassers while minimizing or avoiding severe and potentially damaging risks of abusing such mechanisms.

5.3 Limitations

It is important to note several limitations of our study, beginning with the relatively small sample size of interviewees (N = 12). However, our studied population - people who have experienced both sides of harassment accusations in social VR - is both specific and difficult to recruit, as the social stigma surrounding being called a harasser discourages potential interviewees from participating. Our sample size of 12 is also still consistent with the typical interview sample size for qualitative HCI works, i.e., 12 [15]. Additionally, we want to highlight that the demographics of our participants differ from general social VR user populations, who are often considered cisnormative and male-dominated [6, 7, 30, 32, 63]. In contrast, all our participants are considered marginalized social VR users due to their varied genders, sexualities, and ethnicities. As such, their perspectives and experiences with both being harassed and being called a harasser by others may not fully reflect how all people experience harassment and harassment accusations in social VR. Yet, we believe that they still provide unique insights about more nuanced cases of harassment that are understudied in existing works that focus on social VR harassment in general. Similarly, this recruitment issue extends to the limitations of self-reported data, as we could not verify people's self-reported demographics (e.g., age). Participants might also be reluctant to disclose more severe, rather than mild, harassment incidents that they had been accused of, especially in comparison

to harassment often seen online or in social VR. Regardless, the insights we gained from our participants are still valuable. Those who self-selected into this part of the study were more likely to be honest about their perceptions since they were already honest about admitting to having been accused of harassment.

6 CONCLUSION

As social VR continues to offer more embodied and immersive experiences to users, so too are concerns over how these spaces modify and amplify online harassment risks growing. Our work takes a new and understudied lens to further delve into the complexity and uncertainty surrounding how harassment is defined and perceived in social VR – by understanding the experiences of marginalized users who have experienced both sides of harassment accusations in social VR. Our findings have revealed several new insights through learning from these individuals' unique experiences, including the importance of unpacking specific motivations behind people's behaviors in social VR and how accusations of harassment can be weaponized against marginalized social VR users rather than protecting them. Building on these insights, we also propose two higher-level design principles for existing and future harassment mitigation approaches by taking the complexity and ethical dilemmas surrounding harassment accusations in social VR into consideration. In sum, our work hopes to offer a more comprehensive image of online harassment involving complex ethical reflections, rather than oversimplifying this issue, to inform the future design of safer social VR spaces.

ACKNOWLEDGMENTS

We thank our participants and the anonymous reviewers. This work was supported by the National Science Foundation under award 2112878.

REFERENCES

[1] Carolina Are. 2023. Flagging as a silencing tool: Exploring the relationship between de-platforming of sex and online abuse on Instagram and TikTok. *new media & society* (2023), 14614448241228544.

[2] Shaowen Bardzell and Jeffrey Bardzell. 2011. Towards a feminist HCI methodology: social science, feminism, and HCI. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 675–684.

[3] Nicole A Beres, Julian Frommel, Elizabeth Reid, Regan L Mandryk, and Madison Klarkowski. 2021. Don't You Know That You're Toxic: Normalization of Toxicity in Online Gaming. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[4] Lindsay Blackwell, Tianying Chen, Sarita Schoenebeck, and Cliff Lampe. 2018. When online harassment is perceived as justified. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 12.

[5] Lindsay Blackwell, Jill Dimond, Sarita Schoenebeck, and Cliff Lampe. 2017. Classification and its consequences for online harassment: Design insights from heartmob. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–19.

[6] Lindsay Blackwell, Nicole Ellison, Natasha Elliott-Deflo, and Raz Schwartz. 2019. Harassment in Social Virtual Reality: Challenges for Platform Governance. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–25.

[7] Lindsay Blackwell, Nicole Ellison, Natasha Elliott-Deflo, and Raz Schwartz. 2019. Harassment in social VR: Implications for design. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 854–855.

[8] Lindsay Blackwell, Mark Handel, Sarah T Roberts, Amy Bruckman, and Kimberly Voll. 2018. Understanding" bad actors" online. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–7.

[9] Jason T Bowey, Ansgar E Depping, and Regan L Mandryk. 2017. Don't talk dirty to me: How sexist beliefs affect experience in sexist games. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 1530–1543.

[10] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[11] Virginia Braun and Victoria Clarke. 2012. *Thematic analysis.* American Psychological Association.

[12] Jie Cai and Donghee Yvette Wohn. 2019. What are effective strategies of handling harassment on twitch? Users' perspectives. In *Conference companion publication of the 2019 on computer supported cooperative work and social*

computing. 166–170.
[13] Jie Cai and Donghee Yvette Wohn. 2022. Coordination and Collaboration: How do Volunteer Moderators Work as a Team in Live Streaming Communities?. In *CHI Conference on Human Factors in Computing Systems*. 1–14.
[14] Jie Cai, Donghee Yvette Wohn, and Mashael Almoqbel. 2021. Moderation visibility: Mapping the strategies of volunteer moderators in live streaming micro communities. In *ACM International Conference on Interactive Media Experiences*. 61–72.
[15] Kelly Caine. 2016. Local standards for sample size at CHI. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 981–992.
[16] Stevie Chancellor, Jessica Annette Pater, Trustin Clear, Eric Gilbert, and Munmun De Choudhury. 2016. # thyghgapp: Instagram content moderation and lexical variation in pro-eating disorder communities. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*. 1201–1213.
[17] Eshwar Chandrasekharan, Shagun Jhaver, Amy Bruckman, and Eric Gilbert. 2022. Quarantined! Examining the effects of a community-wide moderation intervention on Reddit. *ACM Transactions on Computer-Human Interaction (TOCHI)* 29, 4 (2022), 1–26.
[18] Aiyu Chen. 2023. Unveiling Gender Power Dynamics in Chinese Digital Discourse: A Case Study of the Discourse of Mutual Accusations in Allegations of Sexual Harassment. *Journal of Education, Humanities and Social Sciences* 23 (2023), 669–683.
[19] Fatih Çömlekçi. 2019. Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media. *Communication Today* 10, 1 (2019), 165–166.
[20] U.S. Equal Employment Opportunity Commission. 2021. *Harassment*. https://www.eeoc.gov/harassment
[21] 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*. 1–13.
[22] Cristina Fiani, Robin Bretin, Shaun Alexander Macdonald, Mohamed Khamis, and Mark Mcgill. 2024. " Pikachu would electrocute people who are misbehaving": Expert, Guardian and Child Perspectives on Automated Embodied Moderators for Safeguarding Children in Social Virtual Reality. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–23.
[23] Cristina Fiani, Robin Bretin, Mark McGill, and Mohamed Khamis. 2023. Big Buddy: Exploring Child Reactions and Parental Perceptions towards a Simulated Embodied Moderating System for Social Virtual Reality. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference*. 1–13.
[24] Cristina Fiani and Stacy Marsella. 2022. Investigating the Non-Verbal Behavior Features of Bullying for the Development of an Automatic Recognition System in Social Virtual Reality. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces*. 1–3.
[25] Jesse Fox and Wai Yen Tang. 2014. Sexism in online video games: The role of conformity to masculine norms and social dominance orientation. *Computers in Human Behavior* 33 (2014), 314–320.
[26] Guo Freeman and Dane Acena. 2021. Hugging from A Distance: Building Interpersonal Relationships in Social Virtual Reality. In *ACM International Conference on Interactive Media Experiences*. 84–95.
[27] Guo Freeman and Dane Acena. 2022. "Acting Out" Queer Identity: The Embodied Visibility in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–32.
[28] Guo Freeman, Dane Acena, Nathan J McNeese, and Kelsea Schulenberg. 2022. Working Together Apart through Embodiment: Engaging in Everyday Collaborative Activities in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP (2022), 1–25.
[29] Guo Freeman and Divine Maloney. 2021. Body, avatar, and me: The presentation and perception of self in social virtual reality. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–27.
[30] Guo Freeman, Divine Maloney, Dane Acena, and Catherine Barwulor. 2022. (Re) discovering the Physical Body Online: Strategies and Challenges to Approach Non-Cisgender Identity in Social Virtual Reality. In *CHI Conference on Human Factors in Computing Systems*. 1–15.
[31] Guo Freeman and Donghee Yvette Wohn. 2020. Streaming your identity: Navigating the presentation of gender and sexuality through live streaming. *Computer Supported Cooperative Work (CSCW)* 29 (2020), 795–825.
[32] Guo Freeman, Samaneh Zamanifard, Divine Maloney, and Dane Acena. 2022. Disturbing the Peace: Experiencing and Mitigating Emerging Harassment in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–30.
[33] Nitesh Goyal, Leslie Park, and Lucy Vasserman. 2022. " You have to prove the threat is real": Understanding the needs of Female Journalists and Activists to Document and Report Online Harassment. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–17.
[34] Kishonna L Gray, Bertan Buyukozturk, and Zachary G Hill. 2017. Blurring the boundaries: Using Gamergate to examine "real" and symbolic violence against women in contemporary gaming culture. *Sociology Compass* 11, 3 (2017), e12458.

[35] Kishonna L Gray and Krysten Stein. 2021. "We 'said her name'and got zucked": Black women calling-out the carceral logics of digital platforms. *Gender & Society* 35, 4 (2021), 538–545.
[36] Jessica Guynn. 2019. Facebook while black: Users call it getting 'Zucked,'say talking about racism is censored as hate speech. *Usa today* 24 (2019).
[37] Oliver L Haimson, Daniel Delmonaco, Peipei Nie, and Andrea Wegner. 2021. Disproportionate removals and differing content moderation experiences for conservative, transgender, and black social media users: Marginalization and moderation gray areas. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–35.
[38] Shagun Jhaver, Darren Scott Appling, Eric Gilbert, and Amy Bruckman. 2019. " Did you suspect the post would be removed?" Understanding user reactions to content removals on Reddit. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–33.
[39] Shagun Jhaver, Amy Bruckman, and Eric Gilbert. 2019. Does transparency in moderation really matter? User behavior after content removal explanations on reddit. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–27.
[40] Shagun Jhaver, Sucheta Ghoshal, Amy Bruckman, and Eric Gilbert. 2018. Online harassment and content moderation: The case of blocklists. *ACM Transactions on Computer-Human Interaction (TOCHI)* 25, 2 (2018), 1–33.
[41] Jialun Aaron Jiang, Charles Kiene, Skyler Middler, Jed R Brubaker, and Casey Fiesler. 2019. Moderation challenges in voice-based online communities on discord. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–23.
[42] Haesoo Kim, HaeEun Kim, Juho Kim, and Jeong-woo Jang. 2022. When Does it Become Harassment? An Investigation of Online Criticism and Calling Out in Twitter. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–32.
[43] Yubo Kou. 2021. Punishment and Its Discontents: An Analysis of Permanent Ban in an Online Game Community. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–21.
[44] Song Mi Lee, Cliff Lampe, JJ Prescott, and Sarita Schoenebeck. 2022. Characteristics of People Who Engage in Online Harassing Behavior. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–7.
[45] Calvin A Liang, Sean A Munson, and Julie A Kientz. 2021. Embracing four tensions in human-computer interaction research with marginalized people. *ACM Transactions on Computer-Human Interaction (TOCHI)* 28, 2 (2021), 1–47.
[46] Renkai Ma, Yao Li, and Yubo Kou. 2023. Transparency, Fairness, and Coping: How Players Experience Moderation in Multiplayer Online Games. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–21.
[47] Daniel Madden and Casper Harteveld. 2021. "Constant Pressure of Having to Perform": Exploring Player Health Concerns in Esports. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.
[48] Daniel Madden, Yuxuan Liu, Haowei Yu, Mustafa Feyyaz Sonbudak, Giovanni M Troiano, and Casper Harteveld. 2021. "Why Are You Playing Games? You Are a Girl!": Exploring Gender Biases in Esports. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.
[49] Divine Maloney and Guo Freeman. 2020. Falling asleep together: What makes activities in social virtual reality meaningful to users. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. 510–521.
[50] Divine Maloney, Guo Freeman, and Andrew Robb. 2020. It is complicated: Interacting with children in social virtual reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 343–347.
[51] Divine Maloney, Guo Freeman, and Andrew Robb. 2020. A Virtual Space for All: Exploring Children's Experience in Social Virtual Reality. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. 472–483.
[52] Alice E Marwick. 2021. Morally motivated networked harassment as normative reinforcement. *Social Media+ Society* 7, 2 (2021), 20563051211021378.
[53] Joshua McVeigh-Schultz, Anya Kolesnichenko, and Katherine Isbister. 2019. Shaping pro-social interaction in VR: an emerging design framework. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.
[54] Narayanamoorthy Nanditha. 2022. Exclusion in# MeToo India: rethinking inclusivity and intersectionality in Indian digital feminist movements. *Feminist Media Studies* 22, 7 (2022), 1673–1694.
[55] Jessica Outlaw and Beth Duckles. 2018. Virtual Harassment: The Social Experience of 600+ Regular Virtual Reality (VR) Users. https://virtualrealitypop.com/virtual-harassment-the-social-experience-of-600-regular-virtual-reality-vr-users-23b1b4ef884e
[56] Benjamin Paaßen, Thekla Morgenroth, and Michelle Stratemeyer. 2017. What is a true gamer? The male gamer stereotype and the marginalization of women in video game culture. *Sex Roles* 76, 7 (2017), 421–435.
[57] Jessica A Pater, Moon K Kim, Elizabeth D Mynatt, and Casey Fiesler. 2016. Characterizations of online harassment: Comparing policies across social media platforms. In *Proceedings of the 19th international conference on supporting group work*. 369–374.
[58] Whitney Phillips and Ryan M Milner. 2021. *You are here: A field guide for navigating polarized speech, conspiracy theories, and our polluted media landscape*. MIT Press.
[59] Sarah T Roberts. 2016. Commercial content moderation: Digital laborers' dirty work. (2016).

[60] Nazanin Sabri, Bella Chen, Annabelle Teoh, Steven P Dow, Kristen Vaccaro, and Mai Elsherief. 2023. Challenges of Moderating Social Virtual Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–20.

[61] Ari Schlesinger, W Keith Edwards, and Rebecca E Grinter. 2017. Intersectional HCI: Engaging identity through gender, race, and class. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 5412–5427.

[62] Sarita Schoenebeck, Yu Yin Shen, and Jill Davidson. 2023. Evaluating the Social Media Profiles of Online Harassers: An Experimental Study of Attention and Attitudes. *Proceedings of the ACM on Human-Computer Interaction* 7, GROUP (2023), 1–14.

[63] Kelsea Schulenberg, Guo Freeman, Lingyuan Li, and Catherine Barwulor. 2023. "Creepy Towards My Avatar Body, Creepy Towards My Body": How Women Experience and Manage Harassment Risks in Social Virtual Reality. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 236 (oct 2023), 29 pages. https://doi.org/10.1145/3610027

[64] Kelsea Schulenberg, Lingyuan Li, Guo Freeman, Samaneh Zamanifard, and Nathan J. McNeese. 2023. Towards Leveraging AI-based Moderation to Address Emergent Harassment in Social Virtual Reality. In *Proceedings of the 2023 CHI Confrence on Human Factors in Computing Systems*.

[65] Kelsea Schulenberg, Lingyuan Li, Caitlin Lancaster, Douglas Zytko, and Guo Freeman. 2023. "We Don't Want a Bird Cage, We Want Guardrails": Understanding & Designing for Preventing Interpersonal Harm in Social VR through the Lens of Consent. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 323 (oct 2023), 30 pages. https://doi.org/10.1145/3610172

[66] 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.

[67] Ketaki Shriram and Raz Schwartz. 2017. All are welcome: Using VR ethnography to explore harassment behavior in immersive social virtual reality. In *2017 IEEE Virtual Reality (VR)*. IEEE, 225–226.

[68] Mel Slater, Daniel Pérez Marcos, Henrik Ehrsson, and Maria V Sanchez-Vives. 2009. Inducing illusory ownership of a virtual body. *Frontiers in neuroscience* (2009), 29.

[69] Weilun Soon. 2022. A researcher's avatar was sexually assaulted on a metaverse platform owned by Meta. https://www.businessinsider.com/researcher-claims-her-avatar-was-raped-on-metas-metaverse-platform-2022-5

[70] Hannah Sparks. 2021. Woman claims she was virtually 'groped' in Meta's VR metaverse. https://nypost.com/2021/12/17/woman-claims-she-was-virtually-groped-in-meta-vr-metaverse/

[71] Tim Squirrell. 2019. Platform dialectics: The relationships between volunteer moderators and end users on reddit. *New Media & Society* 21, 9 (2019), 1910–1927.

[72] Wai Yen Tang and Jesse Fox. 2016. Men's harassment behavior in online video games: Personality traits and game factors. *Aggressive behavior* 42, 6 (2016), 513–521.

[73] Wai Yen Tang, Felix Reer, and Thorsten Quandt. 2020. Investigating sexual harassment in online video games: How personality and context factors are related to toxic sexual behaviors against fellow players. *Aggressive behavior* 46, 1 (2020), 127–135.

[74] Jirassaya Uttarapong, Jie Cai, and Donghee Yvette Wohn. 2021. Harassment Experiences of Women and LGBTQ Live Streamers and How They Handled Negativity. In *ACM International Conference on Interactive Media Experiences*. 7–19.

[75] Kathleen Van Royen, Karolien Poels, Heidi Vandebosch, and Philippe Adam. 2017. "Thinking before posting?" Reducing cyber harassment on social networking sites through a reflective message. *Computers in human behavior* 66 (2017), 345–352.

[76] Jessica Vitak, Kalyani Chadha, Linda Steiner, and Zahra Ashktorab. 2017. Identifying women's experiences with and strategies for mitigating negative effects of online harassment. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 1231–1245.

[77] Amy K Way, Robin Kanak Zwier, and Sarah J Tracy. 2015. Dialogic interviewing and flickers of transformation: An examination and delineation of interactional strategies that promote participant self-reflexivity. *Qualitative Inquiry* 21, 8 (2015), 720–731.

[78] Bob G Witmer and Michael J Singer. 1998. Measuring presence in virtual environments: A presence questionnaire. *Presence* 7, 3 (1998), 225–240.

[79] 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*. 1–13.

[80] Qingxiao Zheng, Shengyang Xu, Lingqing Wang, Yiliu Tang, Rohan C Salvi, Guo Freeman, and Yun Huang. 2023. Understanding Safety Risks and Safety Design in Social VR Environments. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–37.

[81] Jennifer Zimbroff. 2007. Cultural differences in perceptions of and responses to sexual harassment. *Duke J. Gender L. & Pol'y* 14 (2007), 1311.

[82] Douglas Zytko and Jonathan Chan. 2023. The Dating Metaverse: Why We Need to Design for Consent in Social VR. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (2023), 2489–2498.

A APPENDIX

A.1 Interview Questions Specifically Related to This Study

Contextualizing Social VR Use:

- How long have you been engaging in social VR?
- What social VR platforms do you use?
- How often do you use social VR, and how many hours a week?
- What VR device(s) do you use?
- How does using social VR benefit you, if at all?

Apprehending the Phenomenon of Harassment in Social VR:

Defining Harassment:

- We have discussed how social VR may benefit you, and now we want to discuss how to make social VR better and safer. Before we go further, please explain how you define harassment in social VR, such as what types of behaviors in social VR you would call harassment.
- What is an example of something that you might consider harassment but someone else might not?
- What is an example of something that you would **not** consider harassment but someone else would?
- Do you feel some types of harassment are more severe and damaging than others? Why are they more (or less) severe and damaging?

Experiences with Being Harassed

- Please recall for me a time when you felt that you were being harassed or witnessed someone else being harassed.
  - How did that experience make you feel about yourself? About social VR?
  - How did others nearby react to the situation, if at all?
  - Do you feel the harassment was linked to any particular feature about yourself or your avatar? If yes, please explain further.
  - If possible, what would you have done to prevent this harassment from happening?

Experiences with Being Accused of Harassment

- Can you describe a time when someone called you a harasser in social VR?
  - In your opinion, why did other people accuse you?
  - How did that experience make you feel about yourself? About social VR?
  - How did others nearby react to the situation, if at all?
  - Did you change your behavior because of this accusation? Why or why not?
  - Do you feel that the accusation was linked to any of your behaviors or features of your avatar? If yes, please explain further.
  - If possible, what would you have done so others would not accuse you as a harasser?