Culturally justified hate: Prevalence and mental health impact of dark participation in games

Rachel Kowert
Take This
rachel@takethis.org

Elizabeth Kilmer
Take This
elizabeth.k@takethis.org

Alex Newhouse
Middlebury Institute of International Studies
anewhouse@middlebury.edu

Abstract

Hate, harassment, and other forms of so-called "toxicity" are colloquially discussed as normalized activities in gaming spaces. However, there are several challenges that have limited researchers' ability to assess this normalization in terms of the prevalence, nature, and embeddedness of these deviant practices. This work addresses those challenges directly and assesses the rates of dark participation, their mental health impact, player mitigation strategies, and player perceptions around the cultural normalization of these actions within gaming communities. The results provide empirical support for high rates of dark participation in games, a range of mental health consequences to these actions, as well as the endorsement of the culturally justified acceptance of these behaviors within gaming spaces.

Keywords: toxicity, dark participation, video games, mental health, harassment

1. Introduction

Hate, harassment, and other forms of so-called "toxicity" have come to be nearly synonymous with video game communities, with recent work placing the prevalence of hate, harassment, and other actions considered toxic within gaming spaces alarmingly high. For example, the Anti-Defamation League (ADL; 2022) reported that four out of five adults (86%) reported experiencing harassment in online multiplayer games, representing over 67 million adult gamers. The ADL also found that harassment is increasing among those in the gaming community, with 77% of adults having experienced severe harassment in 2022 (defined as physical threats, stalking, and/or sustained harassment), up from 65% in 2019.

This is just one of many attempts that have been made to document toxicity in gaming spaces. However, attempts to document the landscape of these behaviors in games have faced many challenges, prohibiting the creation of a cohesive foundation of knowledge. Specifically, there is no shared language regarding this behavior, nor is there a clear understanding of the contexts wherein these behaviors occur, the mental health impacts of these behaviors on the individual or society, or the prevalence of player-led mitigation efforts. Further, there is a current lack of evidence-based knowledge pertaining to the impact of the cultural normalization of such behaviors. Each of these challenges are discussed in more detail below.

1.1 Challenge 1: Creating a shared language

Researchers have largely failed to create and work from a shared language or singular taxonomy of toxicity in video games, making it difficult to determine prevalence of toxic actions across behaviors, time, and spaces. Taxonomies have ranged from five (Saarinen, 2017) to nine (Komac & Cagiltay, 2019) to ten (Cook et al., 2018; Kwak, Blackburn, & Han, 2015), or even as few as three (Ghosh, 2021; Thacker & Griffiths, 2012) items. There is also often no clear differentiation between what is considered toxic and what is considered something else, such as trolling, making it difficult to pinpoint exactly what behaviors would be encapsulated under these respective categories. In fact, most of the work in this space does not enlist a formal taxonomy at all, but rather conceptualizes these actions under the umbrella term of "toxic" or "trolling" broadly (Herring, Job-Sluder, Scheckler, et al., 2002; Kordyaka et al., 2019; Schachaf & Hara, 2010; Zsila, Shabahang, Aruguete et al., 2022).

In the last few years, there have been several formal attempts to overcome this challenge and develop a more robust taxonomy to create a shared language. For example, in 2020 the Fair Play Alliance released their *Disruption and Harms Online Gaming Framework*. Within it, they list eleven different kinds of "disruptive behaviors" in online games. This framework also provides a new and novel categorization system to organize these behaviors based on the intentionality of the perpetrator: unintended disruption, aggravation, anti-social actions, and abuse of play/antagonistic play.

URI: https://hdl.handle.net/10125/106708
978-0-9981331-7-1

Taking a different approach, Kowert (2020) collated many of the previously used taxonomies and developed a list of 17 different behaviors that would fall under the overarching umbrella of "dark participation." She notes that dark participation is the broadest category that encapsulates all deviant verbal and behavioral actions. Each of the actions within it lie on two axes that range from verbal to behavioral and transient to strategic, outlining the spectrum of these behaviors to more clearly delineate differences in the perception and impact of these actions on their victims. For example, doxxing (sharing someone's personal information online), a behavioral and strategic action, is more likely to have a severe, long-term negative impact on the victim than a verbal, transient action such as trash-talking (Allegra, 2017; Turkay et al., 2020). Within this system, Kowert (2020) considers toxicity as the *outcome* of dark participation. That is, any behavior that causes harm to another is considered toxic, regardless of the intent. Understanding the specific actions that are occurring within these spaces, and the rates at which they are being directly experienced as well as witnessed within gaming communities, is critical for moving the discussion of toxicity as a colloquial term and towards an understanding of the very real and serious actions that are occurring within digital gaming and adjacent spaces.

1.2 Challenge 2: Location, location, location

While research has made significant strides in the last few years to generate a better understanding of the scope and impact of dark participation in games, it remains unclear where these behaviors are taking place. While many interactions happen within the game itself, there are also third-party platforms, such as *Discord* and *Twitch*, where gaming communities gather. In the last few years, the focus has been placed on the latter.

It is worth noting that in the last few years, game adjacent platforms Discord and Twitch, as well as Microsoft Xbox, have released transparency reports outlining moderation actions such as removing content proactively or after a user report, warning users, or banning users are noted as utilized across forms of dark participation (Discord Safety, 2023; Microsoft, 2023; Twitch, 2022). While these reports provide some insight into the efforts around combating dark participation, little remains known about the frequency or prevalence of these behaviors across environments.

1.3 Challenge 3: Mental health impact

Generating a robust foundation of prevalence data is important for understanding the landscape of dark participation in games; however, understanding the mental health impact of these behaviors is equally important to better appreciate the cost of dark participation on consumers, studios, and society.

Being directly targeted or indirectly exposed to any form of dark participation, even if brief, has been found to have a significant, negative impact on players both on- and offline (de Mesquita Neto & Beker, 2018). According to the ADL (2022), a significant amount of players report short and long-term mental health impacts of this behavior including feeling more isolated and alone (14%) and depressive/suicidal thoughts (11%). Significant behavioral impacts were also reported, including taking steps to reduce one's risk to their personal safety (21%), contacting the police (13%), and disrupted personal relationships (11%). The 2020 Bryter reported that one in four female players noted that the "widespread toxicity" in games made them feel upset, intimidated, and made them not want to play games anymore. Increases in anxiety and reductions in self-esteem have also been reported as a result of victimization within online gaming spaces (Ewoldsen, Eno, Oldie, et al., 2012) with female players reporting more emotional consequences (Zsila, Shabahang, Aruguete et al., 2022).

While this work provides some insight into the mental health impact of these actions, more information is needed to better understand the nature and scope of mental health impact across actions (i.e., transient versus strategic) as well as across populations.

1.4 Challenge 4: Experiences across gender

Understanding the likelihood different groups of players are to be the target, witness, or perpetrator of dark participation is vital to develop effective mitigation strategies for communities. Previous studies have noted differences both in the likelihood of different genders to experience or perpetuate dark participation, as well as differential impacts on their mental health and subsequent behaviors.

Previous research has found that men are more likely to engage in harassment than women (Ballard & Welch, 2017; Kowert & Cook 2022), while women are more likely to experience sexual harassment than men in online gaming and game adjacent spaces (Ruvalcaba et al., 2018; Ballard & Welch, 2017; Todd & Melancon, 2019; Trudgett-Close & McLinton, 2023). Notably, these differential experiences are

likely driven by cultural beliefs and societal pressures about gender that exist outside of the game space. For example, men who endorse higher levels of hostile sexism (i.e., men's superiority over women) and social dominance orientation are more likely to engage in online harassment behavior - both general and sexual harassment (Tang & Fox, 2016). Men and women may differ in what they consider dark participation in the first place, for instance in relation to what is considered sexual harassment (Quinn, 2002).

It is important to note that much of the previous research on the impact of dark participation in games has been focused on the impact on women and has identified both negative mental health and behavioral impacts. Women report feeling isolated, anxious, and unsupported in the face of online harassment (McLean & Griffiths, 2019). Though both the experience of non-gender related harassment and sexual harassment can predict women's withdrawal from gaming spaces, sexual harassment was found to predict rumination as well (Fox & Tang, 2017). Notably, players who found the studio or platform to respond appropriately following sexual harassment (i.e., responsive, initiated action against the offender) were less likely to withdraw, highlighting the importance of formal mitigation strategies (Fox & Tang, 2017).

1.5 Challenge 5: Player mitigation efforts

To mitigate dark participation, and consequently its impact on players, a heavy focus has been placed on the role of bottom-up intervention via player reports. Nearly every gaming console and platform has some reporting tool available; however, how often these tools are utilized can vary. While Cook and colleagues (2018) found that reporting tools were the least-used recourse by bystanders and victims when faced with dark participation, a later report by Kowert and Cook (2022) noted that most participants who witnessed forms of dark participation reported it.

Reporting is just one tool in the toolkit as we know that players often employ other strategies. For example, research has found that women utilize a range of avoidance strategies, including playing alone, switching between groups frequently, and playing anonymously (Cote, 2016; McLean & Griffiths, 2019). Other strategies can include ignoring the behaviors, confronting the perpetrator publicly, contacting in-game support, and utilizing whisper networks (i.e., an informal chain of information passed privately between people about harassers or abusers). Who is utilizing what mitigation efforts and to what extent remains largely unexplored.

1.6 Challenge 6: Cultural normalization

Regardless of whether people are utilizing reporting tools, it is clear that the experiences of dark participation have come to be so ubiquitous with gaming experiences that some have argued this behavior has become normalized (Adinolf & Turkay, 2018; Beres, Frommel, Reid, Mandryk, et al., 2021; Hilvert-Bruce & Neill, 2020; Kowert & Crevoshay, 2022). Kowert and Crevoshay (2022) go so far as to situate this cultural justification within their definition of toxic gamer culture as "*a set of culturally justified behaviors within gaming communities that are harmful to others within and outside of gaming spaces*." Qualitative studies in this space have also found that players themselves discuss dark participation as an accepted part of gamer cultures (Beres, Frommel, Reid, Mandryk, et al., 2021) and rationalize it as part of competitive game culture (Adinolf & Turkey, 2018).

An examination of why these actions have become normalized within gamer culture is not within the scope of the current project (e.g., environmental and community factors, personality, gender socialization, etc.); however, many researchers have found competitive gameplay to have higher rates of dark participation behavior in prevalence studies (Adachi & Willoughby, 2011; Shores et al., 2014; Zubek and Khoo, 2002). Understanding how the toxic gamer cultures may vary based on design elements could provide some insight into mitigation strategies and approaches across cultural contexts.

2. Current Study

The aim of this paper is to address the challenges that have prohibited the creation of a cohesive, shared knowledge base around topics of hate and harassment in games. This will be done by assessing prevalence rates of dark participation (witnessed and directly experienced), their mental health impact, player mitigation strategies, and player perceptions around the cultural normalization of these actions within gaming communities.

2. Measures and Methods

Participants were asked to complete an online survey and were treated in accordance with ethical and Middlebury University IRB guidelines. Participants were recruited via snowball sampling across social media and the authors' personal and professional networks. Data was collected for six weeks in Q1 of 2023.

2.1 Demographic information

Participants were asked to report their age, gender identity, and country of residence. Due to IRB constraints, only participants over the age of 18 residing in North America and the United Kingdom were eligible to participate. English proficiency was also assessed to ensure participants understood the content of the questions within the survey. Participants were also asked about play frequency habits.

2.2 Dark participation

To assess the various forms of Dark Participation, participants were asked to report which behaviors from a list of 18 actions they had experienced (as a direct target), witnessed, and/or perpetrated within gaming spaces. The foundation of this list was drawn from Kowert (2020) and Kowert and Cook (2022), with the addition of a new category of incitement to violence, as recent work has identified this as an action that can occur within social gaming spaces (Koehler, Fiebig, & Jugl, 2022). The full list of behaviors are shown in Table 1.

In addition to reporting whether they had experienced these specific actions, participants were asked to report where they had primarily experienced them, in gaming spaces themselves (i.e., in-game chat, on a gaming platform), in gaming adjacent spaces (i.e., Discord, Twitch chat), or whether that they had experienced it equally in gaming and gaming adjacent spaces.

2.3 Mental health impact

To assess the impact of these behaviors, participants were asked how being a direct target and/or witnessing dark participation of any kind impacted their mental health. They chose from a list of 10 options: felt angry, felt uncomfortable/upset, been less social (e.g., muting mic, limiting interactions with only guild members, etc.), felt isolated/alone, increased depression (e.g.,, feeling sad, low self-esteem, worthless), increased anxiety (e.g., feeling more worried, on edge, tense, nervous), symptoms related to post-traumatic stress disorder (e.g., avoidance or hyperviligance), had suicidal thoughts, my mental health has not been impacted by online harassment, or other. Participants were able to check all that applied.

2.4 Mitigation efforts

Participants were also asked how they have responded to being a direct target and/or witnessing dark participation. They chose from a list of five options and were asked to check all that apply: ignore the comment, block or report the person, confront or challenge the comment publicly, share the information with others in my network, and/or contacted support (through the game platform or system) on how to respond and/or report.

Table 1. Types of dark participation

	Definition
Trash talking	Putting down or making fun of other players
Misinformation	Repeatedly sharing game-unrelated chat
Contrary play	Playing outside of what it is intended by most players
Inhibiting team	Inhibiting your team from being successful in winning
Aiding the enemy	Strategically aiding the opposing team
Inappropriate roleplaying	Pretending to be a different person to obtain a specific reaction or not abiding by the norms of the community
Verbal spamming	Sending the same verbal message or using the same in-game move
Griefing	Irritating and/or harassment other players by using the game in unintended ways
Sexual harassment	Insults or comments based on gender, including threats, the criticism, or stalking.
Hate speech	Insults based on religion, ethnicity, nationality, or other personal information
Threats of violence	Threats of physical abuse, vandalism, possession or use of weapons, or other dangerous action
Incitement of violence	Speech, words, or behaviors that encourages the immediate risk of harm to another person
Flaming	Presenting emotionally fueled or contrary statements
Behavioral spamming	Using the same in-game move, often to the consternation of others
In-game cheating	Using methods to create advantage beyond normal gameplay in order to make the game easier for oneself
Hate raiding	Purposefully infiltrating the game space of another with

	the intention of spreading hate or harassment
Doxxing	Publicly sharing and/or publishing another player's identifying information
Swatting	Calling emergency services in an attempt to dispatch armed police officers to a particular address

2.5 Culture assessment

To assess the perception of dark participation as a normalized part of gamer cultures, participants answered several questions relating to gamer culture and the perception of them being normalized actions within gaming communities. As several prevalence studies have noted higher rates of dark participation behavior in competitive versus cooperative games (e.g., Adachi & Willoughby, 2011; Shores et al., 2014; Zubek & Khoo, 2002), we also asked participants to report their perception of this phenomena and report whether they believed that the competitive nature of games is a potential driver of these actions.

3. Results

In total, 423 respondents completed the survey. As participation was limited to residents of the United States and United Kingdom, 56 participants had to be removed from the analyses. We further removed four participants who failed an embedded validity question, and two participants who had extreme responses on the fusion and willingness to fight/die metrics, as twitter screenshots revealed these were an attempt to disrupt the data. The final dataset resulted in 361 observations.

Most participants identified as men (49.6%, n=179), with eight of the male participants identifying as trans (2.2% of the total sample). Women comprised 30.2% of the sample (n=109), with 11 women identifying as trans (3% of the total sample). Additionally, 19.7% of participants identified as non-binary, gender non-conforming, or questioning (n = 71), and .6% of participants did not report a gender (n = 2). Participants ranged in age from 18 to 63, with an average age of 33 (SD = 7.71). The majority of participants resided in the US (n = 308, 85.3%), while a smaller proportion resided in the UK (n = 53, 14.7%).

3.1 Frequency of dark participation

Among all the participants, 82.3% reported being a victim of some form of dark participation, while 88.1% reported witnessing at least one form of dark participation. Notably, 31% of participants reported participating in at least one form of dark participation, while 10% reported they were "not sure" if they had participated.

The most common witnessed experiences were trash talking (87%), hate speech (73%), griefing (73%), verbal spamming (72%), and sexual harassment (70%). The most common direct experiences (i.e., being a direct target of) were trash talking (78.9%), griefing (60.4%), inhibiting team (57.6%), verbal spamming (54.8%), and in-game cheating (50.4%). Prevalence rates across categories can be found in Table 2.

Table 2. Prevalence of witnessing (W) and being a direct target (DT) of dark participation

	W (n = 361)	DT (n = 361)
Trash talking	87.0	78.9
Misinformation	54.8	32.4
Contrary play	63.4	45.2
Inhibiting team	67.3	57.6
Aiding the enemy	59.8	46.0
Inappropriate roleplaying	42.7	25.5
Verbal spamming	72.0	54.8
Griefing	73.1	60.4
Sexual harassment	70.6	40.2
Hate speech	73.1	47.9
Threats of violence	57.1	35.7
Incitement of violence	39.9	23.3
Flaming	61.5	45.4
Behavioral spamming	57.6	44.9
In-game cheating	56.0	50.4
Hate raiding	34.1	12.5
Doxxing	20.8	7.5
Swatting	8.6	.8

Most of the witnessed experiences of dark participation occurred within gaming spaces (51.9%). Just over one-third of participants (37.4%) reported witnessing these incidents equally in gaming and gaming adjacent spaces and only 10.7% of participants reported witnessing dark participation exclusively in gaming adjacent spaces.

Most incidents of being the target of dark participation were reported to have occurred within gaming spaces themselves (72.4%); however, a significant number of participants reported that they had experienced dark participation equally in gaming and game adjacent spaces (21.9%). Only 5.7% of

participants reported that their experiences with dark participation were limited to game-adjacent spaces exclusively (i.e., Discord, Twitch).

For those who disclosed that they perpetrated dark participation, 89.9% of participants reported these incidents were in gaming spaces, 7.4% reported they were equally in gaming and game adjacent spaces, and 2.7% reported they were exclusively in gaming adjacent spaces.

Gender analyses were conducted to evaluate differences in the experiences and impact of dark participation across gender categories. These analyses were limited to the categories of male and female as there were too few participants (N= 71) across the other gender categories. Significant gender differences were found in several categories of dark participation. Men were more likely to report being the target of trash talking (X^2 (1, N = 239) = 5.65, p = .017), contrary play (X^2 (1, N = 239) = 6.09, p = .014), inhibiting team (X^2 (1, N = 239) = 11.61, p < .001), aiding the enemy (X^2 (1, N = 239) = 21.14, p <.001), verbal spam (X^2 (1, N = 239) = 7.72, p = .005), griefing (X^2 (1, N = 239) = 8.93, p = .003), and cheating (X^2 (1, N = 239) = 15.06, p <.001). Women were more likely to report being the target of sexual harassment (X^2 (1, N = 239) = 61.36, p < .001). Men were more likely than women to report witnessing several forms of dark participation, including contrary play (X^2 (1, N = 255) = 16.54, p < .001), inhibiting team (X^2 (1, N = 255) = 11.95, p < .001), and aiding the enemy (X^2 (1, N = 255) = 21.76, p < .001). In line with the direct target data, women were more likely to report witnessing sexual harassment (X^2 (1, N = 255) = 9.55, p = .002).

3.2 Mental health impact of dark participation

The most common response for witnessing dark participation was feeling angry (74.53%), followed by feeling uncomfortable/upset (71.70%), and being less social (65.09%). For direct targets, the most common response was feeling uncomfortable/upset (78.45%), feeling angry (75.42%), and feeling less social (71.04%).

More severe outcomes were reported as well, including nearly half of all witnesses (46.86%) and direct targets (50.84%) reporting increased anxiety, and nearly a quarter of witnesses (23.90%) and direct targets (58.93%) reporting post-traumatic stress disorder symptomatology. Around 1 in 10 of all witnesses (11.64%) and direct targets (10.44%) reported no impact on their mental health due to dark participation.

To examine the relationship between reported mental health impacts of being the target of dark participation and gender, chi-square tests of independence were conducted for participants who reported their gender and mental health impacts. The relationship between the likelihood participants felt uncomfortable/upset when the target of dark participation and gender was significant (X^2 (1, N = 239) = 16.31, p < .001), with women more likely to feel upset when being a direct target. Women were also more likely to report being less social (X^2 (1, N = 239) = 13.14, p <.001), higher anxiety (X^2 (1, N = 239) = 10.28, p = .001), and PTSD symptoms (X^2 (1, N = 239) = 7.12, p = .008), when they are the target of dark participation. In contrast, men were more likely to report no mental health impact due to being a direct target of dark participation (X^2 (1, N = 239) = 15.68, p < .001). Women were more likely to report experiencing mental health impacts after witnessing dark participation than men, including feeling uncomfortable (X^2 (1, N = 255) = 18.32, p < .001), less social (X^2 (1, N = 255) = 16.86, p < .001), anxiety (X^2 (1, N = 255) = 12.00, p < .001), and PTSD symptoms (X^2 (1, N = 255) = 9.15, p = .002). Men were more likely to report no mental health impacts of witnessing dark participation (X^2 (1, N = 255) = 10.00, p = .002).

3.3 Responses to dark participation

The most common response by participants for witnessing (89.62%) and being a direct target of (90.57%) was to report or block the offending user. Overall, direct targets and witnesses reported nearly the same frequency of actions across the options, with the second most common action being ignoring the comment, followed by contacting support, confronting or challenging the comment publicly, and lastly, sharing information with others in their network. These findings can be seen in Table 3.

Table 3. Responses to dark participation

	W (n = 318)	DT (n = 297)
Report or block	89.62	90.57
Ignore comment	72.01	89.23
Contacted support	46.86	46.80
Confront or challenge the comment publicly	44.88	45.79
Share information with network	36.48	35.69

To examine the relationship between types of responses and gender for those who witnessed dark participation, chi-square tests of independence were conducted for participants who reported their gender and responses. Women were found to be more likely to share information about these events within their network than men (X^2 (2, N = 239) = 6.43, p = .011).

All other analyses were non-significant (p's > .05). For those who were direct targets, the relationship between likelihood to ignore the comment and gender was significant, with men being more likely to ignore comments than women (X^2 (1, N = 255) = 5.83, p = .016). Women were more likely to contact support than men (X^2 (1, N = 255) = 4.50, p = .034). No other significant relationships across response options were found (p's > .2).

3.4 Culture assessment

Most participants hold the belief that toxic behavior (65.37%) is embedded (i.e., fixed firmly and deeply within) gaming cultures. The majority of participants also endorsed the belief that toxic behavior (79.50%) and hateful behavior (71.19%) have become culturally justified experiences (i.e., normalized experiences) within gaming spaces. Women were more likely than men to report beliefs that toxic behavior (X^2 (1, N = 288) = 9.23, p = .002) is embedded within gaming cultures. Women were also more likely to report a belief that toxic (X^2 (1, N = 288) = 9.00, p = .003) and hateful (X^2 (1, N = 288) = 13.58, p < .001) behavior have become culturally justified experiences within games.

Most participants (72.85%) also reported that they believe the competitive nature of games contributes to the level of toxicity within gaming communities. No differences were found across gender (p > .05). That is, men and women were equally likely to report a belief that competitive games contributed to environments of toxicity.

3.5 Discussion

Dark participation is incredibly prevalent in gaming spaces, with most forms of dark participation emerging as the norm more than the exception, being witnessed by more than half of all participants assessed in 13 of the 18 categories identified. Being the direct target of these actions was also commonplace, with over a third of players disclosing they have been directly targeted with several severe forms of harassment, including hate speech (47.9%), sexual harassment (40.2%), and threats of violence (35.7%). Supporting previous work in this space, women were more likely to be witness to and the direct target of sexual harassment (Ruvalcaba et al., 2018; Ballard & Welch, 2017; Trudgett-Close & McLinton, 2023) and men were found to be more likely to be perpetrators of dark participation of all kinds (Kowert & Cook, 2022), reinforcing the gendered nature of gaming spaces (Kowert & Oldmeadow, 2012; Paaßen, Morgenroth, & Stratemeyer, 2017).

Notably, most incidents of direct encounters with dark participation were reported to have occurred within gaming spaces themselves (72.4%). This is an important consideration, as recent discussions around incidences, mitigation, and transparency efforts have focused on third-party platforms (Discord Safety, 2023; Microsoft, 2023). These findings highlight the need for more effective in-game and in-platform tools and strategies.

The mental health impact of dark participation varied from transient to more long-term consequences. Nearly a third of all witnesses and direct targets reported feeling increased depression and just under a half of all witnesses and direct targets reported increased anxiety. In addition, nearly one quarter of participants noted they had experienced symptoms associated with post-traumatic stress disorder due to witnessing and/or being a direct target of dark participation. It is worth noting that 5% of witnesses and direct targets reported suicidal ideation because of these actions. Only one in ten participants said that these behaviors had no impact on their mental health. These findings indicate that there are substantial mental health repercussions to consider when discussing "toxic gamer cultures." The actions that are occurring within gaming spaces are not constrained to the four walls of our digital playgrounds. For 90% of players, these actions are impeding their mental well-being in deleterious and impactful ways.

Perhaps unsurprisingly, men were more likely to report no mental health impact from witnessing or being a direct target than women. This likely reflects the disproportionate experiences of men and women in this space. Whereas women were less likely to report being the target of dark participation overall, they were found to be more likely to be the direct target of more severe forms, such as sexual harassment. It is worth noting that sexual toxicity has been linked to particularly deleterious consequences, including eventual withdrawal from games (Fredman, 2018; Fox & Tang, 2017). It is possible these gender differences reflect greater psychological resilience to dark participation among male players, as evidenced by reporting less mental health impact as well as being more likely to "ignore" the action as a mitigation strategy. Additionally, the continued perception of gaming as "male-dominated" space could provide male players with an increased sense of safety in the space, reducing the impact of dark participation. These are areas that could be explored in the future.

Interestingly, the most common responses to dark participation were similar regardless of whether someone was a witness or a direct target, with the most common strategies being to report/block or ignore. However, this did differ across gender. Notably, women were more likely than men to utilize the

strategy of whisper networks, informal chains of conversation among women (typically) about men who need to be watched because of known misconduct, harassment or incidents (Meza, 2017). These whisper networks tend to form in spaces that are perceived to be risky or have normative sexual harassment cultures (Dougherty, 1999) and have been found to protect women in cultures of harassment, help them make sense of their experience, and help informally identify harassers in a space (Johnson, 2022). While whisper networks can be effective at identifying harassers and abusers, they are often created and utilized due to a fear of retaliation (Johnson, 2022) and serve to fill the gaps in ineffective or problematic structural organizational systems. Whisper networks may help women mitigate the impact of harassment or avoid future harassment from perpetrators. However, due to their clandestine nature, they may be ineffective at holding perpetrators accountable or changing the public culture around these actions. The prevalence of this strategy among women in games highlights a lack of formal support and effective response by more official reporting strategies.

Overall, most participants agreed that toxic behavior is embedded within gaming cultures and reported that both toxicity and hateful behavior are *culturally justified experiences within gaming spaces*. This demonstrates a disheartening lack of helplessness and apathy when it comes to these experiences and further supports this idea of normalization within gaming spaces (Adinolf & Turkay, 2018; Beres, Frommel, Reid, Mandryk, et al., 2021; Hilvert-Bruce & Neill, 2020; Kowert & Crevoshay, 2022).

When extreme thoughts and behaviors are normalized, it can reduce an individual's resilience by creating a vulnerability to the ingrained nature of these thoughts and behaviors (Lamphere-Englund, Hamonangan, & Puti, 2022). Research on adolescent bullying (deLara, 2012) and sexual assault (Holland & Cortina, 2017) have identified that expectations of such abuse as ubiquitous and learned helplessness reduce victim willingness to report. When individuals see a behavior as ingrained in a group's culture, they may be less likely to believe their response to any one incident will make a difference. This may be why we found "ignoring the comment" to be such a common strategy for all participants when witnessing or being a direct target of these behaviors. This also supports previous research that has found that players abstain from reporting dark participation because they view it as acceptable, typical of games, as banter, or not their concern (Beres et al., 2021). More exploration is needed to explore this normalization across communities, as most participants also agreed that the competitive nature of games contributes to dark participation within these environments.

3.6 Limitations and future directions

This work provides a good foundation for understanding the nature and impact of dark participation in games; however, there are several limitations to consider. While it drew from Kowert (2020) and Kowert and Cook (2022) to build upon previous efforts of creating a shared language, the addition of a new category, incitement of violence, was included. This further points to the ongoing issue of identifying a shared set of categories and definitions for this work to be built upon. This work also focused specifically on experiences in game with no distinction between in-game text and in-game audio. We know that in-game audio plays a key role in communication in gaming spaces (Reid, Mandryk, Beres, et al., 2022; Kowert & Woodwell, 2022). Future work should dig deeper into the experiences across modalities to understand the landscape more concretely and how to best tailor mitigation and moderation efforts. Future research may benefit from examining gender differences in dark participation across different games and game types. Though the reported time spent playing multiplayer and single-player games did not significantly differ by gender, prior research has found that players may abandon games or types of games after experiencing or witnessing abuse, as well as utilize other mitigation strategies such as not using mic, playing anonymously, that are intended to reduce the amount of harassment they experience (Fox & Tang, 2017). Examining the experiences of dark participation in participants who are playing the same and engaging in the space in a comparable way (i.e., chat only, mic and chat), will help to identify the ways in which gender and gender presentation may impact dark participation experiences.

3.7 Concluding thoughts

"Toxic gamer culture" is a colloquial term that indicates an acceptance of hate, harassment, and other forms of dark participation to be the norm rather than the exception in gamer cultures. In this work, these claims are empirically supported not only by the prevalence of "toxic" acts themselves but the endorsement of the culturally justified acceptance of these behaviors within digital gaming spaces. These behaviors come with a cost, as both being a direct target and witnessing these behaviors are related to mental health consequences that range from relatively mild and transient to potentially more severe and long

term. The experiences of women may be particularly important to studios, as it highlights a driver of churn in a significant demographic in the market. Though the factors supporting the culture of exclusion and toxicity are complex, many gaming spaces need a cultural shift. More effective in-game and in-platform moderation is one starting point for transitioning these spaces away from those where hate is normalized, culturally justified, and sustained through inaction. Additional intervention strategies, such as design efforts for prosocial behavior (Cook, Lau, Tan, Burgess, et al., 2022) should also be considered in the future. Our digital playgrounds are important environments for connection, interpersonal growth, and learning. It is imperative to amplify our efforts to shift the culture away from one where hate is normalized and justified and towards one where everyone is welcome to participate.

3.8 References

Adachi, P. J., and Willoughby, T. (2011). The effect of video game competition and violence on aggressive behavior: which characteristic has the greatest influence? Psychol. Violence 1, 259–274.

Adinolf, S., & Turkay, S. (2018, October). Toxic behaviors in Esports games: player perceptions and coping strategies. In Proceedings of the 2018 Annual Symposium on computer-human interaction in play companion extended abstracts (pp. 365-372).

Anti-Defamation League (ADL; 2022). Hate is no game: Hate and harassment in online games 2022. Retrieved from https://www.adl.org/resources/report/hate-no-game-hate-and-harassment-online-games-2022

Ballard, M. E., & Welch, K. M. (2017). Virtual warfare: Cyberbullying and cyber-victimization in MMOG play. *Games and Culture: A Journal of Interactive Media*, 12(5), 466-491. https://doi.org/10.1177/1555412015592473

Beres, N. A., Frommel, J., Reid, E., Mandryk, R. L., & Klarkowski, M. (2021, May). 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 (pp. 1-15).

Bryter (2020). Female Gamer Survey. Retrieved from https://pages.bryter-research.co.uk/hubfs/003-FGS-1603/Bryter%20-%20Female%20Gamers%20Survey%202020.pdf

Cook, C., Schaafsma, J., and Antheunis, M. (2018). Under the bridge: an in-depth examination of online trolling in the gaming context. New Media Soc. 20, 3323–3340. doi: 10.1177/1461444817748578

Cook, D., Lau, J. C., Tan, M. L., Burgess, J.Moriwaki, T., & Kajioka, E. D. (2022). Kind games: Designing for prosocial multiplayer. *Polaris game design retreat*. Retrieved from https://polarisgamedesign.com/2022/kind-games-designing-for-prosocial-multiplayer/

Cote, A. C. (2017). "I can defend myself" women's strategies for coping with harassment while gaming online. Games and culture, 12(2), 136-155.

DeLara, E. W. (2012). Why adolescents don't disclose incidents of bullying and harassment. *Journal of School Violence*, *11*(4), 288-305.

de Mesquita Neto, J. A., & Becker, K. (2018). Relating conversational topics and toxic behavior effects in a MOBA game. Entertainment computing, 26, 10-29.

Discord. (2022). Discord Transparency Report: October – December 2022. Retrieved from https://discord.com/blog/discord-transparency-report-q4-2022

Dougherty, D. S. (1999). Dialogue through standpoint: Understanding women's and men's standpoints of sexual harassment. Management Communication Quarterly, 12(3), 436468. https://doi.org/10.1177/0893318999123003

Ewoldsen, D. R., Eno, C. A., Okdie, B. M., Velez, J. A., Guadagno, R. E., & DeCoster, J. (2012). Effect of playing violent video games cooperatively or competitively on subsequent cooperative behavior. Cyberpsychology, Behavior, and Social Networking, 15(5), 277-280.

Fox, J., & Tang, W. Y. (2017). Women's experiences with general and sexual harassment in online video games: Rumination, organizational responsiveness, withdrawal, and coping strategies. New media & society, 19(8), 1290-1307.

Fredman, L. A. (2018). Not just a game: sexual toxicity in online gaming hurts women. [Doctoral Dissertation, The University of Texas at Austin]. Texas Scholar Works. Retrieved from https://repositories.lib.utexas.edu/handle/2152/75004

Gosh, A. (2021). Analyzing toxicity in online gaming communities. Turkish Journal of Computer and Mathematics Education 12(10). Doi: 10.17762/turcomat.v12i10.5182.

Johnson, C. A. K. (2022). Whisper networks in organizations: Sexual harassment protection through informal connections. [Doctoral Dissertation, Iowa State University]/ ProQuest Dissertations Publishing.

Herring, S., Job-Sluder, K., Scheckler, R., and Barab, S. (2002). Searching for safety online: managing "trolling" in a feminist forum. Inf. Soc. 18, 371–384.

Hilvert-Bruce, Z., & Neill, J. T. (2020). I'm just trolling: The role of normative beliefs in aggressive behaviour in online gaming. Computers in Human Behavior, 102, 303-311.

Holland, K. J., & Cortina, L. M. (2017). "It happens to girls all the time": Examining sexual assault survivors' reasons for not using campus supports. *American journal of community psychology*, *59*(1-2), 50-64.

Koehler, D., Fiebig, V., & Jugl, I. (2022). From gaming to hating: Extreme-right ideological indoctrination and mobilization for violence of children on online gaming platforms. Political Psychology. doi: 10.1111/pops.12855

Kordyaka, B., Klesel, M., and Jahn, K. (2019). "Perpetrators in league of legends: scale development and validation of toxic behavior," in Proceedings of the 52nd Hawaii International Conference on System Sciences, Hawaii, 2486–2495.

Kowert, R., & Cook, C. (2022). The toxicity of our (Sim)

Cities: Prevalence of dark participation in games and perceived effectiveness of reporting tools. Proceedings of the 55th HAwaii International Conference on System Sciences.
Kowert, R., & Crevoshay, E. (2023). Harassment of game makers: Prevalence and impact. F1000Research, 11(1518), 1518.
Kowert, R., & Oldmeadow, J. (2012, June). The stereotype of online gamers: New characterization or recycled prototype. In Nordic DiGRA: Games in Culture and Society conference proceedings. Tampere, Finland: DiGRA.
Kowert, R. & Woodwell, L. (2022). Moderation challenges in digital gaming spaces: Prevalence of offensive behaviors in voice chat. *A white paper by TakeThis.* Retrieved from www.takethis.org/wp-content/uploads/2022/12/takethismodulatereport.pdf
Kwak, H., Blackburn, J., & Han, S. (2015, April). Exploring cyberbullying and other toxic behavior in team competition online games. In Proceedings of the 33rd annual ACM conference on human factors in computing systems (pp. 3739-3748).
Lamphere-Englund, G., Hamonagan, A., & Putri, F. (2022). Pathways of resilience to violent extremism in indonesian higher education: A mixed method study using the building resilience against violent extremism (BRAVE) approach. USAID Harmoni. Jakarta, Indonesia. Retrieved from http://lovefrankie.link/BRAVE_reports
Meza, S. (2017, November 22). What is a whisper network? How women are taking down bad men in the #MeToo age. Newsweek. https://www.newsweek.com/what-whisper-networksexual-misconduct-allegations-719009
McLean, L., & Griffiths, M. D. (2019). Female gamers' experience of online harassment and social support in online gaming: A qualitative study. International Journal of Mental Health and Addiction, 17(4), 970–994. https://doi.org/10.1007/s11469-018-9962-0
Microsoft. (2022). Xbox transparency report. Retrieved from https://www.xbox.com/en-US/legal/xbox-transparency-report
Paaßen, B., Morgenroth, T., & Stratemeyer, M. (2017). What is a true gamer? The male gamer stereotype and the marginalization of women in video game culture. Sex Roles, 76, 421-435.
Quinn, B. A. (2002). Sexual harassment and masculinity: The power and meaning of "girl watching". *Gender & Society*, 16(3), 386-402.
Reid, E., Mandryk, R. L., Beres, N. A., Klarkowski, M., & Frommel, J. (2022). "Bad Vibrations": Sensing Toxicity From In-Game Audio Features. *IEEE Transactions on Games*, *14*(4), 558-568.
Ruvalcaba, O., Shulze, J., Kim, A., Berzenski, S. R., & Otten, M. P. (2018). Women's experiences in eSports: Gendered differences in peer and spectator feedback during competitive video game play. *Journal of Sport and Social Issues*, 42(4), 295–311. https://doi.org/10.1177/0193723518773287
Saarinen, T. (2017). Toxic Behavior in Online Games. Master's Thesis, University of Oulu, Oulu, FL.
Shachaf, P., and Hara, N. (2010). Beyond vandalism: wikipedia trolls. J. Inf. Sci. 36, 357–370.
Shores, K. B., He, Y., Swanenburg, K. L., Kraut, R., and Riedl, J. (2014). "The identification of deviance and its impact on retention in a multiplayer game," in Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, New York, NY: ACM, 1356–1365.
Tang, W. Y., & Fox, J. (2016). Men's harassment behavior in online video games: Personality traits and game factors. *Aggressive Behavior*, 42(6), 513–521.
Fox, J., & Tang, W. Y. (2017). Women's experiences with general and sexual harassment in online video games: Rumination, organizational responsiveness, withdrawal, and coping strategies. *New Media & Society*, 19(8), 1290-1307. https://doi.org/10.1177/1461444816635778
Thacker, S., and Griffiths, M. D. (2012). An exploratory study of trolling in online video gaming. Int. J. Cyber Behav. Psychol. Learn. 2, 17–33. doi: 10.4018/ijcbpl.2012100102
Todd, P. R., & Melancon, J. (2019). Gender differences in perceptions of trolling in livestream video broadcasting. *Cyberpsychology, Behavior, and Social Networking*, 22(7), 472-476.
Trudgett-Klose, L. H., & McLinton, S. S. (2022). "Pro Gamers" & cyberbullying: Workplace bullying & sexual harassment in professional video gaming. *SSRN Electronic Journal*.
Twitch. (2022). Moderation in channels: Coverage, removals, and enforcements. Retrieved from https://safety.twitch.tv/s/article/H1-2022-Transparency-Report?language=en_US#3ModerationinChannels:Coverage,Removals,andEnforcements
Quinn, B. A. (2002). Sexual harassment and masculinity: The power and meaning of "girl watching". Gender & Society, 16(3), 386-402.
Zsila, A., Aruguete, M. S., Shabahang, R., & Gabor, O. (2022). Toxic behaviors in multiplayer games: Prevalence, perception, risk factors of victimization, and psychological consequences.
Zubek, R., and Khoo, A. (2002). Making the Human Care: On Building Engaging Bots (No. SS-02-01). Evanston, IL: Northeastern University Computer Science, 1–5.