The toxicity of our (Sim) Cities: Prevalence of dark participation in games and perceived effectiveness of reporting tools

Rachel Kowert
Take This
rachel@takethis.org

Christine L. Cook
New Jersey Institute of Technology
christinelcook@outlook.com

Abstract

Dark participation in games (i.e., trolling and toxic behavior) have been gaining ever-increasing academic attention as a negative aspect of online gaming. Much of the literature in this area has focused on the personality and identity of the perpetrators, but this has been largely outside of the gaming context. The present study aims to explore the prevalence rates of dark participation in the online gaming community, the reporting function to punish deviant players, and the importance of dual identities (troll and gamer) in the perpetration of deviant in-game behaviors. Our results indicated that nearly all players in our sample had been victims of dark participation or witnessed in-game victimization, suggesting that it is a major problem in the community, but that many players also use the reporting function. Troll identity was predictive of these behaviors. Theoretical and practical implications are discussed.

1. Introduction

Gamer cultures emerged in the 1970s alongside the popularization of arcades. These dimly lit buildings became safe havens for those with a shared interest in experimenting with new, digital toys called video games. Gamer identities originally developed over a shared love for what was once a niche activity. However, as games have grown in scope so have their associated "gamer" identities. The term gamer and its associated identity are no longer reserved for the obscure group of individuals who ventured to arcades but rather has become its own subcultural movement a ubiquitous part of our vernacular, and widely integrated into personal and cultural identities [1].

However, over time gamer identities have become less associated with a welcoming and open group bonded by universal love of games and more associated with exclusion than inclusion [2]. These patterns of exclusion and hostility have come to be generally housed under the phrase "toxic gamer cultures" [3]. This phrase not only refers to the prevalence of deviant behaviors within games but also the dismissal of one's responsibility for them under the shared idea that it is just part of the "anonymous and toxic gamer" collective identity. Both of these characteristics - anonymity and toxic gamer collective identity - are key points to discuss in the context of dark participation.

The Social Identity Model of Deindividuation Effects (SIDE model) posits that deindividuation or depersonalization of group members emphasize the similarities of members within a group and encourage behaviors consistent with group norms [4]. That is, the more anonymous a person is, the more deindividuated they are, and the more likely they are to adhere to group norms. Research has already found that individuals in online games experience high levels of "disinhibition" online; that is, a sense of unrestrained freedom because of the reduction of concerns about being judged or suffering real-world repercussions because of the relative anonymity afforded by online interactions [5]. This sense of disinhibition is often more formally discussed as the Online Disinhibition Effect [6]. It is worth noting that there are two components of this effect, benign disinhibition (i.e., promotes openness, kindness, and generosity) and toxic disinhibition (i.e., promotes rude language, hatred, and threats) [6], [7]. These components are interrelated and have both been associated with dark participation online, such as flaming and cyberbullying [7]–[9]. However, toxic disinhibition has been found to be more influential at predicting dark participation over time [7]. More recent work has also noted the potential roles of victimization, attitude, and behavioral control in perpetrating dark participation within these spaces [10]. While the Online Disinhibition Effect was originally developed to discuss inhibition effects online generally, it has been extensively discussed within the context of games [10]–[12]

The impact of the online disinhibition effect on toxic behavior is not unique to games. However, what is unique to games is the intermingling of this behavior within the collective gamer cultures and identities [13]. While the term gamer is often used as a shorthand to organize the world into people who play video games and people who do not, self-identifying as a gamer signifies a shared identity with other members of the broader gaming community and culture and denotes an alignment with the group's idiosyncrasies, traditions, and social practices. The gamer identity is a key component because it is cultural capital. As discussed by Grooten and Kowert [1]: "[…] *Being*

URI: https://hdl.handle.net/10125/79724
978-0-9981331-5-7

a 'gamer' is more than just a label given from the outside; it is a part of one's self-conception and an expression of one's affiliation with a group of society" (para 5).

If the social practice within any particular social group favors deviant or derogatory behavior, then a member of that group is more likely to exhibit that kind of behavior [14], [15] to avoid social ostracization [16]. For example, Amiot and colleagues [14] found that derogatory behaviors against an opposing ice hockey team reduced when these behaviors were punished by other fans (i.e., their "in-group") but were reinforced when they were not punished by other members of the in-group (for a more thorough discussion of in-group out group dynamics and its impact on behavior, see Turner [17]). Thus, if you consider yourself part of the social group of gamers and the in-group norm is toxic behavior towards others (e.g., griefing, harassing, doxxing), you may be more likely to engage in that kind of behavior.

For example, several scholars have noted sexist beliefs and practices within gaming cultures [3], [18], [19] as well as the perception of gaming spaces as male-dominated spaces [20], [21] and "boy's toys" [22]. As noted by Consalvo [3], these beliefs and attitudes persist through fan and player networks who magnify (rather than diffuse) these elements of "toxic gamer cultures" (i.e., members of the "in-group"). Other factors unique to games that may also contribute to the perpetuation of these kinds of behaviors in gaming culture, such as othering (i.e., "us" vs "them" dynamic of many games) and the social reinforcement of game play [23]–[25].

These kinds of toxic cultures within games have come to be so commonplace that some researchers have postulated that the prevalence of these behaviors could lead to their acceptance of their normalization [26], [27]. In fact, there is evidence to suggest this shift has already begun in games [28], and has been at work on the internet more generally for years [29]–[31].

1.1 Dark participation and toxic gamer cultures

While many people have discussed toxic gamer cultures and specific facets of dark participation (i.e., trolling) and behaviors, there remains a lot of confusion about what precise behaviors are considered "toxic", as different researchers have historically used different criteria to describe the same things). This is, at least partially, due to the fact that it is a relatively new field of study. There are only a handful of studies on toxic gamer cultures and many of them have been atheoretical [32]–[34]. This is further confounded by the multiplicity and diversity inherent to gaming; different types of toxicity can exist in different gaming genres [35], [36], over and above the various affordances of consoles, computers, and mobile phones [37]. With all this variety, it has been difficult for researchers trying to pin down a hard and fast definition of toxicity and the precise actions that constitute toxic behavior.

A further confound to consider in regard to negative online behaviors is cyberbullying. Personal insults can also fall under this category of negative online interaction, and the specifying a possible repetition criterion for trolling overlaps with the criteria used to classify something as cyberbullying [38]. Altogether, though this gives researchers plenty of resources with which to discuss and categorize online interactions as trolling or otherwise, it has left the field of trolling research largely amorphous and disjointed, spread across fields, disciplines, platforms, and populations [36].

In an attempt to bring cohesion to this area of research and develop a shared language about toxicity, trolling, and its affiliated concepts, Kowert [39] developed a catalogue of deviant behaviors in games which she discusses under the umbrella of "dark participation". Kowert notes 17 different behaviors that range in scope from transient to strategic, and verbal to behavioral. A verbal action is one that is one expressed verbally (via voice chat or text) from one player to another, whereas a behavioral action is either performed with one's in-game character or triggers an "out-of-game" action [35]. Transient refers to any action committed "in the moment", whereas strategic implies a player took time to formulate a plan before performing the action. Specifically, Kowert [39] identifies the verbal actions of trash-talking, misinformation, spamming (verbal), griefing, sexual harassment, hate speech, threats of violence, and flaming, and the behavioral actions of spamming (behavioral), inappropriate role-playing, contrary play, inhibiting team, aiding the enemy, in-game cheating, hate raiding, doxxing, and swatting.

Kowert [39] also distinguishes between "dark participation" (any deviant action that takes place within online games), toxicity, and trolling. She notes:

Dark participation is any deviant action that takes place in online spaces, but what constitutes toxic behavior is often culturally defined. Put another way, dark participation is always deviant in the context of the environment but what behaviors that are considered toxic (i.e., behaviors that cause harm to another's health or well-being) in one situation might not be considered toxic in another. Toxicity refers to particular outcomes of dark participation, trolling refers to the intent of the perpetrator (p. 4).

Thus, dark participation refers to any deviant action that takes place in games, whereas toxic behaviors are actions that cause harm to another player and can vary from culture to culture. For example, trash-talking may be considered appropriate in some cultures and contexts (e.g., eSports competition) but not in others. Last, trolling behaviors are those that are done strategically for malintent, such as griefing or doxxing.

1.2 Prevalence rates

Colloquially, toxic cultures are often described as a cornerstone of gaming cultures. Recent research evaluating the prevalence of toxic cultures in gaming spaces supports this idea. A 2020 study by Cary and colleagues found that 80% of players reported that the average gamer makes prejudiced comments while playing online [40]. The 2020 Bryter report [41] indicated that over half of male and female players have experienced abuse in games, and nearly a third (28%) reported they experienced it regularly. This report also noted that 1 in 4 female players reported that the "widespread toxicity" in games made them feel upset, intimated, and made them not want to play again. A 2019 report from the Anti-Defamation League (ADL) reported that 74% of adults who play online multiplayer games in the US experience some form of harassment while online [42]. Cary and colleagues [40] found over half of their surveyed players (53%) said they experienced harassment because of their race/ethnicity, religion, ability, gender or sexual orientation and 65% had experienced some form of severe harassment, including physical threats, stalking, and sustained harassment. They also found that nearly 1 in 3 (29%) of players have been doxxed (which is where personal identifiable information is posted publicly online, such as your address and phone number). Taken together, this suggests that more than half of all players have experienced some form of harassment while playing online, much of which could fall under legal categories of hate speech, racism, and sexism: serious offenses with serious legal consequences when committed outside of an online game [43]–[45].

While many studies have noted the frequency of toxic behavior in general it remains unclear to what frequency specific kinds of behaviors occur in gaming spaces. For example, do the more severe behaviors (i.e., doxxing, swatting) occur less frequently than less severe behaviors (i.e., trash-talking, contrary play) or are more severe behaviors more common? Notably, not all acts of dark participation occur because of the same motivations [35]. Thus, understanding what behaviors occur most frequently is key to developing targeted plans for extinguishing the most common occurrences of dark participation in games.

1.3 Combatting dark participation: reporting tools

Currently, the standard approach from the video game industry to combat dark participation in gaming spaces has been the installation of in-game reporting tools. Every console and gaming platform has some kind of reporting tool available (visit www.ESRB.com for more information on reporting tools); although, it is unclear how often people use these tools and/or their effectiveness for combating deviant behavior in games. In fact, contrary to the aims of creating the tools in the first place, there seems to be a general consensus that reporting tools are ineffective. The ADL [42] reported that 62% of players think companies should do more to make online games safer and more inclusive. In their interviews with online perpetrators (i.e., "trolls"), Cook and colleagues [35] also found that reporting deviant behavior was the least-used recourse by bystanders and victims alike when faced with a toxic interaction. This was further evidenced in their follow-up study - a mixed-methods content analysis of League of Legends (Riot Games) - which showed that victims and bystanders of these behaviors are actually very likely to begin to emulate the very behaviors they purport to despise, using plenty of typographic energy via caps lock and profanity [46]. This would suggest that, not only does the perceived ineffectiveness of reporting tools lead to their extremely limited usage but may also contribute to further reciprocal toxicity on the part of victim gamers.

It is worth noting that some companies have tried different strategies beyond reporting tools, but to limited success (e.g., the tribunal system from Riot games). For example, some game developers have experimented with more in-game-focused methods of peacekeeping, such as Team Fortress 2's (Valve) kicking system, in which players can be voted out of a game instance by the other players themselves. However, these strategies were often criticized due to their putting the onus on the player to police their own game instead of the game developers taking responsibility for protecting their users, and being a source of further trolling and toxicity, as players abuse it to unfairly report inoffensive players they simply do not like or want in their game [47]. Other games relegate trolls, toxic players, and cheaters to their own special server in an attempt to get them all to torture one another without disturbing other paying customers [48]. However, none of these options offer long-term solutions to the issue of toxicity, over and above their ethically dubious nature. In short, it seems as though reporting is the best option players have, but its effectiveness remains questionable.

2. Current study

The current study will aim to elucidate the state of dark participation and toxic behavior in games by assessing the frequency of these behaviors in game spaces and assess the frequency of use and perception of effectiveness of reporting tools. More specifically, we will address the following research questions:

RQ1: What are the prevalence rates of witnessing, being a victim, engaging in, and reporting in-game toxic behavior?

RQ2: What is the role of identifying as a gamer in perpetrating and reporting different forms of toxicity?

RQ3: What is the role of identifying as a troll perpetuating and reporting different forms of toxicity?

With these questions, we aim to provide critical descriptive information to academia and industry alike in order to give direction and priority to efforts to protect victims from particular types of toxic behavior.

2.1 Participants

Of the 454 participants recruited via Twitter and Facebook posts, as well as the website SurveyCircle, 72 did not complete the survey in its entirety, 2 were too young to complete the survey, and 3 did not consent to their results being used, leaving us with 377 participants for our analyses.

In terms of gender identification, 244 identified as men (64.7%), 98 as women (26.0%), 18 as non-binary (4.8%). Their ages ranged from 18 to 59 (M = 31.79, SD = 7.18), and the majority were from the United States (n = 215, 57.0%), then Canada (n = 48, 12.7%), followed by the United Kingdom (n = 28, 7.4%). The majority of the rest hailed from different parts of Europe (n = 57, 15.1%), Asia (n = 8, 2.1%), and Oceania (n = 8, 2.1%). Participants also self-reported their English proficiency; the majority were either fluent or native speakers (n = 339, 89.9%).

In terms of their gaming experience, on average, 92 (24.4%) played less than 5 hours of multiplayer games per week, 187 (49.6%) from 5 to 20 hours a week, 50 (13.3%) played from 21 to 30 hours a week, 25 (6.6%) from 31 to 40 hours a week, and 23 (6.1%) played over 40 hours a week. For single-player games, 85 (%) played less than 5 hours per week, 217 (57.6%) from 5 to 20 hours a week, 49 (13.0%) played from 21 to 30 hours a week, 19 (5.0%) from 31 to 40 hours a week, and 7 (1.9%) played over 40 hours a week. 88.3% (333) professed to having been victims of toxicity, 85.9% (324) had been bystanders to toxicity, and 45.0% (168) admitted to having engaged in toxicity themselves, with four (1.1%) participants neither admitting to nor denying having engaged in toxicity in the past themselves. The specific types of toxicity they had experienced, witnessed, or engaged in were taken from the list of dark participation terms as outlined by Kowert [39].

2.2 Procedure and Materials

After receiving institutional review board approval from the ethics committee of a mid-sized university in the United States, the authors posted a link to the survey – hosted by SurveyGizmo.com (now Alchemer) – in their social networks (*Twitter* and *Facebook*), and encouraged participants to share the link further. The survey was also shared on SurveyCircle, a platform for mutual participant recruitment for survey-based studies, but only one participant was recruited in this way.

Once they had signed the online consent form, participants provided basic demographic details (age, gender, nationality, and English proficiency). They were then given the following definition of toxicity: Toxicity is usually defined as either deviant or antisocial behavior that negatively affects the gaming experience of at least one other player (adapted from the definition of trolling in Cook [46]). Participants then indicated whether or not they had been a victim of, bystander to, or perpetrator of toxicity while gaming. After each of these questions, if they indicated "yes", participants were presented with a list of possible types of toxicity and asked to indicate which of these they had either experienced, witnessed, or performed. This list of behaviors was taken from Kowert [39]). Participants then completed Buckels and colleagues' [49] Global Assessment of Internet Trolling (GAIT) measure, which consists of four items (e.g., "The more beautiful and purer a thing is, the more satisfying it is to corrupt.") designed to capture various attitudes and behaviors associated with trolling, $\alpha = 0.72$.

Participants also answered questions addressing toxicity in gaming more generally by rating on a scale of 1 (very toxic) to 5 (very positive) to rate how toxic/positive they felt the gaming community as a whole was, and how toxic/positive they felt the gaming communities in which they participate are. This segued into the question of whether or not they had ever used reporting tools in games. If they responded "yes", they were asked to rate how effective they felt these tools were, on a scale from 1 (very ineffective) to 6 (very effective). If they responded "no", they were given a series of reasons why they had never used them (e.g., "There has never been a player I wanted to report.") and asked to indicate which applied.

Lastly, participants filled out Doosje an colleagues' [50] four-item (e.g., "I see myself as a

gamer.") gamer identity scale using a 7-point Likert-type scale in which 1 was "strongly disagree" and 7 was "strongly agree" ($\alpha = 0.87$), and provided an estimate for how many hours a week they played multiplayer and single-player games.

3. Results

All of the possible forms of dark participation were experienced, witnessed, and/or performed by participants. The most common form of toxicity experienced, witnessed, and performed was trash talking. The most serious forms of dark participation (i.e., hate raiding, doxxing, and swatting) were the least experienced, witnessed, and performed actions among participants. A summary of participants' experiences and dark participation behaviors is outlined in Table 1.

Table 1. Percentage of participants who have experienced (E), witnessed (W), or performed (P) types of dark participation.

Trolling Type	E	W	P
Trash Talking	94.0	98.1	84.5
Misinformation	38.7	57.7	17.9
Contrary Play	59.5	71.6	30.4
Inhibiting team	72.1	77.5	27.4
Aiding enemy	61.9	72.2	25.6
Inappropriate Role-playing	22.8	41.7	8.3
Verbal spamming	72.7	80.6	18.5
Griefing	79.0	82.3	28.0
Sexual harassment	45.0	71.3	4.2
Hate speech	64.3	82.7	6.0
Violent threats	46.8	67.6	7.1
Flaming	60.1	75.3	31.5
In-game cheating	65.8	69.8	8.9
Hate raiding	16.5	36.4	3.6
Doxxing	11.1	24.1	1.2
Swatting	1.2	9.0	0.0
Other	3.0	1.5	3.0

3.1 Reporting in online games

Results indicated that most participants who witnessed toxicity, reported it. While 85.9% (324) of our participants were bystanders to toxicity, 85.4% (322) reported it, 12.5% (47) did not report it, and 2.1% (8) were unsure if they had reported toxicity or not. Of the 47 people who did not use reporting tools, 46.8% (22) claimed that there were no trolls in the games they play, while 27.7% (13) claimed that reporting functions did not exist in the games they played. Only 17% (8) said that they did not use reporting tools because they were ineffective, which is particularly interesting when one considers that on average most of our sample found reporting mechanisms to be ineffective ($M = 2.81$, $SD = 1.18$). This suggests that players are using the reporting functions in games, but do not believe that this actually does anything to prevent future toxicity. Finally, 8.5% (4) said that they did not know how to use the reporting tools available in the games they play, and 6.4% (3) claimed that the reporting functions in the games they play were too heavily abused, meaning that people were reporting other players without a valid reason.

An additional 23.4% (11) of participants, when asked why they did not use the reporting tools in the games they play, responded with the "other" category. Some of these answers include players turning off the game when trolling occurs as opposed to reporting (P188), curating one's gaming experience to include only known others they have met offline (P121), or because the moderators of a particular community are uninterested in their job (P53). One participant went so far as to call reporting "a nonsensical waste of energy" (P238). However, the most common answer was simply that they did not care if toxicity occurred or not

(P194; P196; P352). Thus, it would appear that alternate strategies and sheer indifference are also important variables to consider when asking why players seem ill-inclined to use reporting tools.

3.2 Gender, reporting, and perpetration

As historically, the vast majority of gaming research has compared outcomes between only men and women, we ran a logistic regression comparing only those who identified as cis men and women, while including other variables of interest. There was no significant difference (p = .99) between men and women's reporting rates. We did, however, find a significant effect for age (β = -.08, $p < .001$), with younger players being more likely to report toxicity than older players (see Table 2).

In terms of perpetration, those who engaged in toxic behaviors were predominantly cis-gendered men (n = 124), followed by cis-gendered women (n = 32). Because there were overwhelmingly cis men and so few other gender identities that admitted to trolling and toxicity, we did not include gender as a predictor in the individual analyses of each toxic behavior.

Table 2. Logistic model predicting reporting behavior

			Confidence Interval	
Model	**β**	**SD**	**Lower**	**Upper**
Intercept	2.99**	0.97	3.07	138.97
Gender	-0.01	0.40	0.47	2.24
Gamer identity	0.36**	0.11	1.15	1.80
Troll identity	-0.23	0.24	0.51	1.30
Age	-0.08**	0.02	0.89	0.97

*Note: **p = 0.1*

3.3 Gamer identification in trolling and reporting

To test whether gamer or troll identification had any bearing on performing different types of trolling behavior, we ran a series of logistic regressions with these two variables as predictors of whether the participant used or did not use reporting tools (those who were unsure were omitted). Because of previously reported connections between a person's age and their trolling style [35], we also included age as a predictor to see if it connected to any particular types of toxicity that may overlap with trolling types. A stronger gamer identification was only significantly predictive of engaging in trash-talking (β = 0.25, $p = .01$, CI [1.08, 1.53]); it was not associated with the perpetration of any of the other types of toxicity listed in the study (all $ps \geq .12$). This would suggest that identifying as a gamer does not really predict toxic behavior in games, although it may be associated with increased levels of behavior associated with competition, like trash-talking is in sports [51]. However, gamer identity was predictive of reporting behavior (β = 0.36, $p = .001$, CI [1.15, 1.80]), with a stronger identification leading to increased reporting behavior (see Table 2). Taken together, these results would suggest that the average gamer is unlikely to exhibit more toxicity than trash-talking and is likely to report toxicity when they see it.

3.4 Troll identification in trolling perpetration and reporting

A stronger troll identification was significantly predictive of every single type of toxicity exhibited in our sample. Not only does this validate the GAIT [49] in a new, gaming sample, it also suggests that there is indeed an element of a person separate from the identity of gamer that makes people exhibit toxicity in games. No matter how strongly a person identifies with gaming culture, that does not appear to be enough to make them a troll, or a toxic user. It is a separate component of a person's identity that is the "troll" within the person. Identification as a troll does not, however, affect the use of reporting tools ($p = .36$). It would seem that, irrespective of a person's own engagement in trolling and troll identity, they still generally use the reporting tools in games.

However, we did find that age was also an important predictor of certain types of toxic behavior. More specifically, the younger a person is, the more likely they are to engage in trash-talking (β = -0.06, $p < .001$, CI [0.91, 0.97]), misinformation (β = -0.06, $p = .05$, CI [0.88, 1.00]), inhibiting one's team (β = -0.09, $p = .001$, CI [0.87, 0.96]), aiding the enemy (β = -0.06, $p = .02$, CI [0.89, 0.99]), verbal spamming (β = -0.12, $p = .001$, CI [0.82, 0.94]), behavioral spamming (β = -0.08, $p = .002$, CI [0.87, 0.97]), and flaming (β = -0.06, $p = .01$, CI [0.86, 1.04]). It is worth noting that these behaviors - unlike more serious ones like threats of violence, sexual harassment, and hate speech - generally represent the more playful side of trolling. Flaming is the exception, but even this has been previously associated with younger, less mature trolls [35]. This result lends further credence to the idea that some, but not all, forms of toxicity are connected to immaturity.

4. Discussion

Toxic gamer cultures have grown to be a distinctive component of gaming spaces. As a relatively new field of study, little remains known

about the overall prevalence of the different kinds of these behaviors in gaming spaces as well as the perception of the effectiveness of the tools to combat them. This study had three primary research questions. The first (RQ1) had four parts: what are the prevalence rates of a) witnessing, b) being a victim of, c) perpetrating, and d) reporting toxic behavior in online games. Our results indicated that, overall, toxicity does appear to be highly prevalent in gaming spaces, with around 80% or more of our participants either witnessing or experiencing toxicity, or both, effectively replicating Cary and colleagues' [40] earlier work on this topic. Reporting appears to be a popular recourse when faced with this toxicity, as over 85% of our participants had previously used the reporting tools available in the games they play. Interestingly, perpetration rates of toxicity were significantly lower in the sample than what the rates of experiencing and witnessing would suggest, as only 44.6% of our sample admitted to engaging in toxic behavior in-game. There could be several possible explanations for this. One possibility is simple social desirability [52]; online toxicity is generally considered negative [35], so admitting to it could be considered socially risky or embarrassing, even with anonymity assured. Another option could be that a very small percentage of the actual players cause the vast majority of the toxic behavior. It is also possible that people may not consider their particular behavior toxic, even when they perceive it to be when other players enact the same behavior (e.g., "I'm not trash talking, but they are"). These kinds of cognitive distortions (i.e., justification and rationalizations) have been found to contribute to deviant social behaviors [53], [54], including cyber bullying [26]. Regardless of the reasoning, however, our results would suggest that toxicity is indeed alive and well in gaming communities, and current measures - including reporting tools - are insufficient to halt its spread thus far.

Our second research question (RQ2) focused on the gamer identity, as described by Doosje and colleagues [50], and whether or not it was connected to perpetrating different kinds of toxic behavior or reporting toxicity when witnessed and/or experienced. The results clearly indicated that one's identification with the social identity "gamer" was unrelated to the perpetration of any kind of toxic behavior listed in the present study. However, the stronger one's identification as a "gamer", the *more* likely they were to engage in reporting, not less. This suggests several things, one being that identifying strongly as a gamer is not enough to make someone engage in toxicity. In Cook and colleagues [35], it is suggested that there is a kind of generational split in trolling (a part of toxicity), with veteran and new players engaging in different kinds of behaviors. The present study's results add nuance to Cook and colleagues' [35] findings, confirming that age has more to do with trolling than just identifying as a gamer. When it comes to reporting behavior, however, it would seem that "being a gamer" motivates players to keep their gaming spaces clean. This could be due to the historically negative perception of gamers, particularly when it comes to issues like #Gamergate and hegemonic masculinity [55]–[57]; people who identify as gamers may want to dispel that stereotype and are thus particularly motivated to report toxicity when they spot it. Another may be simple practicality: players want to win [58]. Toxicity can reduce the chance of winning for the victims of the behavior [36] as several forms of dark participation are tactics to undermine game play, such as contrary play and aiding the enemy [39]. By reporting trolls, players increase their chance of having a better game further down the road. Both of these options require further testing to confirm, but our results thus far would suggest that a strong gamer identity is actually associated with toxicity prevention as opposed to toxicity.

Our final research question (RQ3) was similar to our second, only focusing on the troll identity's role in the perpetration and reporting of toxic behavior. The results were the direct opposite of what we found with the gamer identity: identification as a troll was significantly related to perpetrating all types of toxicity listed in the present study and had no connection to reporting behavior. When taken together with our findings regarding gamer identification, it further suggests that there is a specific personality element that encourages trolling and toxicity. It is worth noting that the scale used to measure troll identification was developed completely independently of the online gaming community [49], and so it would appear that the identity of a troll transcends questions of platform. This idea is further supported by the multitude of personality research that has been dedicated to examining the connection between the Dark Tetrad and trolling on social media [59]–[62] and games specifically [63].

Beyond the specific research questions, we also found three other critical pieces of information: gender is apparently unrelated to reporting toxicity, although age *is* related, and players generally find reporting to be ineffective. All of these could have interesting practical implications for the gaming industry, particularly pertaining to preventing toxic behavior on their platforms. Extant literature would suggest that trolling behavior and toxicity is overwhelmingly perpetrated by men [33], [35], [49], but the same cannot be said of reporting, as the genders in the present study reported toxicity almost equally, and both in great numbers. The result also suggests that although younger players tend to be perpetra trolling behaviors, they are not especially likely to

engage in more serious criminal pursuits like sexual harassment or hate speech and are actually more likely to report toxicity when they see it. However, despite these great numbers, our sample of players generally found that reporting was ineffective. Taken together, this means that gaming companies and platform owners do not need to create gendered interventions to increase reporting behavior to mitigate trolling and toxicity, but rather increase the effectiveness of existing reporting mechanisms, and/or come up with new initiatives to reduce and/or prevent toxicity. These could potentially include options like increased feedback for reporters to let them know what the outcome of their reporting was, or perhaps moving toward a more content moderation-focused strategy instead of an ad-hoc reporting system. These strategies should be compared and contrasted to determine their effectiveness, ideally with independent researchers and platform owners cooperating to ensure high quality data and analysis.

It should also be noted that although this study provides important information regarding prevalence rates of toxicity in gaming today, it has limitations that should be addressed in future work. First, to reduce the burden on participants in an already long survey, we did not ask for specific games or gaming genres that people play. It is possible that the toxicity described by our participants is clustered in specific genres, and this should be investigated further. Second, although we gathered age and gender demographics, as well as identity in terms of 'troll' and 'gamer', we did not ask for anyone's minority or majority group status in their respective countries. Although it was not this study's goal to look at cultural variables, there is evidence in extant literature that minority group members - be they racial, gender-related, or otherwise - can and do experience more and different trolling and toxicity than others [44], [45], [64]. Future studies can address both of these by adding in gaming genre, ethnic background, and sexual identity questions into their study designs to add nuance to what we found in the present work. Finally, this was a self-report study using a convenience sample and therefore can only give an indication of the behaviors happening in the world of online gaming. To grasp exact prevalence rates, researchers working on this topic in the future should aim for nationally representative samples and actual game data to avoid issues of memory and social desirability.

To conclude, there is still work that needs to be done in terms of preventing toxicity in online gaming spaces, but our findings provide some hope for the future. Most players, despite being victimized and watching others get victimized, are doing their best to combat toxicity with the tools available to them, irrespective of gender identity. That said, tools to combat this behavior in games need further refining and innovating to empower players to protect the communities they inhabit. By empowering players, we are confident that we will see them save not only the virtual people in the worlds they love, but the people they share with this world as well.

5. References

[1] J. Grooten and R. Kowert, "Going Beyond the Game: Development of Gamer Identities Within Societal Discourse and Virtual Spaces," *Loading...*, vol. 9, no. 14, pp. 70–87, 2015.

[2] C. A. Paul, *The Toxic Meritocracy of Video Games. Why Gaming Culture is the Worst.* Minneapolis, MN: University of Minnesota Press, 2018.

[3] M. Consalvo, "Confronting Toxic Gamer Culture," *J. Gender, New Media Technol.*, 2012.

[4] T. Postmes, R. Spears, and M. Lea, "Breaching or building social boundaries? SIDE-effects of computer-mediated communication.," *Communic. Res.*, vol. 25, pp. 689–715, 1998.

[5] A. Joinson, "Causes and implications of disinhibited behavior on the Internet.," in *Psychology and the Internet: Intrapersonal, interpersonal, and transpersonal implications*, J. Gackenbach, Ed. Academic Press, 1998, pp. 43–60.

[6] J. Suler, "The Online Disinhibition Effect," *Cyberpsychology Behav.*, vol. 7, no. 3, pp. 321–326, 2004.

[7] R. Udris, "Cyberbullying among high school students in Japan: Development and validation of the Online Disinhibition Scale.," *Comput. Human Behav.*, vol. 41, pp. 253–261, 2014.

[8] N. Lapidot-Lefler and A. Barak, "Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition," *Comput. Hum. Behav.*, vol. 28, no. 2, pp. 434–443, 2012.

[9] A. Görzig and K. Ólafsson, "What makes a bully a cyberbully? Unravelling the characteristics of cyberbullies across twenty-five European countries," *J. Child. Media*, vol. 7, no. 1, pp. 9–27, 2013.

[10] B. Kordyaka, K. Jahn, and B. Niehaves, "Towards a unified theory of toxic behavior in video games," *Internet Res. Electron. Netw. Appl. Policy*, vol. 30, no. 4, pp. 1091–1102, 2020.

[11] H. Kim and Y. Chang, "Managing Online Toxic Disinhibition: The Impactof Identity and Social Presence," in *SIGHCI 2017 Proceedings*, 2017.

[12] R. Kowert, *Video Games and Social*

Competence. New York: Routledge, 2015.
[13] A. Shaw, "Rethinking game studies: A case study approach to video game play and identification," *Crit. Stud. media Commun.*, vol. 30, no. 5, pp. 347–361, 2013.
[14] C. E. Amiot, M. Doucerain, and W. R. Louis, "The pathway to accepting derogatory ingroup norms: The roles of compartmentalization and legitimacy," *Psychol. Sport Exerc.*, vol. 32, pp. 58–66, 2017.
[15] Z. Hilvert-Bruce and J. T. Neill, "I'm just trolling: The role of normative beliefs in aggressive behaviour in online gaming," *Comput. Human Behav.*, vol. 102, pp. 303–311, 2020.
[16] A. H. Hales, D. Ren, and K. D. Williams, "Protect, correct, and eject: Ostracism as a social influence tool," in *The Oxford handbook of social influence*, Oxford: Oxford University Press, 2017, pp. 205–217.
[17] J. C. Turner, "Social categorization and the self-concept: A social cognitive theory of group behavior.," in *Key readings in social psychology. Rediscovering social identity*, T. Postmes and N. R. Branscombe, Eds. Psychology Press, 2010.
[18] J. Breuer, R. Kowert, R. Festl, and T. Quandt, "Sexist games=sexist gamers? A longitudinal study on the relationship between video game use and sexist attitudes," *Cyberpsychology, Behav. Soc. Netw.*, vol. 18, no. 4, pp. 197–202, 2015.
[19] P. Tan, "Hate speech in game communities," 2011.
[20] T. H. Apperley, "Genre and game studies: Toward a critical approach to video game genres," *Simul. Gaming*, vol. 37, no. 1, pp. 6–23, 2006.
[21] M. J. Heron, P. Belford, and A. Goker, "Sexism in the circuitry: female participation in male-dominated popular computer culture," *ACM SIGCAS Comput. Soc.*, vol. 44, no. 4, pp. 18–29, 2014.
[22] K. Lucas and J. Sherry, "Sex Differences in Video Game Play: A Communication-Based Explanation," *Communic. Res.*, vol. 31, no. 5, pp. 499–523, 2004.
[23] R. Rogers, "The motivational pull of video game feedback, rules, and social interaction: Another self-determination theory approach," *Comput. Hum. Behav.*, vol. 73, pp. 446–450, 2017.
[24] R. M. Ryan, C. S. Rigby, and A. K. Przybylski, "Motivation pull of video games: A Self-determination theory approach," *Motiv. Emot.*, no. 30, pp. 347–365, 2006.
[25] N. Yee, "The Demographics, Motivations, and Derived Experiences of Users of Massively-Multi-user Online Graphical Environments.," *Teleoperators Virtual Environ.*, vol. 15, no. 3, pp. 309–329, 2006.
[26] C. D. Ponari and J. Wood, "Peer and cyber aggression in secondary school students: the role of moral disengagement, hostile attribution bias, and outcome expectancies," *Aggress. Behav.*, vol. 36, no. 2, pp. 81–94, 2010.
[27] T. E. Page, A. Pina, and R. Giner-Sorolla, "'It was only harmless banter!' The development and preliminary validation of the moral disengagement in sexual harassment scale," *Aggress. Behav.*, vol. 42, no. 3, pp. 254–273, 2016.
[28] A. Holz Ivory, J. D. Ivory, W. Wu, A. M. Limperos, N. Andrew, and B. S. Sesler, "Harsh words and deeds: Systematic content analyses of offensive* user behavior in the virtual environments of online first-person shooter games," *J. Virtual Worlds Res.*, vol. 10, no. 2, 2017.
[29] W. Phillips, "Meet the trolls," *Index Censorsh.*, vol. 40, no. 2, pp. 68–76, 2011.
[30] W. Phillips, "The House That Fox Built: Anonymous, Spectacle, and Cycles of Amplification," *Telev. New Media*, vol. 14, no. 6, pp. 494–509, 2012.
[31] W. Phillips, "The Oxygen of Amplification. Better practices for reporting on extremists, antagonists, and manipulators online," *Data Soc.*, 2018.
[32] S. Herring, K. Job-Sluder, R. Scheckler, and S. Barab, "Searching for safety online: managing 'trolling' in a feminist forum," *Inf. Soc.*, vol. 18, no. 5, pp. 371–384, 2002.
[33] S. Thacker and M. D. Griffiths, "An exploratory study of trolling in online video gaming," *Int. J. Cyber Behav. Psychol. Learn.*, vol. 2, no. 4, pp. 17–33, 2012.
[34] P. Shachaf and N. Hara, "Beyond vandalism: Wikipedia trolls," *J. Inf. Sci.*, vol. 36, no. 3, pp. 357–370, 2010.
[35] C. Cook, J. Schaafsma, and M. Antheunis, "Under the bridge: an in-depth examination of online trolling in the gaming context.," *New Media Soc.*, vol. 20, no. 9, pp. 3323–3340, 2018.
[36] C. Cook, "Between a troll and a hard place: the demand framework's answer to one of gaming's biggest problems," *Media Commun.*, vol. 7, no. 4, pp. 176–185, 2019.
[37] H. J. Jiow and S. S. Lim, "The evolution of video game affordances and implications for parental mediation," *Bull. Sci. Technol. Soc.*, vol. 32, no. 6, pp. 455–462, 2012.
[38] H. Vandebosch and K. Van Cleemput, "Cyberbullying among youngsters: Profiles of bullies and victims," New Media Soc.,

vol. 11, no. 8, pp. 1349–1371, 2009.
[39] R. Kowert, "Dark Participation in Games," *Front. Psychol.*, vol. 11, 2020.
[40] L. A. Cary, J. Axt, and A. L. Chasteen, "The interplay of individual differences, norms, and group identification in predicting prejudiced behavior in online video game interactions," *J. Appl. Soc. Psychol.*, vol. 50, no. 11, pp. 623–637, 2020.
[41] Bryter, "Female Gamer Survey 2020," 2020.
[42] Anti-Defamation League, "Free to Play? Hate, Harassment, and Positive Social Experiences in Online Games," Washington D.C., 2019.
[43] American Bar Association, "Racial equity in the justice system," 2020. .
[44] K. L. Gray, "Deviant bodies, stigmatized identities, and racist acts: examining the experiences of African-American gamers in Xbox Live," *New Rev. Hypermedia Multimed.*, vol. 18, no. 4, pp. 261–276, 2012.
[45] K. L. Gray and D. Leonard, Eds., *Woke Gaming: Digital Challenges to Oppression and Social Injustice*. Seattle, WA: University of Washington Press, 2019.
[46] C. Cook, R. Conijn, M. Antheunis, and J. Schaafsma, "For whom the gamer trolls: A study of trolling interactions in the online gaming context," *J. Comput. Commun.*, vol. 24, pp. 293–318, 2019.
[47] Shamicide, "Fix kick/ban system," 2015. .
[48] L. Ambalina, "20 video games that give players a hard time for trolling," *The Gamer*, 2018.
[49] E. E. Buckels, P. D. Trapnell, and D. L. Paulhus, "Trolls just want to have fun," *Personal. Indivdiual Differ.*, vol. 67, pp. 97–102, 2014.
[50] B. Doosje, N. Ellmers, and R. Spears, "Perceived intragroup variability as a function of group status and identification," *J. Exp. Soc. Psychol.*, vol. 31, pp. 410–436, 1995.
[51] S. Turkay, J. Formosa, S. Adinolf, R. Cuthbert, and R. Altizer, "See no evil, hear no evil, speak no evil: how collegiate players define, experience and cope with toxicity," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020.
[52] A. . Ryan *et al.*, "In the eye of the beholder: Considering culture in assessing the social desirability of personality," *J. Appl. Psychol.*, 2020.
[53] A. Q. Barriga and J. C. Gibbs, "Measuring cognitive distortion in antisocial youth: Development and preliminary validation of the 'How I Think' questionnaire," *Aggress. Behav. Off. J. Int. Soc. Res. Aggress.*, vol. 22, no. 5, pp. 333–343, 1996.
[54] A. Q. Barriga, J. R. Landau, B. L. Stinson, A. K. Liau, and J. C. Gibbs, "Cognitive distortion and problem behaviors in adolescents," *Crim. Justice Behav.*, vol. 27, no. 1, pp. 36–56, 2000.
[55] A. Braithwaite, "It's about ethics in games journalism? Gamergaters and geek masculinity," *Soc. Media Soc.*, pp. 1–10, 2016.
[56] M. Condis, *Gaming Masculinity: Trolls, Fake Geeks, & the Gendered Battle for Online Culture*. University of Iowa Press, 2018.
[57] A. Massanari, "#Gamergate and the fappening: How Reddit's algorithm, governance, and culture support toxic technocultures," *New Media Soc.*, vol. 19, pp. 329–346, 2017.
[58] M. Mora-Cantallops and M. A. Sicilia, "MOBA games: A literature review," *Entertain. Comput.*, vol. 26, pp. 128–138, 2018.
[59] N. Cracker and E. March, "The dark side of Facebook: The dark tetrad, negative social potency, and trolling behaviours," *Personal. Indivdiual Differ.*, vol. 102, pp. 79–84, 2016.
[60] M. Lyons, A. Messenger, R. Perry, and G. Brewer, "The dark tetrad in Tinder: Hook-up app for high psychopathy individuals, and a diverse utilitarian tool for Machiavellians?," *Curr. Psychol.*, 2020.
[61] E. March, "Psychopathy, sadism, empathy, and the motivation to cause harm: New evidence confirms malevolent nature of the internet troll," *Personal. Indivdiual Differ.*, vol. 141, pp. 133–137, 2019.
[62] K. Masui, "Loneliness moderates the relationship between dark tetrad personality traits and internet trolling," *Personal. Indivdiual Differ.*, vol. 150, 2019.
[63] W. Y. Tang, F. Reer, and T. Quandt, "Investigating sexual harassment in online video games: How personality and context factors are related to toxic sexual behaviors against fellow players," *Aggress. Behav.*, vol. 46, pp. 127–135, 2020.
[64] T. H. Apperley and K. L. Gray, "Digital Divides and Structural Inequalities," in *The Video Game Debate 2: Revisiting the Physical, Social, and Psychological Effects of Video Games*, R. Kowert and T. Quandt, Eds. New York: Routledge, 2020, pp. 41–52.

This research was made possible by the Nati Science Foundation grant #1841354