Bystander Interventions in MOBA Games: Designing and Evaluating Tools to Combat Online Gaming Toxicity through Psychological Insights

Nisha Qiao

n.qiao@students.uu.nl

Game and Media Technology
Utrecht University

Utrecht, Netherlands

Abstract

In the modern age of online multiplayer gaming, toxicity in gaming communities has become a significant issue. This study focuses on the role of bystanders and investigates bystanders' player experiences and decision-making when facing toxic behaviors in the context of Multiplayer Online Battle Arena (MOBA) games. An interview and thematic analysis were conducted to collect the experiences of bystanders who witness toxic behavior and reveal the responses they employ in such challenging situations. Findings from the 10 participants' interview study indicate that 1) bystanders have diverse understanding and perception of toxicity, 2) bystander intervention strategies are affected by incentives as well as inhibitors, and 3) bystanders are dissatisfied with current features in the reporting system. Based on the interview results and 5D intervention theories, in-game tools were designed to encourage bystanders to assist the target and improve the players' gaming experience. The effectiveness of the tools was evaluated through a Likert scale assessment. Results show bystanders have a preference on intervention tool with a direct strategy to point out and denounce toxicity.

Keywords: Online Toxicity, Cyberbullying, Bystander Effect, MOBA (multiplayer online battle arena)

1 Introduction

In the multiplayer online gaming world, especially in the competitive arenas of *MOBA (Multiplayer Online Battle Arena)* games, toxicity has become a severe issue due to its high competitiveness and interactivity, causing a discouraging atmosphere to newcomers. Because of harassment in online games, the number of players that quit playing specific online multiplayer games has been increasing, and over a quarter of young players' online and offline lives are negatively affected [13]. *League of Legends*, one of the most well-known MOBA game from *Riot Games*, has been a mainstay in the gaming world for 15 years and also a mainstay in games being criticized for its negative and abusive in-game player behavior.

A wide range of toxic behaviors in online multiplayer games, including cyberbullying, griefing, hate speech (white supremacist ideology and misogyny), sexism and sexual harassment, intentional feeding, spamming, trolling, cheating, and flaming [13, 14, 23], can ruin the gamer experience. To study and address the issue of toxicity in games, numerous studies have investigated hot topics, for example, the impact of players' toxic behavior on team performance [30], players' perceptions of toxicity and coping strategies [2, 36, 44], and explanation of toxicity normalization [3, 19]. However, while the majority of research concentrates on the perspectives of players, perpetrators of toxicity, and victims [21], only a minor portion, as discussed in [3, 19], examines toxicity from the bystanders' view. Research specifically dedicated to understanding the role of bystanders in online competitive games is exceedingly rare. This research seeks to fill this gap by investigating bystanders' experiences and responses to toxicity in MOBA games.

Therefore, this research turns the spotlight on a group that is often in the background in studies under an online competitive gaming context: the bystanders. Bystanders in games share a similar meaning with bystanders in real life, referring to the players who witness the toxicity but are not its direct targets. Guided by the bystander theories [9, 24, 43] and through qualitative interviews, this study will capture bystanders' perception of toxicity, revealing the process of their decision-making and the range of actions they consider feasible and effective when combating toxicity.

Based on the conclusions from the rich narratives collected through the interview study and grounded in psychological theory, the next phase of this research involved deploying in-game tools designed to motivate bystanders in support of victims of toxicity. The effectiveness of these tools in encouraging bystander intervention were rigorously evaluated through a user survey. This dual-phase approach not only

contributes to the academic understanding of social dynamics in online gaming but also offers practical applications for game developers seeking to create more inclusive and supportive gaming environments.

Three research questions guide this research:
RQ1: How do bystanders perceive and respond to toxicity in online multiplayer games?
RQ2: What informs a bystander's decision to help or support the victim or not?
RQ3: Are in-game tools effective in motivating bystanders to help the victim of toxicity?

By answering these questions, the research aimed to broaden the conversation on online gaming toxicity, giving insights into the potential of bystander intervention as a mechanism for boosting positive player interactions and mitigating negative behaviors in MOBA gaming communities.

2 Related Works

This project involves topics of game community and various branches of psychological concepts, which will be explained and discussed in the following sections. With over 150 million registered players globally, and reaching a peak of 6.4 million viewers in World Championship 2023, breaking the all-time viewership record for a single esports event [7], League of Legends continues to be one of the most-played online games worldwide as of 2023 [41]. This project research is based on the LOL case, which is representative of MOBA games.
League of Legends (LOL) is a match-based team competition game, typically with each match lasting about 20 to 40 minutes. Despite offering various modes, all modes in *LOL* feature competitive gameplay between two teams. LOL's main gameplay mode is a game between two teams of five players in summoner's rift map, competing to destroy the opposing team's "Nexus" located in the base. The "Nexus" is guarded by spawned minions waves and defensive structures known as "turrets" on three lanes and three "inhibitors" in base. The "jungle" area are region between lanes that respawn "monsters" at regular intervals1. Non-player characters including minions and monsters provide gold and XP when getting killed[46]. A significant portion of the strategy centers on the development of individual character skills and cooperative team dynamics in combat [47], therefore, interactions between players, as well as between opposing teams, play a vital role in the game's progression. Research conducted by Märtens et al. [32] reveals that toxic language is evident in approximately 60% of League of Legends matches, a phenomenon that is exacerbated by the intrinsically competitive environment of MOBA games. [31] also mentioned that current research primarily focuses on player experience and toxic behavior.

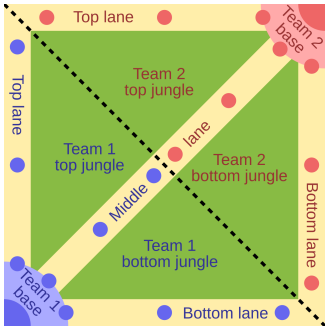

Figure 1. Map of Summoner's Rift: Three lanes, base, jungle area [8]

2.1 Toxicity in Online Gaming

Typology of Toxic Behaviors
The complexity, multifacetedness, and widespread presence of toxicity have consistently posed challenges for game developers [4]. In previous research, toxicity in online games is usually defined as repetitive intentional behavior or abusive negative language that harms other players [4, 22, 23]. The main categories of toxic behavior in MOBA games, referring to the current report system in LOL [14], are as shown in Fig 3 and Fig 2.
Explanations of the typical type of toxic playing behavior [14]:

1. *Griefing*: Griefing's meaning has a broad range. Griefing refers to a form of cyberbullying specific to the online gaming world, where individuals who derive pleasure from hindering or disrupting the gameplay of others are known as "griefers." [12]. It includes sabotaging the game, hindering teammates, spamming in chat, cheating, and using scripts or bugs.

2. *Harassment*: Persistent negative actions or messages aimed at specific individuals. It includes stalking, bullying, or unwanted attention.

3. *Ping/chat spams*: Excessive, repetitive pinging or sending messages without a strategic purpose, often used to criticize, annoy, or distract other players.

4. *Hate speech*: Expressions on targets based on their race, ethnicity, gender, religion, sexual orientation, disability, or any other characteristic that may be subject to discrimination. E.g. homophobia, sexism, racism, and ableism.

5. *Intentional feeding*: The deliberate act of a player repeatedly allowing its character to be killed by the enemy team, providing the enemy with gold or bounty.

Researchers classified toxicity as a form of cyberbullying [4]. Cyberbullying and bystander effects are well-explored

areas in the field of psychology, offering a theoretical foundation for examining similar phenomena within the context of online gaming.

2.2 Conceptualizing Bystander Dynamics

2.2.1 Bystander Definition in Online Gaming. Bystander behavior is not only a real-world phenomenon but also extends to the digital world, where individuals witness others being targeted by cyberbullying or trolling online [28]. Drawing an analogy from the bystander effect in real-life social psychology, where individuals may be less likely to intervene or take action when others are present, bystanders are players who are not directly involved in a particular interaction or conflict between other players within the game. It's important to note that the term "bystander" in the context of online MOBA games doesn't imply a lack of involvement or contribution to the overall game dynamics.

2.2.2 Theoretical Frameworks. *Bystander Effect*: The bystander effect, first proposed after the murder of Kitty Genovese, refers to the social psychological phenomenon that individuals have a lower probability of providing assistance or intervening in an emergency when other people are present, often influenced by a diffusion of responsibility (Latané and Darley, 1968) [9].

Why Does the Bystander Effect Exist?
Latane and Darley [24] identified three different psychological processes in the completion of the bystander effect:

1. *Diffusion of Responsibility*: The first process is a diffusion of responsibility, a phenomenon where individuals perceive their personal responsibility to assist as being divided among the presence of other witnesses. The more bystanders there are, the less personal responsibility any individual bystander will feel.
   Regarding the bystander effect, when the number of passive bystanders grows, individuals are less inclined to assist in an emergency situation [11]. However, in a MOBA game, there are 5 people in a team in general. As the number of bystanders is quite limited (up to 3 players), it naturally forms an environment that easily breaks the diffusion of responsibility in the bystander effect.
2. *Evaluation Apprehension*: The second process is evaluation apprehension, which refers to the fear of being judged by others when acting in public. Research [39] on evaluation apprehension shows that when subjects were required to make a decision, they prioritize acting favorably to researchers–a phenomenon called the *apprehensive hypothesis* or "good subject role." In a competitive gaming context, evaluation apprehension often arises and can be influenced by multiple factors, for example, low performance (which is frequently indicated by a low KDA, KDA: Kill/Death/Assist ratio), causing players to feel "unqualified" to intervene. The intervening action might be perceived as "disrupting victory" by other players within the team. And sometimes, players are worried about the stigmatization that can occur when a player defends a victim of a different gender, e.g., to be mocked as a 'simp' (*Simp* is an internet slang term describing men who submit themselves excessively to women to gain attention, usually in pursuit of affection or a sexual relationship [16]).
3. *Pluralistic Ignorance*: where almost every individual in a group internally rejects what they believe to be the dominant opinions and convictions of the group as a whole. In the competitive game context, pluralistic ignorance could exist in the form that, every bystander player keeps silent assuming other bystanders would prefer to keep silent even if they have some secret intention to help.

Bystander Intervention: The concept of bystander intervention introduced by Darley and Latané [24], posits that a bystander's *degree of responsibility* to act in emergencies is dependent on three factors: 1. Their perception of the victim's need for help; 2. Their ability or competence to provide aid; 3. The relationship between the bystander and the victim. Furthermore, Latané and Darley categorize the means of intervention into two types: 1. *Direct intervention*: offering immediate aid to the victim; 2. *Detour intervention*: this involves seeking help from official authorities, such as the police or fire department. These concepts are foundations of intervention and evolve into *5D* bystander intervention training we know today.

2.2.3 Bystander Intervention Model. Origins: For almost forty years, the Bystander Intervention Model (BIM), introduced by Latané and Darley [24], has been vital in explaining the conditions that determine an individual's decision to either assist or not assist others during emergencies. This model outlines five stages required for a bystander's intervention: 1) notice the event, 2) recognize the situation as an emergency, 3) feel responsible to help, 4) decide how to help, and 5) take action to provide help. Study [10] makes use of the bystander intervention model in the social media context, and its results indicated that bystanders prefer indirect intervention to direct intervention, and users who were exposed to viewers were more likely to increase a sense of personal responsibility.

2.2.4 Existing Tools: 5 D to be The Active Bystander. 5Ds comes from *Green Dot* (renamed as Alteristic), who pioneered the 3Ds of bystander intervention, Distract, Delegate, and Direct. In 2015, *Right To Be*, partnered with *Green Dot*, expanded those to include Delay, and in 2017 they expanded them again to include Document [43].

The 5Ds, namely *Distract, Delegate, Document, Delay, and Direct*, represent five strategies bystanders can employ to aid victims of harassment. These approaches underscore the unacceptability of harassment and convey to individuals the ability they possess to enhance the safety of their community [43]. Crafted to ensure safety and avoid aggravating the situation, these strategies are mostly indirect (4 indirect and 1 direct).

Explanation of the 5D in a general sense:

1. *Direct*: Respond directly to harassment by identifying the inappropriate behavior confronting the person who is responsible for harming.
2. *Distract*: Simply distract the attention of the harassing person by interrupting it. E.g. talk about something completely unrelated.
3. *Delegate*: Delegation is asking a third party (authority) for help with intervening in harassment.
4. *Document*: Documentation involves carefully recording evidence of the harassment while the bystander has to first ensure its safety.
5. *Delay*: After the incident, approach the victim to offer support, showing that they are not alone.

In the context of toxicity in online competitive games, strategies in the "5D intervention tool" can be applied to boost bystander intervention. These strategies empower bystanders to actively address and mitigate toxic behaviors within gaming communities.

2.3 Personal Traits in Bystander Intervention

Big 5 personality traits/Five-Factor model of personality: The Big Five personality traits, known as "the Five-Factor Model of Personality," constitute a framework for analyzing personality through five distinct characteristics [40]. Starting in the 1990s, this model delineates five factors (CANOE) and ten corresponding values. The framework divides each factor into two comparative values as follows [38]:

1. Conscientiousness: (efficient/organized vs. extravagant/careless)
2. Agreeableness (friendly/compassionate vs. critical/rational)
3. Neuroticism (sensitive/nervous vs. resilient/confident)
4. Openness to experience (inventive/curious vs. consistent/cautious)
5. Extraversion (outgoing/energetic vs. solitary/reserved)

Research [37] applied the five-factor model of personality to predict if personality traits affect the probability of a bystander either passively watching cyberbullying online or actively stepping in to protect the victim. Findings revealed that personality traits can predict bystander actions, with higher degrees of agreeableness and extraversion linked to an increased likelihood of defending victims. The study also showed that neuroticism alone was not a significant predictor of bystander intervention among both male and female participants. Further study [20] results show that higher levels of extraversion trait and lower levels of openness, conscientiousness, and agreeableness are related to toxicity perpetration and suggest personalized interventions tailored to individual personalities could be a new approach in combating toxicity.

2.3.1 Social Identity Theory.

Research on altruism indicates that the likelihood of bystander intervention increases when there is a similarity between the helper and the person receiving help. In an experiment in [26], researchers found that bystanders are more willing to assist an injured individual if they are wearing the jersey of a football team the bystander supports, rather than that of a team they dislike. However, when their shared identity as football fans was emphasized, supporters from both teams were more likely to receive significantly more help than an individual wearing a plain shirt.

2.3.2 Passive Observations and Non-Intervention.

In the online competitive game context, literature [3] has discussed the reason why bystanders remain passive to toxic behaviors. Some bystanders who have experienced a high level of toxicity rationalize toxic behaviors as a common thing of games and they are acceptable (experiential factor: Bystanders with experience being a perpetrator are more accepting toward bullying [42]). Some bystanders with high moral disengagement think it is not their obligation to react (psychological factor: moral disengagement).

2.4 Contextual Factors Influencing Bystander Responses

Game design elements, such as reporting systems or in-game moderation, can influence bystander behavior. If these systems are perceived as ineffective or if there's a lack of trust in them, bystanders are less likely to report or intervene in toxicity.

The easiest way for bystanders who are willing to help but do not like chatting is to report the toxic player. Riot has an automated reporting system that is triggered when a specific behavior is detected in the game. For example, a handy *Blocked List* that Riot uses to catch the most blatantly offensive offenses and instantly mute the player in a game [1]. However, the block list is often insufficient, as players can easily bypass it by omitting letters or adding/replacing

numbers or special characters in offensive usernames and messages. [6].

The other part of the reporting system relies on the players themselves. They can report toxic players in and after the game.

Current Shortcomings
1) Reporting System: The reporting system does not have a good reputation among LOL players [36]. Players complain about the reporting system, as it is unclear what happens to the reports after they are sent to Riot. The player who submits the report fails to confirm the status of their reports, sometimes they are notified of the result of their submission saying a penalty is applied while sometimes not.
2) Smart Ping Wheel: Because of the smart ping wheel's question mark shape, the use of Enemy Missing Signal represents different meanings in various contexts now. Except for reporting a missing enemy, the signal means praise for outrageous player skills when someone performs well, but it also signifies players' frustration or anger, or questioning regarding a teammate's absurd actions, ridiculous strategy, or lack of awareness of an approaching enemy. Spam of the enemy missing signal or pinging a player is also regarded as toxic behavior, which is a consensus among nearly all League of Legends players.

- *Reporting System in MOBA games*
  1. League of Legends (current version): The current reporting system in League of Legends allows players to report a player in both ongoing games and post-game. Players can report the player at any time in a game and update their report later. In the post-game report, players can give additional descriptions to the context if toxic behaviors are not included in the selection list, including those that happened in champion select or post-game.

Figure 2. Report system interface in an ongoing game of League of Legends 2023

Figure 3. Post-game report system interface of League of Legends 2023

  2. Old League of Legends Reporting System: The previous League of Legends system employed the *Tribunal* [45], where volunteer players reviewed submitted reports. This system was later replaced by the reporting system as described above.

- *Communication Features in LOL: Smart Ping*
  Smart Ping is a radial menu that includes eight alerts. These can be used to communicate complex strategies within the team faster and more safely than typing.

2.5 Bystander Intervention Research in Online Contexts

Research shows that young people frequently encounter disruptive behavior online, yet only a small portion choose to intervene [25, 34]. Research [18] systematically reviewed the personal and contextual factors that determine bystander actions: "empathy, prior victimization, feelings of responsibility, severity, social norms, relationship with the victim and number of bystanders". Another systematic review examined the use and effectiveness of positive behavior strategies across various environments, aiming to identify ways to adapt these interventions for multiplayer online gaming. It suggests that tootling and good behavior games are the most suitable for application in multiplayer online gaming [29]. Research [17] discussed that emphasizing transparency through public removal explanations intended for the punished users significantly influences the increasing posting

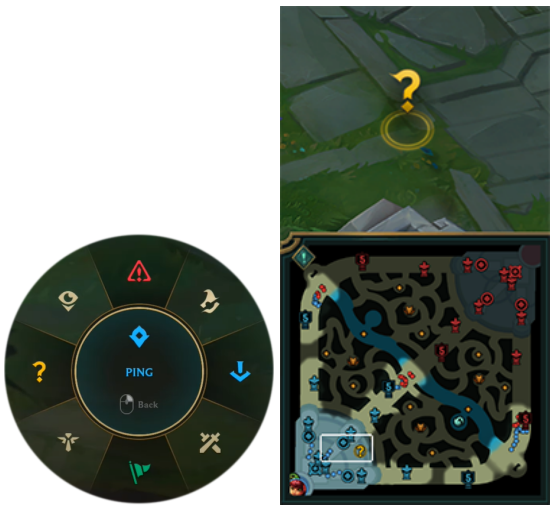

Figure 4. Smart ping wheel and Enemy missing signal

behavior of bystanders. Research [23] highlighted that a clear request to report toxic behavior from a third-party player greatly increases the likelihood of reporting, suggesting that actively encouraging reporting should be integrated into the design of systems to manage toxic behavior.
Research has also shown that individuals must receive bystander behavior education to effectively intervene when an incident occurs [18, 24, 33], so there is a need for Game companies to invest in educating and motivating their communities to raise awareness and encourage players to act as positive bystanders [23], like the "summoner's code" in LOL official game community [15].

3 Research Motivation

While previous studies offer valuable insights, there remains a gap in understanding the specific role of bystanders in gaming environments and how in-game tools can empower them to positively influence the gaming experience. Meanwhile, players have called on a desire for interventions from game developers but so far the success of existing measures is still limited [35]. Researchers in [2] also called out for building tools that encouraging bystanders to take actions in online games. This research aimed to fill that gap by investigating bystander experiences and testing the effectiveness of designed intervention tools.

Here we form the research questions:
RQ1: How do bystanders perceive and respond to toxicity in online multiplayer games?
RQ2: What informs a bystander's decision to help or support the victim or not?
RQ3: Are in-game tools effective in motivating bystanders to help the victim of toxicity?

4 Study 1: Bystander Player Interview

To answer *RQ1: How do bystanders perceive and respond to toxicity in online multiplayer games?* and *RQ2: What informs a bystander's decision to help or support the victim or not?*, this study surveyed how bystanders in a game perceive toxic behaviors. It investigated what kinds of toxic behaviors bystanders consider to be mild, moderate, and severe, whether they will help the victim, and what factors bystanders consider when making decisions about helping and not helping. The insights from the interviews optimized the design of the in-game tools for bystanders in Study 2.

4.1 Methodology

Data Collection:
Interview: Bystander Perception of Toxicity
1) Participants: In the process of registering as an interviewee, before the interview took place, potential participants were contacted through the gaming community and Erasmus Esports Club in the Netherlands, and the competitive gaming social platform (Douban Group) in China. Therefore, 5 interviews were conducted in Chinese and 5 in English. They were selected to ensure they met the predefined recruitment criteria. Participants were those who self-identified as experienced League of Legends players who had been exposed to in-game toxicity. Both gameplay in SR (Summoner's Rift) and HA (Howling Abyss) map count. For the interview, 10 participants over 18 years old were recruited: 6 male and 4 female. On average, participants have a 7.85-year LOL experience (ranging from 2.5 to 11 years). They have a broad variety of game mode preferences (rank(n=6), ARAM(n=4), quickplay/draftpick(n=4)), rank tiers, and preferred positions in the Summoner's Rift map.

2) Interview phase: All interviews were conducted within 2 weeks in late March 2024, which corresponds to Patch 14.6 version in League of Legends. To start with, I verified the participants' information. The selected participants were informed about the purpose of the study, ethical considerations, and the interview process, and signed a written/online consent form to participate in the interview. The interview could be paused or ended at any moment if the participant wished to do so. During the interview, though participants were informed that they could stop at any time if they felt uncomfortable, no interviewees asked to stop, which was either because they were used to the toxicity or because they had a strong desire to talk about the toxicity they were experiencing. The interview took place in a physical location or an online meeting agreed upon by the participant. In the semi-structured interview, I asked questions about the participant's background to collect demographic information and

Table 1. Semi-structured Interview Questions

	Questions
	Participant Background
1.	How long have you been playing League of Legends?
2.	On average, how many hours per week do you spend playing League of Legends?
3.	Which game mode do you usually play?
4.	If you play rank mode regularly, what's your latest rank tier?
	Perception of Toxic Behaviors
5.	In your own words, how would you describe 'toxic behavior' in League of Legends?
6.	How frequently do you encounter what you consider toxic behavior in League of Legends?
7.	In your opinion, what types of behaviors would you classify as mildly toxic in League of Legends? Why?
8.	What behaviors do you believe are severely toxic? How do these differ from mildly toxic behaviors?
	Response to Toxicity
9.	As a bystander in a game, have you ever intervened when witnessing toxic behavior in League of Legends?
10.	If you have intervened, what prompted you to do so, and how did you approach the situation?
11.	If you have intervened, what form did your intervention take?
12.	If you chose not to intervene, could you explain your reasons or concerns at that moment?
	Decision-Making Factors
13.	What factors influence your decision to help or not help a victim of toxic behavior?
	Additional Insights
14.	In your opinion, what could be done to reduce toxic behaviors in League of Legends?
15.	In your opinion, to combat toxicity, what are the drawbacks of the current UI system in League of Legends?

open-ended questions to gain rich insights, as shown in Table 1. Follow-up questions and new topics that emerged during the conversation were asked to obtain deeper insights. Participants were asked open-ended questions about the view of the initial tools design concepts: *"Will you use this tool or not in games when toxicity occurs, why?", "Do you have any suggestions to improve the tools?"*.

3) Post-interview phase: In the post-interview phase, interview recording and notes were backed up securely. Transcription followed, converting the audio recordings into text. Then there was an initial review of transcription, aiming at identifying noteworthy patterns. Then I established a follow-up communication with participants in case of any ambiguities. Each participant was given an 8-euro reward.

4.2 Thematic Analysis

Coding. To analyze the responses, I conducted a thematic analysis proposed by Braun and Clarke [5]. This research used NVivo 14 for the thematic analysis process. After transcription, five transcribed textual materials were linguistically translated from Chinese into English. After reading and getting familiar with the interview data, based on the original BIM (Bystander Intervention Model), I encoded the interview dialogues into several codes and hundreds of subcodes. The thematic analysis was initially carried out deductively, based on the steps in the original Bystander Intervention Model [24]. However, the content of the interviews contains deviations from the BIM, so inductive coding is also used. This qualitative study aimed to gain a rich understanding of the bystanders' perceptions and responses to toxicity, as well as the decision-making process and outcomes of the bystander intervention process in online gaming. Codes addressed both semantic content (i.e., participants' explicit responses) and latent levels (i.e., underlying assumptions within their statements).

Identifying Themes. Regarding the frequently appearing codes, I categorized and re-categorized them to formulate a set of preliminary themes, extracting patterns among them to become the themes and sub-themes below. Overall, the thematic analysis resulted in 109 codes, used to create 4 overarching themes.

4.3 Results

In examining the bystander perception and response to toxicity in MOBA games, the following 2 themes were presented: (1) Perceptions of Toxicity and Definition; (2) Behavioral Dynamics of Responding to Toxicity. Two themes are presented in examining the decision-making process and outcomes of bystander intervention to toxicity in MOBA games: (3) Motivations and Barriers to Intervention; (4) Systemic and Environmental Influence.

Table 2 shows the main themes and sub-themes, and their

components. Note that I coded responses by recognizing that aspects influencing bystander perception (RQ1) are likely to impact the decision-making process for intervention (RQ2). Each theme is discussed below with illustrative quotes labeled by the participant's game mode and LOL experience. Verbatim quotes were lightly edited for readability by removing extraneous fillers and repetitions, preserving the original meaning. The square-bracketed text within the quotes provides contextual supplementation.

4.3.1 Theme 1: Toxicity Perception and Definition. The first theme includes codes related to how toxicity is defined and understood within the gaming community, including what results in toxicity and what is rarely considered to belong to toxic behaviors.

Root of Toxicity. When asking participants about what results in toxic behavior, multiple participants have mentioned that the toxicity comes out of a sense of frustration players gain in games. In general, it is a case when teammates' game skills do not meet their expectations.

> "Toxicity stems from being frustrated. Not really that people are actually mad at each other, but it's the frustration on basically what is expected from another person. E-Sports has become so big and there are a lot of guides out there, right to tell you how to play. It is kind of, that people expect much more from their player base than they should. And then if there's disparity [in matchmaking], then they get really frustrated, and then they become very toxic." (N10, Rank, 9 years)

High-skilled players expect more from high-rank teammates. "In high-rank tier people ping spamming and ask why are you doing this? But in low-rank tier, I would not blame anyone." (N5, Rank, 11 years); "A lot of people can get really frustrated if they don't win and they need to type their frustration to their teammates that are maybe underperforming or not trying their hardest." (N7, Rank, 10 years); "The players that you don't know, they start kicking at you because you're supposedly not as good as they are [rank/level icons show players' tier]. When they start telling you what you're supposed to do, if it's not to their expectations and they lose patience, then it's they who like to become angry about it. It's because you're not learning fast enough." (N8, QuickPlay, 5 years).

Sometimes, the toxicity comes from frustration, and the frustration comes from the high ego (egocentrism) of the toxic player.

> "I think these people are just frustrated, to be honest. The thing about a game is that you like winning. The idea of losing needs to be blamed on someone, that is in their mind, like I have skills, I'm the main character, I'm great. So let's find something to blame or someone to blame as to why I'm losing. It's as if I find a way to release frustration." (N9, ARAM, 7 years)

"In EUW, players pay more attention to the players' enjoyment in games. If you don't satisfy him [the toxic player] in game, he [the toxic player] will probably make some toxic behaviors directly." (N2, QuickPlay, 8 years).

Leaders in the game community could also contribute to environmental toxicity. Participants stated that streamers' toxicity could affect the community and thus affect the environment because normal players may imitate them. "Basically, these big streamers, they're toxic in their streams to other players, so I think that like. It brings the culture into regular games, like people try to imitate toxic streamers" (N6, ARAM, 9 years).

Toxicity Definition. When defining toxic behavior in LOL, most participants give definitions within the range of Riot's official definitions. However, there are some new definitions of toxic behaviors that certain participant experiences frequently. Being too selfish [such as CS/economy stealing] is toxic. "I think if *Graves* [a character in LOL] normally farms, he may win, but if this guy takes three-lane CS and claims that he can carry the game, the chances of winning may increase, but I think he is a selfish person, and this kind of behavior is toxic." (N2, QuickPlay, 8 years).

When arguments occur, a teammate who tries to mediate differences with the sacrifice of principle is also toxic. "For example, when he knows that you have done nothing wrong, but then a teammate starts to verbally attack you for no reason. That is, your other three teammates know that you are innocent. Then they don't say anything at this time, but if you just fight back a word, they just blame you very soon, saying stop it, they don't let the victim explain, that's it. It's not as bad as that kind of direct verbal abuse, but it's actually quite disgusting." (N2, QuickPlay, 8years).

Trying to force others in games is also toxic. "They, like, start being controlling. And then when they get angry at you for not doing what they expected then that's when it becomes severe." (N8, QuickPlay, 5 years).

Toxicity Perception. The perceived severity and frequency of toxic acts have large individual differences. To some participants, verbal abuse is very severe, whilst to others it is toxic but still acceptable compared with hate speech. However, there is an overall trend that verbal abuse, hate speech, cheating, and intentional feeding are severe toxic behaviors, while AFK, offensive name, and negative attitude are less toxic.

Table 2. Themes Structure

Main Themes	Sub-themes	Components
Toxicity Perception and Definition	Root of Toxicity	Frustration, egocentrism, leader effect
	Toxicity Definition	Selfish gameplay style, biasing, coercion
	Toxicity Perception	Severity and frequency
Behavioral Dynamics of Responding to Toxicity	Positive Bystanding	Motivate, support victim, confront perpetrator
	Intervention Outcomes	Toxicity mitigation, escalation, static
Motivations and Barriers to Intervention	Incentives	Intervene to win, worthwhile, enjoyable player experience, out of justice, empathy
	Inhibitors	Distrust, fear of bad consequence, tiredness, Schadenfreuder (players who take pleasure in other's misfortune)
Systemic Shortcomings	Systemic Responsiveness	Penalty insufficiency, unintelligent toxicity detection mechanics
	Systemic Features and Suggestions	Credit system, rewards, Top-down game community management

Severity. Participants reported that verbal abuse is the most toxic behavior. "Well, if it's verbal abuse or afk or something, I think it's quite annoying, but I think verbal abuse is the most annoying." (N3, ARAM, 2.5 years). However, some other participants reported that hate speech and cheating are the most toxic behaviors. "Hate speech is like the worst, and then maybe like griefing and ,like win trading."(N6, ARAM, 9 years). "Cheating is just not fun for everybody that's playing the game. I think it's unfair for everybody, but it's still part of the game, so it's cheating is like the worst thing you can do in the game, but the hate speech is even worse because it can also affect you outside of the game."(N7, Rank, 10 years).

Participants reported a negative attitude, and inappropriate IDs are less toxic. "Least toxic would be negative attitude then offensive name"(N7, Rank, 10 years). "And I personally think offensive or inappropriate names is something that right should deal with it themselves. I feel like this is like the least." (N10, Rank, 9 years).

Frequency. Regarding the frequency, the rank mode, followed by draftpick [normal mode in SR map], contains a higher frequency of toxicity due to its competitive characteristics, whilst the fun modes have less toxicity.

"If I play rank then the the verbal abuse is in almost every game, so it will be or from your teammates or from the enemy team." (N7, Rank, 10 years). "Especially in ranked, it's not in a good state right now. The toxicity is pretty high, I would say. If it's ten rounds, if you rank, you probably have it every game. If you play other modes, it may be less." (N4, QuickPlay, 8 years).

4.3.2 Theme 2: Behavioral Dynamics of Responding to Toxicity.

This theme includes the actions taken by bystanders in response to toxicity, including both direct intervention and other forms of response such as reporting or supporting victims.

Positive Bystanding. Participants reported various strategies of response when encountering toxic behaviors. Some of them would try to give positive feedback to teams, including supporting victims and motivating all team members, while some of them would try to confront the toxic player. 7 participants claimed that if they decided to intervene, they would act positively.

> "My approach of solving this kind of problem is not just to motivate your team, but to give them confidence, to tell them, you don't really need to argue with your own team." (N10, Rank, 9 years)

"Maybe give positive feedback to the guys who are getting flamed on." (N6, ARAM, 9 years). "Now I will seriously tell the victim that there is nothing wrong with his playing, but there is something wrong with the toxic player." (N2, QuickPlay, 8 years). A participant would even send private messages to console the victim after the game. "I think I've done it once where I privately messaged the victim and I asked if he/she was okay." (N8, QuickPlay 5 years)

Meanwhile, 5 participants said they would positively support team players while also taking actions to counter toxic players, as these two types of responses are not mutually exclusive. "It is pointing out that this toxic behavior is wrong." (N1, Rank, 9 years). There are times when bystander intervention form is 'fighting violence with violence', but it's still a way for the bystanders to express their disagreement with the toxic player's behavior. "I will ping that person who is pinging others." (N3, ARAM, 2.5 years). "If that toxic person insists on abusing then I just blame the toxic player. What else can I do?" (N2, QuickPlay, 8 years).

One participant provided an indirect approach: He would suggest that the victim mute the toxic player, which could be categorized as a positive intervention, as it makes the victim aware of possible actions against toxicity. "So most of the time I just mute the person that's typing and being angry. And I advise my team to do the same. I'm trying my best to win the game, and I hope my team is doing the same and not focusing on all of the other people that are maybe not playing." (N7, Rank, 10 years) Another participant suggested convincing the team with reasons. "I mostly try reason. I don't like insulting people. I don't think we need to go through insulting. I go towards reasons and be reasonable. 'This the team game we're playing it together.'" (N9, ARAM, 7 years).

Intervention Outcomes. Being curious about whether bystanders' intervention above affects toxicity interruption or mitigation, I asked about their intervention outcomes. Participants identified 3 potential outcomes following a bystander intervention: 1) Reduced Toxicity: After a bystander intervenes, the toxicity level in the ongoing match decreases, creating a more positive environment. 2) Increased Toxicity: In some cases, intervention leads to an escalation of toxic behavior. 3) Unchanged Toxicity: The intervention has little to no effect, and the toxicity level remains the same throughout the match. The likelihood of each outcome appears fairly random, mainly depending on whether team players are frustrated.

Some participants stated that toxic players may still be arguing and cursing at each other, but to a lesser extent, and eventually, they continue to play the game seriously. "When it actually works, they would still be mad at each other. They would still 'Miss Ping' each other, but they would play the game." *(Miss Ping: spamming question marks to teammates, it is an aggressive way to ask someone "What are you doing?")* (N10, Rank, 9 years).

Participants also mentioned that sometimes intervention works, though the success rate is low. "It works if I say it in a firm manner, but it doesn't have a high rate of being useful, even though I usually say it in a yelling tone of word" (N4, QuickPlay, 8 years). "The toxic person, may shut up and play the game, or they may stop quarreling after persuasion, but the success rate is not high". (N2, QuickPlay, 8 years). Sometimes the toxicity is less because the victim follows bystander's advice to mute the toxic player, "One time where there were multiple other players in our team that were agreeing with me and then we just ended up muting the person." (N8, QuickPlay, 5 years)

In terms of those failed interventions, one of the outcomes is the toxicity gets worse, such as "it escalates in verbal abuse" (N9, ARAM, 7 years) and the intervener starts getting attacked, "they'll start flaming me sometimes" (N6, ARAM, 9 years), suggesting the difficulty of applying the "Direct" strategy in Bystander Intervention Model in a competitive gaming context. A counterintuitive fact mentioned by one participant is that she was attacked by both the toxic player and victim after the intervention, "It's more likely that the two will scold me together, the feeling is I have become the one who joined the arguments by default." (N4, QuickPlay, 8 years).

The other type of failed intervention outcome is the match as toxic as before: "Their result is generally to continue to quarrel, I feel intervention is useless, to be honest." (N3, ARAM, 2.5 years).

4.3.3 Theme 3: Motivations and Barriers to Intervention. This theme covers the internal and external factors that encourage or discourage bystanders from intervening, such as the aim of winning, personal beliefs, fear of retaliation, and perceived effectiveness of the intervention.

Incentives for Intervention: There is a broad range of motivations for intervention. One of the most frequently mentioned reasons is that the bystander is very competitive. Participants, especially Rank players usually try to intervene to ensure everyone can actively play the game thus winning the game.

> "The most important reason is I want to win the game, and if my teammates are being angry or being angry at each other, I think they start to stop focusing on the gameplay and start focusing more on fighting against each other, then I will have a less chance to win." (N5, Rank, 11 years)

Another case for a bystander to intervene actively is when a bystander perceives the toxicity as worth intervention, this could be due to the toxicity being quite severe, or the toxic player is making trouble out of nothing. This motivation is sometimes related to "intervene to win" because bystanders consider the severe toxicity to affect team cooperation.

> "Uh, under what circumstances will I intervene? That is to say, when it is no longer about game skill. For example, if a person insults a person just for playing

> bad, but he curses his parents or families, I think in this situation I will intervene." (N1, Rank, 9 years)

If toxicity within the team becomes excessively severe and constantly affects other teammates' ability to continue playing—such as obstructing their performance or provoking them to join in attacking each other—then it is a situation worth intervening in. "If it's gets really toxic and affects the whole team, I would usually say like, hey, can you cut it out, let's just play the game because maybe it's salvageable to win the game. " (N6, ARAM, 9 years).

One participant mentioned that League of Legends players have a kind of discrimination born of stereotypes against the players who choose to play enchant support charactors, and engage in toxic behavior towards them when at a disadvantage in a game. "Well, just for me, who has no brains and begins to cry for having enchant supports, I will help." (N4, QuickPlay, 8 years); Another participant said that if someone attacks others without provocation, possibly driven by an antisocial mindset, then he will stand out. "If another person keeps insulting the person who didn't do anything, then I must be 100% involved. " (N2, QuickPlay, 8 years);

Bystanders may also intervene because they tend to sustain an enjoyable playing experience.

> "I want to win, but I would like to have a good game of League of Legends. So for example, if I lose a game, right? But every team fight was very drawn out and everything was used correctly and we just lost because someone just played better in the enemy team then it feels really nice because I feel like I've learned something from it or if we've made a really good comeback. Because we played together, we played as a team, and we made the correct choice of calls, right? Yeah. Then this feels very rewarding. And I feel like I've learned something and I enjoyed it because if someone has a dispute in minute 5 and then decides to not play the game correctly, the game is kind of boring. I can't take anything." (N10, Rank, 9 years)

"I used to say, like, just shut up and play the game because everybody queued up and wanted to play the game. Nobody queued up and was like, oh, I'd like to get cursed. " (N7, Rank, 10 years).

In addition, bystanders would intervene out of a sense of justice. "...I think that even in a relatively normal social situation, I will make the same decision, for example, there is a very bad person and bullying a good person. At this time, you feel that you will speak for a good person. It's just moves from a real scene into the game." (N3, ARAM, 2.5 years).
Another intervention reason is the bystanders' empathy for victims, or to themselves. Bystanders sometimes choose to intervene because the situation reminds them of the history of being attacked and helpless. When bystanders put themselves into the victim's situation in the current game, they tend to intervene. "To me, it's like, why would you do this to someone who's learning when you were in the same situation? " (N8, QuickPlay, 5 years); "I think it's because I hate this kind of behavior of watching and then not intervening. So when I see it [toxicity], I definitely don't want to be that kind of person I hate, so I will intervene." (N2, QuickPlay, 8 years);

Inhibitors for Intervention: The reasons and conditions for bystanders to choose not to intervene are complex. A bystander could stay with no action in a game but may intervene actively in the next game. The most frequently mentioned inhibitor is the bystanders' distrust. The distrust includes many aspects, such as distrust of being able to win this game due to the desperate team economy gap or skill gap or banpick diffrence. One participant claimed that if he encounters the toxicity which he would intervene in most cases, but the game situation is like-there is already no chance to win the game, then participant would not intervene. "It depends on the situation of the game, for example, the economy. These things are that I will make a judgment myself, whether there is any hope of winning. If there is any hope of winning, I will intervene, but for the kind of game that it is, maybe at the moment of entering the game. I have already felt you will lose because of the draft diff [differences in character selection before the game starts]. " (N5, Rank, 11 years). Another kind of distrust is the distrust of the system. Some players believe toxic players won't be severely punished or banned even if they are reported for their toxicity. They feel the reporting system prioritizes unauthorized skin-changing behavior (Skin change cheating: using skin changer tools developed by individuals is regarded as a form of cheating by Riot, which could result in account ban punishment.), because it impacts skin sales, which in turn affects Riot Games' revenue. "I feel that the reporting system does not work well. If it is not skin change cheating, it doesn't matter at all. Riot will not punish that." (N1, Rank, 9 years). Another distrust is that players do not believe that their teammates could be convinced to stop being toxic and play in a right way. "I think that quarreling in this kind of game is generally pure emotional venting, so I think if you try to discuss with them and try to comfort them, it seems to be useless. " (N3, ARAM, 2.5 years).

The second often-mentioned inhibitor is the bystanders' fear of negative consequences after intervening, which includes the worry of the carrying person dropping the game and being attacked by the toxic player or putting more fuel into the fire. "If we're winning and he's carrying the game and we like, like, say something to him, he's just gonna throw the game. So it's like it's bad. " (N6, ARAM, 9 years); "If I

jump into that, there's a solid chance you just gonna harass me with." (N9, ARAM, 7 years);

Additionally, participants as bystanders could also keep silent when they are physically or mentally tired in a game. "for example, I am tired like I can't really now you know, I don't have the energy to do so, you know, then I would probably not intervene. But I think mentally a bit more so. For example, if I'm really frustrated. So, for example, if this is the second game in a row where something." (N10, Rank, 9 years).

Two participants stated that they sometimes do not intervene because they like watching people arguing out of an instigating mentality. "If they curse each other, maybe it's fun to watch them like watching a drama. " (N2, QuickPlay, 8 years); "Sometimes the quarrel between these two people is particularly interesting, and I just want to see them quarrel. " (N4, QuickPlay, 8 years). They claimed that it is because the toxicity is so common and everyone is used to it, so there is no need to intervene sometimes. And some people's language when they curse is very eloquent, so it's worth watching.

4.3.4 Theme 4: Systemic Shortcomings. This theme discusses the gaming platform's system, such as policies and community support structures, and player suggestions for system optimization.

Systemic Responsiveness. To delve deeper into the impact of the game system on bystander intervention, we asked participants about their perspectives on the whole system, including toxicity detection, reporting system, and gaming community management system. Some participants highlighted that it is because the penalty system doesn't hit the nail on the head, so the toxic player can unscrupulously spread their toxicity. "Because there used to be something called 'The Tribunal', which was quite serious, but it was removed later. I think Riot should increase the punishment." (N5, Rank, 11 years) *(Tribunal is a system to help discipline players and keep the community in check with the help of their peers.)*

Some participants mentioned that the detection of toxic behavior is not intelligent enough. Players can avoid the low-priority queue penalty by holding 2 accounts and switching between them. "A lot of people have like 2 access accounts that they just switch over to another account and they switch over to another account." (N9, ARAM, 7 years). Participants even stated that it is because Riot is a company focusing too much on profitability and thus unwilling to spend money on research verbal abuse detection and act passively to block many accounts because it would affect its revenue. "Their language monitoring is not good at all. And sometimes I think it's not that they can't do it, but that they don't want to spend money on it." (N2, QuickPlay, 8 years). "You can't really say as a game company for, oh, we're going to ban you or something because they are just trying to make money with their skins." (N7, Rank, 10 years).

Systemic Features and Suggestions. Multiple participants suggested that League of Legends designers should learn from the design of other games' credit systems, such as Dota 2.

> "Do a little bit of like research on the Dota 2 behavior system, or like a credit system in general, so that if you get reported that you are like like certain features are locked, but if you behave well, for example, what I had in mind would be if you are on a Level 5, then you have access to everything. But once you are not on Level 3, you are not allowed to ping certain things. If you are on level 2, you're not allowed to type in chat. And if you're on a level 0, you are not allowed to queue up ranks." (N10, Rank, 9 years)

Participants also suggested redesigning the honor system to provide more useful rewards instead of an annual skin. "I think it makes much more sense than just giving us Grey Warwick. I don't want Grey Warwick. I hate Medieval Twitch, there is no incentive to be honorable in League of Legends." (N10, Rank, 9 years) *(Warwick and Twitch, characters in League of Legends).*

Additionally, participants suggested Riot block toxic streamers or prevent them from being toxic because they set a bad example for the gaming community. "I think you have to ban the.. like [toxicity from] the big streamers, or at least like force them to stop being toxic." (N6, ARAM, 9 years).

5 Study 2: In-game Tools Design

Inspired by the "5Ds" concept discussed in the related work section, I designed in-game tools to mitigate bystander effects, thereby enabling effective bystander intervention for victims of toxic behavior in games. The findings of Study 1 and the limitations of the current UI design in League of Legends informed the design of in-game tools in Study 2.

5.1 Methodology

Implementation

Prototypes of interactive in-game tools were implemented in Unity. In terms of UI, these tools mimicked the UI style of the LOL client, using a high-resolution in-game screenshot as a background image to give participants an immersive experience. All buttons, drop-down boxes, and text entry fields were interactive. These tools only demonstrated design concepts and were not built into the LOL game client. In the Experiment section, I used video recordings of the entire process of using each tool as demonstration videos in the

survey, rather than the interactive prototype, to control for the fact that every participant followed the correct procedure for using the tools.

5.1.1 Theory-Guided In-Game Tools Design.

Distract: This tool allows bystanders to create an interactive shared goals button to distract toxic players and victims from arguments and bullying. It enables bystanders to create a shared goal for the team and invite players to vote for it. For example, in-game events (Drakes, Bashor, River Crab) or strategies (4-1 Split Push, 1-3-1 Skesplit Push, call out teamfight) that require collaboration and communication, as shown in Figure 5. This tool aims to intervene in the ongoing toxicity by interrupting it.

Moderate: The bystander can apply intervention directly by clicking the "moderate" button to address the toxic behavior. This "one-click shout-out" feature allows bystanders to select pre-set messages from a pop-up list and send them in the chat, as shown in Figure 6. The pre-set messages use a diplomatic and gentle tone to emphasize that toxic behavior is unfriendly, harms winning the game, and politely request a stop to toxic behavior. The design concept of this tool is inspired by both 'Direct' and 'Delay' strategies: to confront the toxicity but in a mediating way. A system-vetted message list reduces expressive risks, and the shortcuts provide convenience for interveners, as some participants reported that they had difficulty multitasking in the laning phase or small-scale group fighting.

Broadcast: The Bystander can report the toxic player and click the "Broadcast" button to immediately send all people a notification of the report, as shown in Figure 7. This tool acts as a warning against toxic players by making their toxicity known to everyone to encourage a reduction in toxic behavior. It enables bystanders to call out the name of the toxic behavior and its perpetrator. Simultaneously, it enables victims to perceive positive support through bystander intervention. The 'Direct' strategy in 5D inspires this tool.

Document: Bystander can click the "Export" button to export game replay videos including its in-game chat messages and all players in the same team will receive a notification that the video is exported, as a warning to toxic players that their behavior is being recorded and may be posted on social media, as shown in Figure 8.

PrivateMessage (PM): The private Message (PM) tool allows the bystander to send pre-set friendly messages privately and quickly to the victim, which can console the victim player, as shown in Figure 9. The design concept of this tool is also inspired by 5D's 'Delay' strategy, but focuses on another aspect: supporting the victim. In real life, people don't always have the chance to intervene promptly; however, in the fast-paced progress of the game, players who want to intervene in the form of support for victims can send texts by themselves if they are able to type fast, or send a preset support message with a single click in *PM* when they are busy. Meanwhile, the LOL client has already implemented the function that automatically detects toxic behaviors and sends a comforting message to all players, except the toxic player, at the end of the game.

Delegate: The delegate tool allows bystanders to request help from the team for the victim, as shown in Figure 10. The 'Delegate' strategy in 5D inspires this tool.

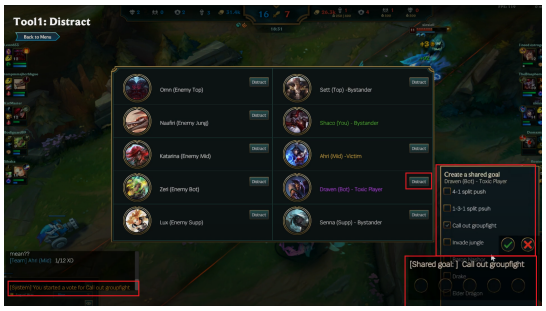

Figure 5. Distract-Screenshot of tool design

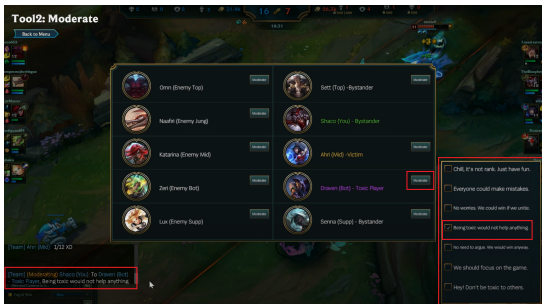

Figure 6. Moderate-Screenshot of tool design

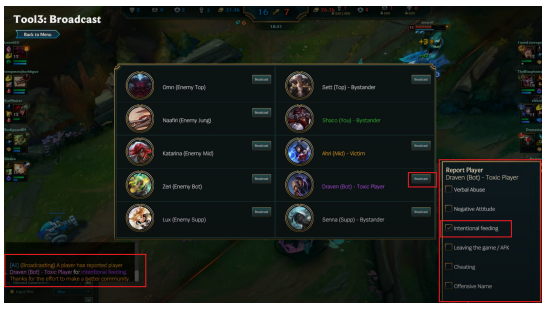

Figure 7. Broadcast-Screenshot of tool design

Figure 8. Document-Screenshot of tool design

Figure 9. PrivateMessage-Screenshot of tool design

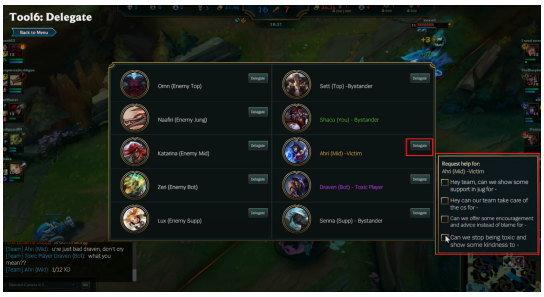

Figure 10. Delegate-Screenshot of tool design

5.1.2 Evaluation of In-game Tools Effectiveness.

In this study, I conducted a quantitative study to evaluate the effectiveness of in-game tools in supporting the victims. The baseline is a simulation of current in-game UI systems in League of Legends, with no support tools.

1) Participants: This study recruited 31 participants for evaluation. The chosen number of participants applied the Central Limit Theorem (CLT), which suggests $n \geq 30$ is a size that enables a normal distribution in the sample group. Participants self-identified as gamers of League of Legends and have at least a basic understanding and experience with in-game toxicity. 11 participants were recruited among the Utrecht University students. 20 participants were recruited from Prolific, which is a platform that allows researchers to easily find vetted research participants. Each participant from Prolific was rewarded 6.70 euros per hour. After examining the data, 2 outliers (one participant gave the same answer to all scales, and one participant gave nonsense answers to open questions) were removed, leaving 29 valid responses (men = 18, women = 11). The remaining 29 participants are between 18 and 44 years old (18-24 years old = 13, 25-34 years old = 15, 35-44 years old = 1). The majority had reported abundant video game-playing experience (over 15 years = 12, 10-15 years = 9, 5-10 years = 3, 3-5 years = 2, 1-3 years = 3).

2) Procedure: After informing the participants of study information and providing informed consent, participants were asked to complete a survey on *Qualtrics*. First, they answered demographic questions to confirm they have enough game experience of League of Legends, and were asked to watch 7 demonstration videos and fill in corresponding Likert scale assessment [27] of each to assess participants' experience about *Baseline* framework (without tools) and 6 levels of tools: *Distract*, *Moderate*, *Broadcast*, *Document*, *PrivateMessage (PM)*, *Delegate*. The order of 6 tools videos and each corresponding Likert Scale assessment followed the Latin Square design. In the videos, a toxic dialogue was presented in the scrolled view in the left corner of the interface to provide context for toxicity to viewers. Within each video, the entire process of using each tool was presented to guarantee participants comprehended how to operate each tool and the nature of the feedback provided by each tool. At the end of the survey, participants answered open questions about the tools.

The Likert scale assessment [27] estimated players' perceptions of the tools' effectiveness in motivating intervention and reducing toxicity from a bystander's perspective. Three statements of the Likert assessment are in the Table 3. Likert scales covered multiple aspects: *Intended Use*, *Willingness to Intervene*, and *Confidence in Intervening*. *Intended Use* measures the overall ease and efficiency of bystanders' interaction with tools. *Willingness to Intervene* measures to what extent tools can motivate bystanders to intervene when encountering toxicity. *Confidence in Intervening* refers to the degree of self-assurance bystanders feel about intervention.

3) Analysis: I conducted repeated measures ANOVA among *Baseline* and 6 tools. Analysis of the 3 items of tools' effectiveness, including *Intended Use*, *Willingness to Intervene*, and *Confidence in Intervening* was reported.

5.2 Results

To answer the *RQ3: Are in-game tools effective in motivating bystanders to help the victim of toxicity?*, I conducted three separate repeated measures ANOVA with 1 factor (six tool levels and a baseline: *Distract, Moderate, Broadcast, Document, PrivateMessage (PM), Delegate*, each ANOVA focused on the 3 items in the 7-point Likert questionnaire: *Intended Use*, *Willingness to Intervene*, and *Confidence in Intervening*. I

Table 3. Likert Scale Assessment

Statements	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
I would use this in-game tool to combat toxicity during my gameplay.	○	○	○	○	○	○	○
This in-game tool increases my willingness to combat negative behavior/assist my teammates when they need help in the game.	○	○	○	○	○	○	○
This in-game tool enhances my confidence in intervening in toxic situations.	○	○	○	○	○	○	○

checked the sphericity and conducted Huynh-Feldt corrections when its sphericity was violated and $\epsilon > 0.75$. When $\epsilon < 0.75$, the Greenhouse-Geisser correction is used. I also conducted post-hoc tests with Bonferroni correction and reported the significant differences between each tool's effectiveness.

Descriptive plots of tools' *Intended Use*, *Willingness to Intervene*, and *Confidence in Intervening* are in Figure 11, 12, 13.

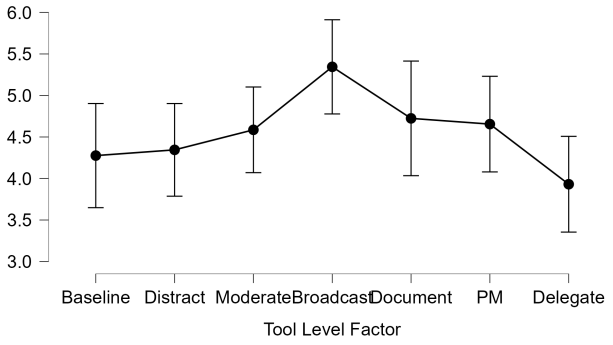

Figure 11. *Intended Use* descriptives plot. Error bars indicate 95% confidence intervals.

5.2.1 Likert Scale Factor 1: *Intended Use*. Regarding the *Intended Use* item of tools, table 4 shows there is a significant difference among tools ($F(4.494, 125.845) = 2.363$, $p = 0.050$, $\eta^2 = 0.078$) after Greenhouse-Geisser correction ($\varepsilon_{GG} = 0.749$). The sphericity is violated (Mauchly's Test $p = 0.005 < 0.05$). For all 6 tools and the baseline condition, Post hoc tests after Bonferroni correction in table 5 shows there is no significant difference between each tool and the *Baseline* in terms of *Intended Use*.

5.2.2 Likert Scale Factor 2: *Willingness to Intervene*. Within subjects effects for *Willingness to Intervene* item in table 6 indicate that there is no difference among different tool levels ($F(3.888, 108.858) = 1.355$, $p = 0.255 > 0.05$, η^2

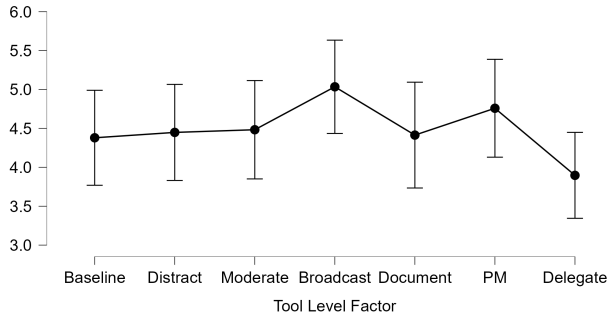

Figure 12. *Willingness to Intervene* descriptives plot. Error bars indicate 95% confidence intervals.

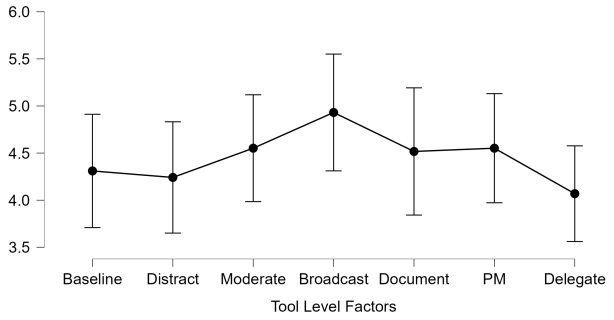

Figure 13. *Confidence in Intervening* descriptives plot. Error bars indicate 95% confidence intervals.

$= 0.046$) after Greenhouse-Geisser correction ($\varepsilon_{GG} = 0.648$). Sphericity is violated (Mauchly's Test $p = < 0.001$).

5.2.3 Likert Scale Factor 3: *Confidence in Intervening*. Within subjects effects for *Confidence in Intervening* item in table 7 indicate that there is no difference among different tool levels ($F(6.000, 168.000) = 0.922$, $p = 0.480 > 0.05$, $\eta^2 = 0.032$). Sphericity is not violated (Mauchly's Test $p = 0.113 > 0.05$).

Table 4. *Intended Use*-Within Subjects Effects

Cases	Sphericity Correction	Sum of Squares	df	Mean Square	F	p	η^2
Tool Level Factor	None	34.069	6.000	5.678	2.363	0.032	0.078
	Greenhouse-Geisser	34.069	4.494	7.580	2.363	0.050	0.078
Residuals	None	403.645	168.000	2.403			
	Greenhouse-Geisser	403.645	125.845	3.207			

Table 5. Post Hoc Comparisons - Tool Level Factor

		Mean Difference	SE	t	Cohen's d	p_{bonf}
Baseline	Distract	−0.069	0.400	−0.172	−0.035	1.000
	Moderate	−0.310	0.418	−0.742	−0.159	1.000
	Broadcast	−1.069	0.468	−2.284	−0.547	0.633
	Document	−0.448	0.402	−1.115	−0.229	1.000
	PM	−0.379	0.386	−0.983	−0.194	1.000
	Delegate	0.345	0.442	0.780	0.176	1.000
Distract	Moderate	−0.241	0.390	−0.619	−0.124	1.000
	Broadcast	−1.000	0.361	−2.768	−0.512	0.208
	Document	−0.379	0.464	−0.817	−0.194	1.000
	PM	−0.310	0.391	−0.794	−0.159	1.000
	Delegate	0.414	0.363	1.140	0.212	1.000
Moderate	Broadcast	−0.759	0.241	−3.143	−0.388	0.083
	Document	−0.138	0.426	−0.324	−0.071	1.000
	PM	−0.069	0.374	−0.184	−0.035	1.000
	Delegate	0.655	0.413	1.586	0.335	1.000
Broadcast	Document	0.621	0.448	1.386	0.318	1.000
	PM	0.690	0.388	1.778	0.353	1.000
	Delegate	1.414	0.445	3.175	0.724	0.076
Document	PM	0.069	0.511	0.135	0.035	1.000
	Delegate	0.793	0.398	1.992	0.406	1.000
PM	Delegate	0.724	0.343	2.108	0.371	0.926

Table 6. *Willingness to Intervene*-Within Subjects Effects

Cases	Sphericity Correction	Sum of Squares	df	Mean Square	F	p	η^2
Tool Level Factor	None	21.478	6.000	3.580	1.355	0.235	0.046
	Greenhouse-Geisser	21.478	3.888	5.524	1.355	0.255	0.046
Residuals	None	443.665	168.000	2.641			
	Greenhouse-Geisser	443.665	108.858	4.076			

Table 7. *Confidence in Intervening*-Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p	η^2
Tool Level Factors	13.478	6	2.246	0.922	0.480	0.032
Residuals	409.094	168	2.435			

Summary of Results In the results of study 2, in terms of *Intended Use*, the descriptive plot showed that the average scale of 5 tools is higher than *Baseline*. *Broadcast* tool is the highest: 1.069 higher than the *Baseline* on average, followed by *Document* (0.448), *PM* (0.379), *Moderate* (0.310), *Distract* (0.069). In contrast, the *Delegate* tool is 0.345 lower than the

Baseline. Repeated-measures ANOVA results showed a significant difference among the tools. Post Hoc test results suggested no statistically significant differences were found between any two tool level factors.
Regarding the *Willingness to Intervene* and *Confidence in Intervening* scale, there is no significant difference between baseline and tools.
Overall, with the current data and results of repeated-measure ANOVA, the tools do not show any significant difference between any tool and the baseline, but the mean difference suggests a promising future for the *Broadcast* tool.

6 Discussion

6.1 Interpretation of Key Findings

In study 1, I found that 1) bystanders have a diverse perception of toxicity, 2) their intervention behaviors were motivated by multiple factors (e.g., competitive drive to win, empathy), but inhibited by barriers such as distrust in reporting systems and fear of escalation. In study 2, it was found that players prefer to have more in-game intervention tools than the current UI design.

The *Broadcast* tool emerged as the most promising in enhancing bystanders' intervention. In contrast, the *Delegate* tool performed poorly, potentially due to players' distrust of third-party mediation, which also corroborates the conclusion from Study 1.

Findings of Study 1 confirm previous literature [22], indicating that toxic behaviors are strongly associated with players' frustration arising from skill gap and perceived loss. It demonstrates previous findings in [44]: Players barely use the reporting function because they consider the reporting system ineffective as well.

This research broads the current theories in two aspects: 1) Beyond the moral motivation 'sense of responsibility' in the traditional bystander intervention model proposed in [9] , bystander intervention could also be driven by practical objectives (e.g., increase the win chance of a game). 2) The unique 'Schadenfreuder' phenomenon: some players take pleasure from spectating chaotic situations or others' misfortune, reveals an unexplored behavioral pattern in gaming contexts. This research provides practical insights for game designers and community managers. First, 1) designers should optimize the current UI by adding intervention tools. 2) Second, developers should optimize the reporting system to give timely feedback to establish players' trust. 3) Upgrade the rewards to encourage more positive behaviors. 4) Build strict norms for streamers as their behavior shapes the community's circumstances.

6.2 Limitations and Future Work

Sample Bias and Generalizability. The small sample size in Study 1 could result in a lack of possible perspectives on research questions. All participants were League of Legends players, which threatens the generalizability to other games (e.g., DOTA2). The gender ratio in Study 2 was not balanced (18 males out of a 29 sample size), which may lead to over-represented results for women.

Tool Design Scope. The current tool designs are conceptual prototypes rather than loadable League of Legends plugins. Meanwhile, the demonstration videos in the questionnaire are video recordings, which makes these interactions not sufficiently immersive.

Future studies can focus on comparing bystanders' perceptions and responses to similar toxicity among different game genres. A larger sample size for experiments will also be great to help explore the significant tool level factor.

7 Conclusion

This project includes two studies. Study 1 aimed to investigate bystander intervention insights and decision-making in MOBA games when encountering toxicity. Study 2 aimed to find ways to boost bystander intervention by designing several in-game tools and examining their effectiveness. This study reveals that bystanders in MOBA games perceive toxicity differently and employ diverse intervention strategies, and in terms of *Intended Use*, the presence of intervention tools is significantly better than the absence of tools. Multiple incentives and inhibitors influence bystanders' willingness to intervene. Notably, game design factors, such as the transparency and real-time feedback of reporting systems, heavily influence players' willingness to intervene.

These findings have implications for reducing and mitigating toxicity in competitive online games, especially in the MOBA gaming area. While this study was limited by its small sample size, it opens avenues for future research into exploring bystander responses and interventions to toxicity in gaming.

References

[1] SeafoamBream . 2022. Chat Moderation - What happens when players use words from the Blocked List. https://support-leagueoflegends.riotgames.com/hc/en-us/articles/10728400692243-Chat-Moderation-What-happens-when-players-use-words-from-the-Blocked-List

[2] Sonam Adinolf and Selen Turkay. 2018. 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 (CHI PLAY '18 Extended Abstracts)*. Association for Computing Machinery, New York, NY, USA, 365–372. https://doi.org/10.1145/3270316.3271545

[3] Nicole Beres, Julian Frommel, Elizabeth Reid, Regan Mandryk, and Madison Klarkowski. 2021. Don't You Know That You're Toxic: Normalization of Toxicity in Online Gaming. 1–15. https://doi.org/10.

1145/3411764.3445157

[4] Jeremy Blackburn and Haewoon Kwak. 2014. STFU NOOB! Predicting Crowdsourced Decisions on Toxic Behavior in Online Games. In *Proceedings of the 23rd International Conference on World Wide Web* (Seoul, Korea) *(WWW '14)*. Association for Computing Machinery, New York, NY, USA, 877–888. https://doi.org/10.1145/2566486.2567987

[5] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa arXiv:https://www.tandfonline.com/doi/pdf/10.1191/1478088706qp063oa

[6] Alessandro Canossa, Dmitry Salimov, Ahmad Azadvar, Casper Harteveld, and Georgios Yannakakis. 2021. For Honor, for Toxicity: Detecting Toxic Behavior through Gameplay. *Proc. ACM Hum.-Comput. Interact.* 5, CHI PLAY, Article 253 (Oct. 2021), 29 pages. https://doi.org/10.1145/3474680

[7] Esports Charts. 2023. 2023 World Championship [Worlds 2023]. https://escharts.com/tournaments/lol/2023-world-championship

[8] Wikipedia Contributors. 2019. League of Legends. https://en.wikipedia.org/wiki/League_of_Legends

[9] John M. Darley and Bibb Latane. 1968. Bystander intervention in emergencies: diffusion of responsibility. *Journal of personality and social psychology* 8 4 (1968), 377–83. https://api.semanticscholar.org/CorpusID:9665680

[10] Dominic DiFranzo, Samuel Hardman Taylor, Franccesca Kazerooni, Olivia D. Wherry, and Natalya N. Bazarova. 2018. Upstanding by Design: Bystander Intervention in Cyberbullying. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173785

[11] Peter Fischer, Joachim Krueger, Tobias Greitemeyer, Claudia Vogrincic, Andreas Kastenmüller, Dieter Frey, Moritz Heene, Magdalena Wicher, and Martina Kainbacher. 2011. The Bystander-Effect: A Meta-Analytic Review on Bystander Intervention in Dangerous and Non-Dangerous Emergencies. *Psychological bulletin* 137 (05 2011), 517–37. https://doi.org/10.1037/a0023304

[12] Chek Yang Foo and Elina M. I. Koivisto. 2004. Defining Grief Play in MMORPGs: Player and Developer Perceptions. In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology* (Singapore) *(ACE '04)*. Association for Computing Machinery, New York, NY, USA, 245–250. https://doi.org/10.1145/1067343.1067375

[13] ADL: Center for Technology and Society. 2022. Hate Is No Game: Hate and Harassment in Online Games 2022 | ADL. https://www.adl.org/resources/report/hate-no-game-hate-and-harassment-online-games-2022

[14] Riot Games. 2023. https://support-leagueoflegends.riotgames.com/hc/en-us/articles/201752884-Player-Reporting-Guide-and-FAQ

[15] HI GEOFFREY. 2009. The Summoner's Code. https://nexus.leagueoflegends.com/en-us/2009/10/the-summoners-code/

[16] Nathan Grayson. 2020. Twitch is cracking down on simp emotes over harassment fears. (5 2020). https://kotaku.com/twitch-is-cracking-down-on-simp-emotes-because-of-haras-1843347454

[17] Shagun Jhaver, Himanshu Rathi, and Koustuv Saha. 2024. Bystanders of Online Moderation: Examining the Effects of Witnessing Post-Removal Explanations. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 191, 9 pages. https://doi.org/10.1145/3613904.3642204

[18] Pierre Voué Konrad Rudnicki, Heidi Vandebosch and Karolien Poels. 2023. Systematic review of determinants and consequences of bystander interventions in online hate and cyberbullying among adults. *Behaviour & Information Technology* 42, 5 (2023), 527–544. https://doi.org/10.1080/0144929X.2022.2027013 arXiv:https://doi.org/10.1080/0144929X.2022.2027013

[19] Bastian Kordyaka, Katharina Jahn, and Björn Niehaves. 2020. Towards a unified theory of toxic behavior in video games. *Internet Research* 30 (04 2020), 1081–1102. https://doi.org/10.1108/INTR-08-2019-0343

[20] Bastian Kordyaka, Samuli Laato, Katharina Jahn, Juho Hamari, and Bjoern Niehaves. 2023. The Cycle of Toxicity: Exploring Relationships between Personality and Player Roles in Toxic Behavior in Multiplayer Online Battle Arena Games. *Proc. ACM Hum.-Comput. Interact.* 7, CHI PLAY, Article 397 (oct 2023), 31 pages. https://doi.org/10.1145/3611043

[21] Bastian Kordyaka, Samuli Laato, Sebastian Weber, and Bjoern Niehaves. 2023. What constitutes victims of toxicity - identifying drivers of toxic victimhood in multiplayer online battle arena games. *Frontiers in Psychology* 14 (2023). https://doi.org/10.3389/fpsyg.2023.1193172

[22] Yubo Kou. 2020. Toxic Behaviors in Team-Based Competitive Gaming: The Case of League of Legends. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Virtual Event, Canada) *(CHI PLAY '20)*. Association for Computing Machinery, New York, NY, USA, 81–92. https://doi.org/10.1145/3410404.3414243

[23] Haewoon Kwak, Jeremy Blackburn, and Seungyeop Han. 2015. 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* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3739–3748. https://doi.org/10.1145/2702123.2702529

[24] Bibb Latane and John M. Darley. 1970. The Unresponsive Bystander: Why Doesn't He Help? https://archive.org/details/unresponsivebyst0000lata/mode/2up

[25] Amanda Lenhart, Mary Madden, Kristen Purcell, Lee Rainie, Aaron Smith, and Kathryn Zickuhr. 2011. Teens, Kindness and Cruelty on Social Network Sites. https://api.semanticscholar.org/CorpusID:149005728

[26] Mark Levine, Amy Prosser, David Evans, and Stephen Reicher. 2005. Identity and Emergency Intervention: How Social Group Membership and Inclusiveness of Group Boundaries Shape Helping Behavior. *Personality and Social Psychology Bulletin* 31, 4 (2005), 443–453. https://doi.org/10.1177/0146167204271651 arXiv:https://doi.org/10.1177/0146167204271651 PMID: 15743980.

[27] Rensis Likert. 1932. *A technique for the measurement of attitudes*. Columbia Univ. 140: 1–55 pages.

[28] P.M Markey. 2000. Bystander intervention in computer-mediated communication. *Computers in Human Behavior* 16, 2 (2000), 183–188. https://doi.org/10.1016/S0747-5632(99)00056-4

[29] Tiago Marques, Sandy Schumann, and Enrico Mariconti. 2024. Positive behaviour interventions in online gaming: a systematic review of strategies applied in other environments. *Crime Science* 13 (05 2024). https://doi.org/10.1186/s40163-024-00208-8

[30] C. K. Monge and T. C. O'Brien. 2022. Effects of individual toxic behavior on team performance in League of Legends. *Media Psychology* 25, 1 (2022), 82–105. https://doi.org/10.1080/15213269.2020.1868322 arXiv:https://doi.org/10.1080/15213269.2020.1868322

[31] Marçal Mora-Cantallops and Miguel Ángel Sicilia. 2018. MOBA games: A literature review. *Entertainment Computing* 26 (2018), 128–138. https://doi.org/10.1016/j.entcom.2018.02.005

[32] Marcus Märtens, Siqi Shen, Alexandru Iosup, and Fernando Kuipers. 2015. Toxicity detection in multiplayer online games. In *2015 International Workshop on Network and Systems Support for Games (NetGames)*. 1–6. https://doi.org/10.1109/NetGames.2015.7382991

[33] Karen M. O'Brien, Elizabeth W. Sauber, Monica S. Kearney, Rachel B. Venaglia, and Jr. Edward P. Lemay. 2021. Evaluating the Effectiveness of an Online Intervention to Educate College Students About Dating Violence and Bystander Responses. *Journal of Interpersonal Violence* 36, 13-14 (2021), NP7516–NP7546. https://doi.org/10.1177/0886260519829769 arXiv:https://doi.org/10.1177/0886260519829769 PMID: 30755066.

[34] Mona O'Moore. 2012. Cyber-bullying: the situation in Ireland. *Pastoral Care in Education* 30, 3 (2012),

209–223. https://doi.org/10.1080/02643944.2012.688065 arXiv:https://doi.org/10.1080/02643944.2012.688065

[35] Susanne Poeller, Martin Johannes Dechant, Madison Klarkowski, and Regan L. Mandryk. 2023. Suspecting Sarcasm: How League of Legends Players Dismiss Positive Communication in Toxic Environments. *Proc. ACM Hum.-Comput. Interact.* 7, CHI PLAY, Article 374 (Oct. 2023), 26 pages. https://doi.org/10.1145/3611020

[36] Antti Pohjanen. 2018. Report please! : a survey on players' perceptions towards the tools for fighting toxic behavior in competitive online multiplayer video games. https://api.semanticscholar.org/CorpusID:169332862

[37] Hanna A. Preston. 2020. *Personality Traits that Influence Bystander Intervention in Cyberbullying*. Ph. D. Dissertation. https://www-proquest-com.proxy.library.uu.nl/dissertations-theses/personality-traits-that-influence-bystander/docview/2414405847/se-2

[38] Sonia Roccas, Lilach Sagiv, Shalom H. Schwartz, and Ariel Knafo. 2002. The Big Five Personality Factors and Personal Values. *Personality and Social Psychology Bulletin* 28, 6 (2002), 789–801. https://doi.org/10.1177/0146167202289008 arXiv:https://doi.org/10.1177/0146167202289008

[39] Ralph Rosnow, Barry Goodstadt, Jerry Suls, and A. Gitter. 1973. More on the Social Psychology of the Experiment: When Compliance Turns to Self-Defense. *Journal of Personality and Social Psychology* 27 (09 1973), 337–343. https://doi.org/10.1037/h0034943

[40] S. Rothmann and E. P. Coetzer. 2003. The big five personality dimensions and job performance. *SA Journal of Industrial Psychology* 29, 1 (2003). https://doi.org/10.4102/sajip.v29i1.88

[41] Orvill Samanta. 2024. League of Legends Player Count 038; Stats 2024. https://prioridata.com/data/league-of-legends/#:~:text=With%20over%20150%20million%20registered,24%20as%20of%20late%202023.

[42] Phillip T. Slee and Ken Rigby. 1993. The relationship of Eysenck's personality factors and self-esteem to bully-victim behaviour in Australian schoolboys. *Personality and Individual Differences* 14, 2 (1993), 371 – 373. https://doi.org/10.1016/0191-8869(93)90136-Q Cited by: 177.

[43] Right to Be. 2022. The 5Ds of Bystander Intervention. righttobe.org/guides/bystander-intervention-training/

[44] Selen Türkay, Jessica Formosa, Sonam Adinolf, Robert Cuthbert, and Roger Altizer. 2020. 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 (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376191

[45] Anonymous FandomWiki User. n.d.. The Tribunal. https://leagueoflegends.fandom.com/wiki/The_Tribunal Accessed: 2024-02-01.

[46] Jacob Wolf. 2020. League 101: A League of Legends beginner's guide. https://www.espn.com/esports/story/_/id/29915901/league-101-league-legends-beginner-guide

[47] Pu Yang, Brent E. Harrison, and D. Roberts. 2014. Identifying patterns in combat that are predictive of success in MOBA games. In *International Conference on Foundations of Digital Games*. https://api.semanticscholar.org/CorpusID:16309552