The Enemy Hates Best? Toxicity in League of Legends and Its Content Moderation Implications

Jesús C. Aguerri[1] **· Mario Santisteban**[2] **· Fernando Miró-Llinares**[1]

Abstract

In recent decades, many sectors of our society have been digitized, and much of our life has moved to cyberspace, especially in terms of entertainment. Users meet, relate, and cooperate in the new public space that is the internet and form digital communities. Video games play a leading role in the formation of such communities. However, these communities also present antisocial behaviors, ranging from disruptive actions to harassment and hate speech. Such behaviors, encompassed under the umbrella term *toxicity*, are a major concern for both users and those in charge of moderating these spaces. This article focuses on toxicity in today's leading online video game League of Legends. Three hundred twenty-eight matches were reviewed using a system of two judges to study the prevalence of these problematic behaviors. We find that 70% of matches were affected by disruptive behavior. Nevertheless, only 10.9% of the analyzed matches were exclusively affected by downright harmful behavior. In our view, the results have relevant implications for content moderation policy that are also addressed in this paper.

Introduction

According to the International Telecommunications Union (ITU) (2021), the percentage of people using the Internet is already at 90% in countries that are considered developed by the United Nations; coverage is 87% in Europe and 81% in North America. Globally, 63% of the world's population has access to the Internet. The Internet has thus become

✉ Jesús C. Aguerri
    j.aguerri@crimina.es

    Mario Santisteban
    mario.santisteban@ehu.eus

    Fernando Miró-Llinares
    f.miro@crimina.es

1   CRÍMINA Research Center for the Study and Prevention of Crime, Miguel Hernández, University of Elche, Avda. de La Universidad S/N. Edif. Hélike, 03201 Elche, Spain

2   Department of Business and Civil Law, University of The Basque Country, Bilibao, Spain

part of our daily lives. While watching television is still the main leisure activity for Americans, the time spent on computers and video games is ever-increasing, especially among the younger population (Bureau of Labor Statistics, 2022). It has been estimated that, on average, North American teenagers spend between 4 and 6 h a day on digital media (web navigation, social media, gaming, texting). This figure has been steadily growing over the past two decades and has been accompanied by a decrease in the consumption of traditional media in the same age bracket (Twenge et al., 2019). This displacement has not only been observed for media consumption. Other areas like shopping have seen a shift from physical interaction to online transactions, thus reducing the time invested in traveling to make purchases (Le et al., 2021).

With digitization, new spaces for human interaction have emerged (Lupton, 2015). Given its degree of adoption, the Internet can be considered a new type of public space, which is in turn composed of different cyberplaces (Miró-Llinares & Johnson, 2018) where subjects meet and form communities. Some authors suggest that these communities can act as third places (Steinkuehler & Williams, 2006), neutral grounds where individuals are free to come and go as they please with no obligations and little involvement with other participants (Oldenburg, 1999). Indeed, the creation of these spaces, including those that are part of video games, may be protected under freedom of expression (Brown v. Entertainment Merchants Assn. et al., 564 U.S. 786 2011). It has been argued that players' speech may be safeguarded by this right as well (Balkin, 2004; Jørgensen & Mortensen, 2022).

However, the Internet also facilitates the dissemination of harmful content (ECHR, Delfi As v. Estonia § 110). The Internet does not discriminate between users' expressions (ECHR, Engels v. Russia § 30) and can be used to offend and harm others, just as any other technological tool. Indeed, online harassment, revenge porn, and defamation have all been matters of concern and have triggered discussions about how to tackle malicious speech without hampering the benefits of online communication (Citron, 2014).

The perturbing history of "a rape in cyberspace" shows that such behaviors have existed in online game communities for a long time (Dibbell, 1994; Suzor, 2019). Other controversies, such as Gamergate, exemplify how video game communities can create systematic harassment toward individuals that attempt to make video games more inclusive for women (Salter, 2017). Professional players also exhibit problematic behaviors. A relevant number of renowned players in the Esports sphere have been suspended for toxic behavior (Tseng, 2020). Riot Games, the owner of League of Legends (LOL), has banned entire LOL teams for similar reasons (Plunkett, 2013). Such problematic behaviors in video game communities have received growing academic attention over the last years (see: Mora-Cantallops & Sicilia, 2018a; Shen et al., 2020; Beres et al., 2021; Canossa et al., 2021).

In this paper, we will attempt to contribute to this body of knowledge by measuring the prevalence of toxic behavior on 328 competitive LOL matches. To this end, we offer a definition of toxic behavior, also known as disruptive behavior, creating a tailored taxonomy for LOL that discriminates between conducts that affect different values or interests. The results show that 70% of the games were affected by some kind of toxic behavior and that 30% of all players who participated in them committed some kind of toxic behavior. However, the most serious behaviors (hate speech, threats, etc.) are very rare, with most of the toxicity coming from insults or complaints about the performance of teammates. As addressed in the discussion, the prevalence and characteristic of the analyzed behaviors might have important implications regarding the moderation of these spaces by game developers.

Background

Online Social Gaming: League of Legends

The video game industry has benefited the most from the digitalization of society. In fact, in 2020, this industry generated around 177.8 billion dollars (Wijman, 2021). Video games have also changed, despite being eminently digital. They have gravitated toward the Internet and have, in turn, become more social (Rimington et al., 2016). The gamer's experience has ceased to be marked by isolation and has now got an eminently social character (Yee, 2014). This means that players experience various forms of in-game interaction (i.e., within the game itself): They are part of digital communities, travel through different platforms (Twitch, Youtube, specific forums, etc.), and engage in professional competitive environments (Esports). All video games past a certain threshold of popularity generate fan communities. However, games with high in-game interaction tend to generate especially rich communities and have drawn the interest of researchers for more than a decade. This is most notably true of massively multiplayer online role-playing games (MMORPG) (Cole & Griffiths, 2007; Johnson et al., 2009; Schiano et al., 2014; Shen et al., 2020; Yee, 2006).

In recent years, multiplayer online battle arena (MOBA) games have received a lot of attention, both among players and academics. This particular gaming genre is a subgenre of real-time strategy games in which two teams, usually formed by five players, face each other (Mora-Cantallops & Sicilia, 2018a). These games emphasize competitive gameplay and rank players according to match results, demanding a high level of strategic and mechanical skill (Johnson et al., 2015; Kou et al., 2018). These elements require players to get both in-game experience and acquire knowledge outside the game (e.g., forms of play, terminology). Players also agree on the most efficient tactics available, also known as the *metagame* or *meta* (Donaldson, 2017), resulting in a rich social exchange taking place both inside and outside the game.

League of Legends is the main exponent of contemporary MOBA. Developed by Riot Games, the game exceeded 180 million active users in October 2021, according to developer data (Zaragoza, 2022). As stated above, this type of video game transcends its medium. In fact, League of Legends is the most-watched video game on the streaming platform Twitch since 2019, both in terms of users and watch time (TwitchTracker, 2022). In addition, LOL has an intensely competitive landscape. The World Tournament achieved an average audience of around 30 million viewers in 2021, and its final match peaked at around 73 million viewers (Liber, 2021).

The Issue of Toxicity

As mentioned above, League of Legends is a team game where a matchmaking system assigns users to one of two teams (Mora-Cantallops & Sicilia, 2018b). Players are selected based on a numerical score, called MMR (i.e., matchmaking rating). The MMR considers previous game results in order to balance the gameplay experience for players. However, the algorithm on which the MMR is based is not publicly available, and players cannot consult their MMR score (Kou et al., 2018).

In addition, the game has a strong tactical component, so players must communicate with each other. This necessity increases as players' skill levels progress (Monge & O'Brien, 2022). There are two main tools to satisfy this need: during character selection,

a waiting-room style chat is used to discuss strategic choices; during the actual match, a small chat window on the game interface allows players to give tactical commands. While these tools may allow for cooperation between team members and even rivals, they are also the vehicle for toxic behavior. These can range from complaints about other players' performance to insults and harassment (Neto & Becker, 2018).

League of Legends users frequently complain about other players' behavior, especially their teammates' (Grandprey-Shores et al., 2014; Neto & Becker, 2018). Complaints occur most often in a competitive game mode called *ranked*, where game results earn or lose player points (LP). These points determine their invisible ELO[1] and their visible player rank (iron, bronze, silver, gold, platinum, diamond, master, grand master, and challenger) (Kou et al., 2018). In the community, a player's skill level is expressed publicly by their rank. Moreover, players who reach the challenger rank are considered hirable by semi-professional teams. In this competitive context, players are concerned about the toxic behavior of their peers, due to player experience deterioration and its connection to negative results (Grandprey-Shores et al., 2014). In addition, Riot Games has expressed its concern about these problematic behaviors on several occasions (Burrell, 2020; McWhertor, 2012).

Although complaints are numerous, there is no consensus on the definition of toxicity in the literature (Kordyaka et al., 2020; Kou, 2020). Toxicity is mainly used as an umbrella term, describing a wide range of negative behaviors, including harassment, *griefing*, and cheating (Adinolf & Turkay, 2018). These behaviors, in addition, highly depend on the immediate and cultural context (Beres et al, 2021). Toxicity has been defined as "the use of profane language by one player to insult or humiliate a different player in his own team" (Märtens et al., 2015) and "a denominator for aggressive and abusive interactions or relationships, both online and offline" (Deslauriers et al., 2020). As stated above, toxicity has been linked to the term *griefing*. A *griefer* can be described as a "player who derives their enjoyment from performing actions that detract from the enjoyment of the game by other players instead of just playing the game" (Mulligan et al., 2003). The Fair Play Alliance (henceforth FPA) has focused on analyzing "disruptive behavior" and abandoned the term toxicity. This term comprises a range of different conducts that mar a player's experience or a community's well-being (FPA, 2020). In addition, disruptive behavior "refers to conduct that does not align with the norms that a player and the community have set". In a similar way, some scholars (Boudreau, 2019) support the use of the term *transgressive* gameplay, to define the acts against the norms of the community.

Toxic (or disruptive/transgressive) behavior is a broad term that encompasses disapproved actions affecting the gameplay. However, the definitions above show that in-game mechanics and norms shaped by developers and players are key to concretizing this issue (Deslauriers et al., 2020; Foo & Koivisto, 2004; Jørgensen & Mortensen, 2022; Kou & Nardi, 2014). While common expressions like insults or taunts are presumably disapproved in most communities as well as in the offline world, actions can also disrupt other players' experience in specific games (Busch et al., 2015). For instance, abandoning a match before it has ended affects the other users' enjoyment of MOBA, because teams are likely to underperform if a member is absent (Canossa et al., 2021). In comparison, the log-out of any given player in an MMORPG does not represent a considerable gameplay disruption, since players are free to interact with the game world alone (e.g., Fallout 76, World of

---

[1] While Riot Games introduced the MMR as their specific player performance score, ELO is often used synonymously. Originally a ranking system for chess, it has become a popular term for player scores across different gaming communities.

Warcraft). Hence, a viable definition of toxicity for its detection in LOL must be tailored to this game specifically.

LOL's End User License Agreement (EULA) broadly defines conducts that are deemed unacceptable by Riot Games. The EULA punishes a wide spectrum of behaviors. Some of them could be categorized as "violent communication" (Miró-Llinares, 2016) and even be subject to criminal prosecution in most European criminal legislations.[2] This is true, for instance, of the rules sanction "harassing, stalking or threatening other players or Riot Games employees" and "transmitting or communicating any content which Riot finds offensive to players" (Riot Games, 2021).

While most rules echo norms of state criminal codes, some behaviors sanctioned in the EULA are tailored to the necessities of the game. Riot forbids any technical tool that enables players to cheat, for example by installing third-party mods or hacks. Furthermore, LOL players are punished for *throwing* (i.e., ostensibly giving up on the match), disconnecting, or leaving a game.

The EULA is detailed and expanded upon in the reporting system, which Riot establishes to help players notice deviant behaviors among users (Riot Games, 2013). Players can currently be reported for the following reasons:

- Insulting, harassing, or offensive language directed at other players
- Hate speech such as homophobia, sexism, racism, and ableism
- Intentionally ruining the game for other players with in-game actions such as *griefing*, *feeding*, or purposely playing in a way to make it harder for the rest of the team
- Leaving or going AFK at any point during the match being played
- Unnecessarily disruptive language or behavior that derails the match for other players
- Inappropriate summoner names

Besides, Riot clarifies that the following actions should not be considered deviant behavior:

- Playing poorly but still trying to win
- Strong language that does not insult or demean other people
- Choosing unusual champions, building unusual items, or experimenting with new ideas that do not match the current *meta*

Reporting is implemented inside the game through seven categories that are rather similar in their definition. These are displayed to players on the reporting screen with little to no context, leaving it up to them to interpret or research their definitions on the Riot Games support website. As Table 1 shows, three of these categories refer to in-game, non-expressive behaviors (i.e., leaving the game, intentional *feeding*, and cheating). Verbal abuse, hate speech, and offensive or inappropriate names, on the other hand, are behaviors based on verbal (i.e., written) expression. Finally, negative attitudes can be understood as a hybrid

---

[2] The global nature of the internet makes it difficult to find conduct that both occurs on the internet and is punishable in all national contexts where it may occur. This issue is particularly problematic with regard to expressive conduct, which is regulated in a variety of ways, for example, between European and Anglo-Saxon legislation. Although there are common elements between the different systems of regulation of expression, in this paper, when we refer to criminal frameworks, we refer to the European context.

Table 1 Report categories

Report category provided by RIOT	Type of conduct: expression or action	Affected interest
"Negative attitude (griefing, giving up)"	Hybrid	Disruptive behavior
"Verbal abuse (harassment, offensive language)"	Expressive	Harmful behavior
"Leaving the game/AFK"	Action	Disruptive behavior
"Intentional feeding (feeding is griefing not just having a bad game)"	Action	Disruptive behavior
"Hate speech (racism, sexism, homophobia, etc.)"	Expressive	Harmful behavior
"Cheating (unapproved third-party programs)"	Action	Disruptive behavior
"Offensive or inappropriate name"	Expressive	Harmful behavior

category that encompasses both in-game actions (e.g., *griefing*) and expressions related to the poor course of the game, demeaning a team's performance.

Considering the possible consequences of players' actions, we can distinguish "disruptive"[3] from "harmful" behavior. Actions such as going "AFK," "intentional feeding," "negative attitude," and "cheating" are disruptive in the sense that they can negatively affect a player's gameplay, that is "the actions performed by the player when involved in a challenge" (Guardiola, 2019). Since gameplay is the result of the "emotionally-charged interaction between the player and the game components" (Guardiola, 2019), the highlighted behaviors can result in making the game too challenging for the affected player, breaking the balanced sense of struggle that games should pose to be enjoyable (Costikyan, 2002).

Harmful behavior, on the other hand, is defined as behavior that can cause "significant emotional, mental or even physical harm to players or other people in the player's life such as family and friends" (FPA & ADL, 2020). Some of these, like hate speech or insults, are already sanctioned by state criminal codes or tort law and are not protected by the principle of freedom of expression contemplated in the European Convention of Human Rights (ECHR, *Feret V. Belgium*, §§ 75–78). Even if their particular characterization as harmful can be discussed (Feinberg, 1985), the term, as introduced by the FPA, helps to highlight a range of conducts that surpass the degree of seriousness that disruptive behaviors entail. This term, therefore, encompasses speech or acts that are (a) likely to offend others (Feinberg, 1985), (b) constitute a direct or indirect incitement to violence or (c) "any behavior that may be considered offensive or demeaning to society even if it is not directed at a specific person" (Miró-Llinares, 2016).

The degree of prevalence of this wide range of conducts is unknown despite being problematic for Riot. According to a recent update by Riot, "only 5% of players are consistently disruptive." This means that this cluster of players is the most complained about and the remaining users only "get tilted every once in a while" (Timttamoster, 2022). This statement contradicts Riot's lead designer's previous claims, who had stated that most toxic behavior was attributed to average players instead of *trolls* (Maher, 2016).

---

[3] For the purpose of analyzing LOL's players actions we will mainly used the term disruptive behavior, aligning with the FPA framework that rejects the concept of toxicity. Nevertheless, we still believe that the term toxic behavior is still adequate to represent the characteristics of the studied conducts and accordingly is used in several parts of this paper. That does not preclude the need for further research, which builds a more consistent and complete taxonomy, a task that this contribution is not capable to fulfill.

In addition, the relevant literature has not been able to paint a precise picture of toxicity in LOL. Based on datasets extracted from the game DOTA, a MOBA and LOL predecessor, Martens et al. (2015) developed an automated system for the detection of toxicity in chats of multiplayer online games. The study focused mainly on the detection of toxic verbal expressions, excluding other antisocial behavior such as leaving the game or more complex behaviors. Despite this restriction, the authors find that at least one toxic expression is present in 63% of the analyzed games' chats.

A similar methodology was used by Kwak & Blackburn (2015) to approximate the characteristics and distribution of toxic expressions throughout matches, in this case, based on data from LOL. Their automatic detection system allowed them to appraise how certain toxic terms were more likely to appear as the game progressed. Moreover, they observed the importance of complaints and insults regarding the performance of other players on the team. This last conclusion coincides with what Neto and Becker found (Neto & Becker, 2018) using a similar database but through their own topic modeling system. Their database, however, consisted of games in which a player had previously reported another user.

In 2019, ADL carried out a US national survey about the social interactions and experiences of video game players. The study found that "nearly three quarters (74%) of online multiplayer gamers have experienced some form of harassment in online multiplayer games." In the case of LOL, ADL reported that "three-quarters of League of Legends players had also experienced in-game harassment, with 36 percent experiencing frequent harassment" (ADL, 2019).

Content Moderation in Videogames: Riot Facing Toxicity

Online communities are governed by large platforms and intermediaries, which detect, assess, and intervene in users' speech in an effort to provide content moderation (Gillespie et al., 2020). Content moderation requires an infrastructure that coordinates human moderators and artificial intelligence systems to enforce platform rules. This complex infrastructure allows the members of these communities to steer away from harmful or illegal content. However, it has sparked a debate on the types of content that should be removed from platforms and the overall accountability of these measures (Busch et al., 2015; Gillespie, 2018; Keller, 2018; Suzor, 2019).

Content moderation is present in online multiplayer games. Following Balkin's early thoughts, game developers control user speech in at least two ways (2004), through game code and an End User License Agreement (EULA). Firstly, in accordance with Lessig, the architecture of virtual space is a factor that shapes users' behavior (1999). This particularly informs verbal expressions in gaming. Game design can affect user choice by preventing players from performing malicious acts, like murdering NPC children (e.g., The Elder Scrolls: Skyrim, Red Dead Redemption II) (Jørgensen & Mortensen, 2022). Furthermore, when chats are deployed to enable users to text each other, filters can be introduced to prevent players from using certain words[4] (e.g., LOL).

Secondly, games affect player attitude through a binding agreement that must be signed before entering, known as the End User License Agreement (EULA). EULAs often contain

---

[4] The effectiveness of these filters is a controversial issue and very difficult to assess in general terms (Gorwa et al., 2020), as there is a wide range of tools available (from mere word lists to constantly updated Artificial Intelligence algorithms) which, moreover, perform differently.

community guidelines that identify the community's values. They constitute legal grounds for account termination (Suzor, 2019). Moreover, the enforcement of EULA protects developers from vicarious liability due to potential tort claims regarding content posted by their users (Fairfield, 2009).[5] In addition, keeping their spaces safe helps sustain and grow communities (ADL, 2019; Sparrow et al., 2021). In this sense, if players experience harassment or other variants of toxic online behavior, they may abandon the game, which goes against developer interests (ADL, 2019). That said, enforcement of content moderation policies also decreases the player base, at least affecting the players that the measures target for being toxic (Sparrow et al., 2021). Content moderation involves the application of a rule system and a series of sanctions for those who infringe on community rules. The most known sanction is the suspension of a player's account, also dubbed *banning*. A LOL account may be suspended when the EULA is breached. However, these rules are not defined through an exhaustive list of examples of inappropriate behavior (Riot Games, 2021). While banning may be necessary in extreme cases, the industry has highlighted some problems associated with this measure, regarding disproportion and possible damages to freedom of speech (Balkin, 2004; Chelsea, 2017; Meehan, 2006). Consequently, there are alternatives to banning, such as temporary bans or cooldowns before being able to join a match again, which have proven to be remarkably effective in other contexts (Matias, 2019; Lewington & Committee, 2021).

Defining, establishing, and enforcing punitive measures is challenging and once a game manages to build a substantial player base, the scale of necessary moderation activities increases significantly. This can pose an unmanageable workload to Trust and Safety teams. The industry often claims that it is impossible for human moderators to control chat feeds and other behaviors due to the quantity of content that needs reviewing (Sparrow et al., 2021). This was one of the reasons why Riot deleted the LOL Tribunal, as they considered it to be "slow and inefficient" (Draggles et al., 2018). Instead, Riot has focused on developing AI tools that can carry out most moderation tasks. Even though Riot's terms of service are not explicit on this matter, we can presume that AI systems conduct all moderation tasks, and human moderators focus on reviewing player reports and appeals. However, it is not likely that these tools will replace human intervention completely (Zachary, 2019). AI systems to require human feedback to be improved and continuously trained (Lewington & Committee, 2021). In addition, ethical concerns are voiced from outside the industry, claiming that AI systems are not suitable to detect the more ambiguous offenses, which require a deep knowledge of the context (Duarte & Llansó, 2017; Shenkman et al., 2021).

Data and Method

During April and May 2022, ElmilloR and Riot Games organized the "SoloQChallenge," a League of Legends competition that involved Spanish streamers familiar with LOL. Seventy-four players across two categories partook in the competition. Categories were based

---

[5] According to Sect. 230 of the Communication Decency Act from the USA, game developers would be exempt from this type of liability in most cases. But according to the legal framework of the European Union, game developers can be held liable for user comments if they do not comply with the E-commerce Directive. See Regulation 2021/784 of the European Parliament and of The Council of 29 April 2021 on addressing the dissemination of terrorist content online and Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.

on the players' previous experience in the video game. Two player categories were created: low Elo for streamers with relatively little experience in the videogame (max. platinum division) and high Elo for professional and ex-professional players and streamers specialized in League of Legends who play in the four highest divisions, that is, the 98th percentile of players in Europe (OP.GG, 2022). Riot provided each streamer with a new account, without an MMR or previous history, but at the necessary level to be able to play matches competitively. Participants had to use these accounts to play and stream ranked games to gain LP and move up through the divisions. The player with the highest score within their category was declared the winner of the competition.

Participants in the competition did not play against each other, but against other players playing ranked games. Their level of experience can be assumed to be comparable due to the division system. The equilibrium of matchups makes the SoloQChallenge matches very similar to those that players encounter normally. In addition, the absence of MMR and account history in the competition eliminated matchup bias and fostered free player encounters. Given the availability of the matches in VOD format, the SoloQChallenge was a unique opportunity to measure the prevalence of toxic behaviors in LOL.

Our sample is composed of game evaluations from the low Elo category, based on the assessments of two judges. Judge number 1 reviewed the first 10 games of 31 participants and the first 9 games of 2 of the participants, totaling 328 games. Judge number 2 followed the same selection criteria and analyzed 198 games in total.

Both judges had previous experience with the video game and watched the matches independently.[6] They annotated verbal behavior according to the coding scheme (Table 2) using the left-column descriptors as category tags. To perform this task, they were provided with a table of behavioral categories as a template and a list of videos to watch. They were also provided with a definition of each of the categories and were instructed on their application. The judges also had to collect the results of the game and information regarding users who had exhibited any categorizable behavior, as well as the minute of the video in which the behavior had occurred. This allowed the researchers to review and verify annotations. To determine inter-judge agreement for the 198 games analyzed by both judges, we calculated Cohen's Kappa index (Cohen, 1960) for both the presence of problematic behavior in a given match ($K = 0.74$, p-value < 0.001), and whether the game was affected only by harmful behavior, only disruptive behaviors, or both types of behavior ($K = 0.64$, p-value < 0.001). Both coefficients yielded substantial inter-judge agreement (Altman, 1999) — in the first case, close to excellent agreement (Fleiss et al., 2003). We posit that this high level of agreement speaks to the quality of the evaluations, even in cases where only assessments from judge 1 are available. To improve data reliability, one of the researchers acted as a third judge, deciding in case of disagreement between the judges. Twenty-two of the actions noted were recategorized by the third judge, generating the final dataset on which data analyses were executed.

The categories were created by the researchers based on Riot Games' reporting categories and the work of Neto & Becker (2018), who identified the following main categories of negative behavior in the LOL game chat through topic modeling: complaints, arguments, insults, and taunts. Consequently, modifications were made to the company's report categories (Table 2, right column). Firstly, the category "inappropriate or offensive names" was not analyzed, as it was not possible to find a solid definition of what should be understood as

---

[6] Both judges had extensive experience in the game, had a university degree, were between 20 and 25 years old, and were of different genders.

Table 2 Annotated verbal behavior according to the coding scheme

Analysis category		RIOT report category
Disruptive behavior		
AFK	Abandoning the game	"Leaving the game/AFK"
Complaints about other players	Negative comments regarding teammates' performance. This includes abusive pinging on repeated occasions and asking teammates to end the game as well	"Negative attitude"
Voluntary death	Overtly voluntary actions leading to the player's death, both at the hands of the other team's players or the enemy team's towers	"Intentional 'feeding'"
Other disruptive behavior	Other disruptive behaviors not considered in the aforementioned categories	
Harmful behavior		
Insults	Offensive expressions that affect players' reputations and most likely involve the use of swear words and slurs, which are NOT considered hate speech under the RCRI General Policy Recommendation N° 15	"Verbal abuse"
Wish for death or serious harm	Expressions that imply wishing for another player to die or to suffer a serious illness or injury in real life (not in-game)	Verbal abuse
Hate speech and sexist language	Expressions considered hate speech under the ECRI General Policy Recommendation N°15[a]	"Hate speech"

[a]Hate speech is defined as "the use of one or more particular forms of expression – namely, the advocacy, promotion or incitement of the denigration, hatred or vilification of a person or group of persons, as well any harassment, insult, negative stereotyping, stigmatization or threat of such person or persons and any justification of all these forms of expression – that is based on a non-exhaustive list of personal characteristics or status that includes 'race', colour, language, religion or belief, nationality or national or ethnic origin, as well as descent, age, disability, sex, gender, gender identity and sexual orientation"

The Enemy Hates Best? Toxicity in League of Legends and Its Content…

447

such. Similarly, the category cheating was not contemplated as it was considered highly difficult to determine whether a player was cheating by mere visual inspection of the matches. For the same reason, the category "griefing" was not used for content evaluation. The category negative attitude was specified to refer only to complaints about other players' performances. The updated report categories are summarized in Table 2 (left column).

We included a category labeled "other disruptive behaviors" so that judges could record obvious disruptive behavior that was not stipulated by the coding scheme. Secondly, relying on the taxonomy of violent and hateful communication developed by Miró-Llinares (2016), the report category "verbal abuse" was divided into two categories: "Insults" are defined as expressions that affect players' reputation and involve the use of swear words, while "wishes for death or serious harm" elevates the threshold of the seriousness of the conduct considering the reference to physical harm.

Results

We detected some type of problematic behavior in 229 out of 328 matches (70%). No behaviors of interest were found in the remaining 30% of matches. The analysis of different types of behavior shows that 45.9% of matches are affected exclusively by disruptive behaviors, while 43.2% are affected by both disruptive and harmful behaviors. We find that only 10.9% of the analyzed matches are affected by harmful behaviors exclusively (Table 3, right column).

Zooming in on the individual players, we find that 29.6%, that is 398 out of 1343, players committed some kind of disruptive behavior during the matches. Note that 72% of reported users lost the game and only 28% ended up winning. The analysis of different types of behavior shows that most of the reported players (59%) only engage in disruptive behavior. However, 25.1% engage in harmful behavior and 15.8% engage in both types of behavior (Table 3, left column).

Disaggregating these categories into specific behaviors according to our coding scheme (Table 2), we find that complaints about teammates' performance, as a form of mere disruptive behavior, are the most frequent problematic behavior, both across all matches (52.4%) and all players (18.3%) (Table 4). More serious harmful behaviors, such as hate speech or death wishes, have relatively low frequencies. They occur in 3.4% and 7% of all matches respectively and are uttered by 0.8% and 1.8% of all players respectively. Insults as a form of harmful behavior, on the other hand, have the second highest frequency overall. Insults occur in 34% of all matches and are uttered by 10.2% of all players analyzed.

Players can not only perform different types of toxic behavior within the same game but can also perform the same behavior on multiple occasions. AFKing a match is an exception to this because players can only leave the match once. Also, behaviors that have to be reiterated to be considered toxic cannot be covered by this analysis (see "Limitations"). The corpus contains a total of 805 individual instances of toxic behavior (complaints, insults). Complaints are the most frequent type of behavior ($n = 463$), followed by insults ($n = 218$). On average, reported players utter 2.05 toxic expressions, resulting in a low count of toxic behavior per player per game. The average number of insults uttered per reported player is 1.59, with only 10% of users exceeding 3 insults in a given match, indicating that the proportion of players who perform multiple toxic behaviors is small. Toxicity seems to be evenly distributed across games. Twenty-five percent of matches contain more than 5 toxic actions, while 3 toxic behaviors

Table 3 Analyzed matches affected by harmful behaviors

	Players			Matches		
	n	Prop. over reported	Prop. over total	n	Prop. over reported	Prop. over total
Disruptive and harmful behavior	63	15.8%	4.7%	199	43.2%	30.2%
Only harmful behavior	100	25.1%	7.4%	25	10.9%	7.7%
Only disruptive behavior	235	59%	17.5%	105	45.9%	32%
Total	398	100%	29.6%	229	100%	70%

Table 4 Complaints about teammates' performance, as a form of mere disruptive behavior, are the most frequent problematic behavior, both across all matches (52.4%) and all players (18.3%)

	Players			Matches		
	n	Prop. over reported	Prop. over total	n	Prop. over reported	Prop. over total
AFK	50	12.6%	3.7%	47	20.5%	14.3%
Complaints about other players	246	61.8%	18.3%	173	75.1%	52.4%
Voluntary death	7	1.8%	0.5%	7	3%	2.1%
Other disruptive behaviors	11	2.8%	0.8%	10	4.4%	3%
Insults	137	34.4%	10.2%	112	48.9%	34.1%
Wish for death or serious harm	24	6%	1.8%	23	10%	7%
Hate speech and sexist language	11	2.8%	0.8%	11	4.8%	3.4%

per match are typically not exceeded, averaging around 3.52 toxic actions for affected matches and 2.45 for all the games analyzed.

Limitations

Although our study aims at the highest epistemic and methodological rigor, certain limitations in its design must be acknowledged. Firstly, Riot Games do not provide detailed definitions for reportable behavior in their community guidelines. The academic literature presents a similar problem as it is not unanimous on clear-cut distinctions between different types of behavior. While we have tried to define our categories for analysis as solidly and transparently as possible, these terminological indiscretions negatively impact the internal validity of our findings, because we have had to establish our own boundaries following the definitional criteria explained above. Consequently, changes in conceptualizations or the coding scheme are probable to lead to slightly different research results. We hope to mitigate these shortcomings by grouping the different categories into two more coarsely defined, and thus more easily distinguishable, macro-categories.

Secondly, it should be acknowledged that the data obtained is highly dependent on the sensitivity and specificity of the judges regarding toxic behavior. While we introduced inter-judge agreement measures to control for stringency in the scores, some variables present more categorical problems regarding classification. For example, behaviors such as abusive pinging (in the category of complaints) have to occur repeatedly to count as such and be reportable. Some users complain throughout the matches in such a continuous and repeated manner that it becomes unviable for judges to annotate each isolated behavior and to define where one complaint ends and where another starts. In these cases, we, therefore, chose to primarily report the proportion of players reported and the proportion of games affected by such behavior instead of analyzing individual instances. Regarding the overall value added of this study, however, it should be noted that this bias mainly results in the underestimation of behaviors of lesser severity, such as repeated complaints. Our coding scheme is still viable for the collection and analysis of severe toxicity, namely, hate speech and wishes of death and serious harm, since the underlying legal definitions are clearer and utterances can be faithfully counted as discrete occurrences.

Finally, it should also be recognized that the external validity of our study is compromised by the streamers, that is, the participants in our study, being aware that they are being watched while playing. This may, consciously or unconsciously, lead to behavioral modifications, most notably a higher degree of compliance with the community guidelines. In defense of our data collection decisions, however, we argue that the high degree of habituation and immersion of players in the game and their environment, as well as the absence of constant reminders of the recording process, somewhat mitigate the risk of extreme behavioral modifications. This bias can, nevertheless, never be perfectly controlled.

Discussion and Conclusions

Riot Game claims that most disruptive behaviors are exhibited by a small percentage of players (5%). Using categories that are similar to reportable behavior in Riot Games' EULA, our judgment task shows that 24.3% of players engaged in toxic behavior during

the analyzed competition matches. It is hard to assert which players exhibit these behaviors more frequently, and for whom they are exceptional events. Without the disclosure of Riot's data, we will hardly be able to verify this number.

Nevertheless, our research shows that a significant number of the reviewed matches are affected by disruptive behaviors (70%). Taking into account that our definition of disruptive behavior encompasses actions that are sanctioned by Riot's EULA, the evidence gathered suggests that those terms and conditions are consistently breached. Thus, and taking into account the particularities of the SoloQChallenge (see "Limitations"), we can safely affirm that toxicity (or disruptive behaviors in the context of LOL) is indeed a phenomenon that it is likely to affect the majority of LOL matches.

Having said that, disruptive conducts are prevalent over more serious behavior, which we have classified as harmful. Furthermore, among instances of harmful behavior, insults are the predominant category. In the context of the game and the possible scope of attention of such comments, it is not likely that they pose a significant threat to users' reputations and well-being. By contrast, hate speech and wishes for death or serious harm, constitute an exception among the reported behaviors in our sample.

From the perspective of the game company's liability, this is positive. Intermediaries fulfill the function of keeping virtual spaces safe from harmful expressions as part of their services (Gillespie, 2018). Indeed, the European Union legal framework makes platforms responsible for the curation of these types of wrongdoings, a responsibility that is going to be expanded in the upcoming Digital Service Act. Thus, a significant prevalence of instances of harmful conduct over disruptive behavior would call for a stronger content moderation response. However, it appears that this is not the case in our sample, that is, the LOL chat. This does not mean that other disruptive behaviors are not concerning to players. The categories that we have utilized to measure phenomena of toxicity characterize behaviors that are sanctionable by criminal or tort law as more harmful to players. However, in a competitive landscape such as LOL, where insults and disrespectful complaints are likely to be normalized (Beres et al., 2021), players might be more concerned about behaviors that ruin matches for their team (i.e., going AFK or voluntary death). Further research is required to assert if this is indeed the case, considering players' perceptions to strengthen the detection of disruptive behavior in the future based on situational requirements rather than legal standards alone.

According to the results of this research, the concern for toxicity should be understood primarily as a concern for disruptive behaviors, acts that negatively affect player's gameplay. Consequently, the actions of the managers of these spaces should be at least partially oriented towards establishing systems that serve the interests of their community (Busch et al., 2015), and not exclusively to comply with external legal frameworks. Legal frameworks are designed to address the behaviors that society as a whole is most concerned about, but they do not fully cover behaviors that negatively affect a community.

In consequence, the results of the study raise the issue of the enforcement of Riot's policy regarding disruptive, not harmful conduct. Judging whether a teammate goes AFK or voluntarily kills their champion to ruin the game requires a review of matches by moderators that understand the game mechanics more deeply. The experience of this study shows that these are complex decisions in which numerous variables must be considered (the presumable intention of the player when commenting, the previous acts of their teammates, and the overall state of the game). If these are the more prevalent conducts, and the duration of a LOL game is typically 30 min, a considerable investment must be made by Riot to detect and act against this content. If human moderators are only tasked with the review of complaints, many more harmful conducts may remain unsanctioned. Automation

of in-game moderation, on the other hand, does not seem to be a viable solution considering the risks that these systems pose in the face of complex judgments that could lead to wrong decisions and sanctions against honorable players. Such systems are not only unjust but could, in the long term, lead to frustration with the game and negatively affect the size of the population of players.

Bans and terminations of accounts on the grounds of disruptive behaviors should thus be carefully considered, particularly to avoid systematically unfair or possibly even discriminatory responses. Banning could be a disproportionate sanction in some situations, especially in games in which players purchase skins or other virtual goods which cannot be used once the account is terminated (Chelsea, 2017; Meehan, 2006). Suspending player accounts can hamper both economic interests and freedom of expression (Balkin, 2004). Moreover, its effectiveness has also been called into question because players can open other accounts or use tools to hide their digital identity. In this sense, the FPA has warned that "banning a player reduces their attachment to the game and sense of responsibility for their actions" and could lead players to act carelessly because they do not have a consistent account or identity to protect (FPA & ADL, 2020). Thus, considering more appropriate measures to deal with less worrying conduct is desirable. Temporary bans, or the establishment of cooldowns before being able to join a match again, may be appropriate measures. In addition, the effectiveness of nonpunitive actions like user warnings should not be overlooked. Giving users feedback about their actions after a match promotes awareness of community rules and decreases the likelihood of repeating offenses (Lewington & Committee, 2021). Studies outside the industry support these findings, pointing out that information about social norms positively influences newcomers' behavior, at least in some online communities (Matias, 2019). Soft approaches, like warning users about allegedly disruptive actions and the promotion of honest behavior through reward systems, seem more appropriate, with the added benefit of avoiding damage that a wrongful account suspension represents.

Finally, it should be noted that having found some kind of toxic behavior in 70% of the matches analyzed, it is safe to say that League of Legends has a problem with toxicity. Moreover, toxicity seems to be more or less evenly distributed across all the matches it affects, and there is no evidence that toxicity is the responsibility of a small group of players. In fact, 24% of the players in the games analyzed committed some kind of toxic behavior. However, most of this toxicity stems from disruptive behavior, mainly complaints about the performance of teammates. The harmful behavior is rare, although this does not mean that they are irrelevant. Consequently, although we can affirm that there is a problem with toxicity, we must also stress that it should not be understood primarily in legal terms or in terms of the need for regulation external to the video game. The issue is primarily about how toxicity harms members of the community and how it may be driving certain players out of the game or even excluding them.

Declarations

Competing Interests The authors declare no competing interests.

References

Adinolf, S., & Turkay, S. (2018). Toxic behaviors in esports games: Player perceptions and coping strategies. In *CHI PLAY '18 Extended abstracts: Proceedings of the 2018 annual symposium on computer-human interaction in play companion extended abstracts. New York: Association for Computing Machinery*, pp. 365–372.

ADL. (2019, Jul 18). *Free to play? Hate, harassment and positive social experiences in online games*. ADL. Retrieved September 7, 2022, from https://www.adl.org/resources/report/free-play-hate-harassment-and-positive-social-experiences-online-games

Altman, D. G. (1999). *Practical statistics for medical research*. Douglas G. Altman, Chapman and Hall.

Balkin, J. M. (2004). Virtual liberty: Freedom to design and freedom to play in virtual worlds. *Virginia Law Review, 90*(8), 2043–2098. https://doi.org/10.2307/1515641

Beres, N. A., Frommel, J., Reid, E., Mandryk, R. L., & Klarkowski, M. (2021). Don't you know that you're toxic: Normalization of toxicityin online gaming. In *CHI conference on human factors in comput-ing systems (CHI'21), May 8–13, 2021, Yokohama, Japan.* NewYork, NY: ACM. p. 15. https://doi.org/10.1145/3411764.3445157

Boudreau, K. (2019). Beyond fun: Transgressive gameplay—toxic and problematic player behavior as boundary keeping. In K. Jørgensen & F. Karlsen (Eds.), *Transgression in games and play*. MIT Press

Bureau of Labor Statistics. (2022, June 23). *American Time use survey - May to December 2019 and 2020 results*. BLS. Retreived 25 October, 2022, from https://www.bls.gov/news.release/pdf/atus.pdf

Burrell, J. (2020, Dec 16). *The disruption and harm in gaming framework*. Riot Games News. Retrieved January 9, 2023, from https://www.riotgames.com/en/news/the-disruption-and-harm-in-gaming-framework

Busch, T., Boudreau, K., & Consalvo, M. (2015). Toxic gamer culture, corporate regulation, and standards of behavior among players of online games. In S. Conway, J.de Winter (Eds*.), Video game policy: Production, distribution, and consumption*, (pp. 176–190). Routledge.

Canossa, A., Salimov, D., Azadvar, A., Harteveld, C., & Yannakakis, G. (2021). For honor, for toxicity: Detecting toxic behavior through gameplay. In *Proc. ACM Hum.-Comput. Interact. 5, CHI PLAY,Article 253 (September 2021)*, p. 17. https://doi.org/10.1145/3474680

Chelsea, K. (2017). Forcing players to walk the plank: Why End User License Agreements improperly control players' rights regarding microtransactions in video games. *William & Mary Law Review, 58*(4), 1365–1399.

Citron, D. (2014). *Hate crimes in cyberspace*. Harvard University Press.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cole, H., & Griffiths, M. D. (2007). Social interactions in massively multiplayer online role-playing gamers. *Cyberpsychology & Behavior, 10*(4), 575–583. https://doi.org/10.1089/cpb.2007.9988

Costikyan, G. (2002). I have no words & I must design: Toward a critical vocabulary for games. In D. F. Mäyrä (Ed.), *Proceedings of Computer Games and Digital Cultures Conference.* Tampere University Press, pp. 9–23.

Deslauriers, P., Lafrance St-Martin, L. I., & Bonenfant, M. (2020). Assessing toxic behaviour in dead by daylight perceptions and factors of toxicity according to the game's official subreddit contributors. *Game Studies, 20*(4), 21.

Dibbell, J. (1994). A rape in cyberspace; or, how an evil clown, a Haitian trickster spirit, two wizards, and a cast of dozens turned a database into a society. In M. Dery (Ed.), *Flame wars: The discourse of cyberculture* (pp. 237–262). Duke University Press. https://doi.org/10.1515/9780822396765-012

Donaldson, S. (2017). Mechanics and metagame: Exploring binary expertise in League of Legends. *Games and Culture, 12*(5), 426–444. https://doi.org/10.1177/1555412015590063

Draggles, Fearless, Bioluminescence & Scathlocke. (2018, Dec 1). *Will the tribunal return?*. Ask riot. Retrieved October 25, 2022, from https://nexus.leagueoflegends.com/en-au/2018/08/will-tribunal-return/

Duarte, N., & Llansó, E. (2017, 28 Nov). *Mixed messages? The limits of automated social media content analysis*. CDT. Retrieved July, 2022, from https://cdt.org/insights/mixed-messages-the-limits-of-automated-social-media-content-analysis/

Fairfield, J. (2009). The god paradox. *Boston University Law Review, 89*, 1017–1068.

Feinberg, J. (1985). *Offense to others (the moral limits of Criminal Law, Vol 2)*. Oxford University Press.

Fleiss, J. L., Levin, B., & Cho Paik, M. (2003). The measurement of interrater agreement. In W. A. Shewart & S. S. Wilks (Eds.), *Statistical methods for rates and proportions* (pp. 598–626). John Wiley & Sons.

Foo, C., & Koivisto, E. (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 (ACE '04). Association for Computing Machinery, New York, NY, USA*, pp. 245–250. https://doi.org/10.1145/1067343.1067375

FPA, & ADL. (2020, Dec 3). *Building a penalty and reporting system*. ADL. Retrieved October 25, 2022, from https://fairplayalliance.org/wp-content/uploads/2020/12/FPA-Framework-Building-Penalty-Reporting-System.pdf

FPA. (2020, Dic 1). *Disruption and harms in online gaming framework*. ADL. Retrieved October 25, 2022, from https://www.adl.org/fpa-adl-games-framework

Gillespie, T. (2018). *Custodians of the internet : Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

Gillespie, T., Aufderheide, P., Carmi, E., Gerrard, Y., Gorwa, R., Matamoros-Fernández, A., Roberts, S. T., Sinnreich, A., & Myers West, S. (2020). Expanding the debate about content moderation: Scholarly research agendas for the coming policy debates. *Internet Policy Review*, *9*(4), 30. https://doi.org/10.14763/2020.4.1512

Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1), 1–15. https://doi.org/10.1177/2053951719897945

Grandprey-Shores, K., He, Y., Swanenburg, K. L., Kraut, R., & Riedl, J. (2014). The identification of deviance Algorithmic content moderation: technicaland its impact on retention in a multiplayer game. In *CSCW '14: Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, February 15-19, 2014, Baltimore, MD,* pp. 2265–2274. https://doi.org/10.1145/2531602.2531724

Guardiola, E. (2019). Gameplay definition: A game design perspective. In *20th International Conference on Intelligent Games and Simulation (*Game-on 2019*)*, Dec 18, Breda, The Neetherlands, pp. 5–10.

ITU (2021, Dec 1). *Measuring digital development: Facts and figures*. ITU. Retrieved 25, October, 2022, from https://www.itu.int/itu-d/reports/statistics/facts-figures-2021/

Johnson, N. F., Xu, C., Zhao, Z. Y., Ducheneaut, N., Yee, N., Tita, G., & Hui, P. M. (2009). Human group formation in online guilds and offline gangs driven by a common team dynamic. *Physical Review E, 79*(6), 1–11. https://doi.org/10.1103/PhysRevE.79.066117

Johnson, D., Nacke, L. E., Wyeth, P., & Assoc Comp, M. (2015). All about that base: Differing player experiences in video game genres and the unique case of MOBA games. In: *Chi 2015: Proceedings of the 33rd Annual Chi Conference on Human Factors in Computing Systems* (pp. 2265–2274). Association for Computing Machinery. https://doi.org/10.1145/2702123.2702447

Jørgensen, K., & Mortensen, T. E. (2022). Whose expression is it anyway? Videogames and the freedom of expression. *Games and Culture, 17*(7–8), 997–1014. https://doi.org/10.1177/15554120221074423

Keller, D. (2018, Oct 31). *Internet platforms: Observations on speech, danger, and money*. Hoover Institution's Aegis Paper Series. Retreived 25 October, 2022, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3262936

Kordyaka, B., Jahn, K., & Niehaves, B. (2020). Towards a unified theory of toxic behavior in video games. *Internet Research, 30*(4), 1081–1102. https://doi.org/10.1108/INTR-08-2019-0343

Kou, Y. (2020). Toxic behaviors in team-based competitive gaming: the case of League of Legends. *Chi 2020: Proceedings of the 38rd Annual Chi Conference on Human Factors in Computing Systems Association for Computing Machinery*. https://doi.org/10.1145/3410404.3414243

Kou, Y., Li, Y., Gui, X., Suzuki-Gill, E., Acm. (2018). Playing with streakiness in online games: how players perceive and react to winning and losing streaks in League of Legends. *In Chi 2018: Proceedings of the 36rd Annual Chi Conference on Human Factors in Computing Systems Association for Computing Machinery*. https://doi.org/10.1145/3173574.3174152

Kou, Y., & Nardi, B. (2014). Governance in league of legends: A hybrid system. *Conference: Foundations of Digital Games, April 3 - 7, 2014, Fort Lauderdale,FL,* p. 10.

Kwak, H., & Blackburn, J. (2015). Linguistic analysis of toxic behavior in an online video game [Proceedings Paper]. *Social Informatics, 8852,* 209–217. https://doi.org/10.1007/978-3-319-15168-7_26

Le, H. T. K., Carrel, A. L., & Shah, H. (2021). Impacts of online shopping on travel demand: A systematic review. *Transport Reviews, 42*(3), 273–295. https://doi.org/10.1080/01441647.2021.1961917

Lessig, L. (1999). *Code: And other laws of cyberspace*. Basic Books.

Lewington, R., & Committee., T. F. P. A. E. S. (2021, April). *Being 'targeted' about content moderation: Strategies for consistent, scalable and effective response to disruption & harm*. FPA. Retrieved 20 July, 2022, from https://fairplayalliance.org/wp-content/uploads/2022/06/FPA-Being-Targeted-about-Content-Moderation.pdf

Liber, C. (2021, Nov 22). *Worlds 2021 final peaked at 73 million viewers. esports-as. 115%*. As. Retreived October 25, 2022, from https://esports.as.com/league-of-legends/final-Worlds-pico-millones-espectadores_0_1522047787.html

Lupton, D. (2015). *Digital sociology*. Routledge.

Maher, B. (2016). Can a video game company tame toxic behaviour? *Nature, 531*(7596), 568–571. https://doi.org/10.1038/531568a

Martens, M., Shen, S., Iosup, A., Kuipers, F., & Ieee. (2015). Toxicity detection in multiplayer online games. *2015 International Workshop on Network and Systems Support for Games (Netgames)*.

Matias, J. N. (2019). Preventing harassment and increasing group participation through social norms in 2,190 online science discussions. *Proceedings of the National Academy of Sciences, 116*(20), 9785–9789. https://doi.org/10.1073/pnas.1813486116

McWhertor, M. (2012, Oct 13). *The League of Legends team of scientists trying to cure 'toxic behavior' online*. Polygon. Retrieved May 25, 2022, from https://www.polygon.com/2012/10/17/3515178/the-league-of-legends-team-of-scientists-trying-to-cure-toxic

Meehan, M. (2006). Virtual property: Protecting bits in context. *Richmond Journal of Law & Technology, XIII*(2), 1–48.

Miró-Llinares, F., & Johnson, S. D. (2018). Cybercrime and place: Applying environmental criminology to crimes in cyberspace. In Bruinsma, Gerben J.N., & Johnson S. D. (Eds.), *The Oxford Handbook of Environmental Criminology* (pp. 883–906). Oxford University Press.

Miró-Llinares, F. (2016). Taxonomy of violent communication and the discourse of hate on the internet. *Idp-Internet Law and Politics,* (22), 82–107. https://doi.org/10.7238/idp.v0i22.2975

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

Mora-Cantallops, M., & Sicilia, M.-A. (2018). MOBA games: A literature review. *Entertainment Computing, 26,* 128–138. https://doi.org/10.1016/j.entcom.2018.02.005

Mora-Cantallops, M., & Sicilia, M. A. (2018). Exploring player experience in ranked League of Legends. *Behaviour & Information Technology, 37*(12), 1224–1236. https://doi.org/10.1080/0144929x.2018.1492631

Mulligan, J., Patrovsky, B., & Koster, R. (2003). *Developing online games: An insider's guide*. Indianapolis Ind.

Neto, J. A. D., & Becker, K. (2018). Relating conversational topics and toxic behavior effects in a MOBA game. *Entertainment Computing, 26,* 10–29. https://doi.org/10.1016/j.entcom.2017.12.004

Oldenburg, R. (1989). *The great good place: Café's, coffee shops, community centers, beauty parlors, general stores, bars, hangouts, and how they get you through the day*. Paragon House.

OP.GG. (2022, July 1). *Game stats by tier*. OPGG. Retrieved July 22, 2022, from https://euw.op.gg/statistics/tiers

Plunkett, L. (2013, January 24). *Entire League Of Legends team disqualified for, yes, "toxic behaviour"*. Kotaku. Retrieved October 25, 2022, from https://kotaku.com/entire-league-of-legends-team-disqualified-for-yes-t-5978832

Rimington, E., Weal, M., Leonard, P., & Acm. (2016). A theoretical framework for online game society: The case of League of Legends. *Proceedings of the 8th ACM Conference on Web Science (WebSci '16). Association for Computing Machinery, New York, NY,* pp. 355–357. https://doi.org/10.1145/2908131.2908194

Riot Games. (2013, May 2). *Player reporting guide and FAQ*. LOL support. Retrieved June 28, 2022, from https://support-leagueoflegends.riotgames.com/hc/en-us/articles/201752884-Player-Reporting-Guide-and-FAQ

Riot Games. (2021 April 30). *Terms of service*. Riot Games. Retrieved June 28, 2022, from https://www.riotgames.com/en/terms-of-service

Salter, M. (2017). From geek masculinity to Gamergate: The technological rationality of online abuse. *Crime, Media, Culture, 14*(2), 247–264. https://doi.org/10.1177/1741659017690893

Schiano, D. J., Nardi, B., Debeauvais, T., Ducheneaut, N., & Yee, N. (2014). The "lonely gamer" revisited. *Entertainment Computing, 5*(1), 65–70. https://doi.org/10.1016/j.entcom.2013.08.002

Shen, C., Sun, Q., Kim, T., Wolff, G., Ratan, R., & Williams, D. (2020). Viral vitriol: Predictors and contagion of online toxicity in World of Tanks. *Computers in Human Behavior, 108*, 106343.

Shenkman, C., Thakur, D., & Llansó, E. (2021, May 20). *Do you see what i see? Capabilities and limits of automated multimedia content analysis*. Centre for democracy and law. Retrieved 25 October, 2022, from https://cdt.org/insights/do-you-see-what-i-see-capabilities-and-limits-of-automated-multimedia-content-analysis/

Sparrow, L., Gibbs, M., & Arnold, M. (2021). The ethics of multiplayer game design and community management: Industry perspectives and challenges. *CHI Conference on Human Factors in Computing Systems (CHI '21),May 08-13, 2021, Yokohama, Japan. ACM, New York, NY*, p. 19. https://doi.org/10.1145/3411764.3445363

Steinkuehler, C. A., & Williams, D. (2006). Where everybody knows your (screen) name: Online games as "third places." *Journal of Computer-Mediated Communication, 11*(4), 885–909. https://doi.org/10.1111/j.1083-6101.2006.00300.x

Suzor, N. P. (2019). *Lawless the secret rules that govern our digital lives*. Cambridge University Press.

Timttamoster. (2022, March 21). *Making behavioral systems matter more*. Riot Games. Retreived 25 October, 2022, from https://www.leagueoflegends.com/en-us/news/dev/dev-making-behavioral-systems-matter-more/

Tseng, Y. S. (2020). The principles of esports engagement: A universal code of conduct. *Journal of Intellectual Property and Law, 27*(4), 209–247.

Twenge, J. M., Martin, G. N., & Spitzberg, B. H. (2019). Trends in U.S. adolescents' media use, 1976–2016: the rise of digital media, the decline of TV, and the (near) demise of print. *Psychology of Popular Media Culture, 8*(4), 329–345. https://doi.org/10.1037/ppm0000203

TwitchTracker. (2022, May 22). *Twitch games statistics*. Twich. Retrieved May 22, 2022, from https://twitchtracker.com/statistics/games

Wijman, T. (2021, May 6). *Global games market to generate $175.8 billion in 2021; despite a slight decline, the market is on track to surpass $200 billion in 2023*. Newzoo. Retreived 25 October, 2022, from https://newzoo.com/insights/articles/global-games-market-to-generate-175-8-billion-in-2021-despite-a-slight-decline-the-market-is-on-track-to-surpass-200-billion-in-2023

Yee, N. (2006). The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments. *Presence-Virtual and Augmented Reality, 15*(3), 309–329. https://doi.org/10.1162/pres.15.3.309

Yee, N. (2014). *The proteus paradox*. Yale University Press.

Zachary, M. (2019, July 2). *Why AI can't fix content moderation*. TheVerge. Retrieved July 20, 2022, from https://www.theverge.com/2019/7/2/20679102/content-moderation-ai-social-media-behind-the-screen-sarah-t-roberts-vergecast

Zaragoza, A. (2022, January 12). *League of Legends promete seguir arrasando en cifras este 2022*. ESPM. Retrieved October 25, 2022, from https://www.esportmaniacos.com/lol/cuantos-jugadores-league-of-legends-2022/