Computers in Human Behavior Reports 19 (2025) 100698

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.sciencedirect.com/journal/computers-in-human-behavior-reports

Review

Defining toxicity in multiplayer online games: A systematic literature review

Bastian Kordyaka [a,*], Sukran Karaosmanoglu [b], Samuli Laato [c]

[a] Åbo Akademi University, Finland
[b] Universität Hamburg, Germany
[c] University of Turku, Finland

ARTICLE INFO

Keywords:
Toxicity
Toxic behavior
Online games
Multiplayer gaming
Definition
Multidimensional definition
Review
Synthesis
eSports

ABSTRACT

Driven by the technological advances of recent years and the opportunity to interact in real-time with others worldwide, toxicity in multiplayer games emerged as a major issue impacting players' well-being and the eSports industry—a lose–lose situation. Despite its urgency, there remains a lack of consensus on the definition of toxicity. To address this, we conducted a systematic literature review (N=853), identifying 32 articles in human–computer interaction databases. Analyzing the identified articles using inductive content analysis, we present (i) a complex picture of various toxicity conceptualizations in the existing literature, (ii) a unifying multidimensional definition for toxicity, (iii) a demonstration of the definition's application as a sequential process relating to widely encountered acts of toxicity, and (iv) recommendations for future research on toxicity, particularly in the growing domain of eSports. Specifically, we provide the following definition for toxicity in multiplayer online games: *"A collective term for acts that are perceived as disruptive by other players that do not occur as a requirement of gameplay that can take different (a) forms of interaction (text and/or speech and/or behavior), (b) targets (teammates and/or opponents), (c) intentions (external and/or internal), and (d) timing (action and/or reaction)*."

1. Introduction

Toxicity has become one of the most significant challenges in popular and commercially successful multiplayer games, such as League of Legends (LoL), Defense of the Ancients 2 (DOTA 2), and PlayerUnknown's Battlegrounds (PUBG). These games are at the forefront of the rapidly growing phenomenon of eSports, the competitive play of video games (Cestino, Macey, & McCauley, 2023; Kordyaka, Pumplun, Brunnhofer, Kruse and Laato, 2023; Scholz, 2020). As a novel form of aggressive behavior, the meaning of the scientific term "toxicity" was brought to the fore by scientists a good ten years ago, such as Blackburn and Kwak (2014) and Kwak (2014). Toxicity is enabled by the need for real-time interaction and competition with other players (e.g., Ma, Li, & Kou, 2023; Mandryk et al., 2023; Sparrow, Gibbs, & Arnold, 2021). Previous research has already shown that toxicity leads to substantial negative consequences in terms of player experience and mental health (e.g., Kordyaka, Laato, Weber, & Niehaves, 2024; Kowert, 2020; Kowert & Cook, 2022) and the resulting economic impact due to player attrition (e.g., Alliance & League, 2020; Kordyaka, Jahn, & Niehaves, 2020). Having recognized and communicated the negative consequences of such behavior some time ago (e.g., Alliance and League (2020) and Kowert, Botelho, and Newhouse (2022)), developers have taken various measures to combat toxicity, including (i) the ability for players to block and mute unwanted interactions (e.g., Kordyaka et al., 2020; Türkay, Formosa, Adinolf, Cuthbert, & Altizer, 2020), (ii) the introduction of honor systems that reward positive interactions (e.g., Marques, Schumann, & Mariconti, 2024; Poeller, Dechant, Klarkowski, & Mandryk, 2023; Poeller & Robinson, 2024) and (iii) recognizing and punishing offenders (e.g., Kordyaka, Klesel, & Jahn, 2019; Kou, 2021). However, despite these measures, toxicity in multiplayer players remains a widespread problem, and how best to address and manage it is unclear.

In this regard, a particular challenge in dealing with toxicity is the lack of a shared understanding of the concept of toxicity and the behaviors associated with it. Although various authors have already attempted to define toxicity, such as Türkay et al. (2020), who describe toxicity as an "umbrella term for various forms of negative behavior by players in online environments", or Frommel and Mandryk (2024), who understand toxicity as "disruptive behaviors that are perceived as harmful by others", it is still unclear when and in which game situations players experience these negative or disruptive behaviors of other players and which aspects should be taken into account. This is partly due to two reasons. Firstly, there are usually general and unspecific definitions (e.g., a collective term of negative behaviors)

* Correspondence to: Åbo Akademi University, Tuomiokirkontori 3, 20500 Turku, Finland.
E-mail addresses: bastian.kordyaka@abo.fi (B. Kordyaka), sukran.karaosmanoglu@uni-hamburg.de (S. Karaosmanoglu), samuli.laato@utu.fi (S. Laato).

https://doi.org/10.1016/j.chbr.2025.100698
Received 11 December 2024; Received in revised form 12 March 2025; Accepted 15 May 2025
Available online 22 July 2025

used by researchers. Secondly, and most importantly, the idiosyncratic characteristics of the environment (e.g., interaction in real-time, disinhibition) and the resulting manifestations of toxicity are often not sufficiently considered in comparison to older constructs of aggressive behavior. To make matters worse, the literature sometimes refers to different and supposedly contradictory aspects in connection with toxicity. For example, the Fair Play Alliance (FPA) describes toxicity as a term that is "ambiguous and imprecise and encompasses very different aspects, such as conflicts caused by poor game design or matchmaking systems, intentionally inappropriate or abusive behavior by a player, or completely accidental misunderstandings without malice" (Alliance & League, 2020). But how can the above-mentioned group of toxic behaviors, such as malicious events be distinguished from accidental misunderstandings? To this end, a closer look at conceptualizations of existing work within human–computer interaction (HCI) research on toxicity seems helpful.

To better understand and develop suitable solutions to address the toxicity challenge, we need a comprehensive understanding that considers the multiple aspects of toxicity. Therefore, this paper asks the following research questions (RQs):

RQ$_1$: How is toxicity described in existing HCI research?

RQ$_2$: How can existing definitions of toxicity be summarized into a unified definition?

To answer these questions, we reviewed existing research on the definition of toxicity within HCI and derived a multidimensional definition (i.e., illustrating an explanation that encompasses multiple aspects or perspectives of a concept, considering various dimensions simultaneously to provide a more comprehensive understanding Vallerand, Deshaies, Cuerrier, BrièRe, & Pelletier, 1996) of toxicity. We followed the PRISMA 2020 checklist (Page et al., 2021) to guide our literature review. On this basis, we conducted an inductive content analysis on the identified articles ($N = 32$) using a codebook and broke down existing definitions into their various facets. Based on this, we propose a multidimensional definition of toxicity that touches on the conditions in which toxicity manifests. To demonstrate the added value of our definition, we used frequently used toxic behavior terms in HCI (such as cheating, flaming, or trolling) and mapped them into our identified dimensions of our definition.

Based on the empirical results of our work, we provide a comprehensive understanding of toxic behavior and its characteristics, laying the foundation for rigorous future research and the development of accurate toxicity measurement tools. In addition, we present empirical findings that can promote human-centered design practice in developing facilitation systems to reduce toxicity. These insights aim to improve the theoretical and practical aspects of addressing toxic behavior in HCI research.

2. Related work

This section summarizes what the term "definition" encompasses in Section 2.1 and highlights previous work related to toxicity in Section 2.2.

2.1. Foundations for clear definitions

To better understand the objectives of our study, it is helpful to explain the basis for clear definitions and understandings of terms. In this paper, we understand a definition as a statement about the meaning of a concept or term (Bell, 2008). Clear definitions are essential in empirical research as they increase precision and clarity and ensure that every reader of a text can understand the variables and concepts used in the same way, thereby reducing ambiguity and confusion and avoiding misunderstandings (Rossi, 2006). They also enable researchers to apply terms consistently, which is crucial for the reliability of research results (Mohamad, Sulaiman, Sern, & Salleh, 2015). In addition, they also facilitate reproducibility, as they allow other researchers to repeat the study and verify the results (Lyons, 1977). In summary, clear definitions enable comparability between different studies, create a coherent body of knowledge and promote understanding in a particular field.

In the context of definitions, it is important to know their building blocks better. Often in the context of definitions, a distinction is made between the two terms "definiendum" (what the definition attempts to explain) and "definiens" (characteristics that the definition encompasses) (Mitchell & Karttunen, 1992). In the context of our study, toxicity in video games is our definiendum, and we are looking for possible components of the definiens to provide clues to the who, what, when, where, why and how of toxicity in video games. Various taxonomies for the classification of definitions can also be found in the literature (Copi, Cohen, & McMahon, 2016; Flowerdew, 1992) (see Table 1). Since one of the central aims of our study is to combine different existing definitions of toxicity, we want to derive a multidimensional definition of toxicity.

2.2. Previous toxicity research

Building on previous research on aggressive behavior on the internet (Balci & Salah, 2015; Lin, 2013; Williams & Clippinger, 2002; Zhen, Xie, Zhang, Wang, & Li, 2011), the concept of toxicity has been increasingly coming into focus within HCI research (Kwak, 2014; Neto, Yokoyama, & Becker, 2017; Türkay et al., 2020) for more than a decade. Toxic behavior has been studied in various settings, including the workplace, where it manifests as actions like verbal abuse, sabotage, and undermining colleagues—creating a hostile environment that harms morale, productivity, and organizational health (Pelletier, 2010). However, for the purposes of our work, we want to focus exclusively on the context of video games, which allows us to control various confounding factors and influences such as workplace hierarchies, face-to-face social cues, and career-driven motivations, and thus derive a definition of toxicity in the unique context of online gaming. Within the gaming context, researchers have repeatedly attempted to define the term or simply used it implicitly, often emphasizing the aspect of a collective term with a negative influence on gamers' lives (Adinolf & Turkay, 2018; Kou, 2020; Martens, Shen, Iosup, & Kuipers, 2015). Researchers often initially equated toxicity with phenomena such as cyberbullying (Blackburn & Kwak, 2014; Kwak, 2014) and emphasized associated behaviors that lead to a negative mood during a game (de Mesquita Neto & Becker, 2018; Neto et al., 2017). Yet, unlike the other terms (i.e., aggressive behavior and cyberbullying), the emergence of toxicity is closely linked to the technological progress of recent years as it compasses distinct features that others do not have: (a) the possibility of interacting with other players in real-time and the sophisticated matchmaking system ensuring that there are no power-imbalances (Cavadenti, Codocedo, Boulicaut, & Kaytoue, 2016; Gong et al., 2020) and (b) the presence of online disinhibition and anonymity of players (Beres, Frommel, Reid, Mandryk, & Klarkowski, 2021; Kordyaka et al., 2020). In addition to this confusion between the term of toxicity and other related ones, we see a lack of clarity within toxicity research defining toxicity. Nevertheless, to this day, no single definition has been able to encompass when, how, where toxicity occurs and what behaviors can be considered as toxic.

Considering relevant antecedents for toxic behavior, several preliminary studies can be found. Online disinhibition, which describes the moral disinhibition in online contexts as well as the associated anonymity and perceived lack of consequences, was particularly influential (Kordyaka & Kruse, 2021; Lapidot-Lefler & Barak, 2012; Suler, 2004). Furthermore, previous work has already shown that toxicity in many games is already normalized and part of the immanent gaming culture (Beres et al., 2021) and shows the influence of variables such as attitude, social norm, and behavioral control (Kordyaka et al., 2020). Also worth mentioning are (a) the dependence of players on each other and the existing competition as favoring factors in primarily multiplayer games (Kordyaka, Pumplun et al., 2023) and (b) the substantial

Table 1
Types of definitions.

No	Definition	Explanation
1	Stipulative	Establishes a new meaning for a term or redefines a term for a specific purpose to bring clarity to a discussion.
2	Lexical	Describes how a term is commonly used in everyday language, aiming to reflect the known meaning of a term.
3	Clarifying	Seeking to make the meaning of a term more precise, especially when that term is ambiguous.
4	Theoretical	Explains a term in the context of a theory attempting to provide a deeper understanding of a term.
5	Persuasive	Influences a particular opinion by framing a term in such a way that a certain judgmental connotation resonates.
6	Multidimensional	Organizes a term by classifying it into dimensions based on specific characteristics.

overlap of roles involved (i.e., perpetrators, victims, and bystanders), suggesting a general susceptibility of players to toxic behavior questioning the existing understanding (Kordyaka, Laato, Jahn, Hamari and Niehaves, 2023).

One aspect that has also been repeatedly emphasized in research is the negative consequences of toxicity for game developers and the eSports industry as a whole (Kordyaka et al., 2020; Kordyaka & Kruse, 2021), which is also one of the reasons why the industry has increasingly founded various initiatives in recent years that have dealt with ways of limiting toxicity, such as the FPA and the Anti-Defamation League (Alliance & League, 2020; League, 2021). Accordingly, game developers have already developed and applied various tools to limit and buffer the challenges related to toxic behavior. Notable in this regard are tools for identifying toxic behavior, such as (a) algorithms and artificial intelligence (AI)-driven systems that automatically detect offensive language and inappropriate behavior in chat (Costa, Drachen, Souza, & Xexéo, 2023; Yang, Grenon-Godbout, & Rabbany, 2024); (b) reporting that allows players to report problematic content or behavior via integrated functions (Kordyaka et al., 2019); (c) filtering mechanisms that automatically filter out profanity and offensive language from chat messages before they are displayed to other players (Adinolf & Turkay, 2018); (d) mute and avoid functions that allow players to mute other players or put them on an ignore list to avoid future interactions (Canossa, Salimov, Azadvar, Harteveld, & Yannakakis, 2021). What all these tools have in common, however, is that toxicity is still an unresolved issue. Game developers also often use sanctions such as temporary bans, permanent bans, or other penalties to hold players accountable who repeatedly violate rules of conduct (Kou, 2021).

For our work, we aim for deriving an unifying definition of toxicity, as it allows us to describe the concept of toxicity by breaking it down into different building blocks. It also allows us to integrate different understandings of toxicity into a coherent whole, facilitating the comparison, and understanding of the relationships between different interpretations of toxicity in a given context. This seems particularly relevant in our study of toxicity in video games, as the term toxicity is already used in other contexts, albeit with different meanings. To mention in this context, without claiming to be exhaustive, are (a) toxic masculinity (e.g., refers to cultural norms and behaviors that promote rigid and harmful stereotypes of masculinity that emphasize dominance, emotional suppression, and aggression) and (b) toxicity in the workplace (e.g., harmful behaviors, attitudes, or environments that create a negative, hostile, or dysfunctional work culture) (Appelbaum & Roy-Girard, 2007; Kusy & Holloway, 2009).

3. Methodology

To contribute to a comprehensive understanding of how the existing literature defines toxicity, we conducted a systematic literature review (Booth, James, Clowes, Sutton, et al., 2021). Systematic literature reviews seek to synthesize existing research, ensuring robust evidence-based conclusions and identifying gaps for future study (Munn et al., 2018; Pollock & Berge, 2018) and have been increasingly performed in the HCI context (Laato, Tiainen, Najmul Islam, & Mäntymäki, 2022; Rogers et al., 2024).

For our systematic review, we followed the recommendations of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) reporting checklists (Moher et al., 2015; Page et al., 2021) and illustrated our review procedure in Fig. 1. Nevertheless, we echo the findings of the recent paper on the use of PRISMA in HCI papers (Rogers et al., 2024); some items of the PRISMA checklist were not applicable to our review, such as the 13e (i.e., causes of heterogeneity), since we did not perform any statistical procedures in our review. Following the best practices (Moher et al., 2015; Shamseer et al., 2015), we supplement our review with a PRISMA protocol (see Appendix B).

Overall, our review consisted of seven steps: (i) search strategy (ii) identification of articles, (iii) screening step, (iv) eligibility step, (v) backward snowballing step, (vi) data extraction, and (vii) inductive content analysis of identified papers.

To specify our search strategy, we conducted several informal (preliminary) searches using basic terms (e.g., toxicity, video games) in the ACM Digital Library (The ACM Full-Text Collection)[1] to identify relevant components of an appropriate search term for our study. After doing so, due to a discussion between three researchers, we specified a search term consisting of the components definiendum, definiens, context, and time interval, whereby all components were connected with an AND. Since a search term is by nature often incomplete or at least limited in scope, as it cannot include all possible relevant terms, variations, or nuances of a term, we adopted a backward snowball approach to iteratively broadening our search. This technique begins with a core set of identified relevant studies or sources and systematically follows up their references (i.e., cited works) to identify additional, possibly missed articles (Jalali & Wohlin, 2012). We list our search terms and details in Table 2. Additionally, following previous studies (Karaosmanoglu, Cmentowski, Nacke, & Steinicke, 2024; Laato et al., 2022), we decided to select a total of six relevant databases to examine in more detail: (i) Association for Computing Machinery (ACM) Digital Library (The ACM Full-Text Collection), (ii) Association for Information Systems (AIS) eLibrary,[2] (iii) Scopus,[3] (iv) Taylor & Francis,[4] (v) Emerald,[5] and (vi) Web of Science[6] as the context for our search.

Overall, we chose Covidence and Excel as software solutions for the following steps of our review as they facilitate efficient screening of studies, data extraction, and collaboration between researchers (Kellermeyer, Harnke, & Knight, 2018).

3.1. Identification of records

After deciding on a search query, we searched the six selected databases in January 2024. We downloaded all the identified papers' bibliographic information in .csv format and then uploaded them to

[1] https://dl.acm.org/.
[2] https://aisel.aisnet.org/.
[3] Scopus indexes many journals and computer science relevant databases, such as IEEExplore https://www.scopus.com/.
[4] https://www.tandfonline.com/.
[5] https://www.emerald.com/insight/.
[6] https://www.webofscience.com/wos/author/search.

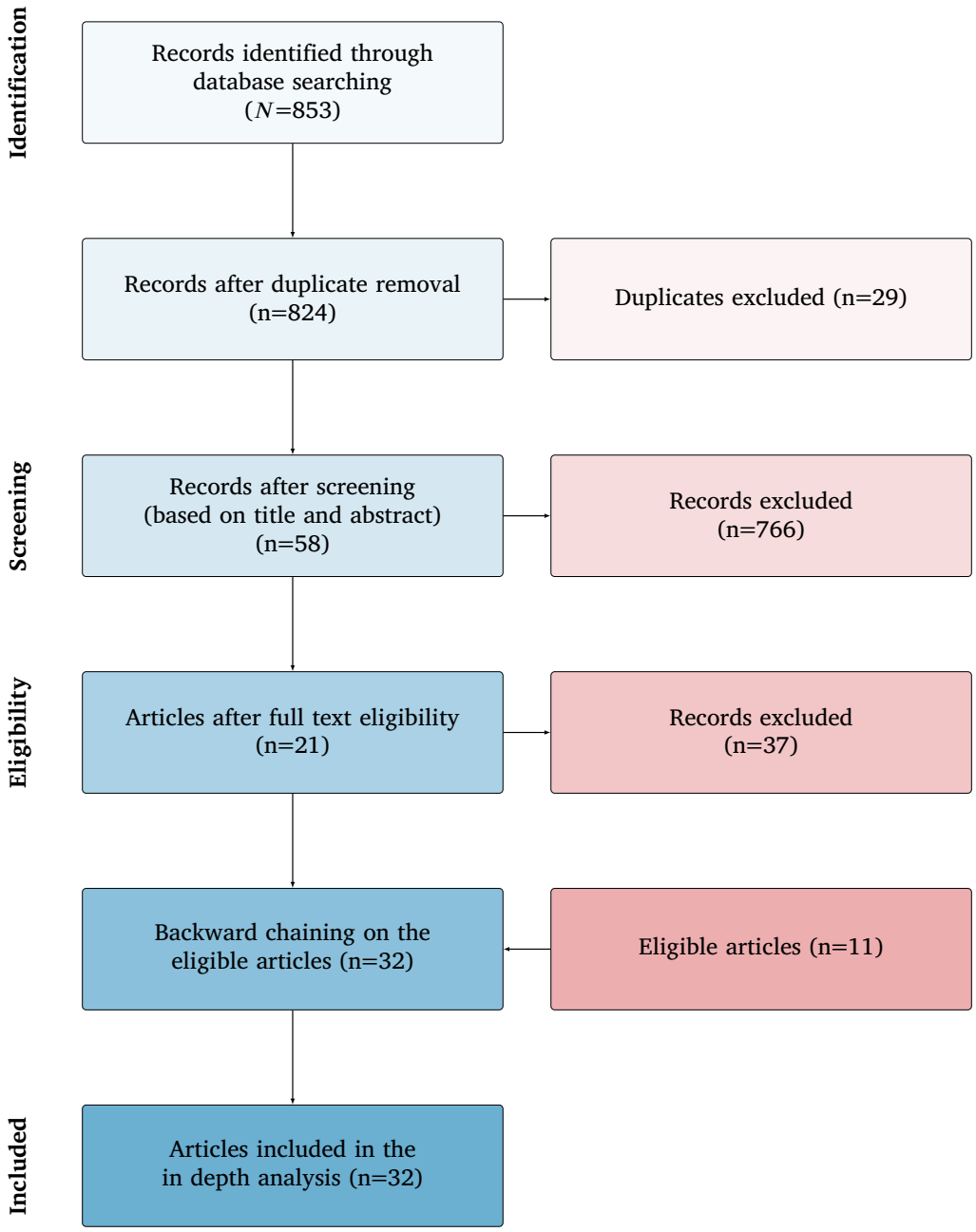

Fig. 1. Literature search with screening and eligibility testing.

Table 2
Search term components.

Component	Search term
Definiendum	(“toxicity” OR “toxic behavior” OR “toxic behaviour”)
Definiens	(“definition” OR “define”)
Context	(“esport” OR “e-sport” OR “dota2” OR “dota 2” OR “cs: go” OR “csgo” OR “counterstrike” OR “counter-strike” OR “pubg” OR “playerunknown*” OR ‘fortnite’ OR ‘lol’ OR ‘valorant’ OR ‘r6 siege’ OR ‘overwatch’ OR ‘call of duty’ OR ‘hearthstone’ OR ‘heroes of the storm’ OR ‘halo 5’ OR ‘moba’ OR ‘game’)
Time interval	01/2014 to 12/2023

Covidence.[7] Our search yielded a total of 853 articles (see Table 3): the ACM Digital Library (n = 161), the AIS eLibrary (n = 112), Scopus (n = 187), Taylor & Francis (n = 300), Emerald (n = 87), and Web of Science (n = 6). After combining databases, we removed 29 duplicates. This left 824 records at the end of the identification phase.

3.2. Screening of identified records

In the screening phase, we reviewed the 824 records based on their titles and abstracts using our inclusion and exclusion criteria (see Table 4). This step was performed by two of the authors independently deciding on either inclusion or exclusion of the identified papers. Consequently, a total of 766 records were excluded, with the most common reasons for exclusion being that they were either not conducted in the context of video games (e.g., toxicity on Twitter Jhaver, Boylston, Yang, & Bruckman, 2021; Kim, Kim, Kim, & Jang, 2022; Maity, Chakraborty, Goyal, & Mukherjee, 2018) and/or described on the outside of video games context (e.g., organizational psychology Bendell,

[7] https://www.covidence.org/.

Table 3
Number of records discovered in the initial search.

Database	Number
ACM Digital Library	161
AIS eLibrary	112
Scopus	187
Taylor & Francis	300
Emerald	87
Web of Science	6
Sum	853
Duplicates	29
Sum after duplicate removal	824

Table 4
Inclusion and exclusion criteria.

Criteria type	Inclusion	Exclusion
Language	English	Not written in English (e.g., Russian or German)
Type	Peer-reviewed full journal or conference papers	Preprints or Short papers (e.g., workshop papers)
Context	About toxicity in video games, including a definition	Toxicity on Twitter without references to a definition
Perspective	Approaches toxicity from an HCI perspective	Approaching the term from a legal perspective
Time frame	01/01/2014 and 12/31/2023	Not in the specified time

Sutherland, & Little, 2017; Morley & Crossouard, 2016; Oleksiyenko, 2018). Overall, the two authors agreed on the decision of 760 of the 824 records reviewed, which corresponds to 92% agreement. The remaining disagreements between the reviewers were resolved in a consensus meeting. At the end of the screening phase, 58 studies remained in our analysis.

3.3. Eligibility of identified articles

In the eligibility phase, the remaining 58 articles were thoroughly reviewed based on full text independently between two coders. In addition to the inclusion and exclusion criteria we used in the screening step, we sequentially applied an eligibility check; we reviewed whether the papers contained a definition of toxicity in the context of multiplayer video games. As a result of this step, a total of 37 records were excluded, with the most common reasons for exclusion being that they were not full papers (e.g., workshop papers or work-in-progress papers) (Martens et al., 2015; Robinson, Hammer, & Isbister, 2019). The two authors agreed on 33 of the 37 records reviewed, corresponding to 89% agreement. The remaining disagreements between the reviewers were also resolved in a subsequent consensus meeting. At the end of the phase, 21 papers remained in our analysis.

3.4. Backward snowballing approach

To ensure that we did not overlook any relevant papers (similar to other HCI papers Rogers, Karaosmanoglu, Wolf, Steinicke, & Nacke, 2021), we conducted a reverse citation search: backward snowballing search (Kitchenham & Brereton, 2013). Accordingly, the lead author checked the reference lists of all the remaining 21 articles. If an article seemed potentially related to the research topic of our study based on its title, we performed the screening and eligibility procedures on it. For the identified articles, we again applied the backward snowballing approach. Overall, this procedure resulted in 11 additional articles, leading to a final set of 32 articles to be considered in the synthesis (see the full list of papers in Table 5).

3.5. Data extraction

In this phase, we extracted data from the final set of 32 articles. For this, we created a data extraction form including the following fields:

1. Publication year of the study,
2. The authors of the study,
3. Outlet in which the study was published,
4. Main contribution of the study (see the extracted contribution texts in Table A.9), and
5. The study's toxicity definition (see the extracted definition texts in Table A.8).

The data extraction process was performed by the two authors. First, the two authors familiarized themselves with the content of the identified papers. Then, using the form described above, both authors extracted the corresponding information. For the definition of toxicity in the papers, the authors checked if there were any (a) definitions or (if no explicit definition was given) an understanding of toxicity was reported (see all the definitions in Table A.8). As with the previous steps, the data extraction was also performed independently. Both authors then met and summarized the results obtained.

3.6. Inductive content analysis

Since we aimed to provide an overview of how existing literature understands toxicity and to establish a definition for toxicity, we did not perform an additional assessment step to exclude papers based on their quality before our final analysis (i.e., appraisal step). Instead, we wanted to consider as many definitions of toxicity as possible in our review. However, we ensured that we only included full papers subjected to a peer-review process. With this, we aimed to create an overarching framework that summarizes the different understandings of toxicity and accounts for potential differences based on existing literature.

To analyze the extracted data from 32 remaining papers, we conducted an inductive content analysis (Elo & Kyngäs, 2008; Harwood & Garry, 2003). Here, following the best practices (Newton, Rothlingova, Gutteridge, LeMarchand, & Raphael, 2012) and other HCI review papers (Karaosmanoglu et al., 2024), we also provide a positionality statement to highlight how our backgrounds might have enhanced the data analysis: Two of the authors, who have extensively researched and published on toxicity in games, engaged with the analysis, leveraging their personal experience as active multiplayer players to bring a deep understanding of the natural conditions of such environments.

To create a unified codebook, two of the authors first reread the papers. Based on this familiarization step, they created inductive codes. To identify the inductive codes, the authors focused on identifying features that could potentially explain differences and similarities in the understanding of toxicity across the papers. Here, both authors relied on the definitions and contributions of the studies extracted in the data extraction phase (see Tables A.8 and A.9). Both authors created their own codes and then met to unify and discuss them. These discussions led to refining, adding, or deleting codes. After two discussion meetings, the following codes were identified: text, speech, behavior, teammate, opponent, external, internal, action, and reaction. Additionally, in these discussion meetings, the authors grouped the codes into overarching dimensions by conducting an affinity mapping. With this, we aimed to identify comprehensive patterns of features that say something important about the data in relation to a multidimensional definition of toxicity. Specifically, our final codes and dimensions were as follows, leading to our codebook:

1. The three codes text, speech, and behavior were grouped together as the category forms of interaction,
2. The two codes teammates and opponents were grouped together as the category target,

Table 5
List of papers that are included in the analysis.

No	Papers	No	Papers
1	Blackburn and Kwak (2014)	17	Sparrow et al. (2021)
2	Kwak, Blackburn, and Han (2015)	18	Kou and Gui (2021)
3	Neto et al. (2017)	19	Reid, Mandryk, Beres, Klarkowski, and Frommel (2022b)
4	Kou and Gui (2017)	20	Kordyaka, Krath, Park, Wesseloh, and Laato (2022)
5	de Mesquita Neto and Becker (2018)	21	Kowert and Cook (2022)
6	Sengün, Salminen, Mawhorter, Jung, and Jansen (2019)	22	Reid, Mandryk, Beres, Klarkowski, and Frommel (2022a)
7	Kordyaka et al. (2019)	23	Lee, Johnson, Tanjitpiyanond, and Louis (2022)
8	Kou (2020)	24	Monge and O'Brien (2022)
9	Türkay et al. (2020)	25	Kordyaka, Park, Krath and Laato (2023)
10	Kordyaka et al. (2020)	26	Kordyaka, Laato, Jahn et al. (2023)
11	Kowert (2020)	27	Poeller et al. (2023)
12	Shen et al. (2020)	28	Kordyaka, Stelter, Laato and Niehaves (2023)
13	Hilvert-Bruce and Neill (2020)	29	Frommel, Johnson, and Mandryk (2023)
14	Canossa et al. (2021)	30	Liu and Agur (2023)
15	Kordyaka and Kruse (2021)	31	Kordyaka, Laato, Hamari , Scholz and Niehaves (2023)
16	Beres et al. (2021)	32	Ma et al. (2023)

Table 6
Publication outlets of the final set of records.

Outlet	Number	Type	Publisher
ACM Conference on Human Factors in Computing Systems	7	Proceedings	ACM
Hawaii International Conference on System Sciences	5	Proceedings	ScholarSpace
ACM Annual Symposium on Computer–Human Interaction in Play	4	Proceedings	ACM
Entertaining Computing	2	Journal	Elsevier
Computers in Human Behavior	2	Journal	Elsevier
International World Wide Web Conference	1	Proceedings	ACM
International Conference on Web Intelligence	1	Proceedings	IEEE
Conference on Hypertext and Social Media	1	Proceedings	ACM
GamiFIN	1	Proceedings	CEUR
Internet Research	1	Journal	Emerald
IEEE Transactions on Games	1	Journal	IEEE
Media Psychology	1	Journal	Routledge
Games and Culture	1	Journal	Sage
Frontiers in Psychology	1	Journal	Frontiers
Transactions in Social Computing	1	Journal	ACM
Safer Communities	1	Journal	Emerald
Computers in Human Behavior Reports	1	Journal	Elsevier
Sum	32		

3. The two codes, external and internal, were grouped together as the category intention,
4. The two codes, action and reaction, were grouped together as the category timing.

Finally, we quantitatively applied the codes and dimensions identified to the 32 papers identified in our systematic literature review.

4. Findings

This section reports the results of our systematic literature review. First, we illustrate a description of the research field based on the identified studies. Following this, we synthesize a definition for player toxicity.

4.1. Description of the research field

Publication numbers & venues. Our results show the number of published studies showed an upward trend over the period between 2014 and 2023, which indicates the increasing importance of the topic of toxicity. While only one study was published at the beginning of our study period in 2014, a total of eight studies were published at the end of our study period in 2023, as shown in Fig. 2. In addition, 63% (20 of 32) of the studies were conference papers. In terms of conference papers, most publications (n = 7) were published in the ACM Conference on Human Factors in Computing Systems, followed by the Hawaii International Conference on System Sciences (n = 5). As journals, Entertainment Computing and Computers in Human Behavior were leading ones (n = 2) (e.g., see Table 6). Most of the studies described toxicity primarily using a lexical definition. However, we note that some articles may contain elements of more than one type of definition; in such cases, we have decided on the most applicable form of definition. Nevertheless, based on our understanding, none of the articles featured a multidimensional definition or attempted to define toxicity as a multidimensional construct (see the articles' toxicity descriptions in A.8).

Methodologies & context. The majority of papers (n = 18) employed quantitative methods. Most of the identified studies (n = 23) dealt specifically with the two MOBA games LoL or Defense of the Ancients 2, which are characterized by a high degree of competition and need for interaction among players and thus provide a narrow ground for toxicity (Zhang & Kou, 2024). However, the remaining ones did not particularly focus on a specific game and referred to MOBA games in general.

Contribution. To gain an overview of the content of the identified studies, we also distilled the central contributions of the identified studies (see Table A.9) in Appendix A covering topics such as:

- Detecting, predicting, and measuring toxic behavior (Blackburn & Kwak, 2014; Canossa et al., 2021; de Mesquita Neto & Becker, 2018; Kordyaka et al., 2019; Kwak et al., 2015; Reid et al., 2022a; Shen et al., 2020).
- Psychological, social, and cultural drivers of toxic behavior (Beres et al., 2021; Hilvert-Bruce & Neill, 2020; Kordyaka et al., 2020, 2022; Kordyaka, Laato, Hamari et al., 2023; Kordyaka, Park et al., 2023; Kordyaka, Stelter et al., 2023; Liu & Agur, 2023; Sengün et al., 2019; Türkay et al., 2020).

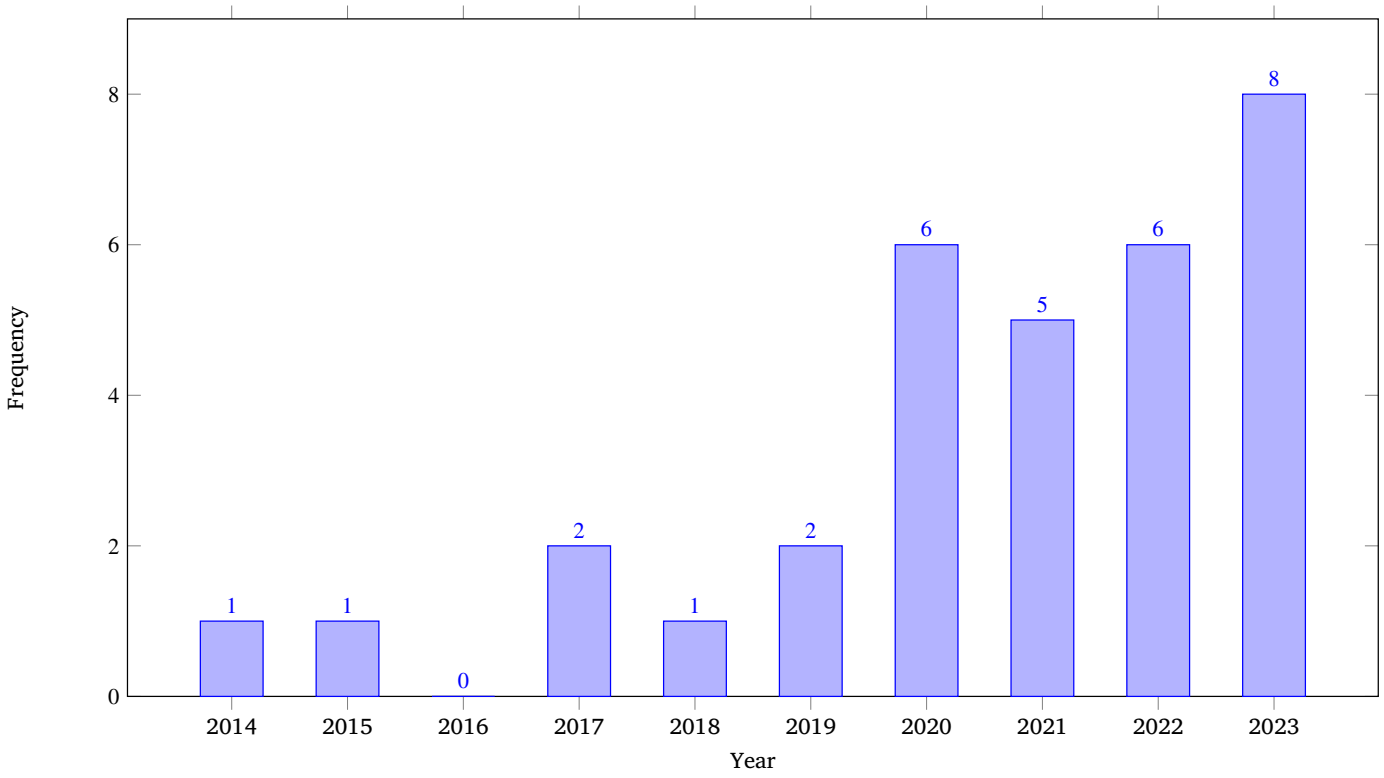

Fig. 2. Histogram of the included studies.

- Impacts and dynamics of toxic behavior (Frommel et al., 2023; Kordyaka, Laato, Jahn et al., 2023; Kou & Gui, 2017; Kowert, 2020; Lee et al., 2022; Monge & O'Brien, 2022; Neto et al., 2017; Poeller et al., 2023).
- Moderation, interventions, and design solutions related to toxic behavior (Kordyaka & Kruse, 2021; Kou, 2020; Kou & Gui, 2021; Kowert & Cook, 2022; Ma et al., 2023; Reid et al., 2022b; Sparrow, Galwey, Jovic, Hardwick, & Butt, 2024; Sparrow et al., 2021).

4.2. Synthesizing a definition for toxic behavior

To propose a comprehensive multidimensional definition of toxicity, we first build on some assumptions of the identified studies (see Table 5), and assume that toxicity is (a) a catch-all term (i.e., a term referring to a collection of behaviors) and (b) perceived as a disruptive act by other actors. Based on our content analysis, we subsequently interpret the quantitative relevance of the inductively determined dimensions (i.e., forms of interaction, targets, intentions, and timing) against the background of potentially explainable differences in understanding of toxicity.

1. *Form of interaction:* The dimension of the form of interaction often described toxic behavior as a mixture of primarily text-based and behavioral toxicity. Only two definitions (6%) explicitly mentioned language-based toxicity, while seven mentioned purely text-based and six purely behavioral toxic behavior (18%). Accordingly, toxic behavior was described at more than one level of interaction and that often by rather general terms such as "trolling, harassment, griefing or cyberbullying", which allow for different forms of interaction at the level of manifestation. For example, Kwak et al. (2015) described "toxicity" as "[...] negative actions such as cyberbullying, teasing, pranks, and cheating, which often occur in groups and may include offensive language, verbal abuse, deliberate feeding, and other harmful activities", potentially encompassing a wide variety of possible forms of interaction. We argue that this is appropriate for the primarily analyzed MOBA games such as LoL, in which the primary forms of interaction are text and pings, and voice is available as a default option within the game with a maximum of one of the other players (Claypool, Decelle, Hall, & O'Donnell, 2015). Considering the frequency with which the target dimension appeared, we recognize it as having notable but relatively low importance in shaping various definitions of toxicity in existing research.
2. *Target:* The contributions to the toxicity target dimension are consistent with the general trends in online gaming, where toxicity is often perceived as a general social problem rather than one that is limited to teammates or opponents. Notably, only five studies (16%) explicitly distinguished between these groups, suggesting that toxicity is often treated as a ubiquitous phenomenon rather than a team-based dynamic. This generalization may overlook important nuances, such as the distinct psychological and strategic implications of toxicity when directed toward teammates versus opponents. Furthermore, the lack of explicit differentiation raises the question of whether interventions to mitigate toxicity should be tailored differently depending on the target audience. Since previous research suggests that many players actively take measures to avoid negative interactions, future studies could examine how this avoidance behavior affects the perception and reporting of toxicity in multiplayer environments, differentiated by in-group or out-group affiliation. Similarly, as the form of interaction dimension, based on how often it was mentioned, we find the target dimension to be

meaningful, yet its impact on differing definitions of toxicity remains relatively limited in the current body of research.

3. *Intention:* The dimension of the intention of toxic behavior showed differences in the understanding of toxicity within the 32 identified studies, which we considered an indication of an important explanation of the dimension with regard to the different perceptions of toxicity. Specifically, 24 of the 32 studies (75%) understood toxicity as an exclusively externally directed intention, while the remaining eight studies also assumed that toxicity can also be understood as an internal attempt (e.g., to cope with frustration and stress) and designated this (in addition to external harm) as a central intention. For example, Kordyaka, Laato, Hamari et al. (2023) understood toxicity as "... a generic term for short-term, non-systematic negative behaviors that can be fueled by situational frustration, anger, and high levels of real-time competition, and that manifest as insults, criticism, resource-hogging, and other actions" that contain both internal and external components. Opposed to the first two dimensions, the intention dimension appeared to hold more weight in influencing different definitions of toxicity, given its frequent occurrence in the literature.
4. *Timing:* The description of the dimension of timing also showed significant differences between the understanding of toxicity in the identified articles. 23 of the identified 32 studies (72%) described toxicity as a proactive action, the remaining 9 studies also described toxicity as a reaction (often in the context of frustration coping attempts during play). With regard to the dimension, we, therefore, also see the potential for existing inconsistencies in the use of the term toxicity. On the one hand, we find authors such as Blackburn and Kwak (2014) who define toxicity as "...a form of cyberbullying, defined as repeated intentional behavior to harm others" as an active act, while Kordyaka et al. (2022) suggested that toxicity describes "...temporary behaviors that generate anger and frustration in players, harming communication and spreading a bad mood, primarily as a reaction to in-game frustration" as a more reactive act. As with the intention dimension before, we understood the timing dimension to play a crucial role in shaping different understandings of toxicity, as indicated by its notable presence in existing studies.

In summary, our analysis underlines that the two dimensions of forms of interaction and targets of toxic behavior have rather little explanatory power for the ambivalent understanding of toxic behavior. In contrast, we found substantial differences between the studies for the two dimensions, intention and timing, which we interpreted as a possible explanation for the ambivalence of the understanding of toxicity. Based on this, we derived a comprehensive definition of toxic behavior based on this categorization. We arranged the dimensions used in ascending order of specificity. Below, we provide a multidimensional definition of toxicity:

Multidimensional Definition of Toxicity

In the context of multiplayer online games, toxicity or toxic behavior is a collective term for acts that are perceived as disruptive by other players that do not occur as a requirement of gameplay that can take different (a) forms of interaction (text and/or speech and/or behavior), (b) targets (teammates and/or opponents), (c) intentions (external and/or internal), and (d) timing (action and/or reaction).

4.3. Application of the definition

To illustrate the added value of using our derived multidimensional definition as a framework and how it can be used to classify toxic behaviors, we demonstrate its application below. To this end, we first describe the framework of the application drawing on previous work, that allows us to illustrate the application of our multidimensional definition based on empirical evidence.

Our structured literature analysis showed that MOBA games such as LoL or DOTA 2 accounted for by far the largest share of the identified studies, at 72%. That is why we will focus on the most common game mode, namely Summoner's Rift Ranked from LoL. This game mode is characterized by high tactical complexity, intense competition, and the need for team interaction to successfully climb the Elo ladder (Kordyaka, Kruse and Niehaves, 2023). Players generally compete in ranked games without malicious intent, as their motivation to perform well and gain points and the sophisticated design of the matchmaking system ensures a fair and competitive environment (Claypool et al., 2015; Kou & Gui, 2014; Zhang, Moradzadeh, Woan, & Kou, 2024). Deliberate losing would run counter to this goal and is prevented by sanctions for repeated misconduct, as earlier work has shown (Kou & Gui, 2021). Furthermore, studies have been able to show that toxic behavior usually arises in the course of the game, often triggered by communication problems or mistakes by other players, which can lead to frustration (Kordyaka, Laato, Jahn et al., 2023; Neto et al., 2017). Due to the prevailing high competitive pressure and the necessity to always make quick and optimal decisions, a multitude of potentially toxic situations arise (Kordyaka, Park et al., 2023; Kou, 2020). To illustrate the application of our multidimensional definition of toxicity, we assume that in these situations both impulsive and reflected toxic actions occur, which can be explained by the dual process model of psychology (Chaiken, 1999; Jahn et al., 2022): While impulsive reactions arise from intuitive processes (system 1), planned toxic actions are the result of analytical processes (system 2). The combination of previous work and assumptions of the process model enables a systematic classification of toxic behavior based on intention and temporal proximity to triggering events. Based on more recent work (Frommel & Mandryk, 2024; Kowert, 2020), we use the following toxic actions as examples, which are listed in Table 7:

Based on our definition, we classify the acts of toxicity listed in the table below concerning a potentially toxic situation in the ranked game of LoL.

- *Cyberbullying:* Since cyberbullying is defined with the criteria (a) repeated behavior (b) an imbalance of power and (c) more than one perpetrator (Kowalski, 2018), there can be no cyberbullying in ordinary ranked games of LoL, since (a) a single behavior is sufficient to be called toxic, (b) players with the same skill level are matched and (c) a single perpetrator is sufficient. We argue that according to this definition of cyberbullying (Kowalski, 2018), cyberbullying requires to fulfill more requirements than toxicity.
- *Smurfing:* This includes toxic actions such as playing on a lower-ranked account or another person's account to improve their rank (Kou, 2020). Therefore, in smurfing, a decision has to be deliberately made to prior to the start of playing a ranked game. Hence, we consider smurfing as a (reflected) planned action.
- *Cheating:* This includes acts such as using hacks or unauthorized programs during gameplay to gain an unfair advantage (Canossa et al., 2021). We argue (as in the case of smurfing before) that corresponding behaviors are (reflective) planned actions by players, often even made before a game starts.
- *Doxxing:* This describes publicly revealing private, personal information about another player without their consent, typically with malicious intent (Kowert & Cook, 2022). However, we consider such behavior not relevant in the context of LoL since only the in-game names of players are available in LoL, making the likelihood of its occurrence very limited.

Table 7
Types of toxic behavior listed in the extant scholarship.

Type	Definition	Source
Cyberbullying	Refers to repeated harm inflicted through the use of electronic devices	Kowalski (2018)
Smurfing	Describes high-level players playing with the accounts of others against less skilled opponents	Kou (2020)
Cheating	Refers to giving an unfair advantage to a player	Canossa et al. (2021)
Doxxing	Refers to personally identifying information about another player that is made public	Kowert and Cook (2022)
Trolling	Refers to the intention to provoke and annoy other players	Hilvert-Bruce and Neill (2020)
Spam pinging	Refers to the short-term, repetitive, and disruptive use of online communication	Kordyaka, Laato, Jahn et al. (2023)
Raging	Refers to aggressive outbursts	Türkay et al. (2020)
Flaming	Refers to the use of aggressive or derogatory language	Kwak et al. (2015)

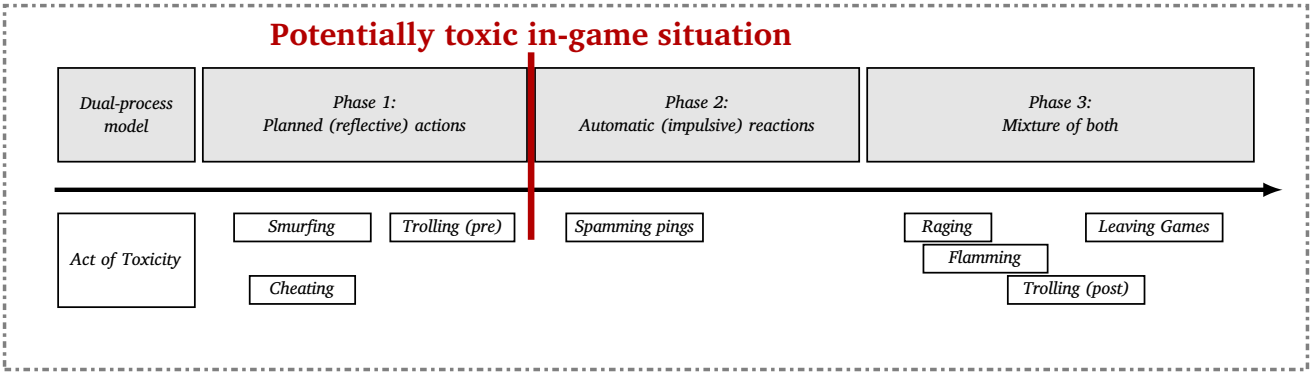

Fig. 3. Sequential classification of toxicity acts.

- *Trolling:* This describes the actions of players who intentionally make bad decisions in the game (Hilvert-Bruce & Neill, 2020). In this case, we assume that this can be a reflective intentional action ("trolling pre") before experiencing a toxic situation in the game (e.g., troll picks of a champion during matchmaking) as well as a more automatic impulsive action ("trolling post") after experiencing a potentially toxic situation (e.g., selling all items after losing a team fight, feeding the opponent).
- *Spamming Pings:* This explains the constant sending of the same signal messages during the game (e.g., the "danger" signal on a teammate who has just died) (Kordyaka, Laato, Jahn et al., 2023). We argue that this is an automatic impulsive form of possible toxic action, which usually occurs automatically as an immediate reaction to a situation in the game.
- *Raging:* Refers to a situation when a player becomes extremely angry after repeatedly dying to an opponent (Türkay et al., 2020). Typically in response to a perceived injustice, failure, or provocation during gameplay and can be considered an automatic impulsive action.
- *Flaming:* This includes toxic actions such as insulting, ridiculing, or belittling other players, including blaming others for mistakes or losses (Kwak et al., 2015). Since the other players in LoL ranked mode are usually unknown when matches start (due to a high number of players and random matchmaking), we assume that this is a rather acquired action during a game as a reaction to in-game situations that contains both automatic impulsive and planned reflective actions.
- *Leaving the game:* This can refer to the cases of quitting the game early or going AFK which puts one's teammates at a significant disadvantage. We think this kind of toxic action falls under both actions. For instance, if a player thinks they will already lose the game in a later stage, this can be seen as planned reflective action. Conversely, it could be simply an impulsive reaction to adverse game events (e.g., losing a Baron fight).

Framework: Our framework encompasses the previous categorization of toxicity acts in relation to the dual process model (see Fig. 3). Overall, our framework is divided into three phases. Phase 1 refers to planned reflective actions. These actions take place before the occurrence of a toxic in-game situation, such as smurfing, cheating, and trolling pre. Conversely, Phase 2 covers the toxic actions that happen after a toxic in-game situation, such as spamming pings. Thus, we call automatic impulsive actions. Lastly, Phase 3 covers both types of actions, i.e., planned and automatic; some examples include raging, flaming, trolling post, and leaving games.

5. Discussion

In this section, we illustrate our key findings (Section 5.1), provide implications for theory (Section 5.2) and practice (Section 5.3), as well as limitations (Section 5.4). Moreover, we explore the broader meaning of the results in the context of the existing literature and contribute to a deeper understanding of toxicity in the field of HCI research.

5.1. Key findings

Based on the findings of our systematic literature review, we empirically answer our two research questions (RQ_1: How is toxicity described in existing HCI research? and RQ_2: How can existing definitions of toxicity be summarized into a unified definition?). In the following, we summarize the key results of our paper in three points:

- Based on a systematic literature search, we identified a total of 32 studies that reveal different understandings of toxicity in the HCI literature and provide a robust snapshot of the phenomenon.
- Through an inductive qualitative content analysis, we identified a total of four relevant dimensions and recorded them in a multidimensional definition, which can be understood as framework conditions under which toxicity manifests itself.
- To demonstrate the added value and application potential of our multidimensional definition, we have shown its application potential with regard to widespread concrete forms of toxicity.

5.2. Implications for theory

Various implications can be derived from the results of our study that are of crucial importance for the theoretical level of existing HCI research related to toxicity. In the following, we discuss three of these that we consider particularly relevant.

First, the identification of 32 studies within our systematic literature review that are relevant to the definition of toxicity provides a solid empirical foundation that can improve the theoretical understanding of the phenomenon. On the one hand, searchers can now use our definition to explain the acts of toxicity they address in their papers; they can clearly communicate, for example, with which kind of intention the form of interaction takes place. On the other hand, researchers can now validate, refine, or challenge — simply classify — their existing theoretical understandings by referring to our definition, leading to more nuanced theoretical advances. In terms of content, most of the identified studies agreed that toxicity is (a) used as a catch-all term and (b) perceived as disruptive and not a requirement of ordinary gameplay. Furthermore, based on our analysis, it is now possible to critically reflect on the terminology used and possible semantic confounding of existing definitions. For example, at the beginning of our study period, the term "cyberbullying" was often used as a synonym for toxicity (e.g., "a form of cyberbullying, defined as repetitive, intentional behavior to harm others through electronic channels" Blackburn & Kwak, 2014), which does not do justice to the complex inner life of toxicity, since, for example, neither a group of perpetrators is needed to generate toxic behavior nor are there significant power differences between players in freely available games with corresponding match-making, both of which are elementary components of the definition of cyberbullying (Kowalski, 2018). Therefore, our systematic literature review provides a solid, granular foundation for future research that makes the phenomenon of toxicity in the HCI context more comprehensible.

Secondly, our inductive identification of the dimensions of our multidimensional definition form of interaction, target, intention, and timing enables us to critically reflect on existing theoretical concepts and to partially modify existing theories. For example, previous studies have already shown that behavioral control is a negative predictor for the occurrence of toxicity (Kordyaka et al., 2020). This shows links with the identification of the two dimensions "intention" and "timing", which indicate that toxicity usually also involves a situational component in dealing with potentially frustrating moments during a game (e.g., coping with stress). We see this as an indication that a theory of toxicity could be developed that considers both the situational and dispositional components of the occurrence of toxicity. This theory could illuminate the dynamic nature of toxicity by demonstrating how situational stressors and individual dispositions interact to promote or inhibit toxic behavior.

Third, the application of the multidimensional definition of toxicity adds considerable value to the theoretical state of the art. Our classification particularly allows toxic behaviors to be sequentially aligned with a game situation. Therefore, this makes it possible to clarify existing contradictions, for example, "malicious events" can be distinguished from "accidental misunderstandings" (Alliance & League, 2020), since malicious events are typically planned actions while accidental misunderstandings usually occur as impulsive, automatic reactions. Further, our classification contributes the efforts of creating a shared understanding around toxicity that eventually support the development of targeted prevention and intervention measures against toxicity (Wijkstra, Rogers, Mandryk, Veltkamp, & Frommel, 2023, 2024). Still, it may also shed new light on the understanding of existing work addressing blocking of other players (e.g., Türkay et al., 2020), honor systems (e.g., Poeller et al., 2023), and punishment (e.g., Kordyaka et al., 2019) in the context of toxicity.

5.3. Implications for practice

Based on the insights of our study, several conclusions can be drawn that promise to add value to existing HCI practice and the eSports industry as a whole as well. We will discuss three of these below.

First, identifying 32 studies through our systematic literature review now provides an expanded and, most importantly, robust empirical basis for developing tailored, human-centered interventions and innovative feedback mechanisms related to the darkside of HCI (Gray, Kou, Battles, Hoggatt, & Toombs, 2018; Kowert, 2020). As a result, interventions can be designed more effectively because they are based on empirically validated knowledge about the causes, dynamics, timing, and, ultimately, the understanding of toxic behavior. Developers and programmers can now use this knowledge as a solid starting point for concrete design implications.

Second, the identification of our inductively determined dimensions provides added value for practitioners and designers as it enables consideration of concrete manifestations of toxicity at a more granular level, which can increase the accuracy of proposed interventions and reduce their ambivalence (Kordyaka & Kruse, 2021). For example, it is now possible to specifically consider and contrast different forms of interaction (e.g., text, language, behavior) and potential targets (e.g., teammates, opponents) when designing interventions, providing expanded opportunities for measurement and evaluation as well as underlying mechanisms.

Third, by applying our multidimensional definition in a sequential classification framework for toxicity, we lay a foundation that allows practitioners and designers to better target specific toxic behaviors based on their sequential occurrence. For example, to reduce the reflexive actions of smurfing or cheating, other forms of intervention (such as educational programs) seem promising compared to more automatic impulsive actions (such as the integration of real-time traffic lights during gameplay that directly affect behavior). In the context of existing measures to combat toxicity, positive interventions and tools from sports psychology, especially real-time stress interventions (Henriksen, Larsen, Storm, & Ryom, 2014; Weinberg & Comar, 1994), appear particularly promising. Such approaches could help not only to identify toxic behavior, but also to counteract it preventively by addressing stress as a central component of toxicity.

5.4. Limitations and outlook

This study has the following limitations. First, as with all systematic literature reviews, our results are influenced by the time frame of the search and the digital databases used in this work. Despite the use of multiple databases, many keywords and the application of a snowballing technique (which we used to minimize potential limitations), it is possible that some relevant papers were overlooked. To address these limitations, future research could expand the time frame, include additional databases, refine keyword strategies, and incorporate more iterative approaches to ensure comprehensive coverage.

Second, our study is subject to the limitations of the research identified in this article. Surprisingly, we found that research on toxicity in the gaming context has so far focused predominantly on MOBA games, often justified by the intense competition and the need for coordinated team interaction. However, there is more recent work that addresses toxicity in other video game genres. For example, Laato, Kordyaka, and Hamari (2024) examine toxic behavior in StarCraft II, a real-time strategy (RTS) game, and identify ten categories of toxic behavior. This study indicates that the harmfulness of toxic acts can be considered a product of severity and frequency, with players' assessment of severity influenced by factors such as direct observation, background, and external influences. This suggests that toxic behavior is not only characterized by direct team interactions and verbal aggression, but in RTS games such as StarCraft II, it also includes strategic sabotage, intentional slowing of the game, or targeted psychological warfare

via in-game chats. Accordingly, we recommend that future studies follow this new line of research and analyze toxicity by considering similarities and differences between different video game genres. Our multidimensional definition could serve as a basis for developing a more comprehensive understanding of the underlying dynamics and influencing factors. A differentiated approach could also help to design targeted countermeasures and prevention strategies for different gaming contexts, for example, through mechanisms to reduce toxic behavior in competitive RTS games or through incentive systems that promote positive behavior.

Third, the inductive search of codes to derive the multidimensional definition involved a significant amount of iterative interpretative work on the part of the authors. While the authors have extensive background in toxicity research and play MOBA games, which can strengthen the data analysis, it is important to note that this could have introduced certain biases into the analysis. To address this, future research could incorporate diverse perspectives through interdisciplinary collaboration and validation by external experts to mitigate potential biases in the interpretative analysis.

6. Conclusion

In this paper, based on a systematic literature review using inductive content analysis, we derived a multidimensional definition of toxicity—an umbrella term for actions in multiplayer games for acts that are perceived as disruptive by other players that do not occur as a requirement of gameplay that can take different (a) forms of interaction (text and/or speech and/or behavior), (b) targets (teammates and/or opponents), (c) intentions (external and/or internal), and (d) timing (action and/or reaction). Building on this, we demonstrated the application of the definition as a sequential classification to better capture

Table A.8
Definitions of toxicity in the identified articles.

No	Definition	Reference
1	A form of cyberbullying, defined as repetitive intentional behavior to harm others through electronic channels.	Blackburn and Kwak (2014)
2	Negative actions such as cyberbullying, griefing, mischief, and cheating that are often grouped together and can include offensive language, verbal abuse, intentional feeding, and other harmful activities.	Kwak et al. (2015)
3	Any kind of behavior that has some negative impact on the game experience of others, including being offensive, making socially unacceptable comments, losing the game on purpose, or leaving the match early.	Neto et al. (2017)
4	A conduct that departs from the norms set for people, encompassing actions such as griefing, flaming, trolling, cheating, and other forms of deviant behavior that are considered detrimental to player experience and the game community.	Kou and Gui (2017)
5	Negative behaviors, including harassment, griefing (gaining enjoyment from intentionally making other players annoyed), trolling, and intentionally helping opposing players.	de Mesquita Neto and Becker (2018)
6	Rude, disrespectful, or unreasonable behavior that is likely to make one leave a discussion, which in the context of gaming means a player leaving the game or turning the chat feature off.	Sengün et al. (2019)
7	Various contextual factors and examples, such as flaming, harassment, and other antagonistic actions in the gaming environment.	Kordyaka et al. (2019)
8	Behaviors performed at others' expense, often driven by competitive gaming contexts and associated with negative social dynamics and emotional states.	Kou (2020)
9	Antagonistic actions such as harassment, cheating, raging, griefing, cyberbullying, and intentionally helping opposing player(s).	Türkay et al. (2020)
10	Encountered when a player comes across a negative event during a game that generates anger and frustration, leading to a harmful, contaminated, and disseminated toxic type of communication.	Kordyaka et al. (2020)
11	Refers to any deviant verbal or behavioral action that takes place on the internet which causes harm to another's health or well-being.	Kowert (2020)
12	Refers to behaviors characterized by incivility, verbal aggression, and antisocial actions, such as trolling and flaming, which degrade the quality of interaction and create a hostile atmosphere.	Shen et al. (2020)
13	The prevalence of hostile and aggressive behaviors, such as insults, threats, and profanity, in online gaming environments, often normalized and justified by the gaming culture.	Hilvert-Bruce and Neill (2020)
14	Behavior that intentionally disturbs another player's experience and well-being, including actions such as flaming, acting nosy, cheating, and illegal behaviors.	Canossa et al. (2021)
15	Aggressive behavior towards teammates that generates a negative atmosphere, which predominantly emerges over the course of a game as a response to negative events.	Kordyaka and Kruse (2021)
16	Various types of negative behaviors, including abusive communications directed towards other players and disruptive gameplay that violates the game's rules and social norms.	Beres et al. (2021)
17	Hostile and disruptive behaviors such as abusive chat, harassment, and cheating.	Kou and Gui (2021)
18	Online behavior that deviates from social norms within a given social context, often manifesting as harassment, hate speech, or other forms of negative conduct.	Sparrow et al. (2021)
19	A set of negative behaviors that disrupt gameplay or game enjoyment, including harassment, cheating, and raging.	Reid et al. (2022a)
20	Temporary behaviors that generate anger and frustration in players, harming communication and spreading a bad mood, primarily as a reaction to in-game frustration.	Kordyaka et al. (2022)
21	Deviant or antisocial behavior in online games that negatively affects the gaming experience of at least one other player.	Kowert and Cook (2022)
22	Abusive communications and disruptive gameplay behaviors directed towards other players, including verbal or textual harassment and actions like griefing that impede others' enjoyment of the game.	Reid et al. (2022b)
23	Antisocial behaviors such as cyberbullying, griefing, mischief, sexism, sexual harassment, trolling, cheating, and flaming.	Lee et al. (2022)
24	Antisocial and offensive behaviors, such as cyberbullying, griefing, mischief, sexism, sexual harassment, trolling, cheating, and flaming, which negatively impact social interactions and team performance.	Monge and O'Brien (2022)

(*continued on next page*)

Table A.8 (*continued*).

No	Definition	Reference
25	Behaviors that are (1) impulsive and of short duration, and (2) a reaction to in-game frustration in real-time.	Kordyaka, Park et al. (2023)
26	Negative behaviors, including criticizing, insulting, and blaming others, that arise from situational frustration and anger during intense real-time competition and typically manifest in short, spontaneous outbursts.	Kordyaka, Laato, Jahn et al. (2023)
27	An umbrella term for various negative behaviors, including transient verbal actions like trash-talking, hate speech, and threats of violence, which are performed in the moment rather than strategically planned.	Poeller et al. (2023)
28	Refers to various forms of negative in-game behaviors such as harassment, flaming, trolling, and cheating, which occur in real-time and are often driven by frustration and the competitive nature of the games.	Kordyaka, Stelter et al. (2023)
29	Encompasses negative and harmful behaviors such as harassment, verbal abuse, and disruptive gameplay that undermine social connectedness and psychological well-being.	Frommel et al. (2023)
30	Deliberate, unfriendly behavior by players that disrupts the gaming experience of others, encompassing actions like flaming, harassment, and intentionally poor play.	Liu and Agur (2023)
31	An umbrella term for short-duration, non-systematic negative behaviors, fueled by situational frustration, anger, and high levels of real-time competition, manifesting as insulting, criticizing, resource stealing, and other actions.	Kordyaka, Laato, Hamari et al. (2023)
32	Refers to disruptive behaviors such as trolling, harassment, and griefing that negatively impact the player community in multiplayer online games.	Ma et al. (2023)

widespread manifestations of toxicity. We were able to show that toxic behaviors can be arranged on a continuum of more planned (reflective) actions and more automatic (impulsive) reactions during a game. Based on our classification, we can now empirically validate the spectrum of toxic acts described by the Alliance and League (2020). This spectrum spans from "intentionally inappropriate or abusive behaviors" to "completely accidental misunderstandings without malice". Such a framework offers substantial value for both research and practical applications. It enables a more precise conceptualization of toxicity, which can guide game developers and influence practices in the eSports industry. The insights gained from this classification have the potential to improve intervention strategies and foster a deeper understanding of toxic behaviors across these domains.

CRediT authorship contribution statement

Bastian Kordyaka: Writing – original draft, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Sukran Karaosmanoglu:** Writing – original draft, Supervision, Methodology, Investigation. **Samuli Laato:** Methodology, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors gratefully acknowledge funding from the Gösta Branders Research Fund, Åbo Akademi University Foundation (Stiftelsen för Åbo Akademi), which supported the article processing charge (APC). ROR: https://ror.org/018fvk877 | Crossref Funder ID: 501100007360.

Appendix A. Appendices

See Tables A.8 and A.9.

Appendix B. Systematic review protocol

This review protocol (Table B.10) is prepared based on Moher et al. (2015) and Shamseer et al. (2015)'s recommended items to address in a systematic review for the paper *Defining Toxicity in Multiplayer Online Games: A Systematic Literature Review*.

Table A.9
Main contributions of the identified articles.

No	Contribution
1	Development of a supervised learning approach to predict crowdsourced decisions on toxic behavior.
2	Large-scale analysis of toxic behavior, providing insights for systems to detect and prevent such behavior.
3	Analysis revealing how toxic behavior affects player performance and communication patterns.
4	Analysis of player and corporate narratives regarding the shift from human judgment to automated systems.
5	Framework identifying conversational patterns in in-game chat on performance and toxicity.
6	Demonstration that game design elements can influence culture-based toxicity among MENA players.
7	Development and validation of instruments to measure toxic behavior.
8	Detailed account of toxic behaviors, providing a basis for future research on moderation approaches.
9	Highlighting the amalgamation of toxicity from physical sports and computer-mediated communication.
10	Proposal and empirical validation of a unified theory explaining factors influencing toxic behavior.
11	Development of a catalog of dark participation in games to better understand and address toxic behaviors.
12	Examination of antecedents and contagion of online toxicity using longitudinal behavioral data.
13	Demonstration that normative beliefs about aggression predict cyberaggressive behaviors in online gaming.
14	Scalable method for detecting toxic behavior based on in-game behaviors, demonstrated in For Honor.
15	Development of design principles to buffer toxic behavior based on the online disinhibition effect.
16	Identification of moral disengagement and toxic disinhibition predicting the normalization of toxicity.
17	Identification of ethical challenges in game design and community management, with design considerations.
18	Flagging practices, highlighting the divergence between designed mechanisms and actual player behavior.
19	Development of in-game tools to support toxicity victims, reducing stress, and increasing positive emotions.

(*continued on next page*)

Table A.9 (*continued*).

No	Contribution
20	Demonstration that national culture influences the occurrence of toxic behavior.
21	Exploration of prevalence rates of dark participation and the effectiveness of reporting tools.
22	Development of a modeling approach using in-game communication and metadata to predict toxicity.
23	Longitudinal analysis between identity and toxicity, whereby habitual play predicts sustained engagement.
24	Demonstration that toxic behavior in League of Legends diminishes team and individual performance.
25	Demonstration that offline cultural environments influence toxic behavior.
26	Examination of the overlap in roles of perpetrators, victims, and bystanders.
27	Exploration of how players perceive and react to positive and toxic in-game chat messages.
28	Identification of physiological and social desires influencing toxic behavior in multiplayer online games.
29	Demonstration that toxicity is associated with lower social capital and increased loneliness.
30	Establishment of psychological mechanisms of in-game toxic behavior in team-based online games.
31	Mapping circumstances relevant to the emergence of toxicity through analysis of gamer essays.
32	Identification of how punishment notifications improve transparency and fairness in behavior moderation.

Table B.10
Review Protocol.

Section & Items	Item No	Note
Administrative Information		
Title:		
Identification	1a	Defining Toxicity in Multiplayer Online Games: A Systematic Literature Review — A Systematic Literature Review Protocol
Update	1b	No update.
Registration	2	Not pre-registered.
Authors:		
Contact	3a	Author 1: Dr. Bastian Kordyaka, Åbo Akademi University, Tuomiokirkontori 3, 20500 Turku. Finland. ORCID: 0000-0003-3495-6855, email:bastian.kordyaka@abo.fi. - Author 2: Sukran Karaosmanoglu, Universität Hamburg, Vogt-Kölln-Straße 30, D-22527 Hamburg, Germany. ORCID: 0000-0002-9624-4258, email: sukran.karaosmanoglu@uni-hamburg.de. - Author3: Samuli Laato, University of Turku, FI-20014 Turun Yliopisto, Finland. ORCID: 0000-0003-4285-0073, email: samuli.laato@utu.fi.
Contributions	3b	Conceptualization: BK, SK, SL - Methodology: BK, SK - Literature Search: BK, SL - Screening and Selection: BK, SL - Data Extraction: BK, SL - Formal Analysis: BK, SL - Writing - Original Draft: BK, SK - Writing - Review Editing: BK, SK - Visualization: BK, SK
Amendments	4	n/a
Support:		
Sources	5a	Gösta Branders Research Fund, Åbo Akademi University Foundation (Stiftelsen för Åbo Akademi)
Sponsor	5b	Åbo Akademi University Foundation (https://ror.org/018fvk877, Funder ID: 501100007360)
Role of sponsor or funder	5c	The funder covered the article processing charges (APC) but had no role in the study design, data collection, analysis, or interpretation.
Introduction		
Rationale	6	Toxicity in multiplayer online games negatively impacts both individuals' well-being and companies. While it is a pressing concern, it is a complex issue as it involves social dynamics. The richness and complexity of toxicity behaviors has led to a lack of conceptual clarity regarding the term.
Objectives	7	A key research gap in the existing scholarship on toxicity in video games is the lack of a precise and comprehensive understanding of toxicity. This is despite the fact that a comprehensive definition of the term would allow researchers and practitioners to better align their work, unify the research field across disciplines, and create a common starting point and point of reflection for future work. We would like to address this issue: • **RQ**$_1$: "How is toxicity described in existing HCI research"? • **RQ**$_2$: "How can existing definitions of toxicity be summarized into a definition"?
Methods		
Eligibility criteria	8	Inclusion Criteria: 1. English 2. Published peer-reviewed full journal or conference papers 3. Needs to deal with toxicity in the context of video games and show implicit/explicit references to a definition 4. Approaches toxicity from an HCI perspective 5. Articles published between 01/01/2014 and 12/31/2023
		Exclusion Criteria: 1. Not written in English (e.g., Russian or German) 2. Preprints or Short papers (e.g., workshop papers) 3. Toxicity on Twitter without references to an implicit/explicit definition 4. Approaching the term from a legal perspective 5. Articles not published until 12/31/2023 or before 01/01/2014
Information sources	9	The ACM Digital Library, the AIS eLibrary, Scopus, Taylor & Francis, Emerald, and Web of Science

(*continued on next page*)

Table B.10 (*continued*).

Section & Items	Item No	Note
Search strategy	10	• Definiendum - (“toxicity” OR “toxic behavior” OR “toxic behaviour”) • Definiens - (“definition” OR “define”) • Context - (“esport” OR “e-sport” OR “dota2” OR “dota 2” OR “cs: go” OR “csgo” OR “counterstrike” OR “counter-strike” OR “pubg” OR “playerunknown*” OR ‘fortnite’ OR ‘lol’ OR ‘valorant’ OR ‘r6 siege’ OR ‘overwatch’ OR ‘call of duty’ OR ‘hearthstone’ OR ‘heroes of the storm’ OR ‘halo 5’ OR ‘moba’ OR ‘game’) • Time interval - 01/2014 to 12/2023
Study records: Data management	11a	Manual download of identified articles and duplicate removal via Covidence. We use Covidence for screening and eligibility and Excel for further steps.
Selection process	11b	Screening: Two coders independently code the identified papers based on their title and abstract. They conduct discussion meetings to resolve the discussion meetings to resolve conflicts. Eligibility: Similarly, two coders independently code the identified papers based on their title and abstract. They conduct discussion meetings to resolve the discussion meetings to resolve conflicts.
Data collection process	11c	Manual data extraction using a data extraction form in identified papers
Data items	12	See the supplementary materials for full list of corpus papers and extracted data.
Outcomes and prioritization	13	The result sought is a multidimensional definition for toxicity based on a systematic literature review. However, to ensure quality, we only consider peer-reviewed full papers.
Risk of bias in individual studies	14	As we aim to create a multidimensional definition for toxicity, we do not perform a critical appraisal step. We would like to include as many existing definitions as possible in our review.
Data synthesis	15a	We provide the quantitative results in the form of figures and tables, and descriptive stats.
	15b	We use frequency reporting for inductive content analysis codes and study features
	15c	n/a
	15d	n/a
Meta-bias(es)	16	n/a
Confidence in cumulative evidence	17	We report a positionality statement for each researcher involved in the review analysis.

Data availability

Data will be made available on request.

References

Adinolf, S., & Turkay, S. (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* (pp. 365–372). Melbourne VIC Australia: ACM, http://dx.doi.org/10.1145/3270316.3271545.

Alliance, F. P., & League, A.-D. (2020). Disruption and harms in online gaming framework. *Fair Play Alliance*, 1–48.

Appelbaum, S. H., & Roy-Girard, D. (2007). Toxins in the workplace: Affect on organizations and employees. *Corporate Governance: The International Journal of Business in Society*, *7*(1), 17–28. http://dx.doi.org/10.1108/14720700710727087.

Balci, K., & Salah, A. A. (2015). Automatic analysis and identification of verbal aggression and abusive behaviors for online social games. *Computers in Human Behavior*, *53*, 517–526. http://dx.doi.org/10.1016/j.chb.2014.10.025.

Bell, M. W. (2008). Toward a definition of “virtual worlds”. *Journal for Virtual Worlds Research*, *1*(1), http://dx.doi.org/10.4101/JVWR.V1I1.283.

Bendell, J., Sutherland, N., & Little, R. (2017). Beyond unsustainable leadership: critical social theory for sustainable leadership. *Sustainability Accounting, Management and Policy Journal*, *8*(4), 418–444. http://dx.doi.org/10.1108/SAMPJ-08-2016-0048.

Beres, N. A., Frommel, J., Reid, E., Mandryk, R. L., & Klarkowski, M. (2021). Don’t you know that you’re toxic: Normalization of toxicity in online gaming. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1–15). Yokohama Japan: ACM, http://dx.doi.org/10.1145/3411764.3445157.

Blackburn, J., & Kwak, H. (2014). STFU NOOB!: predicting crowdsourced decisions on toxic behavior in online games. In *Proceedings of the 23rd international conference on world wide web* (pp. 877–888). Seoul Korea: ACM, http://dx.doi.org/10.1145/2566486.2567987.

Booth, A., James, M.-S., Clowes, M., Sutton, A., et al. (2021). *Systematic approaches to a successful literature review*. London: SAGE Publications Ltd.

Canossa, A., Salimov, D., Azadvar, A., Harteveld, C., & Yannakakis, G. (2021). For honor, for toxicity: Detecting toxic behavior through gameplay. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CHI PLAY), 1–29. http://dx.doi.org/10.1145/3474680.

Cavadenti, O., Codocedo, V., Boulicaut, J.-F., & Kaytoue, M. (2016). What did i do wrong in my MOBA game? Mining patterns discriminating deviant behaviours. In *2016 IEEE international conference on data science and advanced analytics* (pp. 662–671). IEEE.

Cestino, J., Macey, J., & McCauley, B. (2023). Legitimizing the game: how gamers’ personal experiences shape the emergence of grassroots collective action in esports. *Internet Research*, *33*(7), 111–132. http://dx.doi.org/10.1108/INTR-05-2022-0347.

Chaiken, S. (1999). Dual-process theories in social psychology. *Guilford Press Google Schola*, *2*, 206–214.

Claypool, M., Decelle, J., Hall, G., & O’Donnell, L. (2015). Surrender at 20? Matchmaking in league of legends. In *2015 IEEE games entertainment media conference* (pp. 1–4). IEEE, http://dx.doi.org/10.1109/GEM.2015.7377234.

Copi, I. M., Cohen, C., & McMahon, K. (2016). *Introduction to logic*. New York: Routledge.

Costa, L. M., Drachen, A., Souza, F. C. M., & Xexéo, G. (2023). Artificial intelligence in MOBA games: a multivocal literature mapping. *IEEE Transactions on Games*, http://dx.doi.org/10.1109/TG.2023.3282157.

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

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, *62*(1), 107–115. http://dx.doi.org/10.1111/j.1365-2648.2007.04569.x.

Flowerdew, J. (1992). Definitions in science lectures. *Applied Linguistics*, *13*(2), 202–221. http://dx.doi.org/10.1093/applin/13.2.202.

Frommel, J., Johnson, D., & Mandryk, R. L. (2023). How perceived toxicity of gaming communities is associated with social capital, satisfaction of relatedness, and loneliness. *Computers in Human Behavior Reports*, *10*, Article 100302. http://dx.doi.org/10.1016/j.chbr.2023.100302.

Frommel, J., & Mandryk, R. (2024). Toxicity in esports. In *Routledge handbook of esports*. Routledge, http://dx.doi.org/10.4324/9781003410591-57.

Gong, L., Feng, X., Ye, D., Li, H., Wu, R., Tao, J., et al. (2020). Optmatch: Optimized matchmaking via modeling the high-order interactions on the arena. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 2300–2310). http://dx.doi.org/10.1145/3394486.3403279.

Gray, C. M., Kou, Y., Battles, B., Hoggatt, J., & Toombs, A. L. (2018). The dark (patterns) side of UX design. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1–14). http://dx.doi.org/10.1145/3173574.3174108.

Harwood, T. G., & Garry, T. (2003). An overview of content analysis. *The Marketing Review*, *3*(4), 479–498. http://dx.doi.org/10.1362/146934703771910080.

Henriksen, K., Larsen, C. H., Storm, L. K., & Ryom, K. (2014). Sport psychology interventions with young athletes: The perspective of the sport psychology practitioner. *Journal of Clinical Sport Psychology*, *8*(3), 245–260. http://dx.doi.org/10.1123/jcsp.2014-0033.

Hilvert-Bruce, Z., & Neill, J. T. (2020). I’m just trolling: The role of normative beliefs in aggressive behaviour in online gaming. *Computers in Human Behavior*, *102*, 303–311. http://dx.doi.org/10.1016/j.chb.2019.09.003.

Jahn, K., Oschinsky, F. M., Kordyaka, B., Machulska, A., Eiler, T. J., Gruenewald, A., et al. (2022). Design elements in immersive virtual reality: the impact of object presence on health-related outcomes. *Internet Research*, *32*(7), 376–401. http://dx.doi.org/10.1108/INTR-12-2020-0712.

Jalali, S., & Wohlin, C. (2012). Systematic literature studies: database searches vs. backward snowballing. In *Proceedings of the ACM-IEEE international symposium on empirical software engineering and measurement* (pp. 29–38). http://dx.doi.org/10.1145/2372251.2372257.

Jhaver, S., Boylston, C., Yang, D., & Bruckman, A. (2021). Evaluating the effectiveness of deplatforming as a moderation strategy on Twitter. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 1–30. http://dx.doi.org/10.1145/3479525.

Karaosmanoglu, S., Cmentowski, S., Nacke, L. E., & Steinicke, F. (2024). Born to run, programmed to play: Mapping the extended reality exergames landscape. In *Proceedings of the CHI conference on human factors in computing systems*. New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3613904.3642124.

Kellermeyer, L., Harnke, B., & Knight, S. (2018). Covidence and rayyan. *Journal of the Medical Library Association: JMLA*, *106*(4), http://dx.doi.org/10.5195/jmla.2018.513.

Kim, H., Kim, H., Kim, J., & Jang, J.-w. (2022). When does it become harassment? An investigation of online criticism and calling out in Twitter. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW2), 1–32. http://dx.doi.org/10.1145/3555575.

Kitchenham, B., & Brereton, P. (2013). A systematic review of systematic review process research in software engineering. *Information and Software Technology*, *55*(12), 2049–2075. http://dx.doi.org/10.1016/j.infsof.2013.07.010.

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

Kordyaka, B., Klesel, M., & Jahn, K. (2019). Perpetrators in league of legends: Scale development and validation of toxic behavior. In *Proceedings of the annual hawaii international conference on system sciences*, *Proceedings of the 52nd annual hawaii international conference on system sciences, HICSS 2019* (pp. 2486–2495). United States: IEEE, http://dx.doi.org/10.24251/HICSS.2019.299.

Kordyaka, B., Krath, J., Park, S., Wesseloh, H., & Laato, S. (2022). Understanding toxicity in multiplayer online games: The roles of national culture and demographic variables. In *Proceedings of the annual hawaii international conference on system sciences*, *55th hawaii international conference on system sciences* (pp. 2908–2917). Online conference: University of Hawai'i at Mānoa.

Kordyaka, B., & Kruse, B. (2021). Curing toxicity – developing design principles to buffer toxic behaviour in massive multiplayer online games. *Safer Communities*, *20*(3), 133–149. http://dx.doi.org/10.1108/SC-10-2020-0037.

Kordyaka, B., Kruse, B., & Niehaves, B. (2023). Brands in esports–generational cohorts, value congruence and media engagement as antecedents of brand sustainability. *Journal of Media Business Studies*, 1–21.

Kordyaka, B., Laato, S., Hamari, J., Scholz, T., & Niehaves, B. (2023). What drives gamer toxicity? Essays from players. In *CEUR workshop proceedings*, *Proceedings of the 7th international gamiFIN conference 2023 (gamiFIN 2023)* (pp. 86–95). Levi: CEUR-WS.

Kordyaka, B., Laato, S., Jahn, K., Hamari, J., & Niehaves, B. (2023). The cycle of toxicity: Exploring relationships between personality and player roles in toxic behavior in multiplayer online battle arena games. *Proceedings of the ACM on Human-Computer Interaction*, *7*(CHI PLAY), 611–641. http://dx.doi.org/10.1145/3611043.

Kordyaka, B., Laato, S., Weber, S., & Niehaves, B. (2024). Are stress and engagement in toxicity associated with sleep quality? A study with league of legends players. *Proc. ACM Hum.-Comput. Interact.*, *8*(CHI PLAY), http://dx.doi.org/10.1145/3677101.

Kordyaka, B., Park, S., Krath, J., & Laato, S. (2023). Exploring the relationship between offline cultural environments and toxic behavior tendencies in multiplayer online games. *ACM Transactions on Social Computing*, *6*(1–2), 1–20. http://dx.doi.org/10.1145/3580346.

Kordyaka, B., Pumplun, L., Brunnhofer, M., Kruse, B., & Laato, S. (2023). Gender disparities in esports – an explanatory mixed-methods approach. *Computers in Human Behavior*, *149*, Article 107956. http://dx.doi.org/10.1016/j.chb.2023.107956.

Kordyaka, B., Stelter, A., Laato, S., & Niehaves, B. (2023). Dark desires? Using the theory of basic desires to better understand toxic behavior in multiplayer online games. In *Proceedings of the 56th hawaii international conference on system sciences* (pp. 5551–5559). Maui, Hawaii: University of Hawai'i at Mānoa, http://dx.doi.org/10.24251/HICSS.2023.676.

Kou, Y. (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* (pp. 81–92). Virtual Event Canada: ACM, http://dx.doi.org/10.1145/3410404.3414243.

Kou, Y. (2021). Punishment and its discontents: An analysis of permanent ban in an online game community. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 1–21. http://dx.doi.org/10.1145/3476075.

Kou, Y., & Gui, X. (2014). Playing with strangers: understanding temporary teams in league of legends. In *Proceedings of the first ACM SIGCHI annual symposium on computer-human interaction in play* (pp. 161–169).

Kou, Y., & Gui, X. (2017). When code governs community. In A. eLibrary (Ed.), *Proceedings of the 50th hawaii international conference on system sciences*. http://dx.doi.org/10.24251/HICSS.2017.249.

Kou, Y., & Gui, X. (2021). Flag and flaggability in automated moderation: The case of reporting toxic behavior in an online game community. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1–12). Yokohama Japan: ACM, http://dx.doi.org/10.1145/3411764.3445279.

Kowalski, R. (2018). Cyberbullying. In *The routledge international handbook of human aggression* (pp. 131–142). London, UK: Routledge.

Kowert, R. (2020). Dark participation in games. *Frontiers in Psychology*, *11*, 01–08. http://dx.doi.org/10.3389/fpsyg.2020.598947.

Kowert, R., Botelho, A., & Newhouse, A. (2022). Breaking the building blocks of hate: A case study of minecraft servers. *A Report from the Anti-Defamation League.(ADL) Center of Technology and Society*, doi:2022.

Kowert, R., & Cook, C. L. (2022). The toxicity of our (sim) cities: Prevalence of dark participation in games and perceived effectiveness of reporting tools. In *55th hawaii international conference on system sciences* (pp. 3180–3189). Online conference: University of Hawai'i at Mānoa, http://dx.doi.org/10.24251/HICSS.2022.390.

Kusy, M., & Holloway, E. (2009). *Toxic workplace!: Managing toxic personalities and their systems of power*. San Francisco: John Wiley & Sons, doi:2009.

Kwak, H. (2014). Understanding toxic behavior in online games. In *Proceedings of the 23rd international conference on world wide web* (pp. 1245–1246). Seoul Korea: ACM, http://dx.doi.org/10.1145/2567948.2580066.

Kwak, H., Blackburn, J., & Han, S. (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* (pp. 3739–3748). Seoul Republic of Korea: ACM, http://dx.doi.org/10.1145/2702123.2702529.

Laato, S., Kordyaka, B., & Hamari, J. (2024). Traumatizing or just annoying? Unveiling the spectrum of gamer toxicity in the starcraft II community. In *Proceedings of the CHI conference on human factors in computing systems* (pp. 1–18). Honolulu, USA: ACM, http://dx.doi.org/10.1145/3613904.3642137.

Laato, S., Tiainen, M., Najmul Islam, A., & Mäntymäki, M. (2022). How to explain AI systems to end users: a systematic literature review and research agenda. *Internet Research*, *32*(7), 1–31. http://dx.doi.org/10.1108/INTR-08-2021-0600.

Lapidot-Lefler, N., & Barak, A. (2012). Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in Human Behavior*, *28*(2), 434–443. http://dx.doi.org/10.1016/j.chb.2011.10.014.

League, A.-D. (2021). *Hate is no game: Harassment and positive social experiences in online games 2021*. New York: Anti-Defamation League, https://Www.Adl.Org/Hateisnogame.

Lee, M., Johnson, D., Tanjitpiyanond, P., & Louis, W. R. (2022). It's habit, not toxicity, driving hours spent in DOTA 2. *Entertainment Computing*, *41*, Article 100472. http://dx.doi.org/10.1016/j.entcom.2021.100472.

Lin, J.-H. (2013). Do video games exert stronger effects on aggression than film? The role of media interactivity and identification on the association of violent content and aggressive outcomes. *Computers in Human Behavior*, *29*(3), 535–543. http://dx.doi.org/10.1016/j.chb.2012.11.001.

Liu, Y., & Agur, C. (2023). "After All, They don't know me" exploring the psychological mechanisms of toxic behavior in online games. *Games and Culture*, *18*(5), 598–621. http://dx.doi.org/10.1177/15554120221115397.

Lyons, J. (1977). *vol. 2, Semantics: Volume 2*. Cambridge, UK: Cambridge University Press.

Ma, R., Li, Y., & Kou, Y. (2023). Transparency, fairness, and coping: How players experience moderation in multiplayer online games. In *Proceedings of the 2023 CHI conference on human factors in computing systems* (pp. 1–21). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3544548.3581097.

Maity, S. K., Chakraborty, A., Goyal, P., & Mukherjee, A. (2018). Opinion conflicts: An effective route to detect incivility in Twitter. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 1–27. http://dx.doi.org/10.1145/3274386.

Mandryk, R. L., Frommel, J., Goyal, N., Freeman, G., Lampe, C., Vieweg, S., et al. (2023). Combating toxicity, harassment, and abuse in online social spaces: A workshop at CHI 2023. In *Extended abstracts of the 2023 CHI conference on human factors in computing systems* (pp. 1–7). Hamburg Germany: ACM, http://dx.doi.org/10.1145/3544549.3573793.

Marques, T. G., Schumann, S., & Mariconti, E. (2024). Positive behaviour interventions in online gaming: a systematic review of strategies applied in other environments. *Crime Science*, *13*(1), 14. http://dx.doi.org/10.1186/s40163-024-00208-8.

Martens, M., Shen, S., Iosup, A., & Kuipers, F. (2015). Toxicity detection in multiplayer online games. In *2015 international workshop on network and systems support for games (netGames)* (pp. 1–6). Zagreb: IEEE, http://dx.doi.org/10.1109/NetGames.2015.7382991.

Mitchell, R., & Karttunen, S. (1992). Why and how to define an artist: Types of definitions and their implications for empirical research results. In *Cultural economics* (pp. 175–185). Springer, http://dx.doi.org/10.1007/978-3-642-77328-0_18.

Mohamad, M. M., Sulaiman, N. L., Sern, L. C., & Salleh, K. M. (2015). Measuring the validity and reliability of research instruments. *Procedia-Social and Behavioral Sciences*, *204*, 164–171. http://dx.doi.org/10.1016/j.sbspro.2015.08.129.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., et al. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-p) 2015 statement. *Systematic Reviews*, *4*, 1–9. http://dx.doi.org/10.1136/bmj.n71.

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. http://dx.doi.org/10.1080/15213269.2020.1868322.

Morley, L., & Crossouard, B. (2016). Gender in the neoliberalised global academy: the affective economy of women and leadership in south Asia. *British Journal of Sociology of Education*, *37*(1), 149–168. http://dx.doi.org/10.1080/01425692.2015.1100529.

Munn, Z., Peters, M. D., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, *18*, 1–7. http://dx.doi.org/10.1186/s12874-018-0611-x.

Neto, J. A. M., Yokoyama, K. M., & Becker, K. (2017). Studying toxic behavior influence and player chat in an online video game. In *Proceedings of the international conference on web intelligence* (pp. 26–33). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3106426.3106452.

Newton, B. J., Rothlingova, Z., Gutteridge, R., LeMarchand, K., & Raphael, J. H. (2012). No room for reflexivity? Critical reflections following a systematic review of qualitative research. *Journal of Health Psychology*, *17*(6), 866–885. http://dx.doi.org/10.1177/1359105311427615.

Oleksiyenko, A. (2018). Zones of alienation in global higher education: Corporate abuse and leadership failures. *Tertiary Education and Management*, *24*, 193–205. http://dx.doi.org/10.1080/13583883.2018.1439095.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *372*, 1–9. http://dx.doi.org/10.1136/bmj.n71.

Pelletier, K. L. (2010). Leader toxicity: An empirical investigation of toxic behavior and rhetoric. *Leadership*, *6*(4), 373–389. http://dx.doi.org/10.1177/1742715010379308.

Poeller, S., Dechant, M. J., Klarkowski, M., & Mandryk, R. L. (2023). Suspecting sarcasm: How league of legends players dismiss positive communication in toxic environments. *Proceedings of the ACM on Human-Computer Interaction*, *7*(CHI PLAY), 1–26. http://dx.doi.org/10.1145/3611020.

Poeller, S., & Robinson, R. B. (2024). Mute, block, punish, reward? A call to shift the research focus from concealing toxicity in games to promoting genuine positive behavior. In *CHI PLAY companion '24, Companion proceedings of the 2024 annual symposium on computer-human interaction in play* (pp. 282–284). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3665463.3678863.

Pollock, A., & Berge, E. (2018). How to do a systematic review. *International Journal of Stroke*, *13*(2), 138–156. http://dx.doi.org/10.1177/1747493017743796.

Reid, E., Mandryk, R. L., Beres, N. A., Klarkowski, M., & Frommel, J. (2022a). "Bad Vibrations": Sensing toxicity from in-game audio features. *IEEE Transactions on Games*, *14*(4), 558–568. http://dx.doi.org/10.1109/TG.2022.3176849.

Reid, E., Mandryk, R. L., Beres, N. A., Klarkowski, M., & Frommel, J. (2022b). Feeling good and in control: In-game tools to support targets of toxicity. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CHI PLAY), 1–27. http://dx.doi.org/10.1145/3549498.

Robinson, R., Hammer, J., & Isbister, K. (2019). All the world (wide web)'sa stage: A workshop on live streaming. In *Extended abstracts of the 2019 CHI conference on human factors in computing systems* (pp. 1–8). http://dx.doi.org/10.1145/3290607.3299016.

Rogers, K., Hirzle, T., Karaosmanoglu, S., Palomino, P. T., Durmanova, E., Isotani, S., et al. (2024). An umbrella review of reporting quality in CHI systematic reviews: Guiding questions and best practices for HCI. *ACM Trans. Comput.-Hum. Interact.*, http://dx.doi.org/10.1145/3685266.

Rogers, K., Karaosmanoglu, S., Wolf, D., Steinicke, F., & Nacke, L. E. (2021). A best-fit framework and systematic review of asymmetric gameplay in multiplayer virtual reality games. *Frontiers in Virtual Realeague2021hatelity*, *2*, http://dx.doi.org/10.3389/frvir.2021.694660.

Rossi, R. J. (2006). *vol. 82, Theorems, corollaries, lemmas, and methods of proof*. Hoboken, US: Wiley Online Library.

Scholz, T. M. (2020). Deciphering the world of esports. *International Journal on Media Management*, *22*(1), 1–12. http://dx.doi.org/10.1080/14241277.2020.1757808.

Sengün, S., Salminen, J., Mawhorter, P., Jung, S.-g., & Jansen, B. (2019). Exploring the relationship between game content and culture-based toxicity: A case study of league of legends and MENA players. In *Proceedings of the 30th ACM conference on hypertext and social media* (pp. 87–95). Hof Germany: ACM, http://dx.doi.org/10.1145/3342220.3343652.

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., et al. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-p) 2015: elaboration and explanation. *349*, http://dx.doi.org/10.1136/bmj.g7647.

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*, Article 106343. http://dx.doi.org/10.1016/j.chb.2020.106343.

Sparrow, L. A., Galwey, R., Jovic, D., Hardwick, T., & Butt, M.-A. (2024). Towards ethical AI moderation in multiplayer games. *Proc. ACM Hum.-Comput. Interact.*, *8*(CHI PLAY), http://dx.doi.org/10.1145/3677109.

Sparrow, L. A., Gibbs, M., & Arnold, M. (2021). The ethics of multiplayer game design and community management: Industry perspectives and challenges. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3411764.3445363.

Suler, J. (2004). The online disinhibition effect. *Cyberpsychology & Behavior*, *7*(3), 321–326. http://dx.doi.org/10.1089/1094931041291295.

Türkay, S., Formosa, J., Adinolf, S., Cuthbert, R., & Altizer, R. (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* (pp. 1–13). Honolulu HI USA: ACM, http://dx.doi.org/10.1145/3313831.3376191.

Vallerand, R. J., Deshaies, P., Cuerrier, J.-P., BriÈre, N. M., & Pelletier, L. G. (1996). Toward a multidimensional definition of sportsmanship. *Journal of Applied Sport Psychology*, *8*(1), 89–101. http://dx.doi.org/10.1080/10413209608406310.

Weinberg, R. S., & Comar, W. (1994). The effectiveness of psychological interventions in competitive sport. *Sports Medicine*, *18*, 406–418. http://dx.doi.org/10.2165/00007256-199418060-00005.

Wijkstra, M., Rogers, K., Mandryk, R. L., Veltkamp, R. C., & Frommel, J. (2023). Help, my game is toxic! first insights from a systematic literature review on intervention systems for toxic behaviors in online video games. In *CHI PLAY companion '23, Companion proceedings of the annual symposium on computer-human interaction in play* (pp. 3–9). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3573382.3616068.

Wijkstra, M., Rogers, K., Mandryk, R. L., Veltkamp, R. C., & Frommel, J. (2024). How to tame a toxic player? A systematic literature review on intervention systems for toxic behaviors in online video games. *Proc. ACM Hum.-Comput. Interact.*, *8*(CHI PLAY), http://dx.doi.org/10.1145/3677080.

Williams, R. B., & Clippinger, C. A. (2002). Aggression, competition and computer games: computer and human opponents. *Computers in Human Behavior*, *18*(5), 495–506. http://dx.doi.org/10.1016/S0747-5632(02)00009-2.

Yang, Z., Grenon-Godbout, N., & Rabbany, R. (2024). Game on, hate off: A study of toxicity in online multiplayer environments. *ACM Games*, *2*(2), http://dx.doi.org/10.1145/3675805.

Zhang, Z., & Kou, Y. (2024). Casual competition by design: A study of the all random all mid (ARAM) mode in league of legends. *Proc. ACM Hum.-Comput. Interact.*, *8*(CSCW2), http://dx.doi.org/10.1145/3686992.

Zhang, Z., Moradzadeh, S., Woan, A., & Kou, Y. (2024). Toxicity by game design: How players perceive the influence of game design on toxicity. *Proc. ACM Hum.-Comput. Interact.*, *8*(CHI PLAY), http://dx.doi.org/10.1145/3677110.

Zhen, S., Xie, H., Zhang, W., Wang, S., & Li, D. (2011). Exposure to violent computer games and Chinese adolescents' physical aggression: The role of beliefs about aggression, hostile expectations, and empathy. *Computers in Human Behavior*, *27*(5), 1675–1687. http://dx.doi.org/10.1016/j.chb.2011.02.006.