Challenges to Combating Toxicity and Harassment in Multiplayer Games: Involving the HCI Games Research Community

Julian Frommel
j.frommel@uu.nl
Utrecht University
Utrecht, Netherlands

Regan L. Mandryk
regan@cs.usask.ca
University of Saskatchewan
Saskatoon, Canada

Madison Klarkowski
madison@cs.usask.ca
University of Saskatchewan
Saskatoon, Canada

ABSTRACT

Toxicity is and remains a problem in multiplayer games, and can result in harm for players and game environments. Grounded in prior work, we present four challenges that impede solving the problem of toxicity and harassment. We believe that we need to overcome these challenges to ensure safe gaming spaces, and intend to stimulate discussion about how HCI Games research can make substantial contributions toward this goal.

CCS CONCEPTS

• **Human-centered computing → Collaborative and social computing systems and tools**; • **Applied computing → Computer games**; • **Software and its engineering → Interactive games**.

KEYWORDS

games, toxicity, toxic, harassment, griefing, hate, prediction, classification, support

ACM Reference Format:
Julian Frommel, Regan L. Mandryk, and Madison Klarkowski. 2022. Challenges to Combating Toxicity and Harassment in Multiplayer Games: Involving the HCI Games Research Community. In *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '22 EA), November 2–5, 2022, Bremen, Germany.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3505270.3558359

1 INTRODUCTION

Although multiplayer games are often beneficial (e.g., by combating loneliness and improving wellbeing [7, 22, 29, 30]), they are also a platform on which players can experience or witness toxicity, harassment, and other abusive or harmful behaviours. Despite general agreement about the dangers of online toxic behaviour in games [1, 11, 15, 16, 18, 28], these harmful behaviours persist. Toxicity has instead become more problematic and widespread—with recent statistics reporting that 83% of adult gamers experienced harassment in online multiplayer games [20], a number that has grown in recent years [19, 20]. Toxicity has a multitude of adverse effects for game companies (e.g., increased churn [13, 15, 21]) and for the players themselves (e.g., causing distress [9, 10, 20, 23, 27]), and therefore represents a threat to the health of game environments, players, and esports movements.

CHI PLAY '22 EA, November 2–5, 2022, Bremen, Germany

ACM ISBN 978-1-4503-9211-2/22/11.
https://doi.org/10.1145/3505270.3558359

2 CHALLENGES TO COMBATING TOXICITY AND HARASSMENT IN MULTIPLAYER GAMES

Using findings from prior and our own research, we exemplify some of the technical and social challenges of toxicity in multiplayer games. We highlight four challenges that we need to overcome before we can hope to eliminate toxic behaviours.

(C1) *Toxicity is normalized in gaming communities.* Our prior research [2] showed that many players do not report toxic content because they consider it acceptable, inevitable, or typical of games. This normalization of toxicity is further problematic because it begets additional toxicity, as it is often cyclical [2, 15, 16], and leads to unclear boundaries and the problem of subjectivity (that is, what does and does not constitute 'toxicity'). This cyclical normalization of toxic behaviours must be interrupted, or else negative behaviours will continue to be accepted and perpetuated by pursuant generations of players.

(C2) The *subjectivity of what is toxic* represents a huge challenge, and can partially explain why current methods fall short: if we cannot agree on what is acceptable and what is not, how can we implement interventions that enforce this? We know that individuals perceive toxicity at differing levels of severity [3, 11], and that many factors contribute to this (e.g., susceptibility to online disinhibition [2], game expertise [2], and identity [12]). What may be perceived as toxic or harmful by one player, might simply be experienced as acceptable game banter by another [1]. This highlights that we require more tools that can be used to assess experiences of toxicity such as validated questionnaires, while the inherent subjectivity may remain a challenge that requires more comprehensive, dynamic, and individual solutions. Further, this impedes the implementation of automated methods that could help combat toxicity.

(C3) There is still a *lack of widely accessible and effective tools to detect toxic behaviour.* While research and industry have proposed, implemented, and applied toxicity detection methods (e.g., [6, 8, 24]), this is not a solved problem. Promising methods often have limitations, e.g., limited applicability beyond specific communication channels [24, 31], applicability to games contexts and inadequacy to account for context [5, 11], or reliance on objective ground truth ratings, which may not represent the reality of how toxicity is experienced in games (cf. C2: subjectivity). Furthermore, some of the most promising examples are proprietary [4, 6, 14, 26], therefore not publicly accessible and of limited use to other small developers or for tackling toxicity across gaming communities.

(C4) There is a *lack of support for people experiencing or witnessing toxicity* in digital spaces. Most current approaches to address toxicity focus on punishing toxic players—for example, with muting or reporting functions that may eventually lead to game bans. While those approaches are important, they are also sometimes underused or misused [2, 17], impeding their effectiveness. Further, they generally still leave targets and witnesses harmed. Our findings [25] indicate that support mechanisms for targeted players may help them cope with negative experiences. We argue that more support is necessary to ensure that games are safe and healthy environments. However, it is unclear how games can provide access to such support in a way that is effective, unobtrusive, and accepted by a skeptical player base [25]. For example, some players may not use an effective support method that could help them cope if it impedes their performance or experience (e.g., by interrupting gameplay or hindering effective communication). Therefore, HCI games research is necessary to explore how to design support mechanisms that balance effectiveness and player experience.

This list of challenges is by no means comprehensive but highlights four important challenges that researchers, designers, and developers need to tackle.

3 CONCLUSION

Toxicity is and remains a problem in multiplayer games. Combating toxicity is challenging, and requires the concerted efforts of stakeholders involved in games and gaming communities. Importantly, this also includes HCI Games researchers who can contribute in several critical ways. For example: investigating reasons for and how to overcome societal challenges, e.g., with design frictions, game mechanics, or communication features, may help combat normalization and subjectivity (**C1/C2**), while designing and evaluating approaches for prediction and support may fill important gaps currently lacking in the toolset of games and human moderators (**C3/C4**). With this talk, we want to draw attention to the problem of toxicity and what we can do to make gaming environments enjoyable, healthy, and safe spaces for everyone.

REFERENCES

[1] Sonam Adinolf and Selen Turkay. 2018. Toxic Behaviors in Esports Games: Player Perceptions and Coping Strategies. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts (CHI PLAY '18 Extended Abstracts)*. Association for Computing Machinery, Melbourne, VIC, Australia, 365–372. https://doi.org/10.1145/3270316.3271545

[2] Nicole A Beres, Julian Frommel, Elizabeth Reid, Regan L Mandryk, and Madison Klarkowski. 2021. Don't You Know That You're Toxic: Normalization of Toxicity in Online Gaming. In *Proceedings of CHI '21* (Virtual). ACM, 1–15.

[3] Courtney Blamey. 2022. One Tricks, Hero Picks, and Player Politics: Highlighting the Casual-Competitive Divide in the Overwatch Forums. In *Modes of Esports Engagement in Overwatch*. Palgrave Macmillan, Cham, 31–47.

[4] J. Allen Brack and Blizzard Entertainment. 2020. 2020 Blizzard Fireside Chat. https://www.youtube.com/watch?v=CqPdqli0jTs.

[5] Eshwar Chandrasekharan, Umashanthi Pavalanathan, Anirudh Srinivasan, Adam Glynn, Jacob Eisenstein, and Eric Gilbert. 2017. You can't stay here: The efficacy of reddit's 2015 ban examined through hate speech. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–22.

[6] Joel Chapman and FACEIT. 2021. Fighting Abusive Behaviour — Product Update. https://medium.com/faceit-blog/fighting-abusive-behaviour-product-update-d39e490c00ce

[7] Ansgar E. Depping, Colby Johanson, and Regan L. Mandryk. 2018. Designing for Friendship: Modeling Properties of Play, In-Game Social Capital, and Psychological Well-Being. In *Proceedings of CHI PLAY '18* (Melbourne, VIC, Australia). ACM, New York, NY, USA, 87–100. https://doi.org/10.1145/3242671.3242702

[8] Danilo Dessì, Diego Reforgiato Recupero, and Harald Sack. 2021. An Assessment of Deep Learning Models and Word Embeddings for Toxicity Detection within Online Textual Comments. *Electronics* 10, 7 (March 2021), 779. https://doi.org/10.3390/electronics10070779 Number: 7 Publisher: Multidisciplinary Digital Publishing Institute.

[9] Jesse Fox, Michael Gilbert, and Wai Yen Tang. 2018. Player experiences in a massively multiplayer online game: A diary study of performance, motivation, and social interaction. *New Media & Society* 20, 11 (2018), 4056–4073.

[10] Jesse Fox and Wai Yen Tang. 2017. Women's experiences with general and sexual harassment in online video games: Rumination, organizational responsiveness, withdrawal, and coping strategies. *New Media & Society* 19, 8 (2017), 1290–1307. https://doi.org/10.1177/1461444816635778

[11] Julian Frommel, Valentin Sagl, Ansgar E. Depping, Colby Johanson, Matthew K. Miller, and Regan L. Mandryk. 2020. Recognizing Affiliation: Using Behavioural Traces to Predict the Quality of Social Interactions in Online Games. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–16.

[12] Nitesh Goyal, Ian Kivlichan, Rachel Rosen, and Lucy Vasserman. 2022. Is Your Toxicity My Toxicity? Exploring the Impact of Rater Identity on Toxicity Annotation. https://doi.org/10.48550/ARXIV.2205.00501

[13] Kate Grandprey-Shores, Yilin He, Kristina L. Swanenburg, Robert Kraut, and John Riedl. 2014. The Identification of Deviance and Its Impact on Retention in a Multiplayer Game. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Baltimore, Maryland, USA) *(CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 1356–1365. https://doi.org/10.1145/2531602.2531724

[14] Iain Harris. 2020. Toxicity in Overwatch has seen an "incredible decrease" due to machine learning. https://www.pcgamesn.com/overwatch/toxic-behaviour-machine-learning.

[15] Bastian Kordyaka, Katharina Jahn, and Bjoern Niehaves. 2020. Towards a unified theory of toxic behavior in video games. *Internet Research* (2020).

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

[17] Yubo Kou and Xinning Gui. 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* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 437, 12 pages. https://doi.org/10.1145/3411764.3445279

[18] Rachel Kowert. 2020. Dark Participation in Games. *Frontiers in Psychology* 11 (2020), 2969. https://doi.org/10.3389/fpsyg.2020.598947

[19] Anti-Defamation League. 2020. Disruption and harms in online gaming framework. https://www.adl.org/fpa-adl-games-framework

[20] Anti-Defamation League. 2021. Hate is no game: Harassment and positive social experiences in online games 2021. https://www.adl.org/hateisnogame#executive-summary

[21] Matthew Lee, Daniel Johnson, Porntida Tanjitpiyanond, and Winnifred R. Louis. 2022. It's habit, not toxicity, driving hours spent in DOTA 2. *Entertainment Computing* 41 (2022), 100472. https://doi.org/10.1016/j.entcom.2021.100472

[22] Regan L. Mandryk, Julian Frommel, Ashley Armstrong, and Daniel Johnson. 2020. How Passion for Playing World of Warcraft Predicts In-Game Social Capital, Loneliness, and Wellbeing. *Frontiers in Psychology* 11, September (2020). https://doi.org/10.3389/fpsyg.2020.02165

[23] Pew Research Center. 2014. Online Harassment. http://www.pewinternet.org/2014/10/22/online-harassment/.

[24] Elizabeth Reid, Regan L. Mandryk, Nicole A. Beres, Madison Klarkowski, and Julian Frommel. 2022. "Bad Vibrations": Sensing Toxicity From In-Game Audio Features. *IEEE Transactions on Games* (2022), 1–10. https://doi.org/10.1109/TG.2022.3176849

[25] Elizabeth Reid, Regan L. Mandryk, Nicole A. Beres, Madison Klarkowski, and Julian Frommel. 2022. Feeling Good and In Control: In-game Tools to Support Targets of Toxicity. *Proc. ACM Hum.-Comput. Interact.* 6, CHI PLAY, Article 235 (October 2022). https://doi.org/10.1145/3549498

[26] Riot Games. 2022. Valorant Voice Evaluation Update. https://playvalorant.com/en-us/news/announcements/valorant-voice-evaluation-update/?linkId=100000132203363

[27] Kevin C Runions. 2013. Toward a conceptual model of motive and self-control in cyber-aggression: Rage, revenge, reward, and recreation. *Journal of youth and adolescence* 42, 5 (2013), 751–771.

[28] Selen Türkay, Jessica Formosa, Sonam Adinolf, Robert Cuthbert, and Roger Altizer. 2020. See No Evil, Hear No Evil, Speak No Evil: How Collegiate Players Define, Experience and Cope with Toxicity. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, Honolulu, HI, USA, 1–13. https://doi.org/10.1145/3313831.3376191

[29] Dmitri Williams. 2006. On and off the'Net: Scales for social capital in an online era. *Journal of computer-mediated communication* 11, 2 (2006), 593–628.

[30] Dmitri Williams, Nicolas Ducheneaut, Li Xiong, Yuanyuan Zhang, Nick Yee, and Eric Nickell. 2006. From tree house to barracks: The social life of guilds in World of Warcraft. *Games and culture* 1, 4 (2006), 338–361.

[31] Midia Yousefi and Dimitra Emmanouilidou. 2021. Audio-based Toxic Language Classification using Self-attentive Convolutional Neural Network. In *2021 29th European Signal Processing Conference (EUSIPCO)*. IEEE, 11–15.