1828
AKADÉMIAI KIADÓ

An extension of the context of peer aggression
Commentary to the debate "Taxonomy of toxic behaviors in multiplayer gaming environments"

SERGIO PAREDES* and CARLOS VALLS-SERRANO

Department of Psychology, Catholic University of Murcia (UCAM), Murcia, Spain

Journal of Behavioral Addictions

Received: September 30, 2025 • Revised manuscript received: March 19, 2026 • Accepted: March 27, 2026

DOI:
10.1556/2006.2025.00432

DEBATE: TAXONOMY OF TOXIC BEHAVIORS IN MULTIPLAYER GAMING ENVIRONMENTS

ABSTRACT

This commentary is written in response to the paper entitled "Taxonomy of toxic behaviors in multiplayer gaming environments: An extension of the context of peer aggression" by Zsila and Demetrovics (2025). It seeks to support and expand the discussion in two central aspects: first, the conceptual boundaries between toxicity, cyberbullying, and trolling; and second, the structure of the proposed taxonomy, acknowledging its strengths while suggesting clarifications. Conceptual overlaps with cyberbullying and trolling highlight the importance of context and situational features when distinguishing toxic behaviors. We argue that the taxonomy could benefit from refinements that explicitly consider the role of intentionality, repetition, and the interplay between in-game dynamics and mental health consequences. These adjustments may also allow inclusion of behaviors extending beyond direct in-game actions. The taxonomy presented by the authors represents a significant step forward in systematizing research on toxic behaviors in video games. Future work should integrate psychological and contextual factors that shape both the experience of toxicity and the strategies players adopt to cope with it.

KEYWORDS

toxic behavior, cyberbullying, trolling, cyber-aggression, video games, dark participation

In a recent paper published in the *Journal of Behavioral Addictions*, Zsila and Demetrovics (2025) propose a taxonomy and working definition of toxic behaviors in the context of online multiplayer video games. This is a timely contribution as technological and cultural developments continually generate new forms of cyber-aggression, such as toxic behaviors in video games, which require closer and more nuanced scholarly attention. Earlier attempts to classify such behaviors had already laid important groundwork (see Kou, 2020; Quandt, 2018; Thacker & Griffiths, 2012) but taken together frequently overlap with related constructs such as cyberbullying and trolling, without clearly disentangling them. In what follows, this commentary reflects on the conceptual boundaries and the proposed taxonomy, with the aim of complementing the authors' work in the field. We also outline future directions for research that could deepen understanding of these behaviors and implications.

DEFINING TOXICITY: THE ROLE OF INTENT, REPETITION, AND POWER IMBALANCE

The definition of toxic behaviors proposed by Zsila and Demetrovics (2025) —negative behaviors in online multiplayer game environments that can impair individual or team performance, gaming experience, and mental well-being of players— brings a valuable contextual emphasis by situating toxicity within gameplay dynamics and the subjective experience of players.

One of the first questions concerns how toxic behaviors are framed within broader categories of online hostility. At the broadest level, we find what Quandt (2018) calls dark

*Corresponding author.
E-mail: sparedes@ucam.edu

participation, a very wide umbrella that encompasses many kinds of online misconduct, ranging from trolling and flaming to disinformation and media manipulation. Terms more consistently supported in the literature are cyber-aggression and cyberbullying. The former refers to harmful behaviors carried out through electronic means and perceived by the recipient as unwanted (Grigg, 2010). Within this category, cyberbullying is best understood as a sub-type characterized by three specific elements that distinguish it from other constructs: harmful intent, repetition over time, and a power imbalance between perpetrator and victim (Zhang, Huang, Lam, Evans, & Zhu, 2022). Yet in practice, distinguishing between cyberbullying and toxic behavior is not always clear-cut. A closer examination of how intentionality, repetition, and power imbalance operate in the context of toxic behaviour may help refine its conceptualization.

First, it is important to consider the role of intentionality. Whether a behavior is intentional or not often depends on the broader context in which it occurs and how that context is interpreted. Such situational sensitivity is valuable, since players may differ in whether they perceive a given action as disruptive or acceptable (Zsila, Shabahang, Aruguete, & Orosz, 2022). For example, a novice player who performs poorly out of inexperience presents a very different case from one who deliberately underperforms to disadvantage the team. Still, intentionality remains a subjective construct that can only be directly known by the perpetrator, while victims and observers must infer it from their interpretation, which makes its empirical assessment particularly challenging. However, if toxic behavior is to be conceptualized within the framework of cyberbullying, intentionality must be treated as a transversal element inherent to all underlying behaviors.

Second, the repetition of the behavior also deserves consideration. A single act of flaming may be disruptive but situationally forgivable, whereas repeated conduct directed at the same individual or others may move beyond an isolated toxic incident and closer to the domain of cyberbullying. This distinction also depends on how repetition unfolds over time. Toxic behavior may involve repeated acts within a single gameplay session, whereas cyberbullying typically implies a more sustained pattern of behavior that persists across separated encounters. At the same time, repeated behavior does not necessarily imply harmful intent and may instead resemble trolling. Indeed, Zsila and Demetrovics (2025), drawing on prior definitions, describe trolling as relatively benign disruptions that may occur without harmful intent, a view that stands in tension with empirical findings suggesting that trolling frequently involves deliberate attempts to provoke or annoy others (Buckels, Trapnell, & Paulhus, 2014). Separately, Cook, Tang, and Lin (2023) argue that toxicity can be regarded as a specific subset of trolling, namely those instances that produce harmful consequences for players. We are not fully convinced, however, that reducing toxicity to trolling adequately captures the specific ways in which these behaviors are experienced in gaming contexts, particularly given evidence that players themselves draw different boundaries between provocation and harm (Kordyaka, Laato, Weber, & Niehaves, 2023).

We suggest that trolling may at times take the form of playful provocation, whereas toxicity is more closely associated with intentional harm and repetition. Further indicators may help delineate these terms more clearly. In particular, one possible criterion concerns the primary target and function of the behavior: whether it mainly disrupts the flow and objectives of the game or directly targets another player with the intention of causing distress.

Third, the role of power imbalance is also complex within gaming contexts. In our view, players may engage in hostility without any stable or pre-existing relation of dominance between them. One player may insult, ridicule, blame, or humiliate another, yet these behaviors do not necessarily establish a meaningful asymmetry of power in themselves. The imbalance becomes more apparent when toxicity operates through the game itself, that is, when a player uses the mechanics or conditions of play to disadvantage others and reduce their capacity to respond effectively. This may occur when a teammate withholds cooperation, feeds the opposing team, or reveals strategic information. From this perspective, power imbalance may be better understood not as straightforward interpersonal dominance, but as the creation or intensification of situational disadvantage within gameplay. This may be particularly relevant for novice players, who are more likely to face hostility from more experienced teammates or opponents due to differences in skill and game knowledge (Shen, Ratan, Cai, & Leavitt, 2020).

TAXONOMY OF TOXIC BEHAVIORS

A second issue concerns the taxonomy of toxic behaviors presented in Table 1. We believe that the framework represents an important step forward, yet several aspects could benefit from refinement. These include clarifying definitional features, addressing overlaps, and expanding the taxonomy to better reflect the range of behaviors players experience as toxic.

One area of overlap is the relationship between categories such as *flaming* and *hate speech*. Both are defined in terms of offensive language, yet in practice many instances of flaming include discriminatory elements. Similarly, sexual harassment, although listed as a subcategory of harassment, also functions as a form of gender-based discrimination, and in this sense it shares important features with hate speech (Fox & Tang, 2017). The internal structure of categories also leaves room for clarification. For example, griefing is listed at the same level as feeding, even though feeding is usually considered one particular manifestation of the previous (Neto, Yokoyama, & Becker, 2017). Likewise, if cheating is defined broadly enough to include collaborative arrangements such as wintrading, then practices like assisting the enemy or intentional feeding could arguably be treated within that same category rather than placed separately.

As we emphasized earlier, explicit reference to intentionality would strengthen the descriptive clarity of the taxonomy. Smurfing, AFK, and boosting, for instance, all affect competitive balance but may not involve deliberate hostility toward others. Many players perceive them more as unfair than overtly aggressive, which makes their categorization ambiguous. The same ambiguity applies to trolling, which as discussed earlier, oscillates between playful provocation and genuinely harmful conduct. Spamming is also included in the taxonomy, but its disruptive potential derives mainly from its repetitiveness rather than clear hostile intent.

The taxonomy could also benefit from distinguishing between toxic behaviors that stem from violating the norms and basic functioning of the game (e.g., intentionally losing) and those not directly embedded in gameplay itself (e.g., insults). As the authors rightly point out, toxicity has negative consequences both for the game and for players' mental well-being, yet this distinction is not always made explicit in the taxonomy. Practices such as rage quitting, for instance, are absent, even though players often describe them as disruptive to the collective experience of play.

Equally relevant are toxic behaviors that occur before or after gameplay. Repeatedly leaving queues before a match begins is often reported as disruptive, just as verbal aggression may surface during champion selection or in post-game lobbies, where harassment can spill over into unwanted friend requests or direct messages (Kowert & Cook, 2022). False reporting, although listed in the taxonomy, shows how toxicity can manifest outside direct gameplay, with consequences like unjust sanctions that shape how players perceive fairness and community norms.

Recognizing these adjacent contexts would enrich the taxonomy and bring it closer to how players themselves describe the breadth of toxic encounters. This also draws attention to a broader conceptual issue: although toxicity may affect both gameplay and players' emotional well-being, as reflected in existing definitions, the literature does not always assign the same weight to these two aspects. Some approaches emphasize toxicity primarily as a disruption of teamwork, coordination, or the normative functioning of the game (see Kou, 2020), whereas others place greater emphasis on the psychological harm it causes to players, including depression, anxiety, rumination, and even suicidal ideation (see Fox & Tang, 2017; Wong & Ratan, 2023).

We suggest that toxic behavior is ultimately directed at harming another player, but that this harm may be expressed in two broad ways within gaming contexts. In some cases, the harm is enacted through the game itself, by engaging in behaviors that go against its norms or proper functioning in ways that negatively affect other players. In others, the harm is directed at the person more openly, through insults, humiliation, or harassment, and is not mediated by the game's mechanics. The target, in both cases, is the player; what differs is the manner through which the harm is carried out.

FUTURE DIRECTIONS AND CONCLUSIONS

The framework proposed by Zsila and Demetrovics (2025) offers researchers a much-needed starting point for more precise conceptual discussions and provides common ground for comparing findings across studies. The most pressing need, in our view, is to conduct independent empirical studies and generate data that rely on validated measurement tools. Recent attempts to operationalise toxicity, such as the Gaming Toxicity Scale (Díaz-Moreno, Bonilla, & Chamarro, 2025) represent an encouraging process in this direction. Yet the field still lacks robust epidemiological estimates of how common different types of toxic behaviors are, and whether these vary across cultures, platforms or game genres.

A valuable next step would be to explore the perspectives of those directly involved in toxic behaviors, allowing more formal classifications to emerge through qualitative approaches or Delphi panels with experts supported by statistical analysis (Ziems, Vigfusson, & Morstatter, 2020).

We also see merit in expanding the focus beyond perpetrators to include the perspectives of victims, bystanders, and broader communities. It could help explain why some players escalate into hostility while others disengage, reframe, or resist, and provide a basis for interventions that build resilience instead of focusing only on risk reduction. The authors themselves acknowledge prior evidence of a transition from victimization to perpetration, particularly in the cyberbullying literature. Zhang et al. (2022) confirmed that cyberbullying victimization is the strongest predictor of later perpetration. In this regard, coping and protective factors become central, yet research in this area remains scarce. Most studies so far emphasize risks such as moral disengagement or toxic disinhibition, as noted by the authors.

It is also important to consider the consequences of toxicity separately in terms of psychological well-being and in-game performance. This distinction could help bridge work on mental health with perspectives from sports psychology. In professional gaming, for instance, even small decrements in performance may carry significant economic and career implications. Moreover, key psychological dimensions remain understudied. Factors such as educational background, prior exposure to prosocial training, or the presence of psychopathology are rarely examined in relation to toxic behavior. Likewise, motivational aspects have yet to be systematically mapped (Achterbosch, Miller, & Vamplew, 2017). Without integrating these broader psychological and motivational aspects, the field risks reducing toxicity to a narrow set of cognitive and personality variables. Addressing these factors would allow future research to capture the diversity of pathways that lead players toward or away from hostile behavior.

Funding sources: This paper did not receive any financial support.

Authors' contribution: The authors' contributions are according to the authorship order.

Conflict of interest: The authors declare no conflict of interest regarding this commentary.

REFERENCES

Achterbosch, L., Miller, C., & Vamplew, P. (2017). A taxonomy of griefer type by motivation in massively multiplayer online role-playing games. *Behaviour & Information Technology*, *36*. https://doi.org/10.1080/0144929X.2017.1306109

Buckels, E. E., Trapnell, P. D., & Paulhus, D. L. (2014). Trolls just want to have fun. *Personality and Individual Differences*, *67*, 97–102. https://doi.org/10.1016/j.paid.2014.01.016

Cook, C. L., Tang, S. Y. C., & Lin, J. H. T. (2023). Comparing shades of darkness: Trolling victims' experiences on social media vs. online gaming. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1163244

Díaz-Moreno, A., Bonilla, I., & Chamarro, A. (2025). Gaming Toxicity Scale (GTOX): An assessment tool of toxic behaviours within online video games. *Behaviour & Information Technology*, 1–17. https://doi.org/10.1080/0144929X.2025.2546974

Fox, J., & Tang, W. Y. (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), 1290–1307. https://doi.org/10.1177/1461444816635778

Grigg, D. W. (2010). Cyber-aggression: Definition and concept of cyberbullying. *Australian Journal of Guidance and Counselling*, *20*(2), 143–156. https://doi.org/10.1375/ajgc.20.2.143

Kordyaka, B., Laato, S., Weber, S., & Niehaves, B. (2023). What constitutes victims of toxicity – Identifying drivers of toxic victimhood in multiplayer online battle arena games. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1193172

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). https://doi.org/10.1145/3410404.3414243

Kowert, R., & Cook, C. (2022). The toxicity of our (virtual) cities: Prevalence of dark participation in games and perceived effectiveness of reporting tools. In *Proceedings of the 55th Hawaii international conference on system sciences | 2022* (pp. 3180–3189).

Neto, J., Yokoyama, D., & 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). https://doi.org/10.1145/3106426.3106452

Quandt, T. (2018). Dark participation. *Media and Communication*, *6*(4), 36–48. https://doi.org/10.17645/mac.v6i4.1519

Shen, C., Ratan, R., Cai, Y. D., & Leavitt, A. (2020). Do men advance faster? Debunking the gender performance gap in two massively multiplayer online games. *Journal of Computer-Mediated Communication*, *21*(4), 312–329. https://doi.org/10.1111/jcc4.12159

Thacker, S., & Griffiths, M. D. (2012). An exploratory study of trolling in online video gaming. *International Journal of Cyber Behavior, Psychology and Learning*, *2*(4), 17–33. https://doi.org/10.4018/ijcbpl.2012100102

Wong, S., & Ratan, R. (2023). How does exposure to general and sexual harassment relate to female gamers' coping strategies and mental health? *Games and Culture*, *19*(5), 670–689. https://doi.org/10.1177/15554120231177600

Zhang, W., Huang, S., Lam, L., Evans, R., & Zhu, C. (2022). Cyberbullying definitions and measurements in children and adolescents: Summarizing 20 years of global efforts. *Frontiers in Public Health*, *10*. https://doi.org/10.3389/fpubh.2022.1000504

Ziems, C., Vigfusson, Y., & Morstatter, F. (2020). Aggressive, repetitive, intentional, visible, and imbalanced: Refining representations for cyberbullying classification. In *Proceedings of the international AAAI conference on web and social media* 1, (Vol. *14*, pp. 808–819). https://doi.org/10.1609/icwsm.v14i1.7345

Zsila, Á., & Demetrovics, Z. (2025). Taxonomy and working definition of toxic behaviors in online multiplayer video games. *Journal of Behavioral Addictions*. https://doi.org/10.1556/2006.2025.00034

Zsila, Á., Shabahang, R., Aruguete, M. S., & Orosz, G. (2022). Toxic behaviors in online multiplayer games: Prevalence, perception, risk factors of victimization, and psychological consequences. *Aggressive Behavior*, *48*(3). https://doi.org/10.1002/ab.22023