Toxic play: Examining the issue of hate within gaming
by Luke Munn

Abstract
This article examines the problem of hate speech and toxic behavior in gaming. Videogames have risen to become a dominant cultural form, seeing significant increases in players, playtime, and revenue. More people are playing games than ever before, broadening “gamers” into a highly diverse demographic. Yet this rise has been accompanied by a growing recognition of the racism, sexism, xenophobia, and other forms of harassment taking place on these platforms. Hate within gaming creates toxic communities and takes a toll particularly on marginalized groups, raising both ethical and financial issues for the industry, who seek to address this problem in multiple ways. This paper surveys and synthesizes recent research on the topic from both inside and outside academia, laying out the problem, its manifestations, key drivers, and current responses. It concludes with a research agenda that offers a foundation for researchers, policy-makers, and companies to build from.

Contents
Introduction
Gaming’s expansion
Measuring hate, feeling hate
Intense hate, organized hate
Industry acknowledgment
Potential drivers
Responses and challenges
Learnings and future research
Conclusion

Introduction
This article examines the problem of hate speech and toxic behavior in gaming contexts. While videogames have risen to become a dominant cultural form in terms of popularity and play time, there has been a growing recognition of the racism, sexism, xenophobia, and other forms of harassment occurring through them. This abuse takes a toll on its victims, contributes to the normalization of toxic behavior, and raises ethical and financial issues for the industry. This article surveys and synthesizes recent research on the topic from both inside and outside academia, laying out the problem, its manifestations, key drivers, and current responses. In doing so, it aims to provide a foundation for researchers, policy-makers, and companies to build from.

While the definitions of “toxic” vary, toxic behavior is generally considered an umbrella term for a wide variety of negative activities that abuse other players, violate rules, or break social norms (Beres, et al., 2021). Flaming, trolling, griefing, and harassment are commonly given examples of toxic behavior. Certainly such antagonistic or antisocial activities contribute to a climate of toxicity within any particular videogame context. However, by placing “hate” alongside this term, I want to also encompass more explicit and formalized forms of derogation ranging from misogyny and anti-Muslim sentiment to antisemitism and homophobia. Indeed, the rise of videogames as a culture and medium has made them an important new vector for the spread and uptake of these ideologies (Munn, 2019). This means that radicalization and extremism should figure into any discussion of hate and toxicity within gaming contexts.

Methodologically, this paper adheres closely to the traditional literature review. It “extracts and synthesizes the main points, issues, findings and research methods which emerge from a critical review of the readings” [1]. The aim, as with other literature reviews, is the identification of a problem domain, a critical discussion of what has been done, and the identification of knowledge gaps within this domain [2]. This article also takes inspiration from the two S’s approach, seeking to map a broad cross-section of state of the art research and lay out a stimulating set of issues that future research can take forward (Paul, et al., 2021). In doing so, this work offers a platform for knowledge advancement and further research (Palmatier, et al., 2018).
While the method is traditional, I deliberately adopt a more expansive or unorthodox approach of selecting “literature” to survey. When it comes to the topic of hate and gaming, there tends to be a bifurcation of research material. On the one hand, there are articles in academic journals, which primarily draw on other scholarly work. On the other hand, there is research outside academia: studies from gaming companies, industry bodies, independent institutes, and others. Journalistic articles are also in this category and often contain interviews with gamers, developers, and other industry insiders who describe the problem and their experiences in their own words, a viewpoint I have consistently championed in previous work (Munn, 2020, 2019). Both academic and non-academic research thus offer important insights into the problem of hate within gaming contexts. For this reason, this paper draws from both strands, highlighting points where personal experiences and survey statistics resonate strongly with scholarly studies.

The article itself is structured as a kind of meta-case study. The first section stresses the stakes of this issue by highlighting the growth and diversification of gaming. The next two sections unpack the problem of hate and radicalization within gaming contexts by drawing on a mixture of research reports, academic scholarship, and gamers’ experiences. The next three sections examine the recognition of this issue in the gaming industry, lay out potential drivers of toxicity, and discuss current responses to this problem. The final section concludes by summarizing key insights and setting out a research agenda and notional research questions. The aim of this structure is not to suggest any single solution or approach, but rather to provide a robust portrait of toxicity within gaming and highlight issues for the development studios, community groups, policy-makers, and other stakeholders who wish to engage with it.

---

**Gaming’s expansion**

The last two decades have seen gaming ascend as an industry, a cultural form, and a popular activity. Video games can no longer be dismissed as an eschatonic niche, but have proliferated into everyday life in many ways (Bogost, 2011). So while hate to any degree is an issue, the expansion and diversification of gaming shown below makes this problem more acute (Nakamura, 2019). On the one hand, this exposes new populations (women, people of color, queer communities) to damaging forms of racism, sexism, and harassment. On the other hand, it extends the potential reach of hate to a far broader population.

Player counts provide one indicator of gaming’s expansion. Microsoft’s gaming platform, Xbox Live, has over 100 million monthly active users (Lilly, 2021). Steam, a popular game distribution service, has over 120 million monthly active users (Yin-Poole, 2021). Such massive numbers do not even capture the extent of casual and hypercasual games on mobile devices and social media, which are also experiencing unprecedented growth (Koetsier, 2020). The rise of streaming and chat platforms also showcase gaming’s foray into the mainstream. In 2020 Twitch reached 143 million viewers in total (Iqbal, 2023). It attracted 2.58 million concurrent viewers on average in 2022, a figure that makes it bigger than cable channels like Fox, MSNBC, and CNN (Gilbert, 2018). Discord now has 140 million monthly active users and a new peak of 10.6 million concurrent users, with four billion minutes of conversation happening every day on the platform (Curry, 2023).

Playtime is also on the rise. A recent study (Limelight Networks, 2021) surveying 4,000 consumers in China, Vietnam, Germany, India, Indonesia, South Korea, the U.K., and the U.S. found that players spent an average of eight hours and 27 minutes playing per week, an increase of 14 percent from the previous year. The same study (Limelight Networks, 2021) also highlighted a growing trend of “binge gaming,” with a third of respondents stating they have played for more than five consecutive hours. Growth in players and playtime has only been intensified by the recent COVID-19 global pandemic. A week after quarantine orders were issued in the U.S., Verizon reported that video game usage during peak hours had risen 75 percent (Shanley, 2020). Other articles noted how stay-at-home mandates meant that video games were being played at “record levels” (Perez, 2020).

Alongside the growth of gaming is a broadening of the “gamer.” In the early days of gaming, the stereotypical gamer was young, white, and male (Paaßen, et al., 2017). While this association continues to be significant, particularly when it comes to racism, misogyny, and gaming (discussed later), recent demographics of gamers tell a different story. A recent survey (Accenture, 2021) of over 4,000 gamers across China, Japan, U.S., and the U.K. found that there are now nearly as many female gamers (46 percent) as there are male (52 percent). The same survey also noted that one third of “new gamers,” who have only started playing in the last four years, identify as non-white (Accenture, 2021). Another recent survey (Entertainment Software Association, 2020) echoed this demographic shift, reporting that women make up 41 percent of 214 million gamers in the United States, and that their average age range is now between 35 and 44 years old. Such statistics indicate that the once monolithic “gamer” now encompasses a highly diverse population of different ages, races, and genders.

All of these statistics underscore a key point: more people are playing games, a wider variety of people are playing games, and they are playing for longer durations than ever before. As the next section will show, this increase in gaming often means increased exposure to the hate that accompanies it. This also means that such hate within gaming is not a niche problem for a small segment of the population, but a significant issue that impacts broad segments of society.

---

**Measuring hate, feeling hate**

Gamers have long testified to the levels of hate associated with gaming. Whether playing with others on servers or streaming their play on platforms, numerous stories from players document the hate speech, toxic communication, and diverse forms of racism, sexism, homophobia, and xenophobia that they have endured (Lorenz and Browning, 2020). This is particularly the case for those who are already marginalized or seen as members of an out-group. Gaming platforms such as Twitch, for example, have been seen as a particularly harmful
space for women, LGBTQ+ individuals, and people of color (Taylor, 2018). For this reason, game critics have characterized gaming culture as a toxic meritocracy (Paul, 2018).

Recent studies have attempted to empirically measure these anecdotal and theoretical insights. In a nationally representative survey conducted by the Anti-Defamation League in the U.S., 83 percent of respondents ages 18–45 experienced harassment in online multiplayer games in the last six months (Anti-Defamation League, 2021). Of particular concern was that 71 percent of adult online multiplayer gamers experienced severe abuse, including physical threats, stalking, and sustained harassment (Anti-Defamation League, 2021). The same survey also signaled a strongly gendered and racialized dimension to this hate. Compared to the previous year, the largest increases in identity-based harassment occurred among adults who identified as women, Black or African American, and Asian American (Anti-Defamation League, 2021).

Such broader survey-based work has been augmented by more detailed research from computer science and related disciplines. Studies have focused on analyzing the levels of racism, sexism, and general hate speech within particular gaming communities using natural language processing and other computational methods (Poyane, 2018; Ghosh, 2021). Content moderation companies, such as Brand Bastion (https://www.brandbastion.com), have also stressed how pervasive such hate is across multiple gaming platforms, highlighting the toxic comments they encounter on a daily basis, from homophbic slurs to racist insults and encouragements to commit suicide.

One of the key issues with hate in gaming is how taken for granted it is. Gamers have noted how normalized this hate has become, with others dismissing episodes of bigotry and misogyny as ubiquitous and therefore insignificant (lady_haybear, 2019). Others describe being attacked while gaming because of the color of their skin while being told such comments were simply part of gaming culture (Suddi, 2017). Such laissez-faire acceptance within gaming culture may stem partly from broader Internet culture. As Lisa Nakamura (2013) notes, racism on the Internet comes with the infrastructure and is normalized as part of the online experience: it is the signal rather than a glitch. This resignation influences the way in which hate is perceived and responded to. Gamers who are targeted by racism, such as men of color, often cope with it through desensitization, downplaying it as not serious or real (Ortiz, 2019b). This insight echoes other research, which finds that gamers fail to report behavior because they see it as acceptable, typical, or not worthy of flagging (Beres, et al., 2021). As will be discussed later, scholars see this as one of the major drivers of hate within gaming. The acceptance and normalization of toxic abuse creates a kind of feedback loop, perpetuating a culture marked by racism, sexism, and other forms of bigotry.

Taken together, these insights suggest that hate in gaming contexts is often dismissed as part of the culture, a nasty side effect that has to be tolerated. Yet such “casual” racism, sexism, and misogyny online incurs a real-world fallout. In a study of 765 racial minority adults in the U.S., time spent in online gaming predicted greater exposure to online racism, which in turn was linked to higher psychological distress (TaeHyuk Keum and Hearsn, 2022), particularly among Black gamers. Similar forms of physical and emotional toll are described from those who have suffered sexist attacks while gaming. The annual “She plays he says” report has documented the widespread misogyny and rape threats targeted at female gamers, some of them as young as 14, who say they’ve had to significantly change the way they engage with games and communities to cope (Young Gamers & Gamblers Education Trust [YGAM], 2021). Those who stream their gaming have noted that “chat can be ruthless sometimes” and victims of this abuse end up “shaken to their core” (Johnson, 2019). Such testimonies from those who have suffered first hand point to the real-world impact of hate within gaming contexts.

---

**Intense hate, organized hate**

Hate within gaming contexts can manifest in more organized or intense ways, where certain groups are attacked, violence is endorsed, and extremist ideologies are promoted. Radicalization and extremism are not isolated topics, then, but should instead be understood within gaming’s spectrum of hate. In a three-month investigation, BBC researchers found antisemitism, racism, and homophobia on popular gaming platforms like DLive and Odysee and hateful content within popular children’s games, from a concentration camp in Minecraft to a de facto social media platform, yet one without the same level of content moderation and safeguards employed by giants like Facebook and Twitter. The Institute for Strategic Dialogue (Vaux, et al., 2021) noted that Steam has an “entrenched and long-lasting extreme right community” with many groups on the platform dating back to 2016 or even earlier; the report noted that groups used the platform more as a space for community building than deliberate recruitment, though the platform did provide “off-ramps” to podcasts, social media, and articles that groups could use to further their cause. For Vaux, et al. (2021), the platform’s “permissive attitude to this harmful activity means that these communities have a safe haven to promote and discuss extremist ideology and content.”

The same kind of organic development can be seen with Discord. Discord began life in 2015 as a service that provided text and voice-based chat for gamers, allowing them to coordinate and communicate, particularly in team or squad-based games. Discord enjoyed significant success over the next few years, both in terms of venture capital funding and a growing user base. But the platform’s relatively autonomous technical architecture, which allows users to setup “servers” with thousands of users, already began to give it a reputation as a haven for...
hate-based communities. One Discord server was used to organize the infamous 2017 white supremacist rally in Charlottesville, Virginia (Glaser, 2018). Although the platform has fought these communities and shut servers down, white nationalism, antisemitism, and pro-Nazi content still appear (Patterson, 2019). In 2020, the platform announced a new slogan, “your place to talk,” signaling a move beyond gaming and a shift into a more general-purpose communication platform (Chin, 2020). But this expansion has only seemed to increase the platform’s potential for extremist groups and organizations. In 2021, the Institute for Strategic Dialogue released a similar report on Discord, finding that the platform was a “hub for extreme right-wing socializing and community building” (Gallagher, et al., 2021). As with Steam, then, we see a pattern in which “gaming” platforms become a broader communication infrastructure used by a wide variety of people, becoming a potent new recruitment and organizational space for movements associated with hate, extremism, and radicalization.

In recognition of this threat, the U.N.’s Office of Counter-Terrorism (2021) held a roundtable on this topic, stating that “violent extremists are increasingly exploiting these expanding spaces in attempts to recruit, radicalize, and disseminate propaganda.” Similarly, the recently launched Extremism and Gaming Research Network (Royal United Services Institute [RUSI], 2021) noted that “violent extremist organizations are actively exploiting online gaming across the world” but “research into gaming and radicalization is sparse and outdated.”

Gaming, as noted earlier, has expanded to become a highly popular cultural form, with millions of people spending significant periods of time in gaming contexts. Because of this, gaming platforms are natural spaces for extremist groups to engage and mobilize others, making gaming the “latest frontier in radicalization” (Shurin, 2022). And yet, as many organizations suggest, research on this urgent issue — how often it occurs, what forms it takes, and how it might be counteracted — is limited or only just emerging.

Industry acknowledgment

How is this issue perceived in the gaming industry? For the gaming industry, hate and toxic behavior is not just an ethical issue, but a financial and practical one. Receiving harassment in the form of racism, sexism, xenophobia, or other attacks is an unpleasant or even traumatic experience for players. As a result, players may reduce their play or quit the game entirely, seeking more inclusive and welcoming spaces. As Kuipers, et al. (3) note, a negative experience may mean that players “leave for another game operator or even trigger through their social ties departures of large groups of players.” Regardless of a company’s moral stance, then, there are clear economic advantages to avoiding “churn,” reducing harassment, and maintaining safe spaces where play is not disrupted. And yet the industry has often struggled to curb toxicity (Smith, 2019).

Some insiders see a growing acknowledgement in the industry that toxic behavior undermines healthy communities, hurting the gaming business (Figueiredo, quoted in Smith, 2019). For instance, a gaming platform recently aligned with industry bodies to conduct a mental health survey of players, a move that in itself admits the growing role of games and their ability to shape lives in positive and negative ways. The survey, resonating with others presented in this article, unsurprisingly found that significant numbers of players encountered insults (57 percent), trolling (53 percent), and aggressive behavior (52 percent) (MY.Games 2021). After seeing the findings, the Executive Director of the International Game Developers Association responded by openly acknowledging that “toxicity in game communities is still an issue for our industry” (MY.Games, 2021).

This growing recognition can be seen in recent initiatives launched by the gaming industry to address the problem of hate. In Europe, these range from Game Over Hate in Austria to Play Your Role in Portugal, a project aiming to be a “counter-action” to hate speech in gaming contexts. Most recently, Deutsche Telekom (2021) launched a major campaign of no hate speech in gaming, collaborating with an esports foundation and 44 other partners. But the most ambitious of these is undoubtedly the Fair Play Alliance, an association founded in 2017 that describes itself as a “cross-industry initiative of nearly 200 gaming companies working together to encourage healthy communities and player interactions” (Fair Play Alliance, 2021). Members include major studios such as Blizzard, Electronic Arts, Epic Games, and Riot Games.

In 2020, the Fair Play Alliance published their “Disruption and harms in online gaming framework.” One aim of the Framework is to improve fuzzy or subjective terms like “toxic” and articulate more precisely what constitutes disruptive behavior in gaming. For instance, the Framework sees “hate” as a distinct subset of disruptive behavior alongside other practices such as “criminal or predatory conduct,” “cheating,” “harassment,” and “extremism” (4). Such definitional work seeks to establish an industry-wide understanding of this issue, striving to build consensus around key terms. But the Framework also presents a series of concrete steps that companies can take to address and improve disruptive behavior in their products. These steps include “Assessing the behavior landscape” to define problems and goals, “Planning and building a penalty and reporting system,” and “Creating and maintaining community guidelines” (5). Such recommendations include pragmatic in-game mechanisms like reporting while also underscoring the broader sociocultural dynamics that feed into a healthy gaming community.

Potential drivers

What drives the pervasive toxicity and forms of hate that appear consistently in gaming contexts? While there is no single reason, scholarship has identified an array of key issues that seem to contribute to toxic cultures.

The “Gamer”

Both scholars and gamers often point to the history of gaming culture and the hardcore gamer who was consistently catered to (Braegger and
Moeller, 2021). If the demographic of gaming has expanded significantly in recent years, the figure of the white heterosexual male gamer continues to linger in gaming culture (Paaßen, et al., 2017), a legacy defining who is accepted and who is attacked. In fact, one of the most visible forms of hate and gaming culture in the last decade was GamerGate, an online harassment campaign targeting female gaming journalists. GamerGate has been covered extensively, both in academic articles (Phidd, 2019) and the popular press (Dewey, 2014); its complex story will not be repeated here. It suffices to note that the driving force behind this “movement” was a fear of feminization (Vanderhoef, 2013) and a sense that gaming was slipping away from a core white masculine base. Gamers responded with abuse, signaling a refusal of the diversity and progressive values that have encroached on “their” space in recent years. As scholars have argued, these events showcased that, at the heart of gaming culture, there was still a deep connection between geek masculinity and online abuse (Salter, 2018), a connection between antifeminism and toxic technocultures (Massanari, 2017).

**Exposure**

As this paper has already touched on, hate-driven behavior in gaming is so pervasive that it often becomes normalized. This suggests that racist, sexist, or otherwise hateful behaviors are no longer considered taboo, but are framed as commonplace or even inevitable. For this reason, Kordyaka, et al. (2020) suggests that recipients of toxic behavior may come to accept toxic behavior and even emulate it in the future. Such a pattern would certainly align with scholarship on analogous issues like bullying, where there is a consistent overlap between those who are victimized, and those who go on to perpetuate this behavior (Falla, et al., 2022).

Beres, et al. (2021) describe toxicity as cyclical: toxicity in gaming contexts breeds more toxicity. This insight echoes other studies which argue that toxicity is contagious (Shen, et al., 2020). The presence of hateful comments and antagonistic behavior by some players “set the tone” for others and can lead to it being replicated. Players employ a sophisticated array of reasons to justify this amoral, immoral, or otherwise problematic behavior in games (Sparrow, et al., 2019). This scholarship suggests that players internalize a toxic culture of intimidation and abuse through hours and hours of online play, allowing them to reproduce this kind of speech and action when desired. These players then dish out this toxicity to others and, in doing so, contribute to its pervasiveness and persistence in that gaming context.

**Competition**

For some scholars, gaming’s toxicity is connected to its high-stakes context, which encourages antagonism and hostility rather than cooperation and civility. Adachi and Willoughby (2011) suggested that competition rather than violence was the video game characteristic that has the greatest influence on aggressive behavior. Grandprey-Shores, et al. (2014) developed a metric to identify highly toxic players and found that these players tended to play in more competitive game modes.

In a study of millions of instances of toxic behavior perpetrated by hundreds of thousands of accused toxic players, Kwak, et al. (2015) found that competition was key, with in-group favoritism and out-group hostility shaping the level of reporting. Similarly, Shen, et al. (2020) found that teams who are losing or have a high internal skill disparity (i.e., some players are much better than others) tend to breed toxicity. In high pressure situations, lashing out with insults and abuse becomes a way of deflecting blame and dealing with loss.

**Disinhibition**

Other scholars suggest that online gaming environments lack some of the key mechanisms in real world settings that foster prosocial behavior. For instance, Suler (2004) suggests that the anonymity (“you don’t know me”) and invisibility (“you can’t see me”) provided by online spaces fosters a strong sense of disassociation, allowing users to disown their behavior and its consequences. The precise impact of anonymity has been debated. On social media, for instance, Rösner and Krämer (2016) find that anonymity doesn’t have a direct effect, but that users do tend to conform to a more aggressive social norm when commenting in anonymous environments.

Based on these results, we could speculate that anonymity combines with other factors (invisibility, mediation, competition, and so on) to create a sense of distance and erode social contracts in various ways. This is certainly the dynamic witnessed by some gamers in their daily encounters. “Players often berate, belittle, bully, and threaten others with no fear of repercussions,” one gamer wrote [5], “they view their toxic behavior as part of the gaming culture and readily dismiss their wrongdoings with little or no guilt.” More recent research on videogames specifically has echoed this finding. In a survey of 320 participants, (Kordyaka, et al., 2020) found that online disinhibition provided the best explanation for toxic behavior.

**Industry**

While the gaming industry has begun to take steps to address hate, some critics suggest that the industry itself must take part of the blame. For the Anti-Defamation League (2021) there is a connection between the harassment that many gamers experience and harassment that persists within the gaming industry. High-profile episodes of sexual harassment and toxic work cultures within the gaming industry have occurred recently, from Riot Games in 2018 to Ubisoft in 2020 (Dealessandri, 2020). The most recent was an incident at Activision Blizzard in 2021, when a women who experienced intense sexual harassment on a company trip tragically took her own life. For one games scholar, the resulting lawsuit was a “display window into misogyny in the digital-gaming industry,” with toxic behavior extending to the highest places (Kukumbergová, 2021; see also Allsup, 2021). de Castell and Skardzius (2019) draw on testimonies from women in the gaming industry to document the pervasive abuse against them, an all-too-frequent occurrence in an industry with a long history of exclusion, marginalization, and hostility toward women. Together, these incidents and scholarship suggest a pattern rather than isolated anomalies, a deep-seated problem that requires serious and sustained attention from the industry.

A key argument here is that a toxic industry tends to produce toxic products. Based on interviews with industry insiders, Tompkins and Martins (2021) draw a line between game studios dominated by white heterosexual men and the games they produce largely for other young men, where women are hypersexualized and objectified and diversity in character design is seen as risky. Boudreau (2022) also draws a connection between the problematic, toxic, and exclusionary behaviors evidenced in the gaming industry and the kinds of games and
Toxic play: Examining the issue of hate within gaming

and other AI-based approaches still struggle with (Knight, 2016) and a known issue in hate speech recognition (MacAvaney, 2020). Accurately decoding speech into text, a non-trivial task particularly when voices are shouting, clipped, or non-native speakers (Radzikowski, 2016), such customization gestures at the problem of over-filtering and the complexity and nuance of language. Firstly there is the issue of toggle for the N-word (Diaz, 2021). “Body Shaming,” “Racism,” “White Nationalism,” and others. Yet while the product has yet to be released in the final version, it has for harassment and abuse. The product aims to listen for and censor various forms of toxic speech, from “LGBTQ+ hate” to “Misogyny,” “Lolita,” “Rape,” “Schools, and “Satanic,” among others. While a number of companies have employed similar tools for text-based speech, Bleep is the first initiative to tackle voice-chat. Automation and contextualization Alongside notifications and nudging techniques (Thaler and Sunstein, 2009), companies are also experimenting with fully automated solutions. In 2021, Intel launched Bleep, an AI-powered tool that aims to identify and filter out toxic speech on gaming platforms (Porter, 2021). A communal effort to improve civility and respect in a gaming environment, but rather appropriated as a weapon to gain an edge in LoL’s highly competitive environment. Flagging One of the dominant responses to address forms of online abuse has been establishing reporting systems. The seemingly straightforward solution offered through reporting is compelling: flag something and it will be dealt with. Yet as Crawford and Gillespie (2016) noted early on in the context of social media, flagging entails a complex interplay of users, platforms, norms, and regulatory structures — as a mechanism for contestation and complaint, it is both complex and limiting. In addition, the success of reporting is not guaranteed, but instead relies on automated systems to deal with it appropriately or on human moderators to make correct judgements and take effective actions. The flag hides a vast infrastructure required for it to function.

In a gaming context, flagging interacts with gaming norms already discussed. Several studies already cited here found that reporting is uneven, with incidents going unreported or underreported because of the normalization of toxic behavior in gaming contexts (Hilvert-Bruce and Neill, 2020; Beres, et al., 2021). In other words, abuse and vitriol are downgraded to “banter,” and seen as not worthy of flagging. In some games, reporting systems can be taken up in highly strategic and perhaps undesirable ways. In their study on League of Legends, Kou and Gui (2021) found that reporting or “flagging” was used in an instrumental way to punish other players. In this sense, reporting is not a communal effort to improve civility and respect in a gaming environment, but rather appropriated as a weapon to gain an edge in LoL’s highly competitive environment. Flagging then is far from being a straightforward or catch-all solution.

Design nudges Other work stresses that toxic behavior must be dealt with at the level of design. Kordyaka and Kruse (2021) have collaborated with game developers and industry experts, putting forward a range of design interventions that could be beneficial. One suggestion is that player profiles should be more holistic or even tied to real world identities (as is the case in South Korea), counteracting disinhibition by creating social/reputational cues and signaling consequences for in-game activity. A second recommendation is to have interfaces which are more transparent about the balancing/fair mechanisms built into the game, discouraging blaming of others. Another suggestion is to reward prosocial or positive behavior in the gaming context through badges, rewards, or other items. Clearly, none of these interventions is a “silver bullet,” eradicating toxic behavior by itself. These recommendations instead suggest a cumulative effect, in which small design decisions come together to shape the culture of a gaming community in subtle but significant ways.

Riot Games provides one example of this design-centric approach and is often cited as an exemplary case study of addressing hate (Maher, 2016; Campbell, 2021). Riot is the developer of League of Legends, a highly popular game with millions of active players. However, in 2016 the company noted the growing presence of hate speech and toxic comments within the game, and set out to address it. While several techniques were tried, the company experienced most success with a red warning message about harassment, which reduced offensive language by 11 percent and a positive message about player cooperation, which reduced offensive language by 6.2 percent (Maher, 2016). Such results may be promising, but they are far from being a comprehensive solution. In a 2020 player-created survey (Clanaria, 2020) of 3,784 League of Legends players, 98 percent of participants reported they had been flamed during the game and 79 percent said they had been harassed after the game. Yet if hate is but being “solved,” Riot’s work in this space is worth noting for its restorative approach. Rather than simply penalizing or banning players (who may quickly sign up with another account), the company seeks to reform its player base over time, encouraging prosocial activity and discouraging toxic or hateful comments.

Automation and contextualization Alongside notifications and nudging techniques (Thaler and Sunstein, 2009), companies are also experimenting with fully automated solutions. In 2021, Intel launched Bleep, an AI-powered tool that aims to identify and filter out toxic speech on gaming platforms (Porter, 2021). While a number of companies have employed similar tools for text-based speech, Bleep is the first initiative to tackle voice-chat. Voice-chat is a core feature of online gaming used for team communication and built into many popular titles, but it is also a major channel for harassment and abuse. The product aims to listen for and censor various forms of toxic speech, from “LGBTQ+ hate” to “Misogyny,” “Body Shaming,” “Racism,” “White Nationalism,” and others. Yet while the product has yet to be released in the final version, it has already received criticism for sliders which allow users to hear “none,” “some,” or “all” of this hate speech — and even features an on/off toggle for the N-word (Diaz, 2021).

Such customization gestures to the problem of over-filtering and the complexity and nuance of language. Firstly there is the issue of accurately decoding speech into text, a non-trivial task particularly when voices are shouting, clipped, or non-native speakers (Radzikowski, et al., 2019). But secondly and more fundamentally, there is the issue of the context and intent of language, something that machine learning and other AI-based approaches still struggle with (Knight, 2016) and a known issue in hate speech recognition (MacAvaney, et al., 2019). A sentence can be intensely racist or sexist without explicitly using slurs. Similarly, a sentence can endorse xenophobia or homophobia without using flagged phrases. Given these limitations, automatic evaluation of hateful versus benign speech presents high demands. This is
particular case within gaming contexts, where “trash talk” and aggressive behavior is seen as part of the culture (Hilvert-Bruce and Neill, 2020). Banter has always been highly subjective and prone to rapidly shift from playful to hateful, particularly for women, people of color, or anyone whose identities don’t cleanly conform to the legacy “gamer” stereotype (Ortiz, 2019a). Simple dictionary-based techniques are insufficient and a high degree of context awareness is needed to make nuanced distinctions between trolling and intentionally abusive language (Sengün, et al., 2019). Such ambiguity suggests that technical approaches to hate speech in gaming will remain challenging for some time to come.

**From representation to operation**

How can gaming cultures become less hateful and more inclusive? A significant strain of gaming scholarship has focused on the “visual politics” (Murray, 2017) of “gaming representation” (Malkowski and Russworm, 2017). Who is included and who is excluded from games? How are race, sexuality, and gender presented? And how does this representation, skewed representation (e.g., sexualized women), or total lack of representation (e.g., queer people of color) negatively affect players (Smith and Decker, 2018; Gestos, et al., 2018)? Broadly speaking, this work suggests that gaming’s toxic culture derives in part from its privileging of white, heteronormative, hypermasculine characters. The argument, while often implicit, is that more diverse representations in this media would equate to a more tolerant and respectful gaming culture.

However, as Kelley (2020) notes, it’s important to distinguish between games as media and games as social spaces. A game may feature diverse representation or progressive values, but this content or perspective may have minimal impact on the hate that gamers experience on the platform. *Overwatch*, for instance, is an incredibly popular title that features several LGBTQ+ characters, rainbow icons, and celebrates Pride Month in-game. However, Kelley (2020) notes that 75 percent of the *Overwatch* players in their recent ADL survey experienced harassment in the game. In this sense, progressive characters or inclusive icons are largely symbolic, failing to translate into an inclusive and accepting space for players. Such a dynamic runs counter to the representation-in-games scholarship above. In fact, Robinson and Whitaker (2021) argue that existing scholarship has overly privileged the “reading” of in-game representations. While games themselves may feature inclusive elements or support progressive values, racism, sexism, and other forms of hatred may be deeply embedded in the community and culture surrounding them. For those seeking to counteract such hate, inclusive or progressive content needs to be matched with an equally inclusive and affirming community. Alongside representation, we should equally focus on the hard, day-to-day work of community management within gaming contexts, a difficult role that requires experience and emotional labor (Kerr and Keleher, 2015).

**Learnings and future research**

Drawing on the wealth of research and testimony above, we can summarize a set of key learnings:

1. More people (and more diverse people) are playing games, increasing the likelihood of being exposed to toxic behavior, and elevating the impact and significance of hate as a problem
2. Hate includes antisocial behaviors (trolling, griefing) yet also encompasses more organized and explicit forms of hate (racism, radicalization, extremism)
3. Hate within gaming is widespread; and this ubiquity contributes towards normalizing hate as an inevitable aspect of gaming culture
4. The industry is increasingly recognizing and responding to this problem through various initiatives, yet consensus and effective mitigation remains challenging
5. The reasons for hate within gaming are contested but frequently listed are:
   a. a historically homogenous culture (white, male, heterosexual) with its related issues
   b. exposure to hate, which normalizes it and establishes a victim > perpetrator loop
   c. competition that fosters antagonism and out-group hostility
   d. disinhibition which removes social safeguards and allows action without repercussion
   e. overlap between a toxic industry (production) and toxic cultures (consumption)
6. Responses to hate vary from automation to content-moderation and range in efficacy; hate is far from being solved and requires holistic, context-sensitive interventions

Based on these insights, we can identify gaps in current research and suggest some productive questions that might guide future investigations.

**Toxic-hate connections**

What is the relationship between antisocial behaviors such as griefing and cheating and more intense or organized forms of hate such as racism and extremism? Research on “toxicity” within gaming has tended to focus on the former while bracketing out the later. In other words, scholarship has closely examined rule breaking and disruption and partially neglected powerful and historical forms of hate such as white supremacy or antisemitism (Sartre, 1948). However, as some studies gesture (Sengün, et al., 2019), “casual” forms of anti-Muslim sentiment, for example, are precisely the ways in which some players are harassed or trolled. If this is the case, then are antisocial or aggressive players more likely to take up and propagate extremist ideologies?

**Impacts on new gamers**

How are marginalized communities impacted by hate within gaming? There has only been nominal work on this issue to date. The Anti-Defamation League (2021) report, for instance, quantified the prevalence of abuse in gaming on people of color. Some studies have recognized this issue and explored how black men respond to these toxic attacks (Ortiz, 2019b). However, such work remains nascent. It
Toxic play: Examining the issue of hate within gaming

Holistic theories

Is it possible to synthesize the causes of toxicity in gaming into a cohesive theory? Recent papers have pointed to this as a gap in existing research and begun to develop a more unified theory of toxic behavior in videogames (Kordyaka, et al., 2020). However, this work is emergent and the work-to-date around disinhibition has shown less awareness of how powerful historical hate-forms (racism, nationalism, antisemitism, etc.) are being repackage in new ways (Munn, 2019) and contributing to these toxic cultures. Theorizations that brought together key drivers of toxic gaming cultures in a systematic way would be a valuable foundation that helps us understand (and counteract) these drivers.

Resisting normalization

How might we push against the normalization of toxicity in gaming cultures? In other words, what interventions could render hateful behaviors unacceptable? On an individual level, studies have explored the potential of counterspeech, with users speaking back directly against hateful communication (Schieb and Preuss, 2016). On a group level, the use of clear standards which are consistently upheld through moderation and banning have been successful in mitigating some of the most overtly toxic elements within communities (Chandrasekharan, et al., 2017). Such mechanisms may be labor intensive and arguably heavy handed (deletion and removal), but they do seem to draw a line regarding acceptable and unacceptable activity. On a societal level, nascent programs like Game Over Hate (https://gameoverhate.tumblr.com) suggest that there is appetite for educational initiatives which provide young gamers with the tech literacy and skills to recognize and reject particular behaviors. More research is needed to identify the most effective ways to develop a zero-tolerance policy for hate, both at the cultural and technical levels.

Design-centric solutions

How can the design of games reduce toxicity and nurture a respectful culture? Interfaces, prompts, filters, and safety features are all design mechanisms that could contribute to this goal, shaping cultural norms by signaling that some behavior is damaging and other behavior is desirable. Recent studies have started to investigate how design principles could do this more consciously and effectively (Kordyaka and Kruse, 2021). Yet such work is emergent and often narrow in application, with suggestions aimed at particular subgenres of games. This research might draw from more mature work on social media platforms, which have long employed interface prompts and reputation systems as one way to foster civil cultures and communication. Research on gaming that employed a design-centric lens, with concrete case studies demonstrating effectiveness, would be highly valuable.

Conclusion

This article examined the problem of hate speech and toxic behavior within gaming contexts, aiming to provide a starting point for researchers, policy-makers, and companies to build from. In recent years, the video game industry has enjoyed significant growth with gaming rising to become a major cultural form. Such growth means that more — and more diverse — people are playing than ever before. Yet along with this growth, there is a growing recognition of the widespread racism, sexism, xenophobia, and other forms of harassment taking place on these platforms. While such hate is often dismissed as part of the culture, it nevertheless incurs a significant psychological and emotional impact. The most frequently and intensely affected are those from already marginalized groups: women, people of color, queer individuals, and so on. Hate has ethical but also financial implications in terms of creating negative player experiences and hurting business prospects. As a result, the gaming industry has begun to recognize this issue, albeit slowly and unevenly, conducting their own studies, forming alliances, creating safety frameworks, and highlighting some of the toxic culture within studios and the industry itself. Alongside this acknowledgement, initiatives to reduce or remove hate within gaming contexts are also being trialed, from community management to automated software. While some responses have seen marginal success, hate is articulated in complex and context-specific ways, often frustrating both human regulations and technical solutions. Hate is not simply a slur that can be “fixed” through censorship, but is rather deeply tied to emotion and identity, a phenomenon with social, cultural, and political dimensions. For this reason, more research is urgently needed to advance our understanding of the links between gaming and specific forms of racism, sexism, and xenophobia. Such research would not just address the end-products of hate but understand its logics and drivers, laying the groundwork for more systematic and effective interventions.

About the author

Luke Munn is a Research Fellow in Digital Cultures & Societies at the University of Queensland. His wide-ranging work investigates digital cultures, critically combining diverse methods with analysis drawing on media, race, cultural, and environmental studies. This research, spanning 40+ articles and six books, has been published in highly regarded journals and with a variety of academic presses as well as referenced in popular news media.

E-mail: luke [dot] munn [at] gmail [dot] com
Notes

6. Fu, 2019, p. 11.

References


Toxic play: Examining the issue of hate within gaming


Toxic play: Examining the issue of hate within gaming


Toxic play: Examining the issue of hate within gaming


Toxic play: Examining the issue of hate within gaming


Toxic play: Examining the issue of hate within gaming


Wesley Yin-Poole, 2021. “Steam has over 120m monthly active users,” Eurogamer (7 January), st https://www.eurogamer.net/articles/2021-01-14-steam-has-over-120m-monthly-active-users, accessed 11 September 2023.


---

**Editorial history**

Received 18 May 2022; revised 21 September 2022; accepted 11 September 2023

---

This paper is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

Toxic play: Examining the issue of hate within gaming

by Luke Munn.

*First Monday*, volume 28, number 9 (September 2023).

doi: [https://dx.doi.org/10.5210/fm.v28i9.12508](https://dx.doi.org/10.5210/fm.v28i9.12508)