Feeling Good and In Control: In-game Tools to Support Targets of Toxicity

ELIZABETH REID, The University of British Columbia, Canada and University of Saskatchewan, Canada
REGAN L. MANDRYK, University of Saskatchewan, Canada
NICOLE A. BERES, University of Saskatchewan, Canada
MADISON KLARKOWSKI, University of Saskatchewan, Canada
JULIAN FROMMEL, Utrecht University, Netherlands and University of Saskatchewan, Canada

Game developers, researchers, and players recognize the harm of toxic behaviour in online games—yet toxicity persists. Players' coping strategies are limited to tools that focus on punishing toxic players (e.g., muting, blocking, reporting), which are inadequate and often misused. To address the needs of players experiencing toxicity, we took inspiration from research in other online spaces that provide support tools for targets of harassment. We iteratively designed and evaluated in-game tools to support targets of toxicity. While we found that most players *prefer* tools that explicitly address toxicity and increase feelings of control, we also found that tools that solely provide social or emotional support also decrease stress, increase feelings of control, and increase positive affect. Our findings suggest that players may benefit from variety in toxicity support tools that both explicitly address toxicity in the moment and help players cope after it has occurred.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; **Collaborative and social computing systems and tools**; **Empirical studies in collaborative and social computing**; • **Applied computing** → **Computer games**; • **Software and its engineering** → **Interactive games**.

Additional Key Words and Phrases: games, support, toxicity, toxic, control, harassment, griefing, report, mood

ACM Reference Format:
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), 27 pages. https://doi.org/10.1145/3549498

1 INTRODUCTION

Game developers, researchers, and players have identified *toxicity*—a set of negative behaviours that disrupt gameplay or game enjoyment, including but not limited to harassment, cheating, and raging [1, 2, 10, 30, 49, 81, 90]—as a major problem in multiplayer online games. Toxic behaviour in games is sometimes further categorized into 'trolling' (verbal or in-game behaviours intended to provoke and antagonize other players) [10, 24, 90], 'flaming' (aggressive or derogatory language) [10, 49], 'griefing' (play styles that disrupt the gaming experience of other players) [1, 10, 30, 75, 90], or 'spamming' (repeated disruptive use of online communication) [10]. Players who are exposed to toxicity may experience decreased mood or enjoyment [77, 90], lower game performance [57], and

Authors' addresses: Elizabeth Reid, reid42@cs.ubc.ca, The University of British Columbia, Vancouver, Canada and University of Saskatchewan, Saskatoon, Canada; Regan L. Mandryk, regan@cs.usask.ca, University of Saskatchewan, Saskatoon, Canada; Nicole A. Beres, nicole.beres@usask.ca, University of Saskatchewan, Saskatoon, Canada; Madison Klarkowski, madison@cs.usask.ca, University of Saskatchewan, Saskatoon, Canada; Julian Frommel, j.frommel@uu.nl, Utrecht University, Utrecht, Netherlands and University of Saskatchewan, Saskatoon, Canada.

2573-0142/2022/10-ART235 $15.00
https://doi.org/10.1145/3549498

may even withdraw from the game entirely [31, 34]. Further, players from marginalized groups are particularly affected by toxicity, with women [31, 47, 76], players of colour [35, 36], and members of the 2SLGBTQ+ community [7] being disproportionately targeted by toxic players. Yet, despite efforts to identify and combat toxic behaviour in games, toxicity remains highly subjective and difficult to define, partly due to unique game contexts and individual norms within different game communities [10, 34, 90].

In games, current approaches to address toxicity centre around punishing toxic players. For example, most competitive multiplayer games offer some form of reporting tool that allows players to flag perpetrators for punishment, or blocking or muting tools to help players avoid further exposure to toxicity. However, the utility of reporting is limited [65] for various reasons, such as not reporting toxic players because of normalization [10] or reporting non-toxic players who play poorly [45].

In contrast to the common approaches of reporting, blocking, muting, or punishing toxic players in games, research in other online spaces has instead taken the approach of supporting the targets of harassment. For example, similar to hotlines traditionally used to support survivors of domestic violence or sexual assault [92], HeartMob is an online community that helps those exposed to online harassment by providing real-time support [37]. In particular, HeartMob supports targets of harassment by the provision of emotional and instrumental support, such as sending positive messages or helping to guide the process of reporting abusive users. The benefits of this approach are supported by findings from Vitak et al. [93], who identify that sending targets of online harassment positive messages may help mitigate harm. In general, social support and promoting experiences of positivity have been proposed as promising approaches in the design of interactive tools that might mitigate the harm of cyberbullying [5] and racism [88]. Emotional and instrumental social support, along with positivity and humour, are well-established coping approaches that have been demonstrated in a variety of cultures and contexts [20]. However, these resources are often external to the platform on which the harassment occurred, and must typically be actively sought out by those in need of such resources. A more platform-specific, integrative approach to providing these resources could support targets in real-time, potentially further mitigating the harm of online harassment.

Given that toxicity in gaming is—at its core—harassment of other players, and given that interventions supporting targets of harassment have been successful in other online spaces [12, 28, 52], it is surprising that there are no existing game interventions for toxicity that explicitly support the target, rather than aiming to punish the perpetrator. There is evidence that support-based interventions might be welcomed by players; Passmore and Mandryk [62] identified *seeking social support* in a taxonomy of coping strategies among gamers who have been exposed to discrimination within online games. Players who form friendships in a game have also been shown to turn to these friends for offline advice and emotional support [40, 60, 89]. However, options to receive support in games are limited for various reasons. For example, the normalization of toxicity within gaming has led to a culture in which many players feel that toxic behaviours are simply an inextricable element of the competitive video game experience that is acceptable, typical of games, and is not actually harassment, but instead harmless banter [10]. McLean and Griffiths also showed that women players were directly insulted or brushed off when discussing experiences of harassment, and further received no bystander support from other players, not even from other women [55]. For game developers and community managers who may wish to develop and deploy integrated support tools, the problem is that there is little guidance on how tools designed to support targets, rather than punish perpetrators, might work within games: Do gamers value support tools? Do support tools actually help targets of toxicity? And what particular facets of support are players looking for within gaming contexts?

We iteratively designed several in-game tools intended to answer the following research questions:

- RQ1: What do players hope to gain from the use of in-game support tools?
- RQ2: Are in-game support tools effective at providing support for players targeted by toxicity?
- RQ3: Are there differences in effectiveness between women and men?

We conducted two studies evaluating players' perceptions and experiences of these tools. In our first study (n=31), we iteratively designed and qualitatively evaluated low-fidelity prototypes of six tools inspired by research in other online spaces that show the benefit of supporting targets of harassment, aiming to answer RQ1 and to improve initial tool designs.

In our second study (n=132), we refined four of these tools that represent a two-dimensional space: problem- or emotion-focused tools, using either internal or external approaches, to support targets of game-based toxicity (*Blocklist*, *Eyebleach Pictures*, *Send A Message*, *Friendly Messages*). We compared these tools to *Report*, as the industry standard, and *No Tool*, as a baseline condition, by presenting interactive, high-fidelity mockups of each tool. To explore RQ2, these prototypes were embedded in recorded examples of in-game toxicity from the popular online multiplayer game, *Overwatch*, a game chosen because its features provide increased opportunities for toxic interactions amongst players [10] and because its community is suffering from toxicity in general [66].

Study 1 showed that feelings of control and positive emotions were important for players (RQ1), with a reported preference for tools that explicitly address toxicity. Study 2 showed that *all* tools—including those that solely provide social or emotional support—were effective (RQ2), as they significantly decreased stress, increased feelings of control, and increased positive affect after players were exposed to a toxic interaction. A gender-based analysis revealed that the reduction in stress and arousal, alongside an increase in valence and control, provided by the tools was larger for the women in our sample (RQ3). That is, the efficacy of our toxicity support tools was amplified for a group of players who stand to benefit most—the women who are disproportionately targeted by game-based toxicity. In sum, our findings suggest that in-game support tools *are* effective. Players may benefit from more variety in toxicity support tools that both explicitly address toxicity in the moment and help players cope with it after it has occurred. Seeking social support and pursuing positivity are well-established coping approaches for people who have been discriminated against, harassed, bullied, or otherwise mistreated. Integrated tools that provide access to these coping strategies within game play itself may be an effective strategy to help combat the harmful effects of in-game toxicity, making social gaming a safer and more accessible space for all players.

2 BACKGROUND

We contextualize our paper in existing literature on toxic behaviour in games and mechanisms for support.

2.1 The Problem of Toxicity

Harassment and other negative behaviour is present in many online social spaces, with multiplayer video games being no exception, as highlighted by long-standing academic discourse on what is acceptable in games [27, 87]. In-game features intended for positive communication or team strategizing, such as voice or text chat, unfortunately also provide a platform for nefarious or insulting messages, with members of marginalized communities often being targeted by toxic players most often [7, 35, 47]. Additionally, unlike most other online spaces, toxic behaviour in games is not limited to verbal harassment. The interactions between in-game characters within the game space can facilitate negative in-game griefing behaviour such as spatial intrusion [30], where a toxic player violates another player's digital space, potentially by performing lewd actions

or blocking the targeted player from moving freely through the game map. Other forms of non-verbal grief play can include power imposition or greed play, in which toxic players exploit game mechanics to take advantage of other players for their own gain [30]. Ultimately, in all cases these actions serve to disrupt other players' game enjoyment, performance, or both.

The pervasive occurrence of such negative behaviour in online games is widely recognized by both players and game developers. Recent surveys suggest that 83% of adult gamers experienced harassment in online multiplayer games [4]. Despite its prevalence, toxicity remains both difficult to define and hard to mitigate. Both of these factors are likely influenced by the wide range of behaviours that players can perceive as toxic, which is further complicated by the normalization of such behaviours by many players [10]. Moreover, toxicity may even be fostered by the communities' and developers' behaviours, e.g., by popular community figures being bad role models (cf. [48]). Intentionality may also play a role in the somewhat nebulous nature of toxicity: if a player performs badly or unintentionally disrupts the game for another player, they may be viewed as toxic or reported by other players even though they did not intend to cause harm [30, 45]. However, since toxicity can lead to decreased mood [77, 90], lower game performance [57], disruptive effects on day-to-day life [4], or players quitting the game completely [31, 34], it remains a critical area for developers and researchers to address.

2.2 Addressing Toxic Behaviour in Games

In an effort to address toxicity in games, researchers have studied the characteristics and motivations of toxic players. Factors such as player experience [81], playing in competitive game modes [34], and whether or not a player has been the target of toxicity in the past [44] have all been linked to increased toxic behaviour. Players are also more likely to encounter toxicity from teammates rather than their opponents, especially when teams are losing or have high skill disparity between players [54, 81]. Another major factor that contributes to toxicity is the online disinhibition effect [44, 90], in which online anonymity can engender negative behaviour [83]. Toxic online disinhibition has also been linked to a decreased perception of the severity of toxicity, which may contribute to its normalization in games [10]. While the occurrence of harassment and other toxic behaviour is not without precedent in other offline and online spaces [17, 68, 93], the normalization of toxic behaviour in games is still an important issue to address, as it affects existing players and also acts as a barrier to newcomers.

In addition to investigating the characteristics of toxic players, researchers have also studied the ways in which targeted players cope with toxic experiences during play. Many players employ strategies sequentially, especially when toxic behaviour persists, with the most common sequence being endure/ignore, mute/block (if possible), and finally concealing their online identity or switching games [62]. It is important to note that coping strategies are influenced by players' intersectional identities—for example, Black men and Hispanic or white women employ strategies that remove their perception of toxic players, such as muting, blocking, or reporting, whereas mostly women modify their digital self through strategies such as hiding, modifying usernames, and not engaging in game chat [62]. Fox and Tang [31] identified similar responses to toxic behaviour among female gamers, including gender masking, avoidance, seeking help, denial, and self blame. With the exception of seeking help, many of these strategies point to the normalized atmosphere of toxicity in games, in which players may feel as though encountering toxicity is inevitable, and either ignore the problem, blame themselves, or hide their identities to preemptively avoid toxicity.

Of course, the ways that players deal with toxicity depend on the tools offered in games. Most competitive multiplayer games offer some form of reporting tool that aims to flag perpetrators for punishment, though the utility of the tool can vary greatly [65]. For example, *League of Legends (LoL)* [72] players can only report at the end of matches or during the champion select phase at the

beginning, whereas *Valorant* [73] players can report at any time, even from their match history. Similarly, games may offer blocking or muting tools to help players avoid further exposure to toxicity after it occurs. Some games even use reputation systems to reward well-behaved players; for example, Riot uses their honor system to determine who is eligible to receive end-of-season rewards (e.g., cosmetic borders, icons, and banners), providing an incentive for players to behave. Prior work argued that moderation strategies allow for control over spaces [13], suggesting that individual actions, such as muting, may be useful to gain a sense of control. In general though, most games focus on systems to identify and punish toxic players, which is not always effective. First, it can be challenging to moderate player behaviour, as evidenced by prior approaches such as player-based governance in the Tribunal system used in LoL [46], which was ultimately retired but did provide agency [10, 71] and power in governance [46] for players. Another issue with reporting is that players do not always use flags as intended to identify toxic behaviour, and instead either do not report players at all, or report non-toxic players who they believe are playing poorly [45]. Players may not report toxic players because they believe such behaviour is acceptable or normal for games [10]. Players (and users of other online spaces) have also indicated that they want existing report systems to be more transparent, as report feedback in games (if given at all) is typically insufficient [12, 42, 45, 50, 93]. Pohjanen [65] also found that 51.3% of players overall believed the existing tools in games were not enough to deal with toxicity, though specific individual games showed more favourable results. As such, while reporting systems may be beneficial, they are limited because they require good feedback and transparency, and constitute only reactive action, which may not help players who already experienced toxicity.

Game developers themselves have noted some of the challenges that come with moderating toxic behaviour in games. These challenges include unclear boundaries between what is acceptable and unacceptable player behaviour, constraints induced by maintaining game functionality, risks to studio reputation, revenue, or the safety of employees, lack of knowledge or support in making governing decisions, as well as the idea that completely eliminating toxic behaviour is impossible and unreasonable [82]. However, perceived organizational responsiveness has been shown to influence player retention [31], which indicates that players appreciate when game companies attempt to combat toxicity within their communities. To that end, it is useful to explore whether tools that focus primarily on supporting targets of toxicity in games, rather than tools that solely focus on identifying and penalizing perpetrators, can be used to better manage the effects of in-game toxicity. Further, we know that toxic behaviour is directly detrimental to players—but current disciplinary-focused approaches do not immediately help those who are targeted. Alongside penalising toxic players, there exists an obvious and important opportunity to also implement methods that *support* targeted players.

2.3 Social Support

In contrast to systems that primarily punish the perpetrators of harmful behaviours, targets may instead desire and turn to social support for help. People may seek out support in both online and offline spaces, and for a variety of reasons beyond dealing with harassment, including to help cope with bullying, physical or mental health conditions, loneliness, and addictions. While there are many online resources that can provide information about these issues, the sheer number of options can be time consuming to parse, and adolescents in particular may find it difficult to filter through and understand the available information on their own [39]. Therefore, in many cases, direct social support is a more desirable option. This support can originate either online or offline, and may be professionally facilitated (e.g., Alcoholics Anonymous, Kids Help Phone) or more organic and informal (e.g., talking to a friend, help from bystanders). Typically, social support is categorized into the following forms: *informational support*, in which one individual helps another to understand a

stressful event better and/or help them find the right resources and coping strategies to deal with it; *instrumental support*, which involves acquiring tangible assistance such as services, financial aid, or goods; and *emotional support*, which involves providing reassurance or warmth towards another person to help show that they are valued and that others care about them [86].

Researchers have found social support to be both desirable and beneficial in many contexts, from communities about mental health topics [22, 38] to students facing bullying in real life or online [5, 19]. For example, Rothon et al. [74] found that social support from friends and family helped mitigate the negative effects of bullying on educational achievement, though social support alone was not sufficient to prevent mental health difficulties induced by bullying. Other studies have shown that social support can help reduce anxiety during times of stress [86], and Choudhury and Kiciman [23] found that Reddit users who received comments containing esteem or network support were less likely to express suicidal ideation in the future. In online communities about mental health topics, those who receive positive social support are more likely to provide supportive comments for others [22], suggesting that social support can flourish in the right environments. Social support may even benefit physical wellness and longevity within the general population [11, 51, 86]. However, it is important to note that meta-reviews on the benefits of social support reveal methodological flaws and definitional inconsistencies in many existing studies, which make them difficult to compare and replicate [38, 91]. Therefore, while social support interventions can indeed be useful overall, it is not always clear which methods work best for a given situation [38]. Additionally, people may respond differently to the same support strategies based on external factors such as gender [51] or experience within the support community [63]. This points to the importance of understanding how a specific community views support strategies at an early stage of the design process, so that the implemented support systems can best match the desires of the community.

In the context of online games, social support can be fostered through the gain of social capital in group play, particularly if players are motivated to engage with teammates or clans frequently [64, 69, 89]. Players who form friendships with others in the game may turn to them for help completing in-game activities, or for advice and emotional support regarding offline issues [40, 60]. Some games even have specific systems to connect new players with more experienced ones, such as Final Fantasy XIV's novice network. However, support is not always offered equally to all gamers—women, in particular, report being ignored and not given assistance when needed, directly insulted or brushed off when discussing experiences of harassment, and receiving no bystander help from other players, sometimes not even from other women [55]. In general, bystanders may be unlikely to publicly support women in games due to self-protection strategies and/or the norms of the game community overall [9]. Thus, genuine social support from other players may not be enough on its own or may need to be directly encouraged through game design in order to benefit all players. Yet, most games have limited social or other support mechanisms, particularly when it comes to dealing with toxicity. This leaves women and other targeted groups with little in-game support options when they are faced with toxic behaviour and harassment. Further, relying solely on naturally occurring *social* support might not be sufficient because it relies on third parties who might not be aware, available, or willing. As such, in-game approaches that provide support when needed or on-demand could be beneficial for players. To address this lack of universal support in games, it is useful to investigate support tools that have shown promise in other online spaces—these tools may provide a starting point for design considerations regarding support tools in online games.

2.4 Online Support Tools

There are many online tools to support targets of harassment, bullying, racism, or abuse.

A review of such existing approaches may be beneficial to inform the design of in-game support tools for players. For example, a suitable place to get support is provided by hotlines (e.g., Games and

Online Harassment Hotline [33]) and resource pages (e.g., [41]). However, while these approaches can be useful to provide support for those who experience toxicity, they may be less suitable for integration into games, because players have to actively engage with them. Thus, players may need to interrupt gameplay in order to search for and use these tools, which may not be desirable. In particular, externally hosted resources are not very accessible for most players. As noted by researchers studying support tools in other contexts, easy access and clarity are important factors to consider when providing resources: In the context of mental health resources, Williams et al. [95] showed that BIPOC (Black, Indigenous, and People of Colour) university students wanted resources to be more accessible, and desired increased clarity in terms of where and how to access services. In online games, reporting and honour systems (e.g., Overwatch's endorsement system or the honor system in League of Legends) that are only accessible at the end of long matches may result in players forgetting to report someone who was toxic in the first few minutes of the game. Finally, a tool that is seamlessly integrated within the game client itself—in contrast to tools that require external interaction, e.g., a hotline—may enjoy greater uptake, owing to ease-of-access, visibility, and implicit developer endorsement. Other approaches like validation tools may be useful to confirm the presence of toxicity or other negative behaviour, e.g., like with the tool 'Unmochon', which validates screenshots of harassment so that targets of harassment are able to prove that the photos are real [84]. Yet, such concepts may be challenging to implement for game contexts, because toxicity is ill-defined: what might be perceived as toxic to one person will not necessarily be perceived the same way by another player, human moderator, or a machine learning algorithm [32].

However, our review identified four support strategies that seemed suitable as modified approaches for in-game support tools.

Social Support. Social support tools focus on providing social support to users who may not have organic support systems in place, essentially facilitating the support that one might normally receive from friends or peers. Existing tools include HeartMob, in which users can request emotional or instrumental support such as positive messages or help with reporting [37], TrollBusters, which supports female journalists by posting positive messages [28], and even AI chatbots intended to support good mental health, combat loneliness, or act in an everyday supportive capacity for users [6, 39, 85]. Other research has suggested that sending users positive messages might help mitigate the effects of harassment they experience in online spaces [42, 93]. In studies that used participatory design to design theoretical tools to help mitigate the effects of racism or cyberbullying, participants identified wanting social support [5, 88]. Social support often also facilitates positivity and mood improvement.

Positivity and Mood Improvement. Part of the negative effects of toxicity is negative mood. Thus, improving a player's mood through positivity may be beneficial to deal with the aftermath of experiencing toxicity. Social support mechanisms often involve positivity, such as positive messages [28], which may help mitigate the effects of harassment experienced in online spaces [42, 93]. Positivity may help distract from or deal with negative experiences [5, 88]. For instance, one study found that both psychological (anxiety) and physiological (i.e., heart rate, blood pressure) indicators of stress were reduced among university students prior to taking an exam after viewing images of quokkas [29]. Research by Nittono et al. [59] also found that viewing images of cute animals–particularly baby animals–promotes narrowed attentional focus and improved performance in both fine motor tasks and non-motor visual search tasks compared to viewing images deemed as less cute (i.e., adult animals). Additionally, it has been suggested that humour is an effective self-care tool that reduces stress and improves mood by providing perspective on our problems, relieving physical tension, and promoting resilience [96]. However, Samson and Gross [78] found that the

type of humour matters—in particular, positive (good-natured) humour is more effective at down-regulating negative and up-regulating positive emotions compared to negative (mean-spirited) humour. Thus, animal pictures and humour may be useful for improving mood after experiencing toxicity.

Burden Relief. Burden relief tools aim to relieve some of the burden targets of harassment may face when dealing with harassers, such as confrontation or reporting. To some extent, these tools overlap with social support tools, specifically in that they generate instrumental support such as help with content moderation or reporting. For example, HeartMob can be used to get assistance with reporting perpetrators of harassment [37], and the friendsourced email system described by Mahar et al. [52] prevents targets from directly facing or dealing with the negative content in emails sent by harassers. In other cases, people may desire help when it comes to educating others or calling people out on their behaviour. For example, in To et al.'s [88] research, participants suggested that they might appreciate technology that could call someone out on racist behaviour, or educate them on how to be less racist in the future. These tools can help minimize further confrontation with harassers, and allow targets of harassment to quickly deal with and remove themselves from negative situations.

Control. Allowing people more control over the ways they deal with harassers is common to many support systems, but certain tools like muting and blocking are direct examples of this control. Such tools are common on social media platforms as well as in online games, with players generally believing that silencing tools are somewhat effective against toxic players, and blocking tools are very effective [65]. Blocklists have also been shown to be effective in other online spaces—for example, Twitter users felt as though their overall experience on Twitter improved after they started using anti-abuse blocklists [42], and in general people are more likely to think banning users, removing content, and apologies are fair compared to other methods of dealing with harassment online [79]. Some games allow limited forms of blocklists, such as Overwatch's "Avoid as Teammate" or Dota 2's "Avoid Player" options, which are meant to give players more control over who they are matched with. However, players can only avoid a limited number of people, e.g., due to constraints on the matchmaking system, like in Overwatch [65]. Additionally, while strategies such as custom filtering and blocking may be useful for targets of online harassment, Vitak et al. [93] found that women rarely applied strategies to tailor the content and people they interacted with online, possibly because of the effort involved—thus, these systems may not always work for everyone.

2.5 Summary and Research Gap

Toxicity remains a problem in games, and current prevention and mitigation approaches fall short. While research suggests potential benefits of implementing support mechanisms for people targeted by toxicity in other online contexts, it is unclear whether these benefits transfer to game environments that are subject to other norms, constraints, and expectations.

2.6 Aim and Research Questions

In this paper, we address this gap through iterative design and development of in-game support tools and their evaluation in two studies. First, it is important to understand *what do players hope to gain from the use of in-game support tools?* (RQ1). To that end, we conducted Study 1, in which participants evaluated storyboard designs for in-game support tools using qualitative responses, which we distilled into themes about important outcomes. We developed six interactive prototypes of such tools that may facilitate these outcomes—four as iterations of concepts tested in Study 1 and two additional baseline tools. Then, we investigated whether *in-game support tools are effective at providing support for players targeted by toxicity* (RQ2) in Study 2, in which participants could

use these prototypes after exposure to toxicity. Further, we examined if there are *differences in effectiveness between women and men* (RQ3) to assess if women, who are disproportionately targets of toxicity, may benefit more. In addition to answering these research questions, our paper further provides an artefact contribution that may be useful for developers aiming to implement in-game support tools, by reporting on the iterative design and development of our tools in two studies.

3 IN-GAME SUPPORT TOOLS

Informed by online support tool concepts discussed in the related work section, we generated preliminary design concepts for support tools in games. We initially developed six in-game tools (see Table 1 and Figure 3) grounded in the concepts learned from our review of online support tools, aiming to provide control, social support, positivity, and burden relief (some tools used multiple design aims). Four of these tools aimed to generate positivity and mood relief through exposure to cute animal pictures (*Eyebleach Pictures*), supportive message from virtual characters (*Positive Voice Lines*) and other players (*Friendly Messages*), and adding humorous messages to the text chat (*Riddikulus*). Control over the situation and perpetrators was facilitated by muting and blocking (*Blocklist*) and altering a toxic player's messages (*Riddikulus*). Finally, *Send A Message* aimed to relieve the burden of dealing with a toxic player by mediation through the game. We designed the tools in parallel with game-specific design considerations, prioritising the minimization of game disruptions and the avoidance of potential exploitations. These tools were iteratively revised through discussion among the research team and via feedback from internal pilots and implemented in storyboards for a final set of six tool designs. In Study 1, we evaluated these six tools, resulting in four of them—*Blocklist*, *Eyebleach Pictures*, *Send A Message*, and *Friendly Messages*—being further developed, refined, and evaluated based on participant feedback in Study 2 together with two additional baseline tools (*No Tool* and *Report*).

4 STUDY 1

We conducted a primarily qualitative study to assess what outcomes players desired from in-game support tools (RQ1) and to collect feedback to improve the initial concepts of the six tools that we had designed. Both studies in this paper were conducted under ethical approval received from the Behavioral Research Ethics Board at the University of Saskatchewan.

4.1 Methods

Procedure. The study was conducted online using Amazon Mechanical Turk, a crowdsourcing platform which can be used to recruit participants for HCI studies [43] by ensuring validity of responses [53, 61, 80]. As we aimed to recruit participants with multiplayer gaming experience, we conducted a prescreen and contacted those participants who identified as gamers to a moderate degree (>30) on a scale from 1 (= *"not at all"*) to 100 (= *"gamer"*), frequently played multiplayer games (>30) on a scale from 1 (= *"play alone"*) to 100 (= *"play with others"*), and reported playing multiplayer games regularly (*'Every day'* or *'A few times per week'*). Participants were paid $6 USD for their time, and the study took approximately 30 minutes to complete. After participants provided informed consent, they completed questionnaires on demographics, gaming experience, and the characteristics of the people they play with. Following these initial questionnaires, participants viewed the storyboards for the six tool designs in a within-subjects design (order balanced with a Latin square). For each tool design, participants were shown a storyboard outlining the use of the tool (see Figure 1 for an example), and were instructed to imagine that someone was toxic towards them in a general manner (see Figure 1 upper right) in a game and that they had access to the tool shown in the provided storyboard. With this prompt in mind, participants answered the following open ended questions about the tool: *"How would you feel when you used such a tool?"*, *"What are the*

Tool	Description	Design Aim	Study 1	Study 2
Blocklist	Players can mute and block toxic players. Blocked players will be flagged and automatically muted in future matches. Additionally, players can add tags to indicate why they blocked another player.	Control	Yes	Yes
Eyebleach Pictures	Players can open a small popup window that displays cute animal pictures.	Social Support, Positivity	Yes	Yes
Friendly Messages	Players can anonymously request to receive positive messages from other players in the game. Tags can also be added to the request, so the message sender can provide a more personal response.	Social Support, Positivity	Yes	Yes
Send A Message	Players can send an anonymous, system mediated notification to a toxic player notifying them that their behaviour is wrong. Tags can be added to the message to indicate the nature of the toxic behaviour.	Burden Relief	Yes	Yes
Positive Voice Lines	By pressing a key, players can hear a positive, supportive voice line from their in-game character.	Social Support, Positivity	Yes	No
Riddikulus	Players can transform the chat messages of a toxic player into silly phrases. The rest of the players in the match will still see the messages as normal.	Control, Social Support, Positivity	Yes	No
Report	Players can report a toxic player, similar as in many contemporary games.	Baseline	No	Yes
No Tool	Players do not have a dedicated option to deal with toxicity.	Baseline	No	Yes

Table 1. Descriptions of the tools, their corresponding design aims, and whether they were used in Study 1 and/or Study 2.

reasons you would want to use this tool?", *"What are the reasons you would not want to use this tool?"*, *"Are there any problems that could occur when using this tool?"*, and *"Do you have any suggestions to improve the design of this tool?"*. After reviewing all six storyboards, participants completed a final questionnaire about their impressions of the tools overall, and indicated which tools they liked, which tools they did not like, and why. Finally, participants were debriefed, thanked, and given information on how to claim their remuneration for completing the study.

Participants. In total, 40 participants completed the study. Data cleaning was performed, with participant responses analyzed to detect bots, non-diligent responses, and participants who displayed an obvious misunderstanding of tool designs. Additionally, one participant was excluded for having no experience playing multiplayer games. After exclusions, we were left with a final sample of 31 participants (men = 19, women = 12) between the ages of 19 and 66 ($M = 34.4$, $Mdn = 33.0$, $SD = 8.4$). Participants reported playing games either a few times per week ($n = 15$) or every day ($n = 16$). Participants also indicated the frequency with which they played alone vs. played with others on a 0–100 point scale—most played with others at least a moderate amount of time ($M = 53.2$, $Min = 4$, $Max = 88$, $SD = 27.2$). Additionally, participants were asked to self-identify as a gamer on a scale between 0–100, with most identifying as gamers ($M = 78.2$, $Min = 24$, $Max = 100$, $SD = 19.6$). As we were interested in how participants viewed our tools in response to toxic behaviour in games, we also asked participants (using a 0–100 slider) how often they

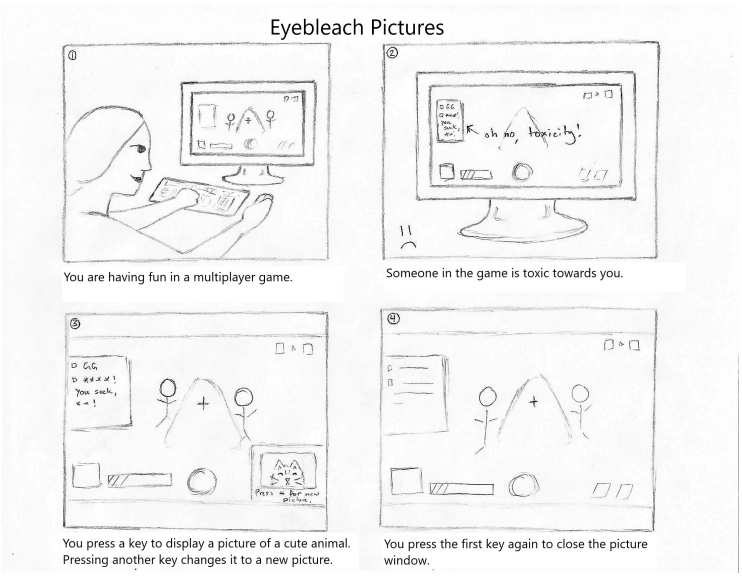

Fig. 1. *Eyebleach Pictures* storyboard used in Study 1.

played with friends ($M = 49.9$, $Min = 1$, $Max = 98$, $SD = 29.6$) and strangers ($M = 50.1$, $Min = 4$, $Max = 93$, $SD = 30.5$).

Data Analysis. We conducted an inductive thematic analysis to generate codes for each tool design following Braun and Clarke [16]. After an initial pass was completed to get familiar with the data, the first author generated preliminary codes for each open-ended question in the study. Codes were generated from both latent and semantic meaning in participant responses. These codes were then reviewed and iterated upon by the first and last authors. The final set of codes were organized by the nine open-ended questions from the study, and data was labelled at the sentence level. Sentences were allowed to map to multiple codes, and some of the same codes appeared across multiple questions. As we intended to iteratively adjust the tool designs based on the results from this study, the codes and tool rankings served as an initial insight into how participants felt about the tool designs at this early stage. These codes and insights for future iterations were discussed amongst all authors.

4.2 Results

Results comprise tool rankings and analysis of qualitative responses.

Tool Rankings. In the final questionnaire, participants were asked to select their most and least preferred tools. Participants were able to select as many or as few tools as they wished for both questions. *Blocklist*, *Send a Message*, and *Ridikkulus* appeared most often in the preferred list, with *Blocklist* outperforming the other tools. *Positive Voice Lines*, *Eyebleach Kitten*, and *Friendly Messages* appeared most often in the least preferred list, though every tool appeared at least once as a preferred and a disliked tool design (see Figure 2).

Qualitative Responses: Outcomes and Reasons to Use. Participants' responses to the question *"How would you feel when you used such a tool?"* revealed different potential outcomes, which we can summarize with a *"feeling in control"* theme. Participants discussed feelings of power, control,

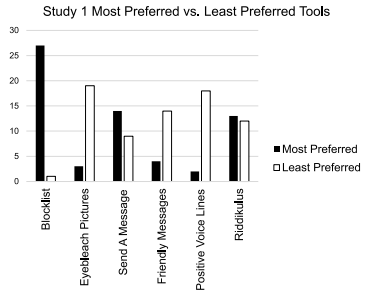

Fig. 2. Counts for selection as most and least preferred tools in Study 1.

protection, and tool effectiveness. They highlighted the importance of being in control: *"[Having access to these tools] would make me feel like I had some control over the toxic situations that arise while playing"*, and the importance of dealing with toxicity directly: *"[The tools I prefer] offer direct methods to deal with toxic people"*. Accordingly, they reacted positively to tools that placed them in control and directly dealt with toxic behaviour, such as the *Blocklist* tool.

On the other hand, the tools that focused mostly on social support or positivity (e.g., *Eyebleach Pictures* and *Positive Voice Lines*) were ranked less preferably favourably than those with direct control (e.g., *Blocklist*) and even considered silly (e.g., *"They all seem silly and superfluous."*), which may imply that they were not liked or considered effective. Despite this, participants' responses suggested that they desired support, mood improvement, a positive distraction from toxicity, and general positive emotions, which we can summarize with a *"positive emotions and mood improvement"* theme. For example, participants stated that such tools would *"... bring a smile to my face for a short while"* and that *"I think it would be cute and good for when you feel down on yourself"*. In stark contrast to those who claimed that social support and positivity tools may be superfluous, some participants indicated that such tools were their preferred options. Thus, even though they were ranked less favourably, tools that aim for positivity and social support may also provide desirable outcomes.

Qualitative Responses: Potential Shortcomings of Tools. We analyzed responses to assess potential shortcomings of tools or reasons why players would not want to use them. Negative codes included perceived gameplay disruption (*"this interrupts my gaming experience"*), tools not addressing the root problem (*"It doesn't solve the issue of the toxic player still being there"*), lack of mood improvement, potential to escalate the situation (*"This could escalate the situation by confronting the person."*), or a belief that the tools were stupid, silly, or childish (e.g., *"The premise is a bit childish"*). Thus, responses suggested that participants considered tools aiming only for positivity as simply not suited for them (e.g., *"This kind of thing ruins immersion in a game for me, personally."*.

4.3 Implications, Design Iteration, and Implementation of Interactive Prototypes

Following the results from Study 1, we concluded that tools may be received quite differently by different players, but they may still provide outcomes that may be beneficial for players and described as desirable by the participants. Regarding RQ1, the participants described two overarching themes as desirable outcomes with *feelings of control* and *positive emotions and mood improvement*. As the next step, we wanted to iterate on the designs and conduct a second study, in which we evaluated effects on participants after both exposure to toxicity and tool use. We chose four of the six prototypes as candidates for Study 2, with two aiming to provide feelings of control (*Blocklist* and

Send A Message) and two aiming for positive emotions (*Eyebleach Pictures* and *Friendly Messages*). We chose not to use *Positive Voice Lines* and *Riddikulus* because we wanted to keep the number of conditions manageable after adding two baseline conditions. Even though they seemingly may have provided desirable outcomes, these two tools were chosen for exclusion because of feasibility of implementation in interactive prototypes—e.g., displaying images vs implementing adjusted audio for voice lines—and because they either had multiple instead of singular design aims (*Riddikulus*) or had conceptual overlaps with other tools like *Eyebleach Pictures* and *Friendly Messages* (*Positive Voice Lines*).

To improve the design concepts, we analyzed responses to the question *"Do you have any suggestions to improve the design of this tool?"*. Participants indicated they wanted tools to have more options, take additional actions (such as muting or reporting), be less disruptive, include automation or AI responses, provide the ability to see non-toxic messages, keep it fresh, and have participation incentives.

Following this feedback, we iterated on the four tools that we wanted to further explore in Study 2. We created a list of specific changes participants requested for each design, and narrowed the scope of possible improvements by eliminating proposed changes that combined the functionality of different tools (e.g., adding a mute feature to the *Eyebleach Pictures* tool), keeping only those changes that would improve the main functionality of each tool. Improvements to these tool designs were relatively minor. *Blocklist* was updated to include an edit window, which allowed participants to remove, add notes, or unblock players after 48 hours rather than leaving them permanently blocked. *Eyebleach Pictures* was updated to include more varied pictures, including photos of nature and other baby animals, which participants could chose from in their settings. *Send A Message* was updated so that participants could optionally send the message after the match was finished in response to anonymity/confrontation concerns, and *Friendly Messages* had no iterations from the initial design.

Interactive Prototypes. Following these iterations, we used refined tool designs and created interactive prototypes for four of the six initial tools: *Blocklist*, *Send A Message*, *Friendly Messages*, and *Eyebleach Pictures*, which aligned with a 2x2 design space in which tools were control-focused or emotion-focused, and internal or external (see Table 2). In addition, we implemented two baseline prototypes for the subsequent study, i.e., *No Tool*, which featured a simple menu, and *Report*, which allowed participants to mute and report a toxic player. See Table 1 and Figure 3 for overviews of these prototypes.

	Control-Focused	**Emotion-Focused**
Internal	*Blocklist*	*Eyebleach Pictures*
External	*Send A Message*	*Friendly Messages*

Table 2. Tool alignment with a 2x2 design space consisting of outcome aim (control vs emotion) and locus (internal vs external).

All prototypes were implemented as interactive overlays in Javascript to be displayed in the context of a toxic interaction. For that purpose, they were embedded after the video of toxic in game-behaviour in an *Overwatch* match. To align with the concept of providing support *after* exposure to toxicity, we overlaid the prototypes on a still image of the end of the video clip. The images were taken from the last frame of a video clip showing a toxic encounter, which participants viewed before interacting with the prototypes. Participants interacted with the prototypes similar to how they would interact with such a tool in a game after being exposed to toxicity, e.g., by being

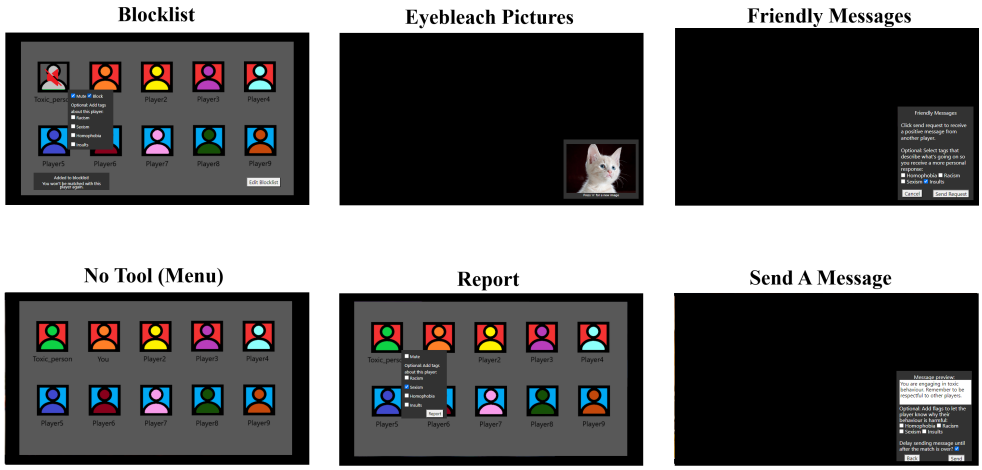

Fig. 3. All prototypes used in Study 2. Each prototype appears as an interactive popup overlaid over the given Overwatch scene.

able to view pictures aimed to induce positive affect in the *Eyebleach Pictures* condition with the option to change pictures and hide the view using keyboard presses.

5 STUDY 2

In this study, we wanted to investigate whether the tools were effective (RQ2) at facilitating mood improvement or feelings of control, i.e., the outcomes considered important in Study 1. We hypothesized that even the tools that were less liked, such as *Eyebleach Pictures* and *Friendly Messages*, can still be effective in providing support, because they *can* provide mood improvements. We also compared effectiveness for women and men (RQ3).

5.1 Methods

We conducted a primarily quantitative study in which we evaluated the effectiveness of six tool prototypes after participants had watched a short video clip containing toxic behaviour in a multiplayer game. We used four prototypes of support tools and two baseline tools, resulting in six conditions, as outlined earlier and visible in Table 1 and Figure 3.

Video Clips. To emulate a toxic encounter in an online game, participants viewed short video clips displaying toxic behaviour in an *Overwatch* match, and were asked to imagine that any toxicity present in the videos was directed towards them. As each video clip acted as a precursor to one of the six interactive prototypes, six different 70s video clips were selected from a larger set of toxic *Overwatch* videos used in previous studies of toxic behaviour in games. This larger set of videos, compiled from *Overwatch* streamers on *Twitch* by the fourth author, had been independently rated by at least two raters per clip, and each two minute subsection assigned a toxicity score between 0–100. The clips chosen for this study were all of a comparable toxicity score, with the lowest score being 63.8 and the highest being 80.0. The content of the clips was similar as well—all six clips contained verbal insults over voice chat, though some featured more general anger or insults to performance, whereas one clip contained specifically sexist comments.

Procedure. The study was conducted online using Prolific (www.prolific.co), which we chose for the second study because it provided additional functionality like built-in prescreening. Respondents were limited to those who self-identified on Prolific as gamers. The study took approximately 30 minutes to complete, and participants were paid $6 USD for their time. After having time to read the study information, including the information that the task involved viewing toxic behaviours and providing informed consent, participants completed a demographics questionnaire, as well as questionnaires on their gaming experience, *Overwatch* experience, the characteristics of the people they play with, and an 8-item scale used in prior work [26], resulting in a single score ranging between 1 and 7 that describes perceived toxicity of their gaming communities. Following these questionnaires, participants were shown one of the video clips of toxic behaviour in *Overwatch*, were asked to complete the Self-Assessment Manikin (SAM) [15] and a Visual Analogous Scale (VAS) for perceived stress [94] to assess experience after exposure to toxicity and before tool usage, and were then given an opportunity to use one of the six tools. After trying out the tool, participants were asked to complete SAM and VAS questionnaires again to assess experience after using the tools. We used these scales to measure whether the tools were able to increase positive affect (SAM Valence), reduce arousal (SAM Arousal), provide feelings of being in control (SAM Dominance), and to reduce perceived stress (VAS Stress), which we considered negative effects that arose from the exposure to toxicity. The six video clips were presented in the same order to all participants, but tool order was balanced across participants using a Latin square. Finally, participants were debriefed, thanked, and given instructions on how to claim their remuneration for completing the study. The debrief included information to get help in case of distress and emotional aftercare in form of baby animal pictures.

Participants. 150 participants completed the study, and after exclusions, we were left with 132 participants (men = 93, women = 31, non-binary = 6, prefer not to disclose = 2). Participants were excluded for high response variance, non-diligent responses, non-English responses, never playing online games or multiplayer games, and clear misunderstandings of the questions, which we assessed based on their responses [18, 56] and discussing candidates for removal. The remaining 132 participants were between 18 and 38 years old ($M = 22.5$, $Mdn = 21.0$, $SD = 4.2$), and the majority reported playing online games often, with 91 playing every day, 38 playing a few times per week, and 3 playing a few times per month. The majority of participants also self-identified as gamers ($M = 79.7$, $Min = 17$, $Max = 100$, $SD = 20.1$), and reported playing multiplayer games with both friends ($M = 61.4$, $Min = 1$, $Max = 100$, $SD = 30.4$) and strangers ($M = 51.9$, $Min = 1$, $Max = 100$, $SD = 31.5$). As the video clips of toxic behaviour featured the game *Overwatch* [14], participants also recorded their familiarity ($M = 34.5$, $Min = 1$, $Max = 100$, $SD = 29.5$) and skill ($M = 21.4$, $Min = 1$, $Max = 100$, $SD = 24.6$) in *Overwatch*. Participants' responses on the toxicity scale indicated that they had ample experience with toxicity ($M = 3.88$, $Mdn = 4.125$, $SD = 1.29$).

5.2 Results

We analyzed the tools' effectiveness in reducing stress, increasing affect, decreasing arousal, and providing feelings of control and provide some insights into differences for men and women.

5.2.1 RQ2: Effectiveness. For our main analysis, we conducted repeated measures ANOVAs with two factors (time: pre vs post tool usage; and tool with six levels: *Blocklist*, *Eyebleach Pictures*, *Send A Message*, *Friendly Messages*, *Report*, *No Tool*) and the dependent variables stress (VAS), positive affect (SAM valence), arousal (SAM arousal), and feelings of control (SAM dominance). Greenhouse-Geisser correction was used when the assumption of sphericity was violated, indicated by adjusted degrees of freedom. We conducted post-hoc tests with Bonferroni correction and report effects for pre-post differences for each tool, which indicates their effect on stress, positive affect,

arousal, and feelings of control. Due to technical problems, responses on the SAM scales were not recorded for 21 participants, resulting in lower sample sizes for the valence, arousal, and dominance measures (n = 111). Figure 4 shows results, which can be summarized as follows: All four tools were effective at providing support for players (RQ2).

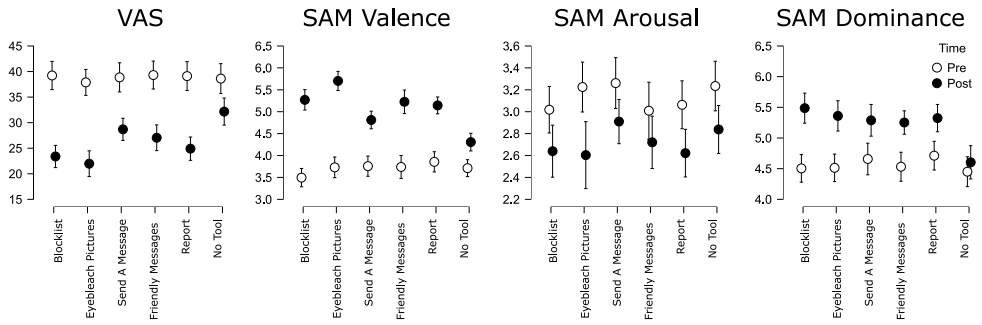

Fig. 4. Stress (VAS), positive affect (SAM Valence), arousal (SAM Arousal), and feelings of control (SAM Dominance) after exposure and before tool usage (Pre) vs after tool usage (Post) in Study 2. Error bars indicate 95% confidence intervals.

VAS Stress. The ANOVA showed significant main effects of tool ($F(4.152, 543.891) = 3.347$, $p = .009$, $\eta_p^2 = .025$), time ($F(1, 131) = 119.566$, $p < .001$, $\eta_p^2 = .477$), and their interaction ($F(5, 655) = 9.060$, $p < .001$, $\eta_p^2 = .065$). In all conditions, stress significantly decreased from pre-tool usage to post-tool usage: *Blocklist* (*mean difference* = 15.846, $t = 9.929$, $p < .001$), *Eyebleach Pictures* (*mean difference* = 15.908, $t = 9.968$, $p < .001$), *Send A Message* (*mean difference* = 10.167, $t = 6.371$, $p < .001$), *Friendly Messages* (*mean difference* = 12.272, $t = 7.690$, $p < .001$), *Report* (*mean difference* = 14.198, $t = 8.897$, $p < .001$), and even in the baseline condition *No Tool* (*mean difference* = 6.457, $t = 4.046$, $p = .004$).

SAM Valence. The ANOVA showed significant main effects for tool ($F(4.181, 459.855) = 8.316$, $p < 0.001$, $\eta_p^2 = .070$), time ($F(1, 110) = 180.452$, $p < 0.001$, $\eta_p^2 = .621$), and their interaction ($F(4.661, 512.701) = 14.881$, $p < 0.001$, $\eta_p^2 = .119$). Positive affect significantly increased from pre-tool to post-tool usage in all conditions: *Blocklist* (*mean difference* = -1.775, $t = -11.395$, $p < .001$), *Eyebleach Pictures* (*mean difference* = -1.973, $t = -12.667$, $p < .001$), *Send A Message* (*mean difference* = -1.054, $t = -6.767$, $p < .001$), *Friendly Messages* (*mean difference* = -1.486, $t = -9.544$, $p < .001$), *Report* (*mean difference* = -1.288, $t = -8.271$, $p < .001$), and *No Tool* (*mean difference* = -0.595, $t = -3.818$, $p = .010$).

SAM Arousal. The ANOVA did not show significant main effects for tool ($F(4.498, 494.818) = 1.493$, $p = 0.197$, $\eta_p^2 = .013$) or the interaction between tool and time ($F(4.224, 464.610) = 0.661$, $p = 0.627$, $\eta_p^2 = .006$), but did show significant effects for time only ($F(1, 110) = 18.663$, $p < 0.001$, $\eta_p^2 = .145$). In all conditions, arousal decreased from pre tool to post tool usage, though this change was only significant for the *Eyebleach Pictures* tool: *Blocklist* (*mean difference* = 0.378, $t = 2.367$, $p = 1.000$), *Eyebleach Pictures* (*mean difference* = 0.622, $t = 3.888$, $p = .008$), *Send A Message* (*mean difference* = 0.351, $t = 2.198$, $p = 1.000$), *Friendly Messages* (*mean difference* = 0.288, $t = 1.803$, $p = 1.000$), *Report* (*mean difference* = 0.441, $t = 2.761$, $p = .393$), and *No Tool* (*mean difference* = 0.396, $t = 2.480$, $p = .889$).

SAM Dominance. The ANOVA showed significant main effects for tool ($F(4.485, 493.299) = 4.044$, $p = 0.002$, $\eta_p^2 = .035$), time ($F(2.803, 144.027) = 51.387$, $p < 0.001$, $\eta_p^2 = .318$), and their interaction ($F(4.619, 508.061) = 3.921$, $p = 0.002$, $\eta_p^2 = .034$). In all conditions except the baseline *No Tool*, feelings of control increased significantly from pre tool to post tool usage: *Blocklist* (*mean difference* $= -0.982$, $t = -6.147$, $p < .001$), *Eyebleach Pictures* (*mean difference* $= -0.847$, $t = -5.447$, $p < .001$), *Send A Message* (*mean difference* $= -0.631$, $t = -3.948$, $p = .006$), *Friendly Messages* (*mean difference* $= -0.721$, $t = -4.512$, $p < .001$), *Report* (*mean difference* $= -0.613$, $t = -3.835$, $p = .009$). In the baseline condition *No Tool*, the difference in feelings of control was not significant (*mean difference* $= -0.153$, $t = -0.959$, $p = 1.000$).

5.2.2 RQ3: Effectiveness for Women and Men. To provide insights into the effectiveness of the tools for groups that are more affected by toxicity, we conducted an additional analysis comparing effects for women and men (RQ3). We conducted an exploratory analysis including participants who reported non-binary gender or preferred not to disclose their gender. With this analysis, we were not able to determine meaningful conclusions due to high variance in responses arising through low group sample sizes (see Limitations for a discussion about this limitation). As a result, we compared effects for women and men only.

To make interpretation more manageable, which was complicated by the addition of gender, we calculated differences between before (pre) and after (post) tool usage for all dependent variables as measures for reduction in stress (VAS), increase in positive affect (SAM Valence), reduction in arousal (SAM Arousal), and increase in feelings of control (SAM Dominance). Then, we conducted repeated measures ANOVAs with tool as within-subjects and reported gender (woman vs man) as between-subjects factors.

The ANOVA with VAS rating differences showed significant main effects for tool ($F(4.620, 549.812) = 7.633$, $p < 0.001$, $\eta_p^2 = .061$), gender ($F(1, 119) = 11.282$, $p < 0.001$, $\eta_p^2 = .087$), but not for their interaction ($F(4.620, 549.812) = 0.434$, $p = 0.811$, $\eta_p^2 = .004$). For SAM Valence, the ANOVA showed significant main effects for tool ($F(5, 520) = 11.329$, $p < 0.001$, $\eta_p^2 = .098$), gender ($F(1, 104) = 11.879$, $p < 0.001$, $\eta_p^2 = .103$), but not for their interaction ($F(5, 520) = 1.571$, $p = 0.166$, $\eta_p^2 = .015$). Regarding SAM Arousal, there was a significant main effect for gender ($F(1, 104) = 8.481$, $p = 0.004$, $\eta_p^2 = .075$), but not for tool ($F(4.363, 453.736) = 0.720$, $p = 0.590$, $\eta_p^2 = .007$) or the interaction ($F(4.363, 453.736) = 1.054$, $p = 0.381$, $\eta_p^2 = .010$). There was a significant main effect on SAM Dominance for tool ($F(5, 520) = 4.168$, $p = 0.001$, $\eta_p^2 = .039$) and for gender ($F(1, 104) = 6.320$, $p = 0.013$, $\eta_p^2 = .057$), but not for the interaction ($F(5, 520) = 0.761$, $p = 0.578$, $\eta_p^2 = .007$).

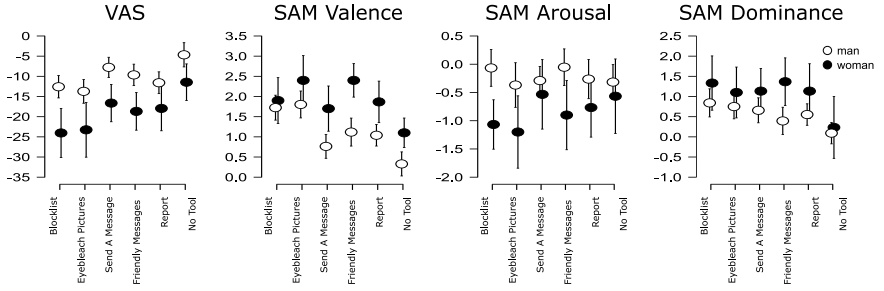

Fig. 5. Reduction in stress (VAS) and arousal (SAM Arousal), and increase in positive affect (SAM Valence) and feelings of control (SAM Dominance) for women and men in Study 2. Error bars indicate 95% confidence intervals.

Figure 5 shows these effects and highlights that women experienced a higher reduction in stress and arousal, and higher increases in positive affect and feelings of control than men.

6 DISCUSSION

We summarize results and situate them in prior work about toxicity in games.

6.1 Summary of Findings

RQ1: What do players hope to gain from the use of in-game support tools? We found that players have quite varied opinions about what they consider important for in-game support tools. The rankings and responses suggested that participants leaned towards tools that they knew (*Blocklist* and *Report*). It is possible that these features may have been more positively received because they are tools most players are already familiar with in games. Thus, these tools already have real-world support backing their efficacy in reducing toxicity. Further, both of these tools represent an actionable solution to the problem: *Blocklist* removes the toxic actor entirely from the player pool the player is exposed to by automatically muting their communication, and *Report* represents a potential path to disciplining the offending player by restricting participation in the game community. These tools provide control, an aspect that was reflected by data about what players desired. This *"feelings of control"* theme suggested that players wanted tools that provided control about the situation and the toxic player. Secondly, there was a *"positive emotions and mood improvement"* theme, indicating that players wanted tools that provided such positive emotions, effectively addressing the negative mood after exposure to toxicity, despite participants indicating a lower preference for these tools compared to the control based approaches.

RQ2: Are in-game support tools effective at providing support for players targeted by toxicity? Overall, our results suggest that all the implemented tools were effective at providing player support. The support approaches *Blocklist*, *Eyebleach Pictures*, *Send A Message*, and *Friendly Messages*, as well as the baseline *Report* all provided positive affect and feelings of control, while decreasing stress. As such, they all facilitated those outcomes that were considered important, suggesting that they are effective at providing support.

RQ3: Are there differences in effectiveness between women and men? Finally, our findings suggest that the benefit provided by the support tools was higher for women than men: after exposure to the support tools, women reported higher increases in valence and feelings of control, and decreases in arousal and stress.

6.2 What Players Want

Participants reported they were most likely to use *Blocklist* and *Report*, i.e., tools that most players are already familiar with in games. Tools that focus solely on emotionally *supporting* the targeted player were shown to be effective—especially as a form of purely supportive aftercare—but comparably perceived as less desirable. Yet, they represent a focused solution for a more intangible problem: the targeted player's emotional wellbeing. One possible explanation is that in gaming communities especially, wherein toxicity is normalised [10], a concession that such support is needed or desired may be regarded with cynicism: for example, as suggested by one participant, that such tools are 'infantilizing'.

Player skepticism of tools that focus more explicitly on emotionally supporting the player—rather than on disciplining toxic offenders, or on moderating and controlling interactions—represents a salient problem, especially in light of how effective these tools were in Study 2. We propose four possible explanations to be evaluated in future work that may contribute to this skepticism.

First, we argue that players are unfamiliar with emotion-focused support tools in games—and that unfamiliarity breeds contempt. In most major commercial titles, tools available to playerbases focus on the *toxic player* rather than the *target*: on punishing the offender (via reporting), or on removing the offender from the player's gameplay (via muting or blocking). There does not exist an analogue for providing players with emotional aftercare following exposure to toxicity. Such aftercare is instead sought externally and informally by the player (for example, by messaging a friend to vent, or taking a break from the game), rather than through integrated game systems or tools. As such, aversion to the formal integration of such tools may be bolstered by a lack of evidence of their success. Further, as players already lack conviction in developers' employment of popular existing tools (such as responding to player reports, or in behaviour flagging systems) [45, 50], they may likewise be cynical about developers' ability to meaningfully integrate emotion-focused support tools.

Second, participants may have been dubious of tools like *Eyebleach Pictures*, which may require appropriate integration with a game in a manner that is less immediately off-putting to a skeptical playerbase, which may be achieved through cogency with game design or aesthetic. While cute animals may be suitable for a Pokémon-style game, titles like *Overwatch* may instead wish to employ endearing art or gifs of their character roster. Many games, such as *Overwatch* and *Apex Legends*, already integrate this art as sprays or holograms (game art players can 'paste' on surfaces in the game world). Another analog is that of the 'emote wheel' in *League of Legends*, which allows players to temporarily flash cute character or champion emojis at other players that reflect specific emotions—often used as tension relief after an erroneous play or in celebration of a victorious play. Developers may also foster community engagement, potentially improving player receptiveness to the tool, by using moderated fan art. However, such emojis may also be used in a toxic manner, e.g., a sarcastic "gg" emote, suggesting the need for careful implementation and future work.

Third, it is possible that as toxicity is normalised [10], players are encouraged to 'have a thick skin' and are ashamed for having negative emotional responses to toxicity. This is evidenced by participant responses in both studies that identified using emotion-focused tools as 'pathetic', 'weak', or 'cowardly'. Thus, some support tools may face challenges due to how players perceive game communities. For example, *Send A Message* operates under the assumption that receiving a message about their anti-social behaviour may prompt toxic players to reflect and self-correct. Players using *Send A Message* may feel less assured about the tool's capacity to prompt behavioural change—especially as toxicity is widely regarded as unassailable in competitive game communities [10]. There may even be concerns that receiving such a message may instead gratify toxic players, thus encouraging or magnifying antisocial behaviour.

Fourth, we suggest that a general aversion to adopting support tools in game may be due to larger social factors at play. The tendency to avoid using support systems despite their effectiveness has been documented for years in mental health related literature [8, 67, 70]. This reluctance to seek care may, in part, be due to the general stigma surrounding mental health and help-seeking behaviours [3, 8]. For instance Affleck et al. [3] found that men underutilise mental health services in comparison to women due to perceived social stigma associated with notions of masculinity (i.e., men must 'tough it out'). As most of the playerbase within competitive multiplayer video games are predominately male [58], members of the community may refuse support in order to look 'tough' in front of their male counterparts, and to avoid being perceived as weak—displaying such weakness may increase the likelihood of being targeted for toxicity. A systematic review by Radez et al. [67] revealed four common barriers to seeking help for mental health problems among children and adolescents—individual factors, social factors, perceptions of the client-professional relationship, and systemic and structural barriers and facilitators—all of which may translate to competitive online gaming environments as well. In online gaming communities, players may have limited

individual knowledge of the resources and support networks available to them, as well as negative pre-conceived notions of those who seek help. Additionally, social factors (e.g., social stigma associated with seeking support) interact with individual factors (e.g., fear of embarrassment) to influence perceptions of help-seeking behaviours. Players may also have a general distrust towards professionals in the gaming industry (e.g., developers), and desire transparency in communication. Lastly, structural barriers and facilitators may influence company action and decision-making. As such, integrating support tools or providing solutions to sanitize toxicity within communities may not be at the forefront of company priorities. However, one solution to remedy this problem may be through reducing the stigma associated with seeking help and improving knowledge of available support systems in both games and real-world contexts. In-game tools may be a suitable solution to offer support for players that is not immediately visible to other players and therefore potentially stigmatized.

6.3 Benefits for Women

Our findings suggest that the benefit provided by the support tools was higher for women than men. That such support tools are more effective for women than men may be influenced by the fact that, as a group, women experience higher exposure to toxicity (and especially gendered or sexualised toxicity) [31, 47]—and feel less supported by gaming communities [9, 55]. For women, these tools may represent a cathartic surrogate for absent community support in the face of increased and targeted toxicity. Further, we understand that women are more likely to seek out therapeutic interventions than men [3], as women are less constrained by the aforementioned social stigmas of masculinity. This comparatively increased openness to therapeutic intervention may also be reflected in the additional benefits our support tools provided to women, as female participants may have been less influenced by negative preconceptions of such support (e.g., that they are 'weak' for using them).

While we provide tools to support players in-game, it is also important to address broader cultural issues within gaming communities and workplace environments. We know that women [31], members of the 2SLGBTQ+ community [7], and players of colour [35], are disproportionately the target of harassment in gaming. The recent lawsuits at Activision Blizzard [25] and Riot Games [21] concerning gender discrimination and sexual harassment further perpetuates this notion. If game companies are unable to evidence support of vulnerable populations under their own employment, it follows that players may doubt their ability to support vulnerable populations within their player communities. As such, it is important to address the effects of these structural issues on the playerbase and gaming communities at large. We argue that prioritizing inclusivity and support within *all* gaming contexts is a crucial step towards cleansing toxicity embedded within broader social structures.

6.4 Integration of Tools in Games

Our results demonstrate variation in player preference for different types of tools, suggesting that players may be more or less inclined to use (or benefit from) specific tools—for example, despite the efficacy of *Eyebleach Pictures*, players that regard the tool as 'infantilising' may be more resistant to its use. However, this should not be taken as an argument against specific methods, but instead as evidence suggesting that a variety of tools should be made available to the player base. An ideal approach would be the provision of a suite of tools that are integrated in the game client. Potentially, there could be a multiple choice support system that offers a selection of tools when players experience toxicity—or, alternatively, players may pre-select the tools available to them in these events. However, we also argue that, based on the results of Study 2, player skepticism of more emotion-focused tools—such as *Eyebleach Pictures* and *Friendly Messages*—is

largely unfounded; as such, extra efforts towards promoting the use of emotion-focused tools may be justified. Transparency around the efficacy of emotion-focused tools may encourage player uptake, per player desire for transparency in reporting tools. Likewise, a trial period wherein players have the full suite of support tools available to them prior to selecting their preferred tools may also allow players to experiment with, and potentially note the benefits of, tools they may have otherwise been resistant to.

Finally, effects for *No Tool* suggests that time is also an important factor in processing toxicity. Without the intervention of a tool, players will experience decreases in stress and increases in valence after the passing of time, implying that they have processed the negative event. Essentially, a negative state after experiencing toxicity will rectify over time. However, it did not significantly increase feelings of control. Compared to the other tools, its effects at providing support were lower—indicating that any intervention is better than no intervention for supporting players after a toxic event.

7 LIMITATIONS AND FUTURE WORK

We compared the effectiveness of the support tools for women and men, and while we conducted an exploratory analysis of the effects for participants who reported non-binary or preferred not to disclose gender, sample sizes precluded us from drawing meaningful conclusions. As a result, our findings in regards to the gender effects are limited to women and men. We recommend future work that explicitly investigates the effectiveness of such tools for more diverse populations of players, including non-binary players, but also players of otherwise marginalized populations (e.g., players of colour [35, 36], and generally members of the 2SLGBTQ+ community [7]) who are also disproportionately targeted by toxicity. It is possible that these players might also benefit more from such support tools, given their increased exposure to and experience with toxicity.

We had participants interact with our tools after toxicity occurred and over a freeze frame of the match video—in a real game, players may feel the tools are too disruptive for their gameplay, potentially impacting their overall effectiveness at providing support. This is especially important for esports and other competitive players, as toxicity support tools that are too disruptive will not be used even if they are effective at supporting players. It is also unknown how effective these tools would be in situations where toxicity persists during a long match, or occurs repeatedly over a number of days—over time, these tools may not provide as much support as they do for the short instances of toxicity we studied here. We also did not consider different levels or types of toxicity in our studies. Future work should explore whether the level or type (e.g., sexism, racism) of toxicity influences how players respond to different types of support tools.

Our findings are limited to the specific sample of participants. First, we used crowdsourcing platforms for recruitment, which can provide valid results. However, findings should be replicated in other samples. Second, we chose to recruit gamers for our studies because we aimed to evaluate tools that benefit those who actively play multiplayer games. However, this means that findings may not transfer to other players, such as those who have already left multiplayer games, e.g., because of their experiences with toxicity. Thus, future work is necessary to evaluate the tools with other samples, including players who are less entrenched with current gaming culture and communities.

Finally, our findings suggest that players have varied preferences for tools, highlighting the potential of an in-game suite that offers access to multiple tools. At this stage, we have not designed and evaluated such a multi-tool suite. We require more work on different design variants, including aspects like effects of autonomy in choosing support mechanisms. Similarly, we also have not yet studied when and how people have access to these tools, e.g., through automated methods that detect toxicity or a button that is always present and offers support on-demand. Effects of such

integration, such as potential harm through false positives or having a button as a ubiquitous reminder of a toxic culture, are still unknown at this stage, warranting further research.

8 CONCLUSION

Toxicity remains a problem in online games with various negative ramifications, including harm for those targeted by toxic behaviour. Existing approaches and tools have not solved the problem and further do not immediately help the targets of toxicity who are left alone after exposure. Inspired by approaches in other online and offline spaces, we proposed in-game support tools to support targets of toxicity. We iteratively designed and evaluated multiple tools. Following insights from Study 1 that suggested the importance of feelings of control and positive emotions, we ultimately ended up with four approaches along a emotion-focused/control-focused and internal/external design space (*Blocklist*, *Eyebleach Pictures*, *Friendly Messages*, *Send A Message*). The evaluation of these tools in Study 2 provided insights into their effectiveness. We found that feeling in control was important for players who accordingly subjectively rated tools highly that provided such control (i.e., *Blocklist* and common *Report* methods). Yet, despite being disliked by some, emotion-focused approaches were also effective at providing support; the *Eyebleach Pictures* tool provided the highest benefits on multiple measures. Further, we found that the tools' effectiveness was higher for women—a group of players disproportionately targeted by game-based toxicity. Generally, we found that in-game support methods work. All four tools were effective, suggesting their value for inclusion in games. Different preferences suggest that players may benefit from more variety in toxicity support tools that both explicitly address toxicity and help players cope with it. To that end, these tools—or others like them—may be part of the ongoing effort to combat the harmful effects of in-game toxicity, leading to safer gaming spaces in which all players can receive the myriad benefits of social gaming.

ACKNOWLEDGMENTS

We thank NSERC and SWaGUR for funding, members of the Interaction Lab for feedback, and our participants.

REFERENCES

[1] Leigh Achterbosch, Charlynn Miller, and Peter Vamplew. 2017. A taxonomy of griefer type by motivation in massively multiplayer online role-playing games. *Behaviour & Information Technology* 36, 8 (Aug. 2017), 846–860. https://doi.org/10.1080/0144929X.2017.1306109

[2] 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

[3] William Affleck, Victoria Carmichael, and Rob Whitley. 2018. Men's Mental Health: Social Determinants and Implications for Services. *The Canadian Journal of Psychiatry* 63, 9 (Sept. 2018), 581–589. https://doi.org/10.1177/0706743718762388

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

[5] Zahra Ashktorab and Jessica Vitak. 2016. Designing Cyberbullying Mitigation and Prevention Solutions through Participatory Design With Teenagers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, San Jose, CA, USA, 3895–3905. https://doi.org/10.1145/2858036.2858548

[6] Petter Bae Bae Brandtzæg, Marita Skjuve, Kim Kristoffer Kristoffer Dysthe, and Asbjørn Følstad. 2021. When the Social Becomes Non-Human: Young People's Perception of Social Support in Chatbots. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, Yokohama, Japan, 1–13. https://doi.org/10.1145/3411764.3445318

[7] Mary Elizabeth Ballard and Kelly Marie Welch. 2017. Virtual Warfare: Cyberbullying and Cyber-Victimization in MMOG Play. *Games and Culture* 12, 5 (July 2017), 466–491. https://doi.org/10.1177/1555412015592473

[8] Makilim N Baptista and Cristian Zanon. 2017. Why not seek therapy? The role of stigma and psychological symptoms in college students. *Paidéia* 27, 67 (2017), 76–83.

[9] Sara Bastiaensens, Heidi Vandebosch, Karolien Poels, Katrien Van Cleemput, Ann DeSmet, and Ilse De Bourdeaudhuij. 2015. 'Can I afford to help?' How affordances of communication modalities guide bystanders' helping intentions towards harassment on social network sites. *Behaviour & Information Technology* 34, 4 (April 2015), 425–435. https://doi.org/10.1080/0144929X.2014.983979

[10] 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 the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, Yokohama, Japan, 1–15. https://doi.org/10.1145/3411764.3445157

[11] Lisa F. Berkman and S. Leonard Syme. 1979. Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. *American Journal of Epidemiology* 109, 2 (Feb. 1979), 186–204. https://doi.org/10.1093/oxfordjournals.aje.a112674

[12] Lindsay Blackwell, Jill Dimond, Sarita Schoenebeck, and Cliff Lampe. 2017. Classification and Its Consequences for Online Harassment: Design Insights from HeartMob. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (Dec. 2017), 24:1–24:19. https://doi.org/10.1145/3134659

[13] 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.

[14] Blizzard Entertainment. 2015. *Overwatch*. Game [PC]. Irvine, USA.

[15] Margaret M Bradley and Peter J Lang. 1994. Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential. *Journal of Behavior Therapy and Experimental Psychiatry* 25, 1 (1994), 49–59.

[16] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[17] Melissa L. Breger, Margery J. Holman, and Michelle D. Guerrero. 2019. Re-Norming Sport for Inclusivity: How the Sport Community Has the Potential to Change a Toxic Culture of Harassment and Abuse. *Journal of Clinical Sport Psychology* 13, 2 (2019), 274–289. https://doi.org/10.1123/jcsp.2019-0004

[18] Erin M Buchanan and John E Scofield. 2018. Methods to detect low quality data and its implication for psychological research. *Behavior research methods* 50, 6 (2018), 2586–2596.

[19] Maria Carmen Cabrera, Elisa Larrañaga, and Santiago Yubero. 2020. The Role of Emotions, Moral Disengagement and Gender in Supporting Victims of Bullying. *Education Sciences* 10, 12 (Dec. 2020), 365. https://doi.org/10.3390/educsci10120365

[20] Charles S Carver, Michael F Scheier, and Jagdish K Weintraub. 1989. Assessing coping strategies: a theoretically based approach. *Journal of personality and social psychology* 56, 2 (1989), 267.

[21] Cecilia D'Anastasio, kotaku.com. 2018. Inside The Culture Of Sexism At Riot Games. https://kotaku.com/inside-the-culture-of-sexism-at-riot-games-1828165483 Accessed: 2021-09-09.

[22] Yixin Chen and Yang Xu. 2021. Social Support is Contagious: Exploring the Effect of Social Support in Online Mental Health Communities. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21)*. Association for Computing Machinery, 1–6. https://doi.org/10.1145/3411763.3451644

[23] Munmun De Choudhury and Emre Kiciman. 2017. The Language of Social Support in Social Media and Its Effect on Suicidal Ideation Risk. *Proceedings of the International AAAI Conference on Web and Social Media* 11, 1 (May 2017). https://ojs.aaai.org/index.php/ICWSM/article/view/14891

[24] Christine Cook, Juliette Schaafsma, and Marjolijn Antheunis. 2017. Under the bridge: An in-depth examination of online trolling in the gaming context. *New Media & Society* 20, 9 (Dec. 2017), 3323–3340. https://doi.org/10.1177/1461444817748578

[25] Daniel Van Boom, cnet.com. 2021. California expands lawsuit against Activision-Blizzard: What you need to know. https://www.cnet.com/news/california-expands-lawsuit-against-activision-blizzard-what-you-need-to-know/ Accessed: 2021-09-09.

[26] 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

[27] Julian Dibbell. 1994. A rape in cyberspace or how an evil clown, a Haitian trickster spirit, two wizards, and a cast of dozens turned a database into a society. *Ann. Surv. Am. L.* (1994), 471.

[28] Michelle Ferrier and Nisha Garud-Patkar. 2018. TrollBusters: Fighting Online Harassment of Women Journalists. In *Mediating Misogyny: Gender, Technology, and Harassment*, Jacqueline Ryan Vickery and Tracy Everbach (Eds.). Springer International Publishing, Cham, 311–332. https://doi.org/10.1007/978-3-319-72917-6_16

[29] First Stop Singapore. 2020. Study: Quokkas can be good for your health. https://www.firststopsingapore.com/en/study-quokkas-can-be-good-for-your-health/ Accessed: 2022-01-31.

[30] Chek Yang Foo and Elina M. I. Koivisto. 2004. Defining grief play in MMORPGs: player and developer perceptions. In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology (ACE '04)*. Association for Computing Machinery, Singapore, 245–250. https://doi.org/10.1145/1067343.1067375

[31] 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 (Aug. 2017), 1290–1307. https://doi.org/10.1177/1461444816635778

[32] 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. https://doi.org/10.1145/3313831.3376446

[33] Games and Online Harassment Hotline. 2020. Games and Online Harassment Hotline. https://gameshotline.org Accessed: 2021-09-08.

[34] 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

[35] K. L. Gray. 2012. Deviant bodies, stigmatized identities, and racist acts: examining the experiences of African-American gamers in Xbox Live. *New Review of Hypermedia and Multimedia* 18, 4 (Dec. 2012), 261–276. https://doi.org/10.1080/13614568.2012.746740

[36] K. L. Gray. 2014. *Race, Gender, and Deviance in Xbox Live: Theoretical Perspectives from the Virtual Margins*. Routledge, New York, NY, USA. https://www.sciencedirect.com/book/9780323296496/race-gender-and-deviance-in-xbox-live

[37] HeartMob by Hollaback! 2021. HeartMob. iheartmob.org Accessed: 2021-09-08.

[38] Brenda E Hogan, Wolfgang Linden, and Bahman Najarian. 2002. Social support interventions: Do they work? *Clinical Psychology Review* 22, 3 (April 2002), 381–440. https://doi.org/10.1016/S0272-7358(01)00102-7

[39] Camilla Gudmundsen Høiland, Asbjørn Følstad, and Amela Karahasanovic. 2020. Hi, can I help? Exploring how to design a mental health chatbot for youths. *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments* 16, 2 (2020), 139–169. https://doi.org/10.17011/ht/urn.202008245640

[40] Omar Bradley II Ictech. 2021. *Virtual Worlds: Social Interactions Among Online Gamers Through Voice Chat*. Doctoral Dissertation. Louisiana State University, Baton Rouge, LA, USA.

[41] International Game Developers Association. 2021. Harassment. https://igda.org/resourcelibrary/harassment/ Accessed: 2021-09-08.

[42] Shagun Jhaver. 2020. *Identifying opportunities to improve content moderation*. PhD diss. Georgia Institute of Technology, Atlanta, Georgia. https://smartech.gatech.edu/handle/1853/62779

[43] Aniket Kittur, Ed H. Chi, and Bongwon Suh. 2008. Crowdsourcing User Studies with Mechanical Turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 453–456. https://doi.org/10.1145/1357054.1357127

[44] Bastian Kordyaka, Katharina Jahn, and Bjoern Niehaves. 2020. Towards a unified theory of toxic behavior in video games. *Internet Research* 30, 4 (April 2020), 1081–1102. https://doi.org/10.1108/INTR-08-2019-0343

[45] 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*. ACM, 1–12. https://doi.org/10.1145/3411764.3445279

[46] Yubo Kou and Bonnie A Nardi. 2014. Governance in League of Legends: A hybrid system. (2014).

[47] Jeffrey H. Kuznekoff and Lindsey M. Rose. 2013. Communication in multiplayer gaming: Examining player responses to gender cues. *New Media & Society* 15, 4 (June 2013), 541–556. https://doi.org/10.1177/1461444812458271

[48] Marc Lajeunesse. 2018. *"It taught me to hate them all.": Toxicity through DOTA 2's Players, Systems, and Media Dispositive*. Ph.D. Dissertation. Concordia University.

[49] Noam Lapidot-Lefler and Azy Barak. 2012. Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in Human Behavior* 28, 2 (March 2012), 434–443. https://doi.org/10.1016/j.chb.2011.10.014

[50] Matthew Lapolla. 2020. *Tackling Toxicity: Identifying and Addressing Toxic Behavior in Online Video Games*. Master's thesis. Seton Hall University, South Orange, NJ. https://scholarship.shu.edu/dissertations/2798

[51] Cara J. Hale Ma, James W. Hannum Phd, and Dorothy L. Espelage Phd. 2005. Social Support and Physical Health: The Importance of Belonging. *Journal of American College Health* 53, 6 (May 2005), 276–284. https://doi.org/10.3200/JACH.53.6.276-284

[52] Kaitlin Mahar, Amy X. Zhang, and David Karger. 2018. Squadbox: A Tool to Combat Email Harassment Using Friendsourced Moderation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, Montréal, QC, Canada, 1–13. https://doi.org/10.1145/3173574.3174160

[53] Winter Mason and Siddharth Suri. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods* 44, 1 (2012), 1–23.
[54] Dave McLean, Frank Waddell, and James Ivory. 2020. Toxic Teammates or Obscene Opponents? Influences of Cooperation and Competition on Hostility between Teammates and Opponents in an Online Game. *Journal For Virtual Worlds Research* 13, 1 (March 2020). https://doi.org/10.4101/jvwr.v13i1.7334
[55] Lavinia McLean and Mark D. Griffiths. 2019. Female Gamers' Experience of Online Harassment and Social Support in Online Gaming: A Qualitative Study. *International Journal of Mental Health and Addiction* 17, 4 (Aug. 2019), 970–994. https://doi.org/10.1007/s11469-018-9962-0
[56] Adam W Meade and S Bartholomew Craig. 2012. Identifying careless responses in survey data. *Psychological methods* 17, 3 (2012), 437.
[57] C. K. Monge and T. C. O'Brien. 2021. Effects of individual toxic behavior on team performance in League of Legends. *Media Psychology* (Jan. 2021), 1–23. https://doi.org/10.1080/15213269.2020.1868322
[58] Newzoo.com. 2017. Male and Female Gamers: How Their Similarities and Differences Shape the Games Market. https://newzoo.com/insights/articles/male-and-female-gamers-how-their-similarities-and-differences-shape-the-games-market/ Accessed: 2021-09-08.
[59] Hiroshi Nittono, Michiko Fukushima, Akihiro Yano, and Hiroki Moriya. 2012. The power of kawaii: Viewing cute images promotes a careful behavior and narrows attentional focus. *PloS one* 7, 9 (2012), e46362.
[60] Erin L. O'Connor, Huon Longman, Katherine M. White, and Patricia L. Obst. 2015. Sense of Community, Social Identity and Social Support Among Players of Massively Multiplayer Online Games (MMOGs): A Qualitative Analysis. *Journal of Community & Applied Social Psychology* 25, 6 (2015), 459–473. https://doi.org/10.1002/casp.2224
[61] Gabriele Paolacci and Jesse Chandler. 2014. Inside the Turk: Understanding Mechanical Turk as a Participant Pool. *Current Directions in Psychological Science* 23, 3 (2014), 184–188. https://doi.org/10.1177/0963721414531598
[62] Cale J. Passmore and Regan L. Mandryk. 2020. A Taxonomy of Coping Strategies and Discriminatory Stressors in Digital Gaming. *Frontiers in Computer Science* 2 (2020). https://doi.org/10.3389/fcomp.2020.00040
[63] Zhenhui Peng, Xiaojuan Ma, Diyi Yang, Ka Wing Tsang, and Qingyu Guo. 2021. Effects of Support-Seekers' Community Knowledge on Their Expressed Satisfaction with the Received Comments in Mental Health Communities. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, Yokohama, Japan, 1–12. https://doi.org/10.1145/3411764.3445446
[64] Ryan Perry, Anders Drachen, Allison Kearney, Simone Kriglstein, Lennart E. Nacke, Rafet Sifa, Guenter Wallner, and Daniel Johnson. 2018. Online-only friends, real-life friends or strangers? Differential associations with passion and social capital in video game play. *Computers in Human Behavior* 79 (Feb. 2018), 202–210. https://doi.org/10.1016/j.chb.2017.10.032
[65] Antti Pohjanen. 2018. *Report please! : a survey on players' perceptions towards the tools for fighting toxic behavior in competitive online multiplayer video games.* Master's thesis. University of Jyväskylä, Jyväskylä. https://jyx.jyu.fi/handle/123456789/59353
[66] Shaun Prescott. 2017. Overwatch's Jeff Kaplan on toxic behavior: 'the community needs to take a deep look inwards'. https://www.pcgamer.com/overwatchs-jeff-kaplan-on-toxic-behavior-the-community-needs-to-take-a-deep-look-inwards/.
[67] Jessica Radez, Tessa Reardon, Cathy Creswell, Peter J Lawrence, Georgina Evdoka-Burton, and Polly Waite. 2021. Why do children and adolescents (not) seek and access professional help for their mental health problems? A systematic review of quantitative and qualitative studies. *European Child & Adolescent Psychiatry* 30, 2 (2021), 183–211.
[68] D.W. Rainey and V. Granito. 2010. Normative rules for trash talk among college athletes: an exploratory study. *Journal of Sport Behavior* 33, 3 (2010), 276–294.
[69] Felix Reer and Nicole C. Krämer. 2019. Are online role-playing games more social than multiplayer first-person shooters? Investigating how online gamers' motivations and playing habits are related to social capital acquisition and social support. *Entertainment Computing* 29 (March 2019), 1–9. https://doi.org/10.1016/j.entcom.2018.10.002
[70] Debra J Rickwood, Frank P Deane, and Coralie J Wilson. 2007. When and how do young people seek professional help for mental health problems? *The Medical Journal of Australia* 187, 7 Suppl (2007), S35 – S39.
[71] Riot. 2019. Ask Riot: Will Tribunal Return? https://nexus.leagueoflegends.com/en-us/2018/08/ask-riot-will-tribunal-return
[72] Riot Games. 2009. *League of Legends.* Game. Los Angeles, USA.
[73] Riot Games. 2020. *Valorant.* Game [PC]. Los Angeles, USA.
[74] Catherine Rothon, Jenny Head, Emily Klineberg, and Stephen Stansfeld. 2011. Can social support protect bullied adolescents from adverse outcomes? A prospective study on the effects of bullying on the educational achievement and mental health of adolescents at secondary schools in East London. *Journal of Adolescence* 34, 3 (June 2011), 579–588. https://doi.org/10.1016/j.adolescence.2010.02.007

[75] Victoria L. Rubin and Sarah C. Camm. 2013. Deception in video games: examining varieties of griefing. *Online Information Review* 37, 3 (Jan. 2013), 369–387. https://doi.org/10.1108/OIR-10-2011-0181
[76] Maria Ruotsalainen and Usva Friman. 2018. "There Are No Women and They All Play Mercy": Understanding and Explaining (the Lack of) Women's Presence in Esports and Competitive Gaming. In *Proceedings of the 2018 International DiGRA Nordic Conference (DiGRA Nordic'18)*. 1–14.
[77] Teemu Saarinen. 2017. *Toxic Behavior in Online Games*. Master's thesis. University of Oulu, Oulu, Finland. http://jultika.oulu.fi/files/nbnfioulu-201706022379.pdf
[78] Andrea C Samson and James J Gross. 2012. Humour as emotion regulation: The differential consequences of negative versus positive humour. *Cognition & emotion* 26, 2 (2012), 375–384.
[79] Sarita Schoenebeck, Oliver L Haimson, and Lisa Nakamura. 2021. Drawing from justice theories to support targets of online harassment. *New Media & Society* 23, 5 (May 2021), 1278–1300. https://doi.org/10.1177/1461444820913122
[80] Danielle N. Shapiro, Jesse Chandler, and Pam A. Mueller. 2013. Using Mechanical Turk to Study Clinical Populations. *Clinical Psychological Science* 1, 2 (2013), 213–220. https://doi.org/10.1177/2167702612469015
[81] Cuihua Shen, Qiusi Sun, Taeyoung Kim, Grace Wolff, Rabindra Ratan, and Dmitri Williams. 2020. Viral vitriol: Predictors and contagion of online toxicity in World of Tanks. *Computers in Human Behavior* 108 (July 2020), 1–6. https://doi.org/10.1016/j.chb.2020.106343
[82] Lucy A. Sparrow, Martin Gibbs, and Michael Arnold. 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 (CHI '21)*. Association for Computing Machinery, Yokohama, Japan, 1–13. https://doi.org/10.1145/3411764.3445363
[83] John Suler. 2004. The Online Disinhibition Effect. *CyberPsychology & Behavior* 7, 3 (2004), 321–326. https://doi.org/10.1089/1094931041291295
[84] Sharifa Sultana, Mitrasree Deb, Ananya Bhattacharjee, Shaid Hasan, S.M.Raihanul Alam, Trishna Chakraborty, Prianka Roy, Samira Fairuz Ahmed, Aparna Moitra, M Ashraful Amin, A.K.M. Najmul Islam, and Syed Ishtiaque Ahmed. 2021. 'Unmochon': A Tool to Combat Online Sexual Harassment over Facebook Messenger. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, Yokohama, Japan, 1–18. https://doi.org/10.1145/3411764.3445154
[85] Vivian Ta, Caroline Griffith, Carolynn Boatfield, Xinyu Wang, Maria Civitello, Haley Bader, Esther DeCero, and Alexia Loggarakis. 2020. User Experiences of Social Support From Companion Chatbots in Everyday Contexts: Thematic Analysis. *Journal of Medical Internet Research* 22, 3 (March 2020), 1–10. https://doi.org/10.2196/16235
[86] Shelley E. Taylor. 2011. Social support: A review. In *The Oxford handbook of health psychology*. Oxford University Press, New York, NY, US, 189–214.
[87] T. L. Taylor. 2006. Does WoW Change Everything?: How a PvP Server, Multinational Player Base, and Surveillance Mod Scene Caused Me Pause. *Games and Culture* 1, 4 (2006), 318–337. https://doi.org/10.1177/1555412006292615
[88] Alexandra To, Hillary Carey, Geoff Kaufman, and Jessica Hammer. 2021. Reducing Uncertainty and Offering Comfort: Designing Technology for Coping with Interpersonal Racism. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, Yokohama, Japan, 1–17. https://doi.org/10.1145/3411764.3445590
[89] Sabine Trepte, Leonard Reinecke, and Keno Juechems. 2012. The social side of gaming: How playing online computer games creates online and offline social support. *Computers in Human Behavior* 28, 3 (May 2012), 832–839. https://doi.org/10.1016/j.chb.2011.12.003
[90] 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
[91] MC van Eenbergen, LV van de Poll-Franse, P Heine, and F Mols. 2017. The Impact of Participation in Online Cancer Communities on Patient Reported Outcomes: Systematic Review. *JMIR Cancer* 3, 2 (2017), e15. https://doi.org/10.2196/cancer.7312
[92] VictimConnect Resource Center. 2021. National Hotlines. https://victimconnect.org/resources/national-hotlines/ Accessed: 2021-09-08.
[93] Jessica Vitak, Kalyani Chadha, Linda Steiner, and Zahra Ashktorab. 2017. Identifying Women's Experiences With and Strategies for Mitigating Negative Effects of Online Harassment. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. Association for Computing Machinery, Portland, OR, USA, 1231–1245. https://doi.org/10.1145/2998181.2998337
[94] Bernadette Von Dawans, Urs Fischbacher, Clemens Kirschbaum, Ernst Fehr, and Markus Heinrichs. 2012. The social dimension of stress reactivity: acute stress increases prosocial behavior in humans. *Psychological science* 23, 6 (2012), 651–660.

[95] Lucretia Williams, Gillian Hayes, Gloria Washington, Rebecca W. Black, Candace Williams, Leah Clements, and Michael Allotey. 2021. Analysis Of Distance-Based Mental Health Support For Underrepresented University Students. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21)*. Association for Computing Machinery, 1–6. https://doi.org/10.1145/3411763.3451708

[96] Patty Wooten. 1996. Humor: an antidote for stress. *Holistic nursing practice* 10, 2 (1996), 49–56.

Received February 2022; revised June 2022; accepted July 2022