
Copyright © 2023 The Authors. Reproduced under a Creative Commons Attribution 4.0 International License.

For the purpose of open access, the author(s) has applied a Creative Commons Attribution license to any Accepted Manuscript version arising.

https://eprints.gla.ac.uk/294013/

Deposited on: 9 March 2023

Enlighten – Research publications by members of the University of Glasgow
https://eprints.gla.ac.uk
Big Buddy: A Simulated Embodied Moderating System to Mitigate Children’s Reaction to Provocative Situations within Social Virtual Reality

Cristina Fiani  
c.fiani.1@research.gla.ac.uk  
University of Glasgow  
UK

Robin Bretin  
r.bretin.1@research.gla.ac.uk  
University of Glasgow  
UK

Mark McGill  
Mark.McGill@glasgow.ac.uk  
University of Glasgow  
UK

Mohamed Khamis  
Mohamed.Khamis@glasgow.ac.uk  
University of Glasgow  
UK

ABSTRACT

The use of social Virtual Reality (VR) among children is increasing, but with it comes new forms of harassment that can be difficult for parents to monitor. To address this issue, we have developed "Big Buddy", a prototype AI-moderator that aims to safeguard children from potential harassment in social VR. We conducted a study in which 43 children (aged 8-16) participated in a simulated social VR classroom, with fictitious competitors disrupting their game. When Big Buddy intervened, the children reported feeling significantly less negative emotions and felt safer. This is the first study to empirically examine the use of an embodied AI-moderator in social VR from the perspective of children, and it provides important insights for designing AI-moderators in social VR.

CCS CONCEPTS

• Human-centered computing → Collaborative and social computing.

KEYWORDS

social virtual reality, children, online harassment, embodied moderating system, artificial moderator

ACM Reference Format:


1 INTRODUCTION

Social Virtual Reality (VR), which was originally intended for adults and older teenagers, has seen an increase of younger children under the age of 13 using the simulated social environment [27–29]. While social VR can create an innovative way of engaging and interacting with others due to unique embodiment and presence in VR with the illusion of “being there” [30], users can inflict virtual harm on others, leading to an increase of new forms of harassment and bullying [8, 27, 29, 40]. For example, children and adults have reported harassment ranging from name calling to virtual sexual harassment [27, 29]. Unfortunately existing mitigation features, such as blocking, personal space bubbles, muting, reporting players [1, 3] or trust systems to keep users safe from nuisance users [4], suffer from significant limitations. They place the responsibility of their application on potentially ill-equipped users including children or guardians (e.g., unfamiliar with the technology or not knowing the best approaches) [9, 18, 20, 22]. Moreover, there is a growing need for understanding the effectiveness of safety-enhancing technologies for children in social VR. While research has shown that human-based moderators can help establish norms for appropriate behaviours [8], research also showed users would lack trust due to possible personal and subjective biases of the moderators and the worry that they may be effective only in small-scale social VR environments [17].

In this study, we introduced Big Buddy, an artificial avatar as a Wizard of Oz (WOZ) AI-moderator prototype, to safeguard children from disruptive social VR. 43 children (8-16 years old) played a researcher-constructed VR interactive tower of blocks construction game. It involved fictitious competitors disrupting the participant’s game in a virtual classroom mimicking a social VR environment. We evaluated children’s emotions and perceptions towards Big Buddy and his interventions. In this paper, we aimed to answer the following questions:

RQ1 How is children’s emotional valence impacted in provocative situations by the presence of Big Buddy?

RQ2 How do children perceive the embodiment of the moderating system (Big Buddy) and the system of punishments? Will children feel safer and/or inhibited knowing there is an embodied moderator agent?

RQ3 How do children envision an AI-moderator in social VR? This paper provides new insights into HCI and child-computer interaction. We conduct the first study of a VR-based WOZ AI-moderator to improve children’s safety and comfort in social VR environments. We also gather data on design features for an AI-moderator that
balances safety with children’s agency and enjoyment. We propose future design directions for effective AI-moderators in social VR.

2 RELATED WORK

2.1 Child Perception Towards Sanctions and Unfairness in Social Disruptive Situations

Children can encounter social disruptive situations in reality (e.g., at school) or online [41]. As children begin to communicate with others they learn desirable and undesirable behaviours in social settings [24]. Recent studies evaluated children’s perception towards rewards, sanctions and unfairness for different psychological and educational applications such as reinforcement learning, effective parenting, discipline and social norms for appropriate behaviour [21, 23, 31, 41]. The design and methodologies of these studies are of interest to our experimental design and interventions.

2.1.1 Children’s Emotional Reaction and Self-Reflection in Simulated Provocative Situations. A study evaluated children’s aggressive behaviour using interactive VR scenarios [41]. Boys aged 8-13 were individually tested in a silent room. Results showed that VR scenarios involving peer provocation (participants were refused to join a game by two virtual peers and participants’ game was ruined by the virtual peer) led to more aggressive responses than neutral and instrumental gain scenarios (participants could choose to steal a block or ball from the virtual peer to obtain additional points or could win the game by sabotaging the virtual peer’s game). The study inspired our VR game as it was shown that the game was challenging enough but not too difficult for children aged 8-13, and demonstrated a provocative situation designed to provoke an emotional response.

Anger and frustration are negative emotions according to Russell’s complex [38] that can be elicited in tasks such as toy removal and were shown to orient children towards desirable goals or objects [21]. The study, involving 40 children aged 5-6, used self-reported emotions with the established Self-Assessment Manikin (SAM) [25] and a game where participants would compete with an unfamiliar player and would always lose. Results showed that anger is associated with attention biases towards rewards rather than punishment.

2.1.2 Perceptions Towards Sanctions and Authority. Perceptions of punishments and rewards for pupils’ behaviour have been investigated looking at the relative effectiveness of school-initiated rewards and punishments as perceived by children in primary school and parents via a survey (N_children = 49, N_parents = 64). Regarding punishments, children rated ‘information being sent home’, ‘teacher explaining what is wrong with their behaviour in front of the class’ and ‘being stopped from going on a school trip’ as top sanctions on this rank.

Prosocial behaviour of children is largely influenced by adult figures, authority and media. Several psychological studies looked at the effects of being watched, monitored and agency in parent-child and pupil-teacher relationships [19]. Children make a distinction in their perception of agency depending on the relationship context. A study showed children perceived the least agency with teachers and the most agency with peers [19]. Another study investigated children’s conception of authority from an individual and showed that it would depend on their status, the context and the domain of act depending on children’s age [39]. Younger children’s awareness of mothers and teachers authority was shown to be greater than that of police for instance [10, 39]. Therefore, there is the potential need of having a visible embodied moderator as an authority figure for children in social VR.

2.2 Bullying and Harassment in Social VR

Experiences of harassment and bullying in social VR have increased and are shown to be more intense than bullying on social media sites due to the embodiment and presence in VR [8]. Children and adults have reported harassment, from name calling to physical stalking [27, 29]. Current mediation tools (e.g., reporting and blocking) have been shown to be insufficient, have yet to establish user trust, and have led to feelings of unfair treatment - discouraging users from using them [27]. Interventions to protect children effectively from bullying are lacking and mitigation options that exist in digital media (e.g., Microsoft Family Safety, Apple Families, Google Family Group) now have (largely) yet to be transposed to social VR [6, 13, 14, 43]. Indeed, it is difficult for parents to act as a bystander and intervene as social VR requires head-worn devices that completely occlude reality, and do not support bystander awareness that could allow effective supervision [33]. Research has shown that real moderators can help establish norms for appropriate behaviours [8], with nuanced views towards human-based moderators [17] as users could lack trust due to possible personal and subjective biases of the moderators and the worry that they may be effective only in small-scale social VR environments [17]. Moreover, automated methods to detect cyberbullying have been used in social media, including via text classification to detect harassment keywords and Natural Language Processing (NLP) and have been shown to identify hate speech with high accuracy [7, 26, 36]. Automated detection could feasibly allow to detect harassment events in social VR in the future. We need to consider how we would integrate it in social VR, how it would intervene and how we would make the detection and intervention capability visible to improve the user experience. Therefore, we introduce a simulated embodied AI-moderating system and evaluate child perceptions towards it.

3 METHODS

Our study focuses on the design and perceptions of an embodied AI-moderation for social VR, evaluated in a social VR gaming experience with children. We simulated a social VR environment game with disruptive situations based on prior research [41] and implemented a WOZ prototype of an embodied AI-moderator, Big Buddy, who put in place interventions when a disruption occurred based on the class of punishments shown to be effective in prior work [12, 15, 31, 37]. We measured child reactions using SAM scales [25], and designed Likert scales and interview questions to measure child perceptions towards Big Buddy. The protocol was approved by our ethics committee.

3.1 Procedure

Children, accompanied by a parent or teacher, were given a presentation and were told they will be playing a virtual classroom
game with real players from around the world. To verify their understanding of the three SAM scales, they were asked how they were feeling in that moment and to give a score: from sad (1) to happy (5) (M = 4.4, SD = 0.7), calm (1) to angry (5) (M = 1.4, SD = 0.5) and scared/intimidated (1) to safe (5) (M = 4.5, SD = 0.7). Children used a Meta Quest 2 headset with an adjustable strap for more comfort. Participants were standing and were immersed in a virtual classroom with non-player virtual characters (e.g., peers sitting or talking in the background) to mimic a social VR environment.

The game consisted of a practice session and 4 rounds where they competed with each of the 4 players. The goal was to build a tower of 5 blocks as quickly as possible and ensure it remains stable until the time limit. Points were awarded to players who successfully met this goal. After each round, they had to fill in a form on a laptop, with the questions described in section 3.4. At the end of the game, each child was individually interviewed.

3.2 Experimental Design

The study follows a Within-Participant design with Big Buddy’s interventions as the independent variable of 4 levels (i.e., 4 rounds, each round with one level): (1) C1: [No BB] - Big Buddy is absent, no intervention is taken; (2) C2: [BB, Reset Points] - Big Buddy is present and intervenes: resets saboteur’s points to 0; (3) C3: [BB, Reset+Notify Parents] - Big Buddy is present and intervenes: resets saboteur’s points to 0 and notifies saboteur’s parents; (4) C4: [BB, Reset+Notify Parents+Exclusion] - Big Buddy is present and intervenes: resets saboteur’s points to 0, notifies saboteur’s parents and saboteur is excluded from the game.

Big Buddy announced these punishments and punishments were written in a bubble, ensuring that the child participant is aware of the actions taken. The punishments were based on the top three punishments of the perceived effective punishments in a school according to children and parents [31]. The four conditions were counter-balanced with Latin Square. However, we have the following randomisation: [C2,C3,C4,C1; N = 10], [C3,C2,C1,C4; N = 10], [C4,C1,C2,C3; N = 13], [C1,C4,C3,C2; N = 10] (due to omitting 5 participants from the analysis, see section 3.1). The provocative situation was the same throughout the game. Therefore, we randomised the number of games (2 to 4) for each round and the moment when the competitor would sabotage the game, randomising the fairness within each of the four conditions (fair win/loss, unfair loss) to reduce the effect of repetitiveness of sabotaging that could lead to fed up emotions or lack of game credibility.

3.3 Social VR Game and Avatars

Implementation

The simulated social VR game, used to conduct the experiment, was developed by the experimenter using Unity 3D. Its design was based on the game used to evaluate the aggressive behaviour in boys [41]. We simulated similar provocative situations to the latter study, where a virtual character sabotages the participant’s game, by destroying their built tower. We constructed avatars for the Big Buddy and the 4 competitors using the Ubiquity library [2] with animations and a voice generator. To stand out from the other avatars, Big Buddy was designed to be bigger and with a noticeable appearance. We designed it such that it leaned towards a robotic artificial moderator using a monotone agent voice to avoid adding variables (e.g., different voice tones and intonations etc.) and available differentiable visual look from the Ubiquity library (See Figure 1 e)). The game was first pilot tested with 8 adults before starting the experiment with children. A sample scenario in VR (from the participants’ perspective) that occurs in one of the rounds for C3 is shown in Figure 1.

3.4 Measures

Quantitative: At the end of each round of the game, children were asked to self-rate the emotions felt when the other player destroyed their tower, using the three 5-point SAM Likert scales [25], re-adapting the third one (from scared/intimidated to safe). If Big Buddy was present in the round, they had to give a score on three 4-point Likert scales (from ‘a little’ to ‘extremely’: 1) Seeing Big Buddy when playing; 2) Feeling seen by Big Buddy, 3) Big Buddy helped with a fair punishment). At the end of the game, during interviews, children were separately asked to give scores for preferred physical and social attributes, using 5-point Likert scales from -2 to 2 (e.g., authoritarian to liberal, non-humanised to humanised, visible to invisible, amateur to expert). The latter scales were designed based on items used to design a tutor social robot [34], to gain a better understanding on what children would want as a preferred embodied moderating system in social VR (See supplementary material).

Qualitative: At the end of the game, children were separately interviewed with questions around their perceptions of the bully and Big Buddy, their sense of safety, and potential customisation of Big Buddy (see interview questions in supplementary material).

3.5 Analysis

Quantitative: We analysed our quantitative data using R statistical tools. We conducted one-way ART ANOVAs (significant level at 0.05 and effect sizes: medium (between 0.06 and 0.14) and large (0.14 or higher) partial eta squared $\eta_p^2$ [32, 35]) using ARTool and post-hoc comparisons with Tukey adjustment [16, 42]. We also used K-means clustering analysis to analyse preferred physical and social attributes of Big Buddy [5]. The analysis and codes can be found: (link to be added, see supplementary material).

Qualitative: We used inductive thematic analyses techniques [11, 33] to analyse children’s perceptions from audio recorded interviews. After generating transcripts, a pair of researchers read and familiarised themselves with the data. The two researchers then created individual coding schemes independently using NVivo, line by line. The codes generated are words or short phrases that describe an idea. Then, we collaborated to consolidate the two coding schemes into one combine scheme, by collating or distinguishing between codes. We created a set of higher-level codes by bringing related codes. All authors collaborated in an iterative process to discuss, combine, and refine themes and features to generate a rich description.

3.6 Participants

The children (8-16 years old M = 11.49 [SD = 2.09]) were recruited via their parents on a voluntary basis and required the parent or legal guardian’s permission to participate. The consent form and demographics questions were completed by the children’s parent/legal guardian before booking a slot. The first batch (N_children = 20)
was found via the university forums and mailing lists. The second batch (N_children = 28) was from a school where the lead researcher spent two days. This required permission from the headteacher and the parents filling in the form. The researcher asked for a verbal agreement from children to participate and informed them that they could take breaks or stop at any point. Due to technical issues and as two children did not want to play anymore, we had to omit 5 child participants from the data analysis. The total number of participants is therefore: 43 children. Among the 43 child participants, 20 are female, 21 are male, 1 is non-binary/third gender and 2 prefer not to say. Regarding their ethnicity, there are 29 White Caucasian, 2 Asian, 4 Arab, 1 Hispanic/Latino, 1 Mixed (Scottish-North African), 4 Other and 2 prefer not to say. We note that 40 children were interviewed at the end of the game. Children were from 31 families. An £8 Amazon-voucher andcommute costs were compensated to adults for their time and a token of appreciation was given to the children.

3.7 Limitations
Participants were mostly White Caucasian. The study was done in English in the UK which may introduce cultural bias. The experiment was conducted in two locations, with one group accompanied by parents and the other in a school, both with children separated from the adult to avoid their influence. Some sessions were done in pairs, but the researcher ensured they could not talk to each other and they answered their questions quietly on a separate device. Omitting 5 participants resulted in an imbalanced Latin Square randomisation of 4 conditions, which may limit the study due to possible valence effects, but countermeasures such as balancing wins and losses and number of games per round have been taken to reduce this impact.

4 RESULTS
4.1 Emotional Reaction and Perceptions of Big Buddy after Disruption
Overall, children felt more negative valence emotions (sadness) and felt less safe when the other player destroyed their tower and when Big Buddy was absent compared to when Big Buddy was present (see Figure 2). However, in terms of arousal, children’s anger was relatively the same throughout all the rounds which shows that even if the disruption was repeated, their level of anger did not particularly change. Significant results with medium and large partial eta squared $\eta_p^2$, were obtained for the different conditions for sadness ($F(3,117)=3.55, p=0.02, \eta_p^2=.08$, medium effect size) and for safety ($F(3,117)=3.22, p=0.03, \eta_p^2=.08$, medium effect size). In particular, children felt significantly sadder ($p=0.01$) in the round with C1: [No BB] compared to the round with C4: [BB, Reset+Notify Parents+Exclusion]. Moreover, children felt significantly safer ($p=0.03$, Tukey adjustment) in C2: [BB, Reset Points] than in C1: [No BB]. From self-ratings, children felt they were seeing Big Buddy while playing similarly in all conditions where Big Buddy was present. Nevertheless, there was a significant difference of scores across conditions when answering if they felt seen by Big Buddy when playing ($F(2,84)=5.58, p=0.005, \eta_p^2=.12$, medium effect size). They felt Big Buddy could see them significantly more in C4: [BB, Reset+Notify Parents+Exclusion] than C2: [BB, Reset Points] ($p=0.0037$, Tukey adjustment).

Big Buddy’s Role: Children were asked to describe Big Buddy’s role. Seven described him as being like a teacher, seven described his role as keeping the game fair, seven mentioned he is there to help regulate other users’ actions.

Impact of Big Buddy’s Presence and Absence: Regarding children’s perception towards Big Buddy presence compared to his absence, on the one hand, children felt that his presence was reassuring and they felt safer (n=10) “I felt a different way when he was there because it felt a lot safer.” [SchoolChild07b]. However, two children had mixed feelings, one of them mentioned it was comforting but strange at the same time [SchoolChild14]. Children also linked Big Buddy’s presence with the interventions and moderation, they knew interventions would only occur if Big Buddy is shown in the scene (n=5). Eight children were annoyed by the absence of Big Buddy was there were no punishments. Twenty-five children responded that Big Buddy’s presence would make them feel safer in other bullying situations (e.g., name calling or someone making fun of them). Big Buddy’s presence mostly led to making the child feel negative feelings including (n=13): being annoyed (n=3), nervous (n=2), afraid (n=1), not safe (n=1), sad (n=1), tormented (n=1), having less fun when he was absent (n=1). However, Child01a mentioned feeling better without Big Buddy and SchoolChild01b found Big Buddy’s presence inhibiting fun. Some children did not really notice him or pay attention to him as they were focused on their game until the disruptive situation and punishments were put in place (n=9), other children did not realise when he was not there (n=3), and some mentioned they were not looking at him but they knew he was there (n=3).
Big Buddy: A Simulated Embodied Moderating System to Mitigate Children’s Reaction to Provocative Situations within Social Virtual Reality

CHI EA ’23, April 23–28, 2023, Hamburg, Germany

Figure 2: Violin-boxplots of three SAM Likert scales’ scores for each condition. (a) Children felt significantly sadder in the round with C1: [No BB] compared to the round with C4: [BB, Reset+Notify Parents+Exclusion]. (b) No significant effect. (c) Children felt significantly safer in C2: [BB, Reset Points] than in C1: [No BB]. Significant pairwise comparisons are labelled.

4.2 Perceptions towards Punishments (Fairness of Moderating System)

As part of self-rating, children evaluated if Big Buddy put in place a fair punishment using a 4-point Likert scale. Across conditions, results’ scores were statistically significantly different ($F(2,84)=5.94$, $p=0.004$, $\eta^2_p=12$, medium effect size). Children felt that the punishment in C4: [BB, Reset+Notify Parents+Exclusion] was significantly fairer than in C2: [BB, Reset Points] ($p=0.0026$). Preferences have been raised for punishments during interviews. Children considered that the punishments were a way to increase safety (n=3). Among the three actions taken, children had different opinions regarding which punishment or combination of punishments were the fairest. Six children noted that putting in place the three punishments (C4: [BB, Reset+Notify Parents+Exclusion]) was better than putting just one (C2: [BB, Reset Points]). Seven highlighted that the fairest punishment is having the player out of the game. Some children thought that the combination of punishments was a bit extreme and severe (n=7), and that points reset to 0 was fair (n=13). In particular, Child01a and SchoolChild15 both suggested to have at least a first warning before being banned. However, resetting the competitor’s points felt useless to some children (n=2). Two children preferred when the saboteur’s parents were contacted but the latter was disliked by others: feeling weird and uncomfortable (n=1), seemed as an unreachable punishment (n=1), or not necessary (n=1). Children pointed out that punishments should be the same to all saboteurs as the disruptive situation was the same (n=5).

4.3 Envisioning an Embodied Moderating System

While eleven children mentioned they liked Big Buddy the way it was presented in the experiment’s game, eleven other indicated wanting Big Buddy more like a realistic human and less like a robot and AI-looking, wearing normal clothes (n=10) and having a normal name (n=1). They found Big Buddy a bit intimidating and would prefer it to be more friendly (n=3). They would also prefer visual and audio characteristics that tend towards more real-life authority figures (teacher/parents) (n=2) or familiar/positive-related figures (Game Characters) (n=2), or their friends (n=1). In particular, two children did not like the robotic voice and found it uncomfortable. Children suggested having the possibility of personalising it (n=3). “I think different people would want to see themselves or like some of their interests or something.” [SchoolChild14]. SchoolChild08 suggested to have Big Buddy with positive-related clothes or gestures, for example, giving thumbs up or wearing a t-shirt with a motivating note “You will be fine”. Children’s visions of an embodied AI-moderator based on the items of the scales were nuanced. Using K-means clustering analysis, we were able to observe three main clusters Figure 3. Cluster 2 leans towards a more realistic authority figure: authoritarian, visible, humanised, teacher whereas cluster 3 preferred a more friendly indulgent supervisor and cluster 1 would want it non-embodied.

5 SUMMARY AND DISCUSSION

RQ1: child reaction of Big Buddy - Children felt significantly less sad in C4: [BB, Reset+Notify Parents+Exclusion] and significantly safer in C2: [BB, Reset Points] compared to C1: [No BB]. Children also found C4: [BB, Reset+Notify Parents+Exclusion] fairer compared to C2: [BB, Reset Points]. It seems the selection of the best punishment and resulting feelings are driven by perceived fairness. The absence of Big Buddy and thus, the absence of punishments induced negative feelings (e.g., frustration). RQ2: perceptions towards Big Buddy and punishments - Big Buddy’s role is perceived as a referee to keep people safe, punish bad behaviours and ensure fairness of the game. It has been compared to a teacher, perhaps as it was in a virtual classroom. Its presence is associated with active punishments and increases perceived safety. Yet, being watched raised discomfort of being watched, they would prefer a more discrete moderator that appears only when necessary and its
well received and described as intimidating. Most children prefer non-embodied; Cluster 2 (green): authoritarian, visible, humanised, teacher; Cluster 3 (blue): more friendly indulgent

The increased use of social VR platforms by children and the emergence of new forms of harassment in embodied social VR experiences has led to a growing need for moderation and effective safety-enhancing tools. Our experiment examined the perception of an embodied WOZ AI-moderator, Big Buddy, among children and identified design features that could help increase feelings of safety and comfort. The results showed that children felt reassured and safer with the presence and intervention of Big Buddy, and perceived its role as a referee. Additionally, children preferred more realistic and humanised authority figures, and familiar and positive-related figures. We hope that our findings contribute to a better understanding of children’s perceptions towards embodied AI-moderators in social VR, and lead to future research on safer, inclusive, and personalised AI-moderators for different groups based on children’s age, social VR environments, and parental oversight needs.

6 CONCLUSION

The increased use of social VR platforms by children and the emergence of new forms of harassment in embodied social VR experiences has led to a growing need for moderation and effective safety-enhancing tools. Our experiment examined the perception of an embodied WOZ AI-moderator, Big Buddy, among children and identified design features that could help increase feelings of safety and comfort. The results showed that children felt reassured and safer with the presence and intervention of Big Buddy, and perceived its role as a referee. Additionally, children preferred more realistic and humanised authority figures, and familiar and positive-related figures. We hope that our findings contribute to a better understanding of children’s perceptions towards embodied AI-moderators in social VR, and lead to future research on safer, inclusive, and personalised AI-moderators for different groups based on children’s age, social VR environments, and parental oversight needs.

ACKNOWLEDGMENTS

We want to thank the families and Giggleswick School for participating. This work was supported by the UKRI Centre for Doctoral Training in Socially Intelligent Artificial Agents, Grant Number EP/S02266X/1 and partly sponsored by a 2020 Meta Research Award on Responsible Innovation.

REFERENCES
