Fighting Bullying with the Help of Autonomous Agents in a Virtual School Environment

Carsten Zoll (carsten.zoll@ppp.uni-bamberg.de)
Institut für Theoretische Psychologie, Universität Bamberg,
D-96047 Bamberg, Germany

Sibylle Enz (sibylle.enz@ppp.uni-bamberg.de)
Institut für Theoretische Psychologie, Universität Bamberg,
D-96047 Bamberg, Germany

Harald Schaub (schaub@iabg.de)
Institut für Theoretische Psychologie, Universität Bamberg,
D-96047 Bamberg, Germany

Ruth Aylett (ruth@macs.hcw.ac.uk)
MACS, Heriot-Watt University, Riccarton, Edinburgh EH10 4AS, UK

Ana Paiva (ana.paiva@inesc-id.pt)
Instituto Superior Tecnico and INESC-ID, Av. Prof. Cavaco Silva, IST, Taguspark
Porto Salvo, Portugal

Abstract

Using empathic virtual agents for Personal and Social Education is a powerful and immersive way to sensitise students for the problems and dangers related to persistent aggressive behaviour among students, also known as bullying. The work reported here relates to the evaluation of applying such agents in schools in the UK, Portugal, and Germany as well as to consequences that can be drawn from these evaluation results.

The Bullying Problem

The problem of bullying is well-known to educators and parents across Europe. Among others, the enduring victimisation of one student by one or more other students may cause health problems, depression, truancy, and even suicide in the victim. It has been defined as "repeated action that occurs regularly over time, and usually involves an imbalance in strength, either real or perceived" (Olweus, 1999). Bullying means various types of aggressive behaviour including punching, kicking, hitting (direct physical bullying), teasing and cruel name-calling (direct verbal bullying) and rather indirect actions like social exclusion or deliberate withdrawal from friendship (indirect relational bullying). To fight the serious long-term effects, the EU sponsored project VICTEC (Virtual ICT with Empathic Characters) intended to deliver a software product based on episodic virtual drama to be used in schools with the aim to enhance students' insight in the destructive nature of bullying and sensitise them for the harmful consequences for the victim. By using a 3D virtual environment for education purpose, students are encouraged to explore the problem of bullying in a save and exiting environment. The basic idea is to encourage empathic processes towards the virtual victims of bullying and hence create sensitivity towards the bullying problem.

Modelling Empathy

Titchener (1909) was the first to use the term "empathy" as a translation of the German term "Einfühlung" ("feeling into" something or somebody). Originally the term "Einfühlung" was introduced by Lipps (1903/1906) to describe the process of feeling into an art object, and has been subsequently transferred to interpersonal processes between a target person (who displays overt behaviour in a specific situational context) and an observer (who perceives the target's behaviour and the situational context and reacts empathically). Contemporary psychologists widely agree on the fact that there are two perspectives on the empathic process, focusing on its' affective and/or cognitive nature. Hence, definitions of empathy vary to a certain degree, depending on the underlying theoretical assumptions. For the authors of this paper, empathy includes both affective an cognitive processes, with "affective" meaning feeling something due to the perceived feelings/thoughts of a target person, and "cognitive" focusing on understanding feelings and thoughts of a target person. These processes produce affective (parallel affect, reactive affect) as well as non-affective (perceptual accuracy, attribution styles, behavioural tendencies) outcomes (see also Davis, 1996), all related to an observer's internal simulation of a target person's internal states.

Empathy may be elicited by two different classes of stimuli, coming from the emotional expression of the target person or from the situational context: the former occurs whenever

expressional behaviour is interpreted, e.g. a person who smiles is happy, a person who frowns is sad; the latter takes place if the observer concludes the emotional state of the target from the situation the target is dealing with, e.g. a friend who is hit and hustled in the schoolyard might be frightened, ashamed and/or angry.

One goal of the project is to use the empathic process to get students to empathise with bullying victims. To achieve this goal, the software focuses on the issue of empathy as one crucial aspect of the affective interaction between the user and the characters. These characters should be able to evoke empathy in the user, but also they should be able to empathise with each other. The first issue mainly concerns decisions related to character design, especially to the perceived similarity between the user and the character; the second issue tackles the inner life of the character, the agent architecture.

The software application "FearNot!"

To achieve the goal of sensitising students for the dangers of bullying, a software called "FearNot!" (Fun with Empathic Agents to Reach Novel Outcomes in Teaching) was developed. The software features a virtual school inhabited by virtual students which are equipped with an agent architecture that allows the generation of "autonomous behaviour". These agents engage in bullying episodes with the user watching. The agent's architecture is parameterized so that different personalities can be created for different characters and different episodes, thus allowing for a diversity of characters in the system.

Narrative structure An episodic approach was chosen since bullying is episodic in nature, being defined as "repeated action that occurs regularly over time" (Olweus, 1999). The user acts as a spectator while the virtual bullying is happening, mainly to provide emotional distance and security for those who are bullies or victims of bullying in the real world, and because offering the possibility to intervene in the virtual bullying poses the danger of building a bullying game rather than intervening on the cognitive and affective level in order to prevent bullying.

Figure 1: Interaction with the victimized character.

In between the bullying episodes, the victim turns to the user, asking him/her for help and advice how to deal with being bullied. By interacting with the victimised character and trying to offer appropriate coping suggestions, the user acts as the 'invisible friend' of the victimised character. The approach was inspired by the Forum Theatre approach (Boal, 1979), another method that involves groups (not individuals) in discussing the strategies that are then communicated to and performed by an actor on stage (in reality, not in a virtual world). This interaction sequence between the user and the virtual bullying victim is meant to trigger empathic reactions in the user, trying to understand the thoughts and feelings of the victim as well as feeling something due to the fact that he/she perceives the inner state of the victim. The content of the dialogue influences the character's behaviour. Since this behaviour is triggered autonomously by the characters themselves, there is a certain danger that the episodes do not match up to a believable storyline.

For the sake of a user-centred experience, a stage-manager organizes certain characteristics of the episodes and offers some possibility to take into account an educational message. The software is meant to be used in schools, thus some attention must be drawn to the fact that there has to be some sort of educational message, even though this poses a certain limitation to emergent narrative. The problem is, that in the case of bullying, there is no clear and simple solution to the problem that will work in any situation whatsoever; instead, the only advice education experts agree on is that staying inactive and doing nothing will not improve a victim's situation. These considerations resulted in an educational message "Don't suffer in silence, tell someone you trust!". Apart from the educational impact, the location of an episode as well as the characters that are involved in it need to fit the previous as well as the following episode, which is also provided by the stage manager. Despite this regulatory influence of the stage manager, characters act – within their means – autonomously and provide a variety of behaviours in different situations and hence allow for the user to take over different perspectives on the bullying problem.

Similarity Similarity between observer and target is able to enhance empathy and helping behaviour within the observer (Eisenberg & Miller, 1987; Davis, 1996). The FearNot! characters therefore need to be perceived as similar by the user group in order to evoke empathic processes on the side of the user. These empathic processes rely on representations within the observer of the virtual bullying (user). Affective as well as cognitive processes involved in empathy build on the former experiences of the observer regarding the target person and the current situation. Whenever the user observes for example John being hit by Luke in the virtual school, representations are activated within the user connected with what they already know about John or the situation of being hit themselves as well as how one generally feels whenever being hit. The intensifying influence of similarity on empathic processes is based on the similarity of the underlying representations that are shared by observer and target person: shared representations between observer and target person facilitate an empathic experience within the observer because they – when activated – allow for a more accurate and faster inner

simulation of the target person's inner state within the observer.

Similarity in FearNot! Keeping in mind that empathy is mediated via emotional expression as well as via situational context information, believable character behaviour, design and the development of believable, familiar episodes contribute to the subjective impression within the user that he/she is similar to the characters. To ensure that both character design and story content are accepted and liked by the target group, students and teachers were involved throughout the project in the development of story content and the design decisions regarding characters and environment. Regarding the physical features of characters, studies with the target group of users clearly indicated that the target group preferred the cartoon characters, which led to the design of the FearNot! 3D cartoon-like characters (see figure 1), taking into account age, gender and culture related specifications (physiognomy, dress, school uniform, etc.).

Apart from these features, the actions carried out by the characters are able to foster or undermine their believability. Moreover, as the FearNot! characters are empathic synthetic characters, they need to be able to empathise with each other, in terms of cognitive reactions (modelling others' inner states), affective reactions (reacting emotionally towards perceptions of others' inner states), and behavioural reactions (expression in overt behaviour).

In FearNot!, characters act autonomously, their action regulation embracing the following components.

Agent Architecture. The agent's behaviours (the characters in the story) are generated from an agent architecture that perceives all the objects, events, and other agents that exist in the world and appraises the significance of these perceptions, activating emotions accordingly.

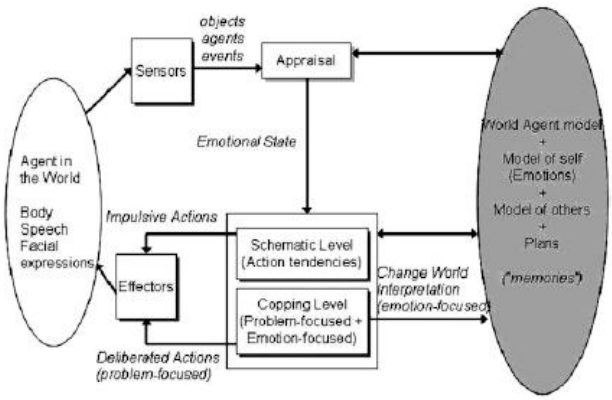

Figure 2: Agent Architecture.

The appraisal mechanism is based on the model proposed by Ortony, Clore and Collins (1988), and can be described as "a subjective evaluation of a given stimulus according to the character's goals, standards and beliefs" (Aylett, Louchart, Dias, Paiva, Vala, 2005). The OCC model devises active-pursuit goals and interest goals which are fed in a goal hierarchy using goal links (necessary, sufficient, facilitative, and inhibitory) that specify the relations among goals. After a goal update phase the character appraises the importance of its goals, appraises aspects of objects, action of agents (regarding standards of behaviour), and consequences of events. As a result of the appraisal, character's emotions are activated depending on a set of emotional thresholds and decay rates that are pre-set for different personality profiles. The threshold for an emotional state has to be reached for the emotion to be activated, and the decay rate decides for how long the emotion will remain activated before fading out. Then, the action selection mechanism that decides on which action will be performed, includes two stages: on the first stage, action is directly triggered by activated emotional states; the second stage is based on existing research on coping (Lazarus, 1991), discriminating problem-focused coping (planning to reach goals) from emotion-focused coping (changing the appraisal that produces the emotion). The currently most intense emotional state is associated with an intention that is selected and stored. Actions are performed accordingly using effectors. These actions are then transformed into graphical animations that are shown to the user as an episode of the system.

Evaluating the Prototype The pedagogical aim of FearNot! is to change students' behaviour and cognition regarding bullying via social immersion which is promoted by using believable agents and stories on the one hand and autonomous character behaviour in a non-scripted drama (emergent narrative) on the other hand. The prototype of FearNot! that was evaluated within the scope of VICTEC did not feature emergent narrative; a scripted version including one physical and one relational scenario was used instead. This prototype model also lacked a language system. The interaction of the user with this prototype software proceeds as follows: After having initially provided their personal information (name, gender and age) the user gets introduced to the characters, school and situation. After that, the first bullying episode occurs, followed by the victimised character seeking rescue in the school library and addressing the user. A dialogue starts with the user selecting answers from a predefined drop-down menu that covers several coping strategies (tell a friend, avoid bully, tell parents/ teacher, run away, fight back, etc.) and justifying his/her selections by typing them in. Being a scripted application, the content of the final episode directly depends on the choices made by the user concerning the coping strategies. The aim is to react to either functional as opposed to dysfunctional coping strategies suggested by the user. Having selected primarily functional strategies (e.g. "telling someone"), the final episode offers a rather positive ending with one of the bystander characters offering help to the victim. In case the user's selection of coping strategies is considered overall dysfunctional (e.g. "run away"), in the final episode the victim might be offered help, but he/she rejects it. At the very end, the educational message is displayed saying that "telling someone" is always a good choice.

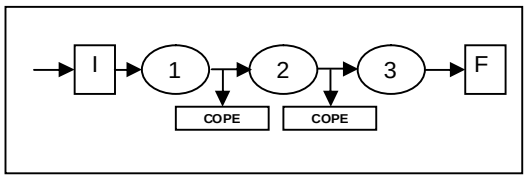

Figure 3: Structure of a bullying scenario.

Figure 3 details in a flow chart the structure of one scenario. Symbols indicate the following:

- I: introduction of characters and school, typing in code, name, age and gender
- 1-3: bullying episodes
- COPE: interaction with victim character in the school library in between episodes
- F: educational message after end of episode 3

To sum up, the running model of FearNot! simulates real bullying situations with autonomous agents acting as students in a virtual school building with the user watching the bullying and acting as an "invisible friend" to the victimised character in between the episodes. Even though the final product of VICTEC features emergent narrative, the application used for evaluation purpose is a scripted one that uses pre-scripted storylines and relates the outcome of the story directly to the choice the user has made from a selection of coping strategies.

Extensive qualitative usability evaluation escorting the developmental process of the software indicated that the target group appreciates the key idea to create an interactive software application on bullying which they seem to prefer to a teacher-led instruction, as they can approach the matter individually and privately. Usability evaluation also revealed that even though ratings of characters' physical appearance and emotional expression were quite negative, empathic engagement was high, indicating that children seem to infer the emotions from situational cues rather than to rely on the characters' facial and body expression.

Evaluation of "FearNot!"

The prototype running model was evaluated in a school setting as well as in the laboratory. The evaluation was extensive (approximately 400 students contributed to the laboratory evaluation, 252 students took part in the classroom evaluation); hence, in this paper, we will focus on a specific section of the data that stems from the within-classroom evaluation. This evaluation included qualitative as well as quantitative data collection and analyses and offers a good impression of the acceptance and benefit of the software as part of Personal and Social Learning within the classroom. With this extensive evaluation, we aimed at assessing cognitive effects of the software interaction (change in attitudes towards bullying, increased understanding of victims' reactions), affective effects (affective empathy as inhibitor of bullying), and behavioural effects (rated by peers and teachers).

Procedure For the evaluation, four questionnaire instruments have been designed to assess general cognitive and affective empathic reactions (Empathy Questionnaire), actual empathic reactions to a given bullying situation (Picture Story), and bullying status from a self- and peer-perspective (Bullying Questionnaire) as well as from the teacher perspective (Teacher Rating). These questionnaires were completed four weeks prior to (pre-test) and after (post-test) the software interaction. Immediately after the software interaction, students discussed pros and cons of the software in a semi-structured classroom discussion. One class in each country was treated as control group, with the software interaction taking place after the post-test questionnaire session.

Sample The sample embraces 252 students from UK (N=101), Portugal (N=55), and Germany (N=96) aged eight to 14 years (M=9.38; SD=1.32) from rural as well as urban regions.

Results Pre-test questionnaire data were analysed to determine cultural differences between the countries. Regarding trait empathy as measured by the Empathy Questionnaire, German students described themselves as significantly less affectively empathic than Portuguese and UK students whereas Portuguese students showed significantly lower cognitive empathy scores compared to UK and German students. No age effects could be found, but the rather common gender effect that females describe themselves as significantly more empathic could be confirmed, for affective as well as for cognitive empathy. UK students rated the picture-story of bullying as least "bad" compared to students from Germany and Portugal, indicating that they react with less empathy towards the depicted incident. This fact probably reflects the greater awareness of bullying incidents in UK schools and curricula, in the sense that the students somehow get used to being confronted with rather grave scenes of harassment for educational purpose. Analyses on cultural differences regarding bullying status revealed a mean incidence of 15% for relational bullies and 16% of physical bullies across countries, with no cultural differences emerging. However, when looking at the victim incidence, significant differences can be reported for Portugal compared to UK and Germany, with significantly more relational victims in Portugal (31% relational victims) than in the other countries (UK:7% relational victims; D: 11% relational victims). The Portuguese sample also revealed age and gender effects: Boys are significantly more often involved in bullying (as victim and bully) than girls. Also, younger Portuguese students are significantly more often victims to physical bullying than older students.

In sum, meaningful cultural differences could be detected regarding the self-description on empathic dimensions with the Portuguese reacting predominantly with affective empathy and the UK students with highest scores on cognitive empathy. Portuguese students also rated the bullying incident depicted in the picture-story as graver than their colleagues from UK and Germany and named significantly more victims to relational bullying than students from UK and Germany, suggesting that Portuguese students react more sensitively to bullying and empathise

more with the victims of bullying than students from UK and Germany.

Comparing pre- to post-test data did not yield any rises in empathy, neither affective nor cognitive, or any decreases in bullying incidence which remained stable. In contrast to what has been hypothesized, affective empathy scores even decreased, in experimental groups as well as in control groups. This overall trend strongly indicates that factors different from the software application caused the decrease in self-reported empathy; otherwise, control group scores should have remained stable. Factors that might have prevented an increase in self-reported empathy can be found in the experimental design of the study: Both empathic reacting and bullying behaviour are quite stable behavioural styles that develop over months and years and can hardly be changed by a single software interaction that does not exceed 30 minutes. Bullying has to occur by definition on a regular basis over a longer period of time to be distinguished from "normal" aggressive behaviour. Thus, the software needs to be implemented in the curriculum on a regular basis with recurring interactions taking place between the user and the characters for any significant behavioural alterations to occur. The actual decrease in empathy found might be due to a familiarisation of the students with the experimental procedure and the experimenters from pre- to post-test that may also have caused a decrease in social desirable answering in the post-test.

Further development of "FearNot!" within the scope of eCIRCUS

To build up on the software development and evaluation performed in VICTEC, the consortium, along with additional partners from Switzerland, Germany, and Italy, has applied for additional funding. In the framework VI project eCIRCUS (Education through Characters with emotional-Intelligence and Role-playing Capabilities that Understand Social interaction), our goal is to use our developments and experiences from VICTEC to technically improve the software, as well as to expand its' content to problems of refugee integration in schools, and to develop and implement a suitable longitudinal evaluation approach to assess the long-term effects of the software on stable empathic styles and bullying roles.

Acknowledgments

The projects VICTEC and eCIRCUS are funded by the European Union under its framework V and Framework VI programme, respectively.

References

Adams, G. R., Jones, R. M., Schvaneveldt, J. D., & Jenson, G. O. (1982). Antecedents of affective role-taking behaviour: Adolescent perception of parental socialization styles. *Journal of Adolescence, 5*, 259-265.

Aylett, R., Louchart, S., Dias, J., Paiva, A., & Vala, M. (2005). FearNot! – an experiment in emergent narrative. In T. Panayiotopoulos et al. (Eds.), *Intelligent Virtual Agents 2005*. Hamburg: Springer.

Boal, A. (1979). Theatre of the oppressed. New York: Urizen.

Bryant, B. K. (1982). An Index of Empathy for Children and Adolescents. *Child Development, 53* (1), 413-425.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*, 74-101.

Dautenhahn, K., & Woods, S. (2003). *Possible connections between bullying behaviour, empathy and imitation.* Presented at AISB '03, The University of Wales, Aberystwyth.

Davis, M. H. (1996). *Empathy. A Social Psychological Approach*. Boulder, CO: Westview Press.

Dörner, D., & Schaub, H. (2002). Gefühle. Modellierung emotionaler Zustände. In G. Bente (Hrsg.), *Virtuelle Realitäten*. Göttingen: Hogrefe.

Eisenberg, N., & Strayer, J. (1987). *Empathy and its development.* New York: Cambridge University Press.

Eisenberg, N., & Miller, P. A (1987). The Relation of Empathy to Prosocial and Related Behaviors. *Psychological Bulletin, 101*, 91-119.

Enz, S., Zoll, C., & Schaub. H. (2004). FearNot!: The use of empathy towards autonomous agents for education purpose. In H. Schaub, F. Detje & U. Brüggemann, (Eds.), *The Logic of Artificial Life*. Berlin: Akademische Verlagsgesellschaft.

Hoffman, M. L. (2000). *Empathy and moral development.* New York: Cambridge University Press.

Lazarus, R. (1991). *Emotion and Adaptation*. Oxford University Press.

Leibetseder, M., Laireiter, A.-R., & Köller, T. (2001). E-Skala: Fragebogen zur Erfassung von Empathie – Beschreibung und psychometrische Eigenschaften. *Zeitschrift für Differentielle und Diagnostische Psychologie, 1*, 70-85.

Lemerise, E. A., & Arsenio, W. F. (2000). An integrated model of emotion processes and cognition in social informational processing. *Child Development, 71*, 107-118.

Lipps, T. (1903/1906). *Ästhetik. Psychologie des Schönen und der Kunst.* Hamburg & Leipzig: Leopold Voss.

Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality, 40*, 525-543.

Miller, P. A., & Eisenberg, N. (1988). The Relation of Empathy to Aggressive and Externalizing/Antisocial Behavior. *Psychological Bulletin, 103* (3), 324-344.

Olweus, D., & Limber, S. (1999). *Blueprints for violence prevention: Bullying Prevention Program*. Institute of Behavioral Science, University of Colorado: Boulder.

Ortony, A., Clore, G., & Collins, A. (1988). *The Cognitive Structure of Emotions.* Cambridge: University Press.

Paiva, A., Dias, J., Sobral, D., Aylett, R., Woods, S., Zoll, C., & Hall, L. (2004). Caring for Agents and Agents that Care: Building Empathic Relations with Synthetic Characters. In IEEE (Org.), *AAMAS '04* (The Third

International Joint Conference on Autonomous Agents and Multi Agent Systems, 19.-23.July, New York). New York: Springer.

Paiva, A., Dias, J., & Aylett, R. (2005). Learning by Feeling: Evoking Empathy with Synthetic Characters. *Applied Artificial Intelligence, 19*, 235-266.

Schaub, H. (1993). *Modellierung der Handlungsorganisation*. Bern: Huber.

Schaub, H. (1997). Modelling Action Regulation. In J. Brezinski, B. Krause & T. Maruszewski (Eds.), *Idealization VIII: Modelling in Psychology*. Amsterdam: Rodopi.

Sobral, D., Machado, I., & Paiva, A. (2003). Managing authorship in plot conduction. In G. Goos, H. Hartmanis & J. Leeuwen (Eds.), *Virtual Storytelling*. Berlin: Springer.

Sutton, J., Smith, P. K., & Swettenham, J. (1999). Bullying and 'theory of mind': A critique of the 'social skills deficit' view of anti-social behaviour. *Social Development, 8*, 117-127.

Titchener, E. (1909) *Experimental psychology of the thought processes*. New York: Macmillan.

Wolke, D., Woods, S., Schulz, H., & Stanford, K. (2001). Bullying and victimisation of primary school children in South England and South Germany: Prevalence and school factors. *British Journal of Psychology, 92*, 673-696.

Zoll, C., Enz, S. & Schaub, H. (2003). Empathic Virtual Agents. In T. Rist, R. Aylett, D. Ballin & J. Rickel, (Eds.), *Intelligent Virtual Agents – IVA 2002*. Berlin: Springer.