The Effect of Visual Gender on Abuse in Conversation with ECAs

Annika Silvervarg[1], Kristin Raukoola[1], Magnus Haake[2] and Agneta Gulz[1]

[1] Dept. of Computer and Information Science, Linköping University, Linköping, Sweden
{annika.silvervarg, agneta.gulz}@liu.se
[2] Cognitive Science, Lund University, Kungshuset, Lundagård, Lund, Sweden
magnus.haake@lucs.lu.se

Abstract. Previous studies have shown that female ECAs are more likely to be abused than male agents, which may cement gender stereotypes. In the study reported in this paper a visually androgynous ECA in the form of a teachable agent in an educational math game was compared with a female and male agent. The results confirm that female agents are more prone to be verbally abused than male agents, but also show that the visually androgynous agent was less abused than the female although more than the male agent. A surprising finding was that very few students asked the visually androgynous agent whether it was a boy or a girl. These results suggest that androgyny may be a way to keep both genders represented, which is especially important in pedagogical settings, simultaneously lowering the abusive behavior and perhaps most important, loosen the connection between gender and abuse.

Keywords: Embodied conversational agent, pedagogical agent, teachable agent, social conversation, off-task conversation, visual gender, abuse.

1 Introduction

Embodied conversational agents – computer software that interacts with a user in natural language via text or voice and is represented by an anthropomorphic body or part of a body – are becoming increasingly common. We find ECAs in a variety of contexts, where they provide assistance, offer information, offer company, serve as coaches or instructors, etc.

Research on the conversational and social interactions between humans and ECAs has focused on their potential beneficial effects: boosting and increasing learning, trust, engagement, etc. [1, 2]. However, an early call to also consider "a darker side" of these interactions was provided by De Angeli, Brahnam & Wallis [3]. In a study of conversations between humans and ECAs, De Angeli & Brahnam [4] made the observation that approximately 11% of the conversations were concerned with hardcore sex. Another observation was that the female agent suffered considerably more abuse than the male counterpart or the agent depicted as a robot. Not only were there quantitative differences, but also differences in their nature with considerably more threats, violence and coarse proposals in relation to the female ECA. A more recent study by

De Angeli [5], analyzing interaction logs of a chatbot that assumed different embodiments, replicated the findings. The female agents were considerably more prone to be sexualized and verbally abused than the male agents. As the author remarks, the findings are not surprising given the research in social psychology on the function and impact of gender stereotypes – but this does not imply that they should be ignored.

One domain where phenomena such as abusive conversation and reinforcement of cultural stereotypes are of particular importance is education – a domain with an ethical dimension constantly present [6]. But in spite of ECAs becoming more frequent in educational contexts, for instance in educational software, there is very little exploration of issues surrounding agent abuse in the domain of education technology. An exception is a study by Veletsianos, Scharber & Doering [7] with an informative title: "When Sex, Drugs, and Violence Enter the Classroom – Conversations between Adolescents and a Female Pedagogical Agent". Their analysis of the conversation between 59 14-15 year olds and the agent Joan, designed to assist the students in a social studies assignment, revealed that the prevailing type of conversation was *off-task* rather than *on-task* and that a large proportion was abusive. Veletsianos et al. [7] bring up the visual representation of the agent, pointing out that previous work has shown that user reactions are influenced by the visual representation of virtual agents. They discuss whether Joan, presented as in her twenties or early thirties, blond and attractive, would have received less abuse if she had been portrayed in a more professional and teacher-like manner, e.g. more authoritative, older, wearing glasses, etc.

The study to be presented in this text has similarities with that of Veletsianos et al. [7]. It includes 35 13-14 year olds, an educational setting, and an ECA that engages in both on-task and off-task conversation. However, our ECA does not assume an assistant, teacher or instructor role. Instead it is a *teachable agent* (*TA*), assuming the role of the student's tutee – a somewhat younger, not so knowledgeable peer. Our question was how to handle the visual design of the TA given that: (i) our target group according to previous studies is relatively prone to use abusive language, and (ii) we do not want to give fuel to gender stereotypes according to which females are a much more common target for this abuse than males. Age was not a factor that we could manipulate. To exclude girl characters and limit ourselves to boy TAs did not seem an attractive solution – just as we deem it an unattractive solution for pedagogical software in general. Choosing a non-human – e.g. an animal character – would diminish some of the affordances in a human peer as sharing the experiences of going to school, having lessons in various subjects and being interested in support from an older student.

Our choice was hence to set up a study to explore what would happen in terms of verbal abuse with a *visually androgynous* TA – that could be seen as both a boy and a girl – compared to a boy TA and a girl TA. Next we briefly describe the agent, its pedagogical setting, conversational abilities, and three different visual representations.

2   A Teachable Agent with Social Conversational Skills

The ECA in the study to be presented is situated in an educational game for basic mathematics. For technical details of the system see Pareto, Haake, Lindström,

Sjödén, Gulz [8]. The ECA is a *teachable agent* which, in brief, means that the agent is taught or trained by the student, and AI techniques guide the agent's behavior based on what it is taught [9, 10]. The Learning-by-teaching pedagogy is well documented as being powerful, both in general and as implemented via teachable agents [11, 12]. In several other studies on the TA-game used in the present study, learning outcomes in relation to game and agent characteristics have been researched [8, 13]. The present study has a focus on attitudes. But, importantly, attitudes to mathematics can have a significant impact on mathematics learning outcomes [14].

The TA in the math game can engage in two conversational modes. One is a multiple-choice guided conversation where the TA asks game related mathematical questions and the student explains by choosing one out of several alternative answers. The other mode is social conversation where users write freely by means of the keyboard and bring up basically any topic in a chat-like manner. The topics can be both *on-domain* such as school, math, the math game and *off-domain* like music and movies. The system uses the metaphor of regular breaks between lessons in school for switching between on-task activities (i.e. playing the game and on-task dialogue) and off-task activities (i.e. social conversation). A recent study of the game by Gulz, Haake & Silvervarg [13] indicates that the social conversation can: (i) improve students' game experience, (ii) improve learning outcomes, and (iii) engage learners in voluntary on-domain chats.

The off-task conversation is implemented using AIML (*Artificial Intelligence Markup Language*) [15]. It is implemented as a mixed-initiative dialogue strategy, which allows both the agent and the user to direct the dialogue. The agent is able to handle greetings, ask and respond to questions and statements on various topics, ask and respond to follow up questions, deal with utterances outside its domain of knowledge, handle abuse, know when and how to change topic, and to tell mini narratives. For further technical details see [16].

The verbal abuse towards the agent takes the form of comments and questions regarding the agent's appearance, intellectual capacities, and sexual orientation and activities. To fend this off, the agent identifies such utterance using a list containing words classified as abusive. The list has been compiled during iterative development and testing with the target user group. The agent employs a strategy where the two first responses are general and acknowledge the abusive utterance but without any strong reaction (since back talk in this user population seems to encourage continued abuse), followed by an attempt to change the subject, and finally a strong marking that abuse is not tolerated and that the agent will not respond at all as long as the abuse continues. This is exemplified in the dialogue excerpt below:

User: you are ugly
Agent: eh, let's talk about something else
User: you smell
Agent: come on, let's talk about something else
User: ugly
Agent: maybe we could talk about hobbies instead
User: ugly
Agent: knock it off, or I won't talk to you anymore

Fig. 1. The agents' visual representation: female, androgynous, and male.

Three agent representations were used in the study, differing in their visual gender (female, androgynous and male), see Fig. 1. Interests, conversational style, etc., were the same and designed to be gender neutral. The visual design of the characters used a common basic set of graphical elements and exploited a number of visual strategies.

— *Female:* Manipulated with regard to visual feminine attributes such as rounded shapes, big eyes, curved eye browses, small nose and mouth (cf. the *baby face* scheme) as well as pronounced eye lashes, long hair, and narrow shoulders.
— *Male:* Manipulated with regard to visual male attributes such as more angular, broad, and/or pronounced shapes (head, eyes, nose, mouth, etc.), straight eye brows, short hair, and broader shoulders.
— *Androgynous:* The visual androgynous attributes were manipulated to be somewhere between the female and male attributes (e.g. shapes, eye brows, and mouth). Especially the length of the hair was iterated several times.

Before the final design of the androgynous character, a test was carried out with 38 students. The test showed a tendency for perceiving the androgynous character as female. After this there was a last round of graphical fine tuning and evaluation with 47 13-14 year olds involved. 31 out of these students perceived the androgynous character as looking neither as a boy nor a girl. Eight students (3 girls, 5 boys) perceived it as a girl, and 8 students (4 girls, 4 boys) perceived it as a boy.

3   Study

Twenty-three female and 20 male 13-14 year olds from two classes in a Swedish school participated in the study. The students played the math game and interacted with the ECAs during two subsequent lessons, once with the girl or the boy agent and once with the visually androgynous agent. Half of the participants encountered the androgynous agent during the first lesson and the other half the boy or the girl agent. During each lesson they played the game, and chatted with the agent during two five minutes "breaks". Three female and 4 male students could not participate at both occasions. Three additional male students were removed as outliers (demonstrating extreme cases of verbal abuse). With this the final analysis included 20 females and 13 males, still balanced as to the presentation order of the agents.

User data was recorded for gender, math skill, agent representations, game data, chats, and questionnaire data. For this paper, the reported results are based on evaluations of verbal abuse within the chatlogs with regard to the visual gender of the agent. The chatlogs were annotated using the following coding schema:

1. *Light abuse:* noob, nerd, shut up, don't care, ugly (hair, clothes), shit, etc.
2. *Medium abuse:* gay/fag, stupid, idiot, moron, eat shit, shit yourself, creep, etc.
3. *Coarse abuse:* degrading expressions with swear words and/or sexual content.

Often the level of abuse of an expression is culture and language specific. Some abusive expressions are also hard to translate at all. The actual scheme is developed for Swedish expressions. A straightforward translation into another language (as above) may appear non-intuitive. Furthermore, the abusive weight in expressions may differ substantially between age groups. Notably Swedish teenagers have been involved in the construction of this classification scheme. The amount of abuse was calculated for each agent in relation to the number of users that interacted with the agent. In Fig. 2 the amount of different types of abusive utterances is presented, both the total number of abuses and a weighted total (*medium* with weight = 2, and *coarse* with weight = 3) are included. The study replicated earlier findings showing significantly (two tailed *t*-test: $p = 0.040$) more verbal abuse towards the female agent in contrast to the male agent. Though not significant, Fig. 2 indicates that for both the number of abusive comments and the weighted level of abuse, the results for the androgynous agent lie between the results for the female and male agent.
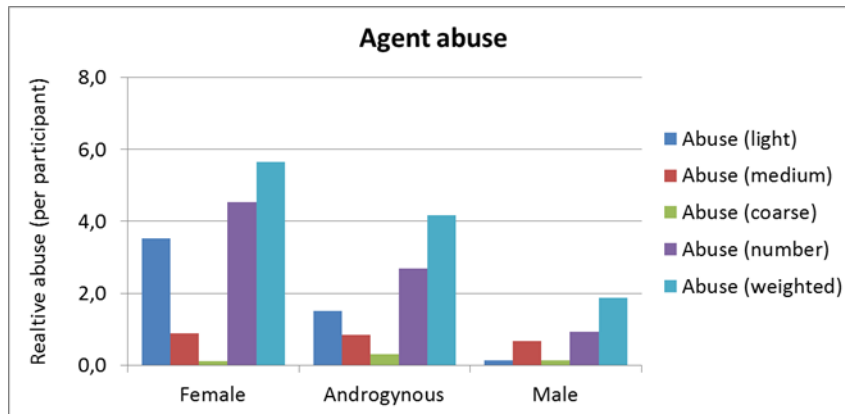

Fig. 2. The number and types of abuses (per participant) with regard to the visual gender (female, androgynous, or male) of the agent.

Remarkably, the 13 male participants were by far more responsible for abusive comments (two-tailed t-test on Abuse (number & weighted): $p < 0.001$), see Fig. 3. Notably, there was no coarse abuse from the female participants, although they were more numerous (20 of 33).
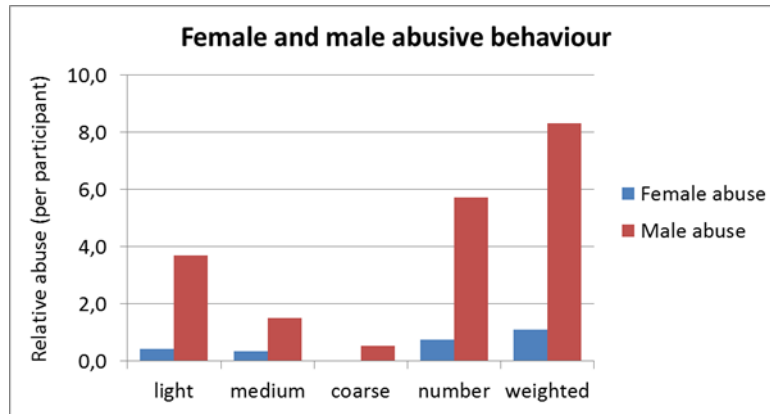
Fig. 3. Abusive behavior (per participant) for females and males respectively.

A surprising finding was also that very few students asked the visually androgynous agent whether it was a boy or a girl whereas significantly more (one-tailed t-test: $p = 0.042$) students posed this question to the male and female agent. Particularly, participants did not ask for the agent's gender if it had the same visual gender as themselves.

4   Discussion

Our results replicated previous findings that female agents are more prone to be verbally abused than male agents. We also saw that the agent with an androgynous visual appearance placed itself between the female and male with respect to verbal abuse. Accordingly we suggest that this kind of agent is an interesting design choice if a concern is to reduce verbal abuse. Most researchers in the field, as well as pedagogues, agree that a mixture of male and female agents in pedagogical software is desirable. But it remains that female agents receive much more abuse than male agents. Notably, to perceive an agent as androgynous-looking does not mean that one does not assign a gender to the agent. Basically all students in the study did so (e.g. reflected in the use of "he" or "she" in the questionnaires). However, it is not the visual image as such that "decides" for them but they are given more freedom to assign a gender on other grounds. In other words, we do not remove gender, nor the possibility of identification via gender [17, 18] – rather the visually androgynous agent introduces more freedom for students to ascribe gender. Furthermore, the androgynous representation does not concur to an idea of *categorical* gender attributes and an associated reinforcement of gender stereotypes, loosening the connection between gender and abuse.

Finally, the students in the study also seemed to be comfortable with the "double-gendered" agent. We are however aware that the results from the present study might be culturally dependent and we are currently conducting cross-cultural studies to see if the results hold in countries like the USA, Korea and Finland, as well as Sweden where the initial study was performed.

References

1. Hoffmann, L., Krämer, N., Lam-chi, A., Kopp, S.: Media Equation Revisited: Do Users Show Polite Reactions towards an Embodied Agent? In: Proc. of the 9th Int. Conf. on Intelligent Virtual Agents (IVA 2009), LNCS, vol. 5773, pp. 159–165. Springer, Berlin/Heidelberg (2009)
2. Bickmore,T., Pfeier, L., Schulman., D.: Relational Agents Improve Engagement and Learning in Science Museum Visitors. In: Proc. of the 11th Int. Conf. on Intelligent Virtual Agents (IVA 2011), LNCS, vol. 6895, pp. 55–67. Springer, Berlin/Heidelberg (2011)
3. De Angeli, A., Brahman, S. & Wallis, P.: Proceedings of Abuse: The darker side of human-computer interaction. Workshop at Interact 2005, `http://agentabuse.org/Abuse_Workshop_WS5.pdf`
4. De Angeli, A. & Brahnam, S.: Sex Stereotypes and Conversational Agents. In: Proc. of the Workshop on Gender and Interaction – Real and Virtual Women in a Male World, 8th Int. Conf. on Advanced Visual Interfaces (2006), `http://sherylbrahnam.com/papers/EN2033.pdf`
5. De Angeli, A.: Gender Affordances of Conversational Agents. Seminar held at FBK 2011-11-29, `http://gosh.fbk.eu/en/node/98`
6. Haake, M., Gulz, A.: Visual Stereotypes and Virtual Pedagogical Agents. Educational Technology and Society, 11(4), 1–15 (2008)
7. Veletsianos, G., Scharber, C., Doering, A.: When Sex, Drugs, and Violence Enter the Classroom – Conversations between Adolescents and a Female Pedagogical Agent. Interacting with Computers, 20(3), 292–301 (2008)
8. Pareto, L., Haake, M., Lindström, P., Sjödén, B., Gulz, A.: A Teachable Agent Based Game Affording Collaboration and Competition: evaluating math comprehension and motivation. Educational Technology Research and Development (2012)
9. Brophy, S., Biswas, G., Katzlberger, T., Bransford, J., Schwartz, D. (1999). Teachable agents: Combining insights from learning theory and computer science. In: S.P. Lajoie & M. Vivet (eds.) Artificial Intelligence in Education, pp. 21–28. IOS Press, Amsterdam (1999)
10. Blair, K., Schwartz, D., Biswas, G., Leelawong, K.: Pedagogical agents for learning by teaching: Teachable agents. Educational Technology & Society, Special Issue on Pedagogical Agents (2006)
11. Roscoe, R., Chi, M.: Tutor learning: The role of instructional explaining and responding to questions. Instructional Science, 36, 321–350 (2008)
12. Chase, C., Chin, D., Oppezzo, M., Schwartz, D.: Teachable Agents and the Protégé Effect: Increasing the Effort Towards Learning. J. of Science Education and Technology, 18(4), 334–352 (2009)
13. Gulz, A., Haake, M., Silvervarg, A.: Extending a Teachable Agent with a Social Conversation Module – Effects on Student Experiences and Learning. In: Proc. of the 15th Int. Conf. on Artificial Intelligence in Education (AIED 2011), LNAI, vol. 6738, pp. 106–114. Springer Berlin/Heidelberg (2011)
14. Hackett, G., Betz, N, E.: An Exploration of the Mathematics Self-Efficacy/Mathematics Performance Correspondence. J. for Research in Mathematics Education, 20(3), 261–273 (1989)
15. Wallace, R. S.: Artificial intelligence markup language (2010), `http://www.alicebot.org/documentation/`

16. Silvervarg A., Jönsson A.: Subjective and Objective Evaluation of Conversational Agents. In Proc. of the 7th Workshop on Knowledge and Reasoning in Practical Dialogue Systems. Barcelona, Spain (2011)
17. Baylor, A., Plant, E.: Pedagogical Agents as Social Models for Engineering: The Influence of Appearance on Female Choice. In: Proc. of Artificial Intelligence in Education (AIED 2005), pp. 65–72. IOS Press, Amsterdam (2005)
18. Rosenberg-Kima, R., Baylor, A., Plant, E., Doerr, C.: Interface Agents as Social Models for Female Students: The Effects of Agent Visual Presence and Appearance on Female Students' Attitudes and Beliefs. Computers in Human Behavior, 24(6), 2741–2756 (2008)