This manuscript is an 'authors copy' (version) of the original article published in LNCS (Lecture Notes in Computer Science) Volume 10948:

Månsson, K., & Haake, M. (2018). Teaching without learning: Is it OK with weak AI? In C. Penstein Rosé et al. (Eds.), *19th Int. Conf. on Artificial Intelligence in Education (AIED 2018). LNCS: Vol. 10948* (pp. 241-245). Cham, Switzerland: Springer.

Nota Bene: There is no guarantee that this manuscript version is identical to the original published article.

Teaching Without Learning: Is It OK With Weak AI?

Kristian Månsson & Magnus Haake

LUCS, Lund University Cognitive Science, Lund University, Sweden

Abstract. Two different learning models for a teachable agent were tested with respect to perceived intelligence, the protégé effect, and learning in Swedish grade 5 and 6 students. A strong positive correlation was found between perceived intelligence and the protégé effect, but no significant differences were found between the two different implementations of the learning algorithm. The results suggest that while the perceived intelligence of the agent relates to the induced protégé effect, this perceived intelligence did not correspond to the implemented learning model. This, in turn, suggest that a simple learning model can be sufficient for a teachable agent system, but more research is needed.

Keywords: teachable agent, learning model, perceived intelligence, protégé effect, learning outcome.

Introduction

Studies have shown that the 'teachable agent'-paradigm, i.e. *learning-by-teaching* using a Teachable Agent (hence TA) in educational software, benefits learning by increasing students' sense of responsibility and supporting metacognition [1, 2]. The *protégé effect* refers to how students make larger learning efforts when the goal is to teach another (social) agent than when the goal is to learn for themselves [3]. Three mechanisms are suggested to underlie this increase in learning effort: a feeling of responsibility towards the TA, an adoption of an incrementalist view of knowledge, and a protection of the ego (*ego-protective buffer*) since it is the agent who is tested for its learning and who potentially fails.

Elaborating on the TA-paradigm, there is a difference between an agent that can learn and an agent that can be taught [4], suggesting different approaches for the design of the AI in corresponding TA-software. In the learning software *Guardians of History* (GoH), the student has a teacher role and the TA appears to be learning – but is in fact only responding to pre-defined actions from the student-teacher. In this sense, the TA is taught but is not actually learning. Alternatively, an artificial neural network can be trained to make, for instance, discriminations between different breeds of dogs by processing a training set of thousands of images with dogs. In this case the system is in some sense learning, but there is not much teaching involved. This spurs the question of how much effort should be put into the development of the underlying artificial intelligence of a TA? It needs to be teachable, but to what extent does it need to learn versus seem to learn? The well-studied TA system *Betty's Brain* (BB) [5] lies somewhere in-between the two examples above. In BB, the student-teacher is tasked with teaching Betty to make inferences by constructing concept maps representing concepts and functions. In BB, the teaching becomes akin to mapping out a visual representation of a knowledge model. This visual representation, in turn, supports the student's teaching of the agent.

If no such clear visual representation is provided, can the underlying learning model of the TA still invoke a sense of intelligence – and does such perceived intelligence affect the positive learning effects of a TA-system? This paper presents a first pilot study to explore whether the underlying artificial intelligence model of a TA can affect the suggested mechanisms of the protégé effect. In order to pursue this question, the pilot study aims to explore the relation between the protégé effect, perceived intelligence of the TA, and students' learning outcomes by comparing two different implementations of a TA in an educational software.

Method

The Teachable Agent Educational Software

The TA software used in the study is called *Guardians of History* (GoH) and targets middle-school history. It is developed by the Educational Technology Group at Lund University and Linköping University. In GoH, the student helps their TA (suffering from time traveling nausea) to pass a set of history tests by going on time travel missions to gather information (Figure 1). Returning from the time travels, the student engages the TA in different so-called classroom activities (Figure 1). The classroom activities consist of tasks such as building concept maps, sorting propositions and organizing a timeline. The student does the task while the TA observes. After being taught, the TA conducts a test consisting of filling in gaps in sentences (without the help from the student) where s/he provides answers depending on facts s/he has learned. For this study, a subset of the available time travel scenarios was selected and used for two missions.

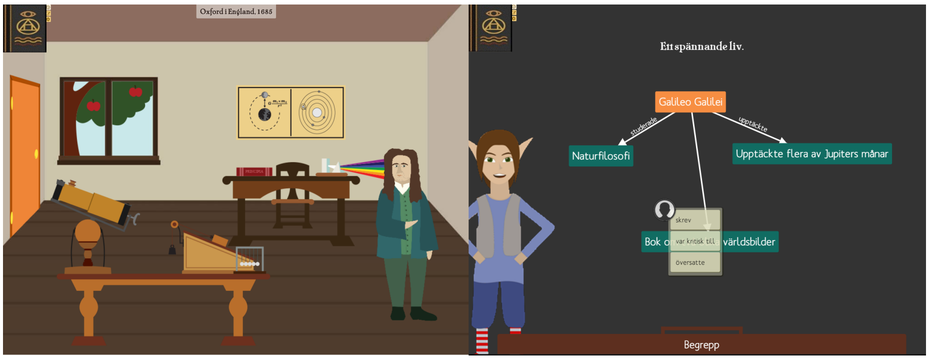

Fig. 1. Example of time-travelling visit at Isaac Newton, Oxford (left) and a classroom activity (right) in the GoH software.

The Learning Models of the Teachable Agent

The TA was implemented with two different learning models: TA_R (as in 'recency') and TA_A (as in 'associative'). The recency setting (TA_R) corresponds to the original implementation of the TA where the agent's learning model reflects the latest facts it has been exposed to in each learning activity, i.e. for every new learning activity, the learning model overwrites all the previous facts learned in previous learning activities – whether they are correct or not. The associative agent (TA_A) [6] is implemented with a basic learning model modulating the agent's certainty of different facts. The certainty of the facts varies depending on the results of the learning activities. Furthermore, the TA_A asks for confirmation of learned facts at random intervals.

Experimental Design

The study involved 94 Swedish grades 5 and 6 students from 5 classes from the same school. The students were randomly assigned to one of two conditions (GoH using the TA_R model and the TA_A model, respectively) of 47 students each. The students were given a short introduction to the game and the characters, whereafter they were instructed to work alone but allowed to ask for help. The whole intervention lasted for two subsequent sessions, each lasting 60 minutes.

A questionnaire addressing perceived intelligence of the TA and the protégé effect was distributed to the students at the end of the second session. Upon finishing the questionnaire, the students were presented with a knowledge assessment. The questionnaire as well as the knowledge test was presented using the same Google Forms format. After everyone in the class had submitted their questionnaires and knowledge assessments, a general group discussion of the experience was conducted.

Prior to the analyses, 9 students were excluded from the dataset due to: language difficulties, not completing the game, or not handing in questionnaires. The resulting data set consisted of $N = 85$ participants (TA_R: 41; TA_A: 44).

Measurements

The protégé effect (PE) was measured with five 5-level Likert items operationalized for studies using GoH [7]. Perceived intelligence (PI) was measured with 6 semantic difference items used for measuring perceived intelligence of a robot [8]. Learning outcome was assessed by a knowledge test consisting of 10 multiple choice questions based on the content of GoH.

Results

All statistical analyzes were performed with the statistical software R version 3.4.3 at an alpha level of 0.05; all effect sizes interpreted according to Cohen [9]. The final dataset consisted of $N = 85$ students. An analysis of grade revealed no significant effects on the measurements and the students were treated as a single population.

An independent samples *t*-test showed no significant difference in PI between the TA_R ($M = 16.4$, $SD = 4.5$) and the TA_A ($M = 17.8$, $SD = 4.3$) conditions ($t(83) = -1.32$, $p = .19$, Cohen's $d = .32$), i.e. there was no significant difference between the TA_R group and the TA_A group with regard to perceived intelligence as measured by the questionnaire items.

A strong positive correlation ($r = .64$, $p < .001$) was found between PI and PE, i.e. the students' scores for perceived intelligence and protégé effect followed each other to a high degree.

A Matt-Whitney's U test displayed no significant difference in learning outcome between the TA_R (*Median* = 6; *Range* = 1-10) and TA_A (*Median* = 6; *Range* = 0-9) conditions ($W = 843$, $p = .60$, Cohen's $d = 0.11$), i.e. neither the TA_R group nor the TA_A group performed better as measured by the knowledge assessment.

No significant correlation between PE and performance score in the knowledge test could be established.

Discussion

The students did not perceive TA_A as more intelligent. The lack of any significant difference between the conditions regarding the perceived intelligence and the protégé effect might, however, point to other factors as eliciting the protégé effect, such as the narrative or the explicitly stated role for the student as the teacher.

The strong positive correlation between perceived intelligence and protégé effect may reflect that one strongly influences the other or that they strengthen each other reciprocally. The strong correlation between how the student either actively ascribe or passively perceive the TA as a thinking and learning agent and the elicited protégé effect, points to one having a strong influence on the other, or to both manifesting an underlying phenomenon. We suggest this is a correlation of interest for researchers as well as designers of TA software.

The lack of a correlation between the protégé effect and the knowledge test scores is somewhat surprising. Yet, as the protégé effect is a theoretical construct aimed at explaining the positive learning outcomes from TAs, it might be too coarse to measure it in the way done in the study, i.e. the result might be sensitive to false negatives. Another possibility is that the measurement does not actually reflect the protégé effect but instead students' general positive – or negative – attitude towards the software.

A long-term study where the students would have more time to interact with the TA might provide further insight towards how the students' perception of the TA vary over time. Validation of the measurements is also of great importance, as they were newly adapted for this study.

References

1. Schwartz, D.L., Chase, C., Chin, D., Oppezzo, M., Kwong, H., Okita, S., Roscoe, R., Jeong, H., Wagster, J., Biswas, G. (2009). Interactive metacognition: Monitoring and regulating a teachable agent. In Hacker, D.J., Dunlosky, J., Graesser, A.C. (Eds.) *Handbook of metacognition in education* (pp. 340–358). New York, NY: Routledge.
2. Biswas, G., Leelawong, K., Schwartz, D., Vye, N., TAG-V (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence*, 19(3-4), 363–392.
3. Chase, C., Chin, D.B., Oppezzo, M., Schwartz D.L.: Teachable agents and the protégé effect: Increasing the effort towards learning. J. Sci. Educ. Technol. **18**(4), 334–352 (2009).
4. Brophy, S., Biswas, G., Katzlberger, T., Bransford, J., Schwartz, D.: Teachable agents: Combining insights from learning theory and computer science. In Lajoie, S.P., Vivet, M. (eds.) Artificial Intelligence in Education, pp. 21–28. IOS Press, Amsterdam (1999).
5. Chin, D., Dohmen, I., Cheng, B., Oppezzo, M., Chase, C., Schwartz, D.: Preparing students for future learning with teachable agents. Educ. Technol. Res. Dev. **58**(6), 649–669 (2010).

6. Bäckström, J., Månsson, K., Persson, M., Sakurai, E., Sundström, F.: Designing and implementing an associative learning model for a teachable agent. In Balkenius, C., Gulz, A., Haake, M., Wallergård, M. (eds.) LUCS 168: Intelligent, socially oriented technology III, pp. 5–13. Lund University, Lund (2017).

7. Kirkegaard, C.: Adding challenge to a teachable agent in a virtual learning environment. Licentiate thesis. Linköping University, Linköping (2016).

8. Bartneck, C., Kuli, D., Croft, E.: Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. Int. J. Soc. Robot. **1**(1), 71–81 (2009).

9. Cohen, J.: Statistical power analysis for the behavioral sciences. Routledge, Oxford (1988).