Hastie, Helen, Lim, Mei Yii, Janarthanam, Srini, Deshmukh, Amol, Aylett, Ruth, Foster, Mary Ellen and Hall, Lynne (2016) I Remember You! Interaction with Memory for an Empathic Virtual Robotic Tutor. In: AAMAS 2016, 9-13 May 2016, Singapore.

# I Remember You! Interaction with Memory for an Empathic Virtual Robotic Tutor

Helen Hastie, Mei Yii Lim, Srini Janarthanam, Amol Deshmukh, Ruth Aylett
School Of Mathematical and Computer Sciences
Heriot-Watt University, Edinburgh
EH14 4AS, UK
{h.hastie, m.y.lim, sc445, a.deshmukh, r.s.Aylett}@hw.ac.uk

Mary Ellen Foster
School of Computing Science,
University of Glasgow, Glasgow
G12 8RZ, UK
maryellen.foster@glasgow.ac.uk

Lynne Hall
Dept. Computing, Engineering and Technology,
University of Sunderland, Sunderland
SR6 0DD, UK
lynne.hall@sunderland.ac.uk

## ABSTRACT

We present a study that investigates the effect of incorporating memory in the interaction for a virtual robotic tutor in terms of helping children achieve a pedagogical goal and the perceived likeability and empathy of the tutor. The domain is a virtual robotic tutor who is guiding and helping learners through a mobile Treasure Hunt exercise that tests their map reading skills. The contribution described in this paper is the discovery that incorporating 'memory' through utterances that recall events from previous interactions significantly increases the learner's ability to perform a pedagogical task. However, the virtual tutor with memory was perceived as less likeable and the instructions given as harder to follow than with a virtual tutor without memory. In addition, there was a significant drop in perceived empathy. This work has a large potential influence in the field of interaction design for agents as one cannot blindly add in human-like features, such as, memory that improve task performance without considering the potential detrimental effects to the perceived empathy and likeability.

## Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

## Keywords

Human-Robot Interaction; Human-Agent Interaction; Empathy; Memory

## 1. INTRODUCTION

This paper describes a study with a robotic and migrating virtual tutor who is able to significantly improve task success of participants on a pedagogical task by recalling previously learned geography skills, experiences and interactions.

**Appears in:** *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016), John Thangarajah, Karl Tuyls, Stacy Marsella, Catholijn Jonker (eds.), May 9–13, 2016, Singapore.*


Effective human teachers are able to engage, motivate and empathise with learners. Part of this effective teaching strategy is building up a long term relationship between teacher and learner, knowing the learner's' strengths and weaknesses and adapting interactions and exercises accordingly. As part of the EMOTE project (http://www.emote-project.eu), we aim to build socially intelligent, empathic tutor robots and virtual agents that could play a long-term role in an educational environment. Therefore, some 'human-like' memory is needed to avoid shallowness of personality [1] and improve the effectiveness of the robot and virtual agent as teaching aids.

Personalised learning environments have been shown to have a positive effect on achievement and attendance rates [2]. Memory, as part of shared social interaction, can help provide such a personalised experience. How this memory should be implemented in an agent and its effect on the perceived overall interaction are research questions addressed in this paper.

### 1.1 Hypotheses

The hypotheses we are testing are fourfold where *with-memory* is the ability of the virtual tutor to remember and mention previously discussed and practised skills and tools.

H1 (TaskSuccess): Participants who interact with the agent with-memory will have higher task *performance* in a pedagogical task than participants who interact with the agent without-memory.

H2 (Likeability): The agent with-memory will be rated higher in terms of general *likeability* than the agent without-memory.

H3 (Empathy): The agent with-memory will be rated more *empathic* than the agent without-memory.

H4 (Empathy Shift): The agent with-memory will cause a *shift* in perceived empathy.

In this paper, we show we can accept H1 and H4 but must reject H2 and H3. This work has a large potential influence in the field of human-agent interaction as it shows that one cannot blindly add in personality features that improve task performance without thinking about the potential detrimental effects to the perceived character of the virtual tutor agent. This is particularly relevant when dealing with vulnerable users such as the young and elderly and may put users off interacting with the system in the long term.

In the remainder of the paper, we will examine previous related work (Section 2); discuss the methodology and experimental set-up and introduce the task (Section 3); we describe both the empathic robot and the touch table map application (Section 4); we then give an overview of the virtual robotic tutor and its interaction behaviour on a mobile Treasure Hunt application (Section 5); results are given in Section 6 and finally we conclude discussing the impact of including memory in the interaction design of both robots and virtual agents for learning environments (Section 7).

## 2. RELATED WORK

Work on declarative memory (both episodic and semantic) for robots and agents has concerned itself with improving planning [3], story telling [4] or improving user experience in domains such as the service domain by remembering user preferences [5]. There is less work with respect to memory for human-robot and human-agent interaction for educational domains and for older children. In this section, we discuss related work including the use of memory for migrating agents and 'social robots', focusing on educational settings.

Memory has been considered for some time as a vital part of successfully migrating agents [6, 7]. [7] showed that a migrating agent that remembers events lived in other embodiments contributes to the user's perception of a consistent identity and showed that those with memory are perceived as more competent. We are not exploring here whether the migrated agent is perceived as the same identity, rather whether using memory adds to the effectiveness of the learning environment or not. [8] shows that there is a trend for people to prefer a companion with selective memory (stores only significant information) as compared to one with absolute memory (stores everything). Similarly, we investigate here the preferences of users towards using declarative memory or not, in an educational setting.

Previous studies on robotic companions in real-world classroom environments [9] have shown that robotic platforms are promising tools for experimental learning. A similar application but with a virtual tutor is described in [10] and [11], where the authors investigated affective vs non-affective feedback. In this study, it was found that the use of a virtual tutor increased the perceived difficulty of the pedagogical task, while the affective virtual tutor's feedback, in particular, made the questions seem more difficult to answer than the non-affective virtual tutor.

One can posit that discussing shared experiences, i.e. having memory, is a 'social behaviour'. How to incorporate social behaviour into robots and virtual agents has been much studied, in particular, for systems aimed at infants (e.g. aged 1.5-2 years of age [12]). There has been less work, however, on older children in learning environments. In some recent work, [13] found that, whilst the presence of a robot can improve learning gain for children aged 7 or 8, this improvement is lost when the robot is 'social', using affective responses, gestures and personalisation. The authors speculate that the affective robot may be a distraction and is viewed more as a teacher in the non-social case, and warn that applying social behaviour to a robot in a tutoring context may have negative effects. Here, we corroborate the essence of their argument that the assumption of adding human-like characteristics, such as affect and memory should not be implicit, rather an in-depth study of how exactly these types of behaviours affect the learning environment is needed. Work described here thus contributes to this much needed study.

As [14] states, older children are less likely to view robots and virtual agents as social actors and therefore would require more complex interactions and social behaviours including memory. However, with this higher level of sophistication comes a greater expectation of capability and if the agent falls short of this then negative feelings may result towards the agent. This may be the case in the study described here with children aged 11-12, where the introduction of a 'social' robot agent (i.e. one that remembers) adversely affects the perceived ease of an educational task and results in a less likeable and empathic agent. As a consequence, this may affect morale and the overall learner experience. In particular, it clearly demonstrates that adding memory to a conversational agent is not a straightforward modification: the agent's behaviour may have an unanticipated, negative effect on the user experience.

## 3. METHODOLOGY

To address the above mentioned hypotheses, we assigned participants to two groups. One group was in the with-memory condition and the second group was in the without-memory condition. The task was in two stages. The first stage, which was the same for both groups, involved doing a short map-based pedagogical exercise on a touch-table with an empathic real-world robot called Susie in the form of an EMYS robot [15] (see Figure 1). In the second stage, Susie was 'teleported' onto a mobile application on a tablet (see Figure 4). For the second stage, the participants went immediately outdoors to perform a real world treasure hunting exercise using the mobile application with the virtual robotic tutor either with or without-memory. During the first stage, the real-world robot is empathic in that it adapts to the user affective state and the current game state. This empathic behaviour is described in further detail below. Post-migration, the virtual robotic tutor responds according to whether the participants answer each question correctly. It does not employ the same empathic strategies as the physical robot in either conditions due to the fact that it is not practically possible to have the relevant sensors during this mobile task.

36 school pupils aged between 11-12 years old participated in this study. Participants were randomly distributed between conditions whilst maintaining a gender balance between groups, however, the majority of participants were boys with 27 boys and 9 girls. The school has a policy for these types of tasks to be done in pairs and therefore the participants were allocated into pairs in consultation with the teacher, maintaining similar overall levels of abilities and compatibility within each pairing. The individuals were told to take it in turns answering questions on the touch table and on the tablet. Participants filled-out individual questionnaires after interacting with the robot and then again after completing the Treasure Hunt.


**Figure 1: Two participants interacting with the physical robot and the touch table**

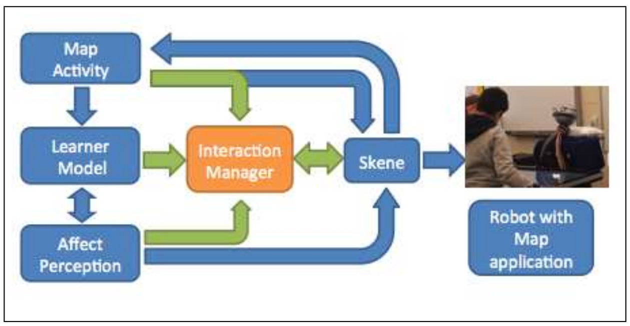


**Figure 2: The Emote Architecture configuration for empathic interaction using the touch table map reading application and the physical robot**

## 4. THE EMPATHIC ROBOT AND THE TOUCH TABLE APPLICATION

Empathy is the psychological processes that make a person feel more congruent with another's circumstances than with his own [16]. In order to embed empathy in learning environments, the tutor needs to be able to perceive, model and reason about the affective states experienced by learners as well as respond emotionally to the situation. As a result, recent research on computer-based learning systems attempts to endow artificial tutors with the capability to perceive the learner's emotional states and incorporate these into pedagogical strategies [17].

Analogously, during the interaction the robotic Susie was empathic, that is, it adapts to the learner's emotional state [18], taking into consideration the learner state and the emotional state of the participant. Figure 2 shows the modules configured for this empathic robotic interaction environment.

The Learner Model functions by checking a learner's answers for correctness and the time to complete and provides an indication of the current skill levels of the learner. The model for the following mapping skills application tracks the skills or competencies: compass reading, map symbol knowledge, and distance measuring. The Perception module takes sensor data from a vision-based facial expression understanding technology (OKAO) [19], which processes information about facial expression into valence/arousal dimensions for affect, giving positive, neutral and negative outputs. The affective state is also inferred by considering the progress of the user in the current tasks, for example many errors may suggest a negative affective state.

These module outputs are used by the Interaction Manager (IM) module, which controls dialogue and other interaction modalities for the robotic tutor at a high level of abstraction [20, 21]. It provides necessary feedback and other helpful pedagogical tactics as the participants proceed through the task. This is the module that deals with the empathic response in which the affective state of the user impacts the actions of the tutor. In consultation with teachers and through examination of human-human pedagogical interactions in the same map reading domain, a set of empathic behaviours including feedback were defined and outlined here.

- High arousal and negative valence suggest a stressed and unhappy user: IM increases pedagogical support;
- Low arousal and negative valence indicate a bored or disengaged user: IM tries to engage and motivate;
- High arousal and positive valence indicate the user is happy and engaged: IM gives less scaffolding support; and
- Low arousal and positive valence indicate a focused user: IM gives less scaffolding support.

Finally, the Skene module translates IM actions into movement of the robot head and eyes, for example, the robotic tutor would look at the relevant point on the map. It also supports semi-automated gaze behaviour, such as gaze at an active speaker or joint gaze at an object of interest [22].

Figure 1 shows a pair of participants interacting with Susie and the touch table application. This interaction was limited to 5 minutes and involved the participants using a map application. The robot presented them with a series of tasks where basic skills relating to the use of compass directions, finding distances, and recognising and using map symbols were involved. The tools used and the skill levels of the pair as they interact are logged. An example of one step in the map reading task would be to find a museum 500 metres north of the railway station.

## 5. THE MOBILE TREASURE HUNT APPLICATION

After the robot interaction activity, the participants are shown the robot 'teleporting' onto a tablet application (see Figure 4) where Susie emulates going to sleep and the virtual robotic tutor on the application 'wakes up', thus following the 'soul-shell' approach [23]. It has been shown in previous studies, as discussed above, that users perceive a single agent across multiple embodiments if the personality of the agent is consistently maintained [24, 25, 26, 27]. Here, the virtual robot's voice and appearance are the same as the physical robot and, therefore, we believe that the participants perceived Susie to be the same character across embodiments.

Information about the pair's performance is also migrated including the level of skill for distance measuring, cardinal directions and their knowledge of ordnance survey symbols.

**Figure 3: Two participants doing a map reading exercise and interacting with the virtual tutor, Susie, on a tablet**


**Figure 4: The Android application for the Treasure Hunt with the virtual robotic tutor**

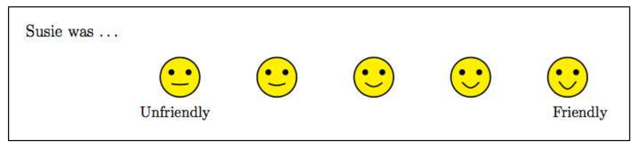


**Figure 5: The smileyometer for the questionnaire**

Whether they used any of the tools (namely map key, distance tool and compass) is also transferred to the tablet.

This real-world Treasure Hunt activity, which has been carried out at a local school for several years using paper, requires the participants to carry out a series of navigation steps in the real world, following a predetermined route on a map (see Figure 3). Each navigation step first requires the participants to walk a few yards while making use of their map-reading skills, and then to answer a series of questions regarding their new location; for example, they might be asked to identify the colour of a door of a house in a specific grid reference. The application includes the virtual Susie head, which presents the navigation instructions, poses the questions, and gives feedback on the correctness of the participants' answers (see Figure 4 for a screenshot of the application).

There are two versions of this application for the two conditions. For the with-memory condition, virtual Susie makes comments with respect to the previous interaction. Example utterances are given here for the with-memory condition: 'Do you remember, we used the compass before on the table to determine which direction you are heading. We can use it again now.' or 'We didn't try using the compass before on the table, but why don't you make use of it now?' and for the without-memory condition 'You can try using the compass to determine which direction you are heading'. The length of the system utterances between conditions and the lexicon used were balanced as much as possible, with the with-memory condition being only slightly longer but not significantly so by an unpaired t-test (mean 10.6 words per utterance for with-memory, compared to 10.1 for without-memory).

## 6. RESULTS AND ANALYSIS

To measure the participants' objective success in the pedagogical task, the following measures were collected:

- **TaskSuccess**: the number of questions answered correctly on the tablet Treasure Hunt;
- **WaypointScore**: the total number of waypoints (11) minus the number of diversions away from the route as calculated using the GPS trace. See Figure 6 for the route and an example GPS trace; and
- **TimeonTask**: the time from starting the real-world Treasure Hunt with the tablet to the final waypoint

The participants' subjective experience was measured through a three-part questionnaire, answered on an individual basis:

- Four questions regarding the participants' opinion of the robotic Susie after interacting with the robot but before the real-world Treasure Hunt;
- Four questions regarding the real-world Treasure Hunt itself after completion of the Treasure Hunt; and
- Ten questions addressing the participants' opinion of the virtual robotic tutor, Susie, asked of the participants after the real-world Treasure Hunt (includes the initial 4 questions repeated).

The items from the first and last part of the questionnaire were taken and modified for children from the Godspeed questionnaire series [28], which is designed to be a standard user measurement tool for human-robot interaction. The items were drawn primarily from the 'likeability' portion of the questionnaire and were rephrased to make them clear for the target age group. In addition, two questions were targeted at determining whether Susie was deemed empathic ('I would describe Susie as soft-hearted' and 'Sometimes Susie felt sorry for me when I was having problems'). These were modified and adapted from the empathy questionnaire devised and reported in [29]. Finally, a question to determine whether Susie seemed to remember things about the participants was included in the final questionnaire.

All questions were presented using a five-point Smileyometer (see Figure 5), which has been shown to be an effective instrument for evaluating child-computer interactions

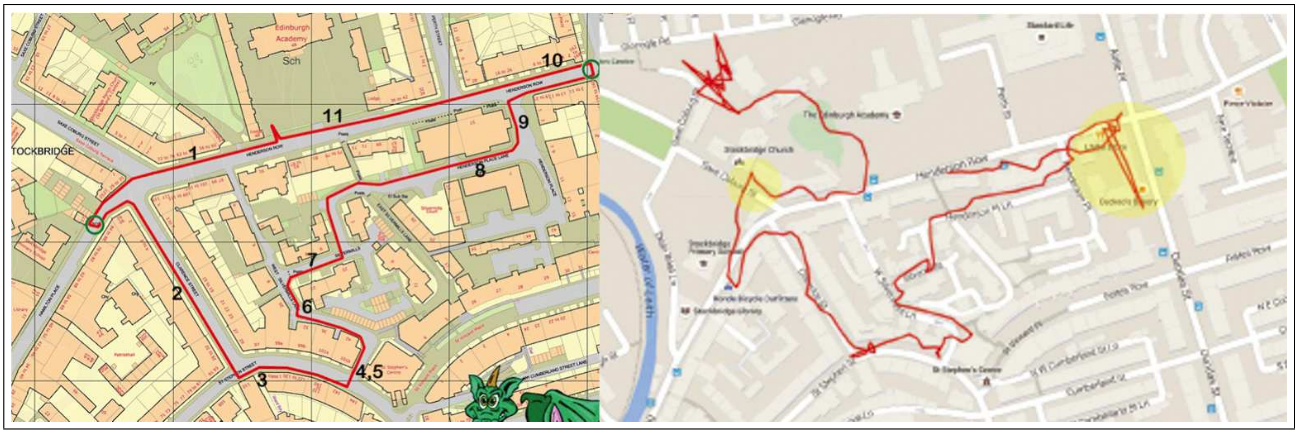


**Figure 6: The intended route of the Treasure Hunt on the left hand side and an example route taken by one pair of participants as logged by the GPS trace on the right hand side with a WaypointScore of 9 out of 11. The yellow areas highlight the two diversions from the Treasure Hunt route**

[30]. Below, we list the exact wording of the questions and a keyphrase for the questions used in this paper (but not given on the questionnaire). The activity was performed in pairs, however the questionnaires were filled out individually.

Although the objective measures, such as task success would be the same for both individuals in the pair, the interaction may have been perceived differently. For example, the subjects may vary in academic ability and confidence and this may affect their perception of the tutor. Indeed, the Kappa coefficient for agreement of the individuals within the pair does vary, with half the groups mostly agreeing with respect to their subjective ratings from the questionnaires (Kappa coefficient of $\geq$0.5) and roughly half the groups disagreeing. Therefore, the following results analysis assumes that each participant is a datapoint with 36 datapoints.

**Questions about Susie BEFORE the real-world Treasure Hunt outside the school**

1. **SusieFriendliness:** Susie was Unfriendly...Friendly
2. **SusieLikeability**: I liked Susie...Not at all...A lot
3. **SusieUnderstandability**: Susie was...Hard to understand...
   Easy to understand
4. **SusieSoftheartedness**: I would describe Susie as softhearted...No, I disagree...Yes, I agree

**Questions about the real-world Treasure Hunt using the tablet**

5. **Fun**: The Treasure Hunt was...No fun at all...Lots of fun
6. **InstructionEase**: The instructions were...Hard to follow...Easy to follow
7. **QuestionEase**: The questions were...Hard to answer...Easy to answer
8. **Groupwork**: I think my group did...Very badly...Very well

**Questions about Susie DURING the real-world Treasure Hunt**

9. **SusieFriendliness:** Susie was Unfriendly...Friendly
10. **SusieLikeability**: I liked Susie...Not at all...A lot
11. **SusieRememberedMe**: Susie remembered things about me...No, I disagree...Yes, I agree
12. **SusieUnderstandability**:Susie was...Hard to understand...Easy to understand
13. **SusieKindness**: Susie was...Unkind...Kind
14. **SusiePleasantness**: Susie was...Unpleasant...Pleasant
15. **SusieAwfulness**: Susie was...Awful...Nice
16. **SusieHelpfulness**: Susie was...Not helpful...Helpful
17. **SusieSoftheartedness**: I would describe Susie as softhearted...No, I disagree...Yes, I agree
18. **SusieSorryforme**: Sometimes Susie felt sorry for me when I was having problems...No, I disagree...Yes, I agree

## 6.1 Task Success Results

With regards H1 (TaskSuccess), we can see from Table 1, in general the with-memory condition results in higher TaskSuccess and WaypointScore, significantly so for the latter ($t(34) = 2.63$, $p<0.05$, unpaired t-test) so we can accept H1. In general, the task took slightly longer using the system with memory, but not significantly so.

## 6.2 Likeability and Empathy Results

For the subjective scores across all of the 14 questions, the without-memory condition was rated higher. This difference was significant for the following questions: InstructionEase, SusieLikeability and SusieRememberedMe ($p<0.05$, unpaired Mann-Whitney U test). Figures 7 and 8 and Table 2 give results for a subset of the questions showing these significant differences. The fact that the ratings for SusieRememberedMe were actually significantly lower for the with-memory condition is counter-intuitive and it was evident by

**Table 1: Objective measures (standard deviation in brackets), where * indicates significance p< 0.05 by a unpaired t-test**

| | TaskSuccess (%) | WaypointScore (out of 11) | TimeOnTask (minutes) |
|---|---|---|---|
| `With Memory` | 88.2(6.4) | **9.89(0.58)*** | 37.33(8.43) |
| `Without Memory` | 83.3(11) | 9.22(1.06) | 34.78(5.98) |

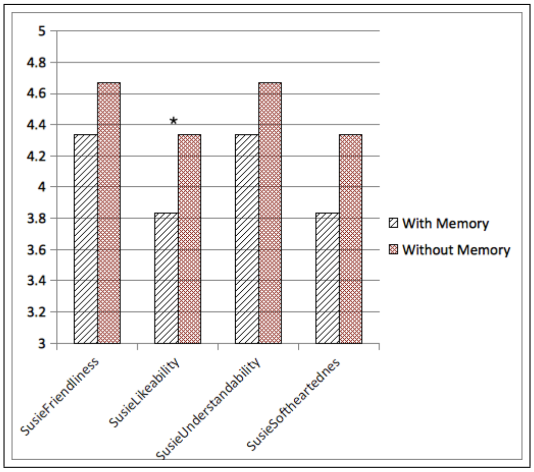


**Figure 7: Graph comparing mean rating on the y-axis for a sub-set of responses for the two conditions. * indicates p< 0.05, unpaired Mann-Whitney U test**

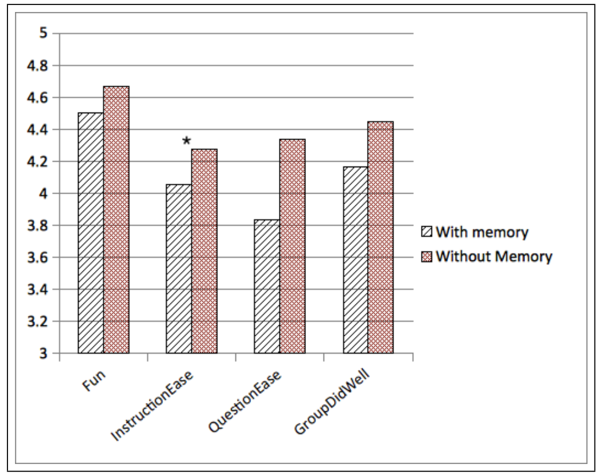


**Figure 8: Graph comparing the mean rating responses on the y-axis to questions asking about the real-world Treasure Hunt activity. * indicates p< 0.05, unpaired Mann-Whitney U test**

interacting with the participants post-experiment that they did not completely understand the question. In fact, some of the participants were referring to the robotic Susie and the fact that the robotic Susie remembered their names rather than the memory of the virtual Susie on the tablet. Therefore, one should not read into this result too much and we will not refer to this measure going forward.

With regards task ease, we can see from Figure 8, the without-memory condition has been rated higher, significantly so for InstructionEase. One reason for this may be the fact that for the with-memory condition, we are asking them to recall facts and memories and therefore the cognitive load may be higher than the without-memory condition.

Recall H3 (Empathy) is "The tutor with-memory will be rated more *empathic* than a tutor without-memory." There are no significant differences between the ratings of SusieSofthearted between the conditions after the real-world Treasure Hunt application (see Table 2), therefore, we have no evidence to accept the H3 hypothesis.

The second empathy question "SusieSorryforme: Sometimes Susie felt sorry for me when I was having problems" was answered once after the entire interaction and rated generally lower than the other questions (mean with-memory is 1.83; mean without-memory is 2.44). There was no significant differences between the conditions. On reflection, this question would be more suited to a long-term study where more problems might arise, which the empathic robotic tutor could help the participants solve. Here, it was observed through qualitative feedback that if the participant did not have any problems then they rated low on the scale for this question which was not the opinion the question was intending to elicit.

With regards H4 (Empathy Shift) "The tutor with-memory will cause a shift in perceived empathy", we are testing whether there is any shift in opinion with regards empathy after the interaction with the virtual robotic tutor. Recall the physical robot, Susie, is designed to be empathic and to test this, we asked the subjects a number of questions immediately after interacting with the robot. The level of empathy is reflected in the subjective scores for questions such as SusieSofthearted where, across all participants (i.e. both conditions), the mean rating is 4.64 out of 5 (with mode and median 5). It is interesting to see if this level of empathy holds when the robot is migrated to a virtual robotic tutor. In fact, the virtual tutor with-memory had a significant *drop in rating* (Wilcoxon Signed-Rank Test, Z=-2.73, p=0.003) for this question post-Treasure Hunt whereas the virtual tutor without-memory did not see a drop (see Figure 9 and Table 3). The inclusion of memory may have a compounded effect on the perception of the empathic tutor and project the tutor as a hard task master, resulting in lower post-experiment ratings. We can, therefore, reject the null hypothesis for H4, having observed a significant, negative effect a tutor with memory has on the perceived empathy of the tutor.

## 7. DISCUSSION AND FUTURE WORK

As expected the use of memory significantly improved the TaskSuccess of participants so we can therefore accept the hypothesis H1. The virtual robot tutor that reminded them of the types of skills that they had learned and the tools they had previously used had a positive effect on the participants in terms of performing the task correctly and getting around the route with fewer diversions.

Table 2: Mean/mode/median rating scores on a 5 point Smileyometer scale for a subset of the questions comparing the answers between the two memory conditions after the real-world Treasure Hunt. * indicates a significant difference (p<0.05 for an unpaired Mann-Whitney U test).

| | SusieFriendliness | SusieLikeability | SusieUnderstandability | SusieSoftheartedness |
|---|---|---|---|---|
| With Memory | 4.33/5/4.5 | 3.83/4/4 | 4.33/5/4 | 3.83/4/4 |
| Without Memory | 4.67/5/5 | **4.33/4/4*** | 4.67/5/5 | 4.33/5/4.5 |

Table 3: Mean/mode/median rating scores on a 5 point Smileyometer scale for the first 4 questions comparing the answers between the two conditions after the interaction with the empathic robot comparing *before* and *after* the real-world Treasure Hunt. * indicates a significant difference comparing consistency of answers for the same question before and after the Treasure Hunt (p<0.01 for Wilcoxon Signed-Rank test)

| | SusieFriendliness | SusieLikeability | SusieUnderstandability | SusieSoftheartedness |
|---|---|---|---|---|
| Before With Memory | 4.5/5/5 | 4.11/4/4 | 3.94/4/4 | **4.5/5/5*** |
| After With Memory | 4.33/5/4.5 | 3.83/4/4 | 4.33/5/4 | 3.83/4/4 |
| Before Without Memory | 4.78/5/5 | 4.33/5/4 | 4.61/5/5 | 4.67/5/5 |
| After Without Memory | 4.67/5/5 | 4.33/4/4 | 4.67/5/5 | 4.33/5/4.5 |

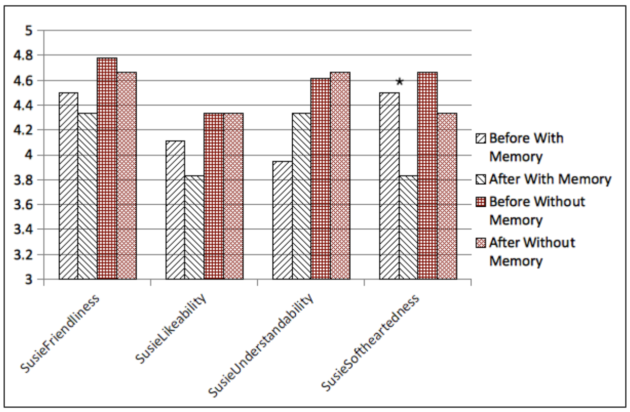


Figure 9: Graph comparing mean rating responses on the y-axis for the same questions before and after the Treasure Hunt. * indicates a significant difference of ratings before/after (p< 0.05 paired Mann-Whitney U test)

Recall hypotheses H2, H3 and H4 are around testing whether memory has a positive effect on general likeability of the tutor (H2) and whether the tutor is deemed as empathic (H3) and if there is any change in perceived empathy (H4). For H2 (Likeability), there is evidence that the participants rated the likeability of using the virtual tutor with-memory significantly lower than the virtual tutor without-memory. As with the study reported in [8] where the users' disprefer absolute memory versus selective memory, it may be that the virtual robotic tutor is almost reminding them too much of what they have done wrong and therefore subjective measures are lower for the memory condition.

With regards empathy (H3 and H4). There is no significant difference between the ratings of the robot between the groups before the participants encounter the virtual tutor. This is to be expected as there is no difference in this part of the experiment across conditions. Post experiment, again there is no evidence of differences between the two conditions, we therefore have to accept the null hypothesis for H3 indicating that adding memory does not increase the level of empathy significantly.

On the other hand, there was a significant drop in the ratings for how 'softhearted' Susie was after interacting with the with-memory agent. As discussed above, it may be that the robot, once perceived as empathic, after the real-world Treasure Hunt was felt to be more of a hard task master, constantly reminding them of skills and the tools that they should know about from the previous interaction, rather than letting them enjoy the experience of being on a Treasure Hunt outside of school.

In summary, we have observed here that there are positive aspects to adding memory in terms of aiding the learners in completing a task. However, the positive effect of empathy observed with the physical robot was lost when the robot migrated to a virtual agent with memory condition. In addition, the with-memory agent was perceived as less likeable and their instructions harder to follow. It is, therefore, clear that one cannot simply add memory to an agent without the possibility of adverse affects in terms of how the agent is perceived to the user. Recent studies have corroborated the idea that by incorporating traits of human interaction such as being social [13] or having memory as discussed here, can have both positive and negative effects on the interaction in learning environments and therefore caution must be taken when developing educational robots with such 'human-like' features.

Our contribution lies in the study of memory as a teaching tool for virtual and robotic tutors, its efficacy and its potential effect on the perceived character of the physical and virtual robot. In the context of the larger research project—which has the overall goal of developing empathic robot tutors—future work includes applying the findings from this study to the other robotic and virtual tutors being developed in the project, taking care to ensure that any affective feedback from the agents has the intended effect on the overall pedagogical and social goals of the interaction.

## 8. ACKNOWLEDGMENTS

This work was supported by the EC, funded by the FP7 ICT-317923 project EMOTE. The authors are solely responsible for the content of this publication. It does not represent the opinion of the EC, and the EC is not responsible for any use that might be made of data appearing therein.

## REFERENCES


[1] K. Dautenhahn, "The art of designing socially intelligent agents - science, fiction and the human in the loop," *Applied Artificial Intelligence Journal, Special Issue on Socially Intelligent Agents*, vol. 12, pp. 12–7, 1998.

[2] Hanover Research, "Best practices in personalized learning environments (Grades 4-9)," 2012. [Online]. Available: http://www.hanoverresearch.com/media/Best-Practices-in-PersonalizedLearning-Environments.pdf

[3] D. Stachowicz and G. Kruijff, "Episodic-like memory for cognitive robots," *Autonomous Mental Development, IEEE Transactions on*, vol. 4, no. 1, pp. 1–16, March 2012.

[4] W. C. Ho, J. a. Dias, R. Figueiredo, and A. Paiva, "Agents that remember can tell stories: Integrating autobiographic memory into emotional agents," in *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems*, ser. AAMAS '07. New York, NY, USA: ACM, 2007, pp. 10:1–10:3.

[5] W. Ho, K. Dautenhahn, M. Lim, and K. Du Casse, *Modelling human memory in robotic companions for personalisation and long-term adaptation in HRI*, ser. Frontiers in Artificial Intelligence and Applications. IOS Press, 2010, pp. 64–71.

[6] M. Lim, R. Aylett, W. Ho, S. Enz, and P. Vargas, "A socially-aware memory for companion agents," in *Intelligent Virtual Agents*, ser. Lecture Notes in Computer Science, Z. Ruttkay, M. Kipp, A. Nijholt, and H. H. Vilhjálmsson, Eds. Springer Berlin Heidelberg, 2009, vol. 5773, pp. 20–26.

[7] R. Aylett, M. Kriegel, I. Wallace, E. M. Segura, J. Mercurio, and S. Nylander, "Memory and the design of migrating virtual agents," in *Proceedings of AAMAS*, 2013, pp. 1311–1312.

[8] M. Y. Lim, R. Aylett, P. Vargas, S. Enz, and W. C. Ho, "Forgetting through generalisation: a companion with selective memory." in *Proceedings of AAMAS*, 2011, pp. 1119–1120.

[9] I. Leite, G. Castellano, A. Pereira, C. Martinho, and A. Paiva, "Modelling empathic behaviour in a robotic game companion for children: An ethnographic study in real-world settings," in *Proceedings of HRI 2012*, 2012, pp. 367–374.

[10] M. E. Foster, A. Deshmukh, S. Janarthanam, M. Y. Lim, H. Hastie, and R. Aylett, "Influencing the learning experience through affective agent feedback in a real-world treasure hunt," in *Proceedings of AAMAS*, 2015, pp. 1711–1712.

[11] M. E. Foster, M. Y. Lim, A. Deshmukh, S. Janarthanam, H. Hastie, and R. Aylett, "Affective feedback for a virtual robot in a real-world treasure hunt," in *Proceedings of the 2014 Workshop on Multimodal, Multi-Party, Real-World Human-Robot Interaction*, ser. MMRWHRI '14. New York, NY, USA: ACM, 2014, pp. 31–32.

[12] F. Tanaka, A. Cicourel, and J. Movellan, "Socialization between toddlers and robots at an early childhood education center," *PNAS*, vol. 104(46), 2007.

[13] J. Kennedy, P. Baxter, and T. Belpaeme, "The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, 2015, pp. 67–74.

[14] R. Ros, M. Nalin, R. Wood, P. Baxter, R. Looije, Y. Demiris, T. Belpaeme, A. Giusti, and C. Pozzi, "Child-robot interaction in the wild: Advice to the aspiring experimenter," in *Proceedings of the 13th International Conference on Multimodal Interfaces*, ser. ICMI '11. New York, NY, USA: ACM, 2011, pp. 335–342.

[15] J. Kedzierski, R. Muszynski, C. Zoll, A. Oleksy, and M. Frontkiewicz, "Emys-emotive head of a social robot," *International Journal of Social Robotics*, vol. 5, no. 2, pp. 237–249, 2013.

[16] M. Hoffman, *Empathy and moral development: Implications for caring and justice*. Cambridge University Press, 2001.

[17] W. Burleson, "Affective learning companions: Strategies for empathetic agents with real-time multimodal affective sensing to foster meta-cognitive and meta-affective approaches to learning, motivation, and perseverance," Ph.D. dissertation, Massachusetts Institute of Technology, Massachusetts, 2006.

[18] A. Deshmukh, A. Jones, S. Janarthanam, H. Hastie, T. Ribeiro, R. Aylett, A. Paiva, G. Castellano, M. Ellen Foster, L. J. Corrigan *et al.*, "An empathic robotic tutor in a map application," in *Proceedings of AAMAS*, 2015, pp. 1923–1924.

[19] S. Lao and M. Kawade, "Vision-based face understanding technologies and their applications," in *Advances in Biometric Person Authentication*. Springer, 2005, pp. 339–348.

[20] S. Janarthanam, H. Hastie, A. Deshmukh, R. Aylett, and M. E. Foster, "A reusable interaction management module: Use case for empathic robotic tutoring," in *Proceedings of SemDial*, 2015.

[21] S. Janarthanam, H. Hastie, A. Deshmukh, and R. Aylett, "Towards a serious game playing empathic robotic tutorial dialogue system," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*, ser. HRI '14. New York, NY, USA: ACM, 2014, pp. 180–181.

[22] T. Ribeiro and A. Paiva, "Make way for the robot animators! bringing professional animators and AI programmers together in the quest for the illusion of life in robotic characters," in *AAAI Fall Symposium Series*, 2014.

[23] P. F. Gomes, E. M. Segura, H. Cramer, T. Paiva, A. Paiva, and L. E. Holmquist, "Vipleo and phypleo: Artificial pet with two embodiments," in *The Proceedings of The 8th International Conference on Advances in Computer Entertainment Technology (ACE)*. Lisbon, Portugal: ACM, November 2011.

[24] M. Kriegel, R. Aylett, P. Cuba, M. Vala, and A. Paiva, "Robots meet ivas: A mind-body interface for migrating artificial intelligent agents," in *Intelligent Virtual Agents - 11th International Conference IVA 2011*, ser. Lecture Notes in Computer Science, H. H. Vilhjálmsson, S. Kopp, S. Marsella, and K. R.

Thórisson, Eds., vol. 6895. Eindhoven, The Netherlands: Springer Berlin / Heidelberg, September 2011, pp. 282–295.

[25] K. Koay, D. Syrdal, K. Dautenhahn, K. Arent, L. Malek, and B. Kreczmer, "Companion migration-initial participants' feedback from a video-based prototyping study," in *Mixed Reality and Human-Robot Interaction*, ser. Intelligent Systems, Control and Automation: Science and Engineering, X. Wang, Ed. Springer Netherlands, 2011, vol. 1010, pp. 133–151.

[26] M. Imai, T. Ono, and T. Etani, "Agent migration: communications between a human and robot," in *Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings. 1999 IEEE International Conference on*, vol. 4, 1999, pp. 1044–1048 vol.4.

[27] K. Ogawa and T. Ono, "Ubiquitous cognition: mobile environment achieved by migratable agent," in *Proceedings of the 7th international conference on Human computer interaction with mobile devices & services*. ACM, 2005, pp. 337–338.

[28] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International Journal of Social Robotics*, vol. 1, pp. 71–81, 2009.

[29] M. H. Davis, "Measuring individual differences in empathy: Evidence for a multidimensional approach," *Journal of Personality and Social Psychology*, vol. 44, pp. 113–126, 1983.

[30] J. Read and S. Macfarlane, "Endurability, engagement and expectations: Measuring children's fun," *Interaction Design and Children*, pp. 1–23, 2002.