Robots Tutoring Children: Longitudinal Evaluation of Social Engagement in Child-Robot Interaction

Citation for the original published paper (version of record):
Serholt, S., Barendregt, W. (2016). Robots Tutoring Children: Longitudinal Evaluation of Social Engagement in Child-Robot
Interaction. Proceedings of the 9th Nordic Conference on Human-Computer Interaction.
http://dx.doi.org/10.1145/2971485.2971536

(article starts on next page)

Robots Tutoring Children: Longitudinal Evaluation of Social Engagement in Child-Robot Interaction

Sofia Serholt
Department of Applied IT,
University of Gothenburg
Gothenburg, Sweden
sofia.serholt@ait.gu.se

Wolmet Barendregt
Department of Applied IT &
The Linnaeus Centre for
Research on Learning,
Interaction and Mediated
Communication in
Contemporary Society (LinCS),
University of Gothenburg
Gothenburg, Sweden
wolmet.barendregt@ait.gu.se

ABSTRACT
This paper explores children's social engagement to a robotic tutor by analyzing their behavioral reactions to socially significant events initiated by the robot. Specific questions addressed in this paper are whether children express signs of social engagement as a reaction to such events, and if so, in what way. The second question is whether these reactions differ between different types of social events, and finally, whether such reactions disappear or change over time. Our analysis indicates that children indeed show behaviors that indicate social engagement using a range of communicative channels. While gaze towards the robot's face is the most common indication for all types of social events, verbal expressions and nods are especially common for questions, and smiles are most common after positive feedback. Although social responses in general decrease slightly over time, they are still observable after three sessions with the robot.

Author Keywords
Human-Robot Interaction; children; implicit social probes; long-term development.

ACM Classification Keywords
H.5.m. Human robot interaction

INTRODUCTION
Advances in the field of robotics in recent years have inspired many researchers to explore the potential applications of robots for education [17, 30]. One such application is that of tutoring robots able to convey learning content to students in a personalized way [20, 23, 27]. When it comes to teacher-student tutoring situations, Bergin and Bergin [7] argue that a secure attachment or social bond between teacher and student is an important prerequisite for academic achievement. This has inspired designers of educational robots to investigate ways to equip robots with empathic capabilities able to recognize, interpret and adapt to students' emotional states [9, 19, 25]. Referred to as *affect sensitivity*, it denotes "the way social affective cues conveyed by people's behaviour can be used to infer behavioural states, such as affective or mental states" [8]. These inferences are then used to carry out a context-appropriate action. It is speculated that this can facilitate a social bond between students and robots similar to that between teachers and students [13].

Indeed, in the field of Child-Robot Interaction (CRI), previous research has suggested that children can ascribe agency to a robot, enabling them to form friendships or a social bond with it even without any empathic component [5, 6, 21, 32]. Unlike other computational objects, Turkle [41] argues that children do not only try to classify robots, they also want to nurture or be nurtured by them. Consequently, "children's focus shifts from cognition to affect, from game playing to fantasies of mutual connection" which she claims causes attachment in children.

One important aspect of establishing a social bond is that of social engagement [35]. "Engagement is the process by which interactors start, maintain and end their perceived connection to each other during an interaction" [38]. In this paper, we set out to explore whether a robot equipped with empathic capabilities is able to elicit and maintain students' social engagement, and how this engagement is expressed.

Furthermore, as social bonding takes time to develop and is considered long-lasting [37], it is important to study whether students' social engagement changes across time. In addition, the novelty effect has been shown to be salient in CRI [20, 26], which means that children have been shown to

NordiCHI '16, October 23-27, 2016, Gothenburg, Sweden

ISBN 978-1-4503-4763-1/16/10…$15.00
DOI: http://dx.doi.org/10.1145/2971485.2971536

lose interest in robots once they grow accustomed to them after some time has passed.

We thus conducted a longitudinal field trial in a Swedish school with a robot tutoring individual students in map reading. By conducting video analysis of socially significant events during the student-robot interactions, we aim to answer the following research questions:

- Do children express social engagement with the robot as a reaction to socially significant events, and if so how?
- Are there different ways of expressing social engagement to different kinds of socially significant events?
- Do children's responses to different socially significant events change over time?

The results of this study can contribute to a better understanding of social bonding between children and robots, and especially of the kind of reactions robots could specifically pay attention to in a social context.

RELATED WORK

In the following subsection we will briefly present research pertaining to how robots with physical embodiments have been shown to differ from virtual agents displayed on computer screens. Signs of engagement in general and towards robots in particular will thereafter be detailed.

From Virtual Agents to Physical Robots

The field of tutoring robots has partly emerged from the success of using virtual agents as a way to provide individualized and personalized support to students within digital learning environments. According to Johnson et al. [18] lifelike virtual characters displayed within virtual learning environments offer the possibility of engaging and motivating students during a learning task through verbal and nonverbal communication. Studies have shown that virtual agents contribute to students' learning experiences in terms of achievement scores, attitudes, retention of learning [44] as well as study outcomes [39].

Research has demonstrated that different levels of embodiment affect users' perceptions of artificial entities. The level of embodiment may range from a static or animated image of a character on a computer screen to a physically present robot. For example, Lusk and Atkinson [29] found that a virtual agent that provided examples and illustrations led to increased learning outcomes for participants compared to static representations of the same agent that did not provide illustrations. Subsequently, experiments comparing virtual agents to physical robots have indicated that robots are preferred in terms of perceived social interaction [3], trust [24], learning gain [28], as well as enjoyment [34].

Signs of Engagement

Both verbal and non-verbal communication are important for social interaction. Interestingly, in Human-Robot Interaction (HRI) researchers have noted that people engage in interactive behaviors although the robot does not necessarily understand or respond to them. For example, Sidner et al. [38] observed that head nodding was a frequently occurring communicative response among adults interacting with a robot although the robot could not react to it. They thus concluded that head nodding occurred naturally and automatically in conversation with a robot in such a way that it was similar to human-to-human conversation.

Yet, people's social responses are not limited to robots that emulate human social behaviors, but rather have been observed in human-computer interaction as well. In line with their media equation theory, Reeves and Nass [36] argued that people tend to treat computers running very simple software programs as social actors too. This, they argued, was made evident in their numerous experiments when people e.g., engaged in polite and reciprocal behaviors towards computers, or applied human stereotypes to computers such as a preconception that a car navigation system using a female voice was not a credible means to acquire directions [31]. The media equation theory postulates that interaction with media is the same as interaction in real life. Nass and Moon [31] claimed that such tendencies to treat computers as social actors should not be confused with anthropomorphism, which entails attributing and rationalizing the behavior of animals, computers or other inanimate objects based on human characteristics [14]; but rather that such social responses occur mindlessly.

In a similar vein, Takayama [40] argued that people's perceptions of agency in robots occur on different levels: There are mindless (or in-the-moment) perceptions of robots, and there are reflective beliefs about robots. How people respond socially to robots may not go hand-in-hand with their stated beliefs about robots' ontological status. Therefore, there is a disconnect between people's claimed beliefs about robots and their social behavior towards robots, much like the disconnect observed by Reeves and Nass [36] when asking their experimental subjects to state their beliefs about the social status of computers. Interestingly the subjects argued that they would never be polite to computers although the evidence suggested otherwise.

Returning to physical signs of engagement, eye-contact is an important part of social interaction. According to Argyle and Dean [2] people look in each other's eyes repeatedly for short periods of 3 to 10 seconds, especially when listening. If there is no eye-contact, people generally do not feel that there is a full communication, and if the eye-contact is longer it makes them feel anxious. There is more eye-contact between two people if they like each other. Eye-contact has been shown to suggest engagement in HRI as well [1, 38]. In the field of CRI, Okita et al. [33] explored how various levels of affective behaviors in a robot affected interactions between children ages 4-10 and an Asimo robot. Through initial pilot studies, the authors concluded that children make eye contact "when they show interest, seek attention, have questions, want approval, and express emotions (e.g., excitement, boredom)".

According to Castellano et al. [10] smiles can be a sign of engagement in CRI. In their study on children's engagement with a chess-playing robot iCat, they found that children tended to smile at the iCat more when they were engaged with it than when they were not engaged with it.

According to Vacharkulksemsuk and Fredrickson [42] behavioral synchrony plays a role in fostering embodied rapport, where pairs of strangers showed more mirroring behaviors in self-disclosure-tasks, and who in turn rated their social interaction more positively, mutually, and vitally. Indeed, mirroring or mimicry can be a sign of empathy [12].

STUDY

The longitudinal field trial took place in a Swedish primary school over a period of 3.5 months. Prior to the study, ethical approval was obtained from the university along with students' parental consent. Students were informed that they could opt out at any time should they not wish to continue with the study. The tasks that the robot was able to perform were, on the one hand, teaching map reading skills to individual students, and on the other hand, playing a sustainability game with pairs of students. For this paper, we analyzed three socially significant events on the individual task where students were to navigate on a map. The setup and task will be described next.

Setup

The robot setup was developed within the EMOTE-project[1] and consisted of an emotionally perceptive Aldebaran Nao T14[2] torso attached to a MultiTaction[3] table, as shown in Figure 1. Various sensors such as Microsoft Kinect[4] and OKAO vision software[5] were used to gather information about students' current emotional state based on valence and arousal [16]. The setup was placed within a small room adjacent to one of the classrooms at the school. Interaction sessions always took place during ordinary lessons.

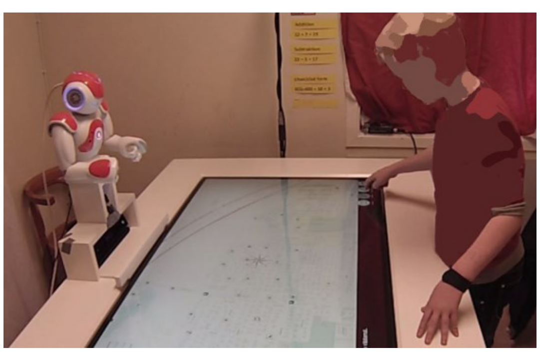

Figure 1. Interaction session displaying set-up of Nao T14 torso attached to the MultiTaction table.

The individual map task was designed as a treasure hunt where the students were instructed to follow a trail by clicking on appropriate map symbols through scaffolding provided by the robot. There were also digital map reading tools within the task in the form of a compass, map key and measuring tool that students were encouraged to use. Depending on the robot's perception of students' current learning or emotional state, the pedagogical strategy varied.

The scaffolding principle was based on Vygotsky's notion of students' zones of proximal development (ZPD) [43], and specifically designed through observations of practicing teachers' scaffolding behaviors on paper-based mock-up studies with students.

In addition to the pedagogical strategy, the robot initiated each session with implicit social probes. A probe is here defined as "a non-intrusive, pervasive and embedded method of collecting informative data at different stages of an interaction" [11]. The probes are defined as social as they function as invitations for social rather than task-based interactions. Specifically, the probes consisted of a **greeting**, e.g., *"Hello [student]! Nice to see you again"*, followed by **feedback/praise** from the previous session, e.g., *"I remember that last time you were very good at [skill]"* (This probe was not present in the first session). Then the robot provided information about the task and the map tools to serve as an introduction or reminder, which was followed by the final probe in the form of a **question** to initiate the task, e.g., *"Are you ready to begin?"* Students were not required to answer this question as the task started after a short time-out anyway. The greeting, feedback and question exemplified here served as three different implicit social probes.

Participants

Although a total of 43 students took part in the field trial, the sample (n=30) of participating students chosen for this paper consisted of those who had completed at least three sessions with the map task. The participants were aged 10-13 (M = 11.4, SD = 0.86, 10 girls, 20 boys).

Procedure

Prior to the study, students were informed that the robot could not understand speech, rather it tried to understand them based on their facial expressions as well as their actions within the educational task. Before each session, students were informed that they could summon the researcher if they needed help or had technical difficulties. They were also asked whether they had any questions.

The researcher always initiated the task from inside the room to make sure that everything was working before stepping outside into the classroom in order to not disrupt the session.

[1] www.emote-project.eu

[2] www.aldebaran-robotics.com

[3] www.multitaction.com

[4] www.microsoft.com

[5] www.omron.com

Each session varied in length depending on how much time the student needed to accomplish the task. Some sessions took only 10-15 minutes whereas others could take up to 40 minutes.

Analysis

To study children's social behaviors towards the robot, we conducted video analysis of three moments in time containing the three implicit social probes conveyed by the robot at the beginning of three consecutive interaction sessions. This means that we had potentially 30x3 greetings, 30x2 feedback utterances, and 30x3 questions, totaling 240 implicit probes to analyze. However, since some implicit feedback probes were not delivered properly due to technical problems, and two students chose to withdraw from the study, the following number of probes are missing: 3 greetings, 9 feedback probes and 3 questions. Thus, our material consists of 225 reactions to implicit probes.

As we were interested in studying students' verbal and non-verbal responses to the robot's social probes, the first author began by viewing all videos in order to develop a preliminary coding scheme based on students' behaviors within the categories eye gaze, facial expression, verbal response and gesture. Thereafter, joint video analysis was undertaken by both authors in order to be able to discuss potential disagreements as they arose. During the joint analysis, the coding scheme was updated when needed.

Finally, within each of these categories it was determined which indicators would signify social engagement with the robot, and which indicators would signify a lack of engagement based on previous literature.

Gaze: Following Argyle and Dean [2] all types of gaze that included the robot's face were considered a sign of social engagement, while all other types of gaze not including the robot's face were considered a sign of no engagement.

Facial expression: Following Castellano et al. [10] smiles were considered to signify engagement with the robot. Timid or flushed smiles as a reaction to the robot's positive feedback were also considered a sign of engagement. All other facial expressions, such as nervous or confused expressions were not considered a sign of social engagement.

Verbal response: If children reacted verbally to the robot's greeting, praise or question this was generally considered a sign of social engagement. The only exception was 'What?' as it signified mainly that the child was unable to hear the robot properly.

Gesture: Greeting gestures such as waving, answering gestures such as nods or headshakes, and victory gestures as a result of feedback/praise were all considered signs of social engagement. Furthermore, mirroring behaviors during the implicit probes were also considered more subconscious signs of engagement, as suggested by Vacharkulksemsuk and Fredrickson [42].

The final coding scheme is presented in Table 1.

	Positive indications	**Negative indications**
Gaze	Robot face Robot face-robot hand alternating Robot face-table alternating Robot face-researcher alternating	Robot hand Researcher Table Elsewhere Table-researcher alternating Robot hand-table alternating
Facial expression	Smile Flushed	Wide-eyed "nervous" expression Raised eyebrows "confused" expression Grimace Serious expression
Verbal response	"Hello" "Thank you" "Yes" or "Okay" "No"	"What?" None
Gesture	Wave Nod Hand movement mirroring robot movement Headshake Victory gesture	None

Table 1. The final coding scheme used during video analysis

RESULTS

In this section we will present our findings with our research questions as starting points in separate subsections. In the subsequent section we will interpret our findings and discuss the implications for the use of social cues by a robot in educational contexts.

Expressions of Social Engagement

The first research question is in some way already answered with our coding scheme. While the main categories of the coding scheme were based on the literature, and the possible subcategories were informed by literature on social engagement, the actual subcategories present were those found in our data. In Figure 2 we show in more detail which expressions of social engagement were used most often for each of the communicative channels: gaze, facial expressions, verbal responses, and gestures.

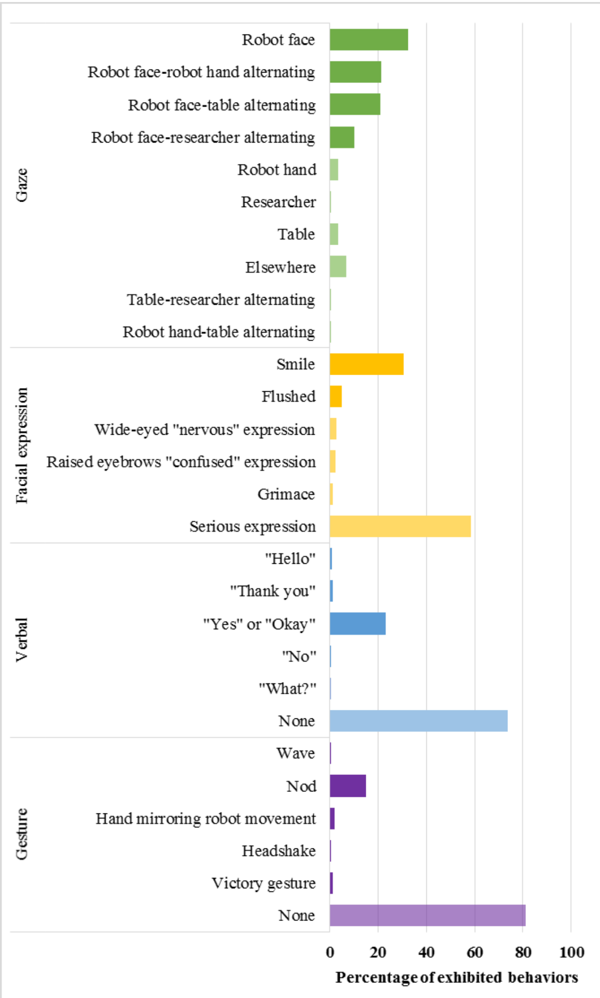

Figure 2. Expressions of engagement exhibited by students for each category displayed in %.

In terms of gaze, students quite often looked at the robot's face. However, we observed that their gaze often temporarily shifted to the robot's hand when it was in motion. This seemed related to the noise produced by the robot's motors or that students were surprised by the robot's waving gesture during the initial greeting.

The brief presence of the researcher at the beginning of the session often caused students to divert their gaze temporarily to the researcher once the robot started speaking. This seemed to be related to sharing expressions of excitement with the researcher as the robot starting speaking, but it could also sometimes be the case that they were simply distracted by the researcher leaving the room. Although it was common that students alternated their gaze between the robot and the task on the table, it was not as common for students to look elsewhere in the room.

When it came to facial expressions, students most frequently smiled or looked serious. When smiling, this was usually accompanied by gazing at the robot's face, or alternating between the robot and the researcher. Serious expressions were interpreted in several different ways depending on the particular student. In some cases, it seemed related to the personality of the student, i.e., some were not quick to smile, in which case they may have nodded towards the robot instead. In other cases, the serious expression seemed to be related to suspiciousness of the robot, i.e., a non-willingness to engage socially with a "machine". In yet other cases, it seemed to be related to the presence of the cameras, not wanting to appear foolish by interacting with a robot.

As explained previously, students sometimes exhibited timid or flushed smiles as though they were emotionally touched by the robot's positive feedback. In some rare cases, students looked nervous or confused at the beginning of the interaction, which was usually accompanied by gazing at the researcher.

Another facial expression which was only observed a couple of times was in the form of a grimace in the general direction of the robot. This was accompanied by a verbal outburst, seemingly in excitement about the fact that the task was about to start.

The most frequent verbal responses were those signaling understanding or agreement, e.g., *"Yes"* or *"Okay"*. It was quite rare that students verbally greeted the robot by saying *"Hello"*. It should be noted, however, that there might not have been a long enough pause between the robot's greeting and its positive feedback, possibly making students more prone to listening to what the robot was about to say next rather than to greet the robot.

Gestures in the form of head-nodding was quite common during the social probes. We found that head-nodding (or headshake) was often times accompanied by a verbal *"Yes"* or in one instance *"No"*. There were furthermore a few instances of mirroring gestures, i.e., when the robot waved at the students, or when it raised its arm over the screen. On these occasions, students sometimes produced a similar gesture themselves in what seemed to be an unconscious manner. Furthermore, some students exhibited victory gestures which seemed to signal excitement about the forthcoming task or that they were proud about being praised for their performance in a previous session.

Social Engagement Behaviors per Type of Probe

Since the three implicit probes are different in nature, we expected children's reactions to these probes to also differ. Figure 3 shows the communicative channels used for each of the different probes.

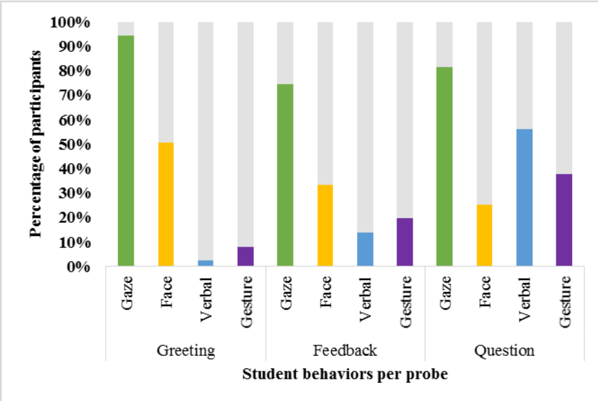

Figure 3. Percentage of positive reactions to different implicit probes across the different communicational channels.

This figure shows that for all types of probes, children most often showed engagement by gazing at the robot, and in several cases they also displayed a smile. Only for the question whether the child was ready to start, a verbal response was more common than a smile, and a nod was also observed regularly.

Changes in Expression of Social Engagement over Time

Figure 4 shows children's reactions to the robot's greeting over the three sessions. For each of the communicative channels we see that the observable indications of engagement decreased slowly, but they were still present.

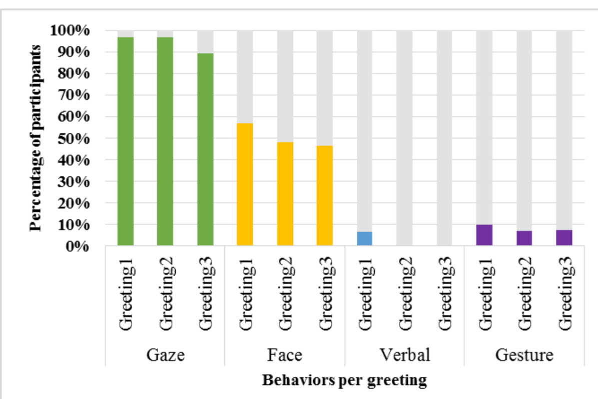

Figure 4. Students' reactions to the robot's greetings over time.

Figure 5 shows children's reactions to the robot's feedback over the two sessions in which this feedback was given. While children still gazed quite often at the robot's face when the robot told them how well they had done the last time, there were several children that only showed a timid smile and kept looking down at the interactive table. In one case where the child showed this timid behavior, a parent later commented about how happy the child had been about this praise.

Although there was some decrease in facial and verbal expressions, both the gazing behavior and the gestures seemed to be at a rather constant or even slightly increasing level.

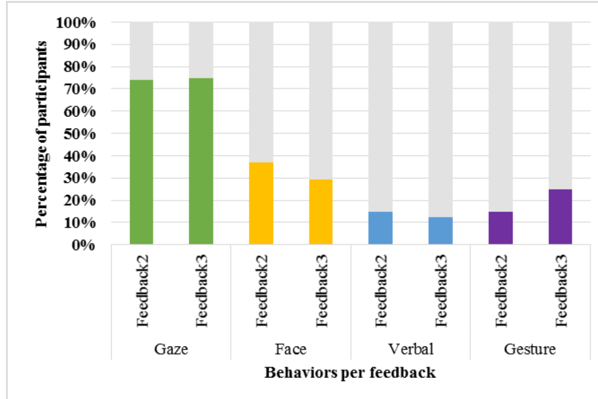

Figure 5. Students' reactions to the robot's feedback over time.

Figure 6 shows children's reactions to the robot's question whether they were ready to start over the three sessions. Although the indications of engagement decreased somewhat from the first to the third session for each of the communicative channels, this was not as obvious as for the greeting probe. It is also clear that the questions consistently provoked indications of engagement over a range of communicative channels. A typical reaction was to look at the robot's face, answer *'Yes'*, and nod slightly.

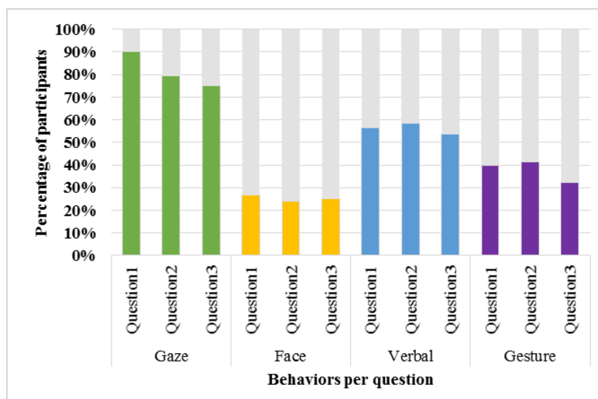

Figure 6. Students' reactions to the robot's questions over time.

DISCUSSION

The results presented above show that children indeed express engagement with a social robot, which could either be understood as a developing social bond [5, 6, 21, 32], or evidence of in-the-moment mindless reactions to the robot as a perceived social actor [31, 40]. They do this through different communicative channels, of which gaze is the most prominent one for all kinds of social probes. Smiles are most often related to greetings, while verbal expressions and gestures (especially nods) are more common as a reaction to a question.

Acknowledging positive feedback through verbal or gestural responses, often coupled with smiling or timid smiling indicates that the students were affected by the robot's praise. Previous research has shown that adults are susceptible to praise and flattery conveyed by a computer even when they are informed that the flattery is not contingent on anything the person in question does: "Praise makes people feel better about themselves, their performance, the interaction and the computer offering the evaluations" [15]. In contrast, during the course of our field trial, students were quite keen on asking the researcher whether the praise they received from the robot was contingent on their performance or not. This was interpreted by the researcher as a way to ascertain that they were singled out from the other students, i.e., that their relationship with the robot was exclusive for them.

While, as suggested by Leite et al. [26], the wearing off of the novelty effect is visible as an overall decrease of behavioral indications in all communicative channels, they do not disappear completely. Indeed, in our material it was obvious that children continued to respond to the implicit probes, although they appeared to become more focused on starting the actual educational task with the robot. This is not necessarily an unwelcome development, as too much focus on social contact could hamper children's learning [22]. Nevertheless, it also suggests that the social interaction with the robot became less important for children as time passed, possibly when they realized that the robot could not respond to their attempts at social interaction. Although several children seemed to hang onto the fantasy despite this knowledge, e.g., as in the case of one student who said to the robot, *"I know you can't hear me, but I'm talking to you anyway,"* there were also children who told the researcher that they still believed that the robot could understand them due to its timely responses. Indeed, Belpaeme et al. [4] argued that "lacks in artificial processing and in generating appropriate responses, often go undetected by young users".

For those few children who went from being highly interactive with the robot to completely ignoring the robot, it seemed to be related to a disappointment due to unmet expectations of what robots should be able to do. As Belpaeme et al. [4] emphasized: "users often expect the robot to have the same perceptual modalities as the user has, and it are these modalities that have proven to be very hard to realise artificially". Although the authors argued that adults' (unmet) expectations of robots can sometimes cause problems when conducting field trials, their experience suggests that children are normally not noticeably troubled by it. In our study, those who became notably less socially engaged were mainly the older children (13-year-olds).

In order to improve social bonding to facilitate learning with robots it could thus be beneficial to develop responsive behaviors for robots to pick up indications in all different communicative channels. However, implementing responsive behaviors in robots is not straightforward. Technical feasibility causes different combinations of modalities to be explored. For example, facial and physiological perception are often employed to interpret children's affective states and reactions [10]. An educational task on e.g. a tablet can also provide information on children's struggles with learning material. Although these modalities provide valuable information *about* the child, they do not allow the child to actively communicate their needs through speech or gestures, which is important for a social connection to take place. Indeed, speech recognition (especially with children) is currently not technically feasible for long-term interactions. Yet, it is nevertheless important for the future of CRI to consider what modes of communication come naturally to children.

We recognize that the uneven distribution of participants' gender (20 boys and 10 girls) may have influenced the results of the study. Although we have not accounted for gender differences in this study, it should be mentioned that our video analysis did not reveal any obvious signs of gender differences in expressions of social engagement towards the robot.

CONCLUSION

In this paper we set out to explore whether a robot equipped with empathic capabilities is able to elicit and maintain students' social engagement, and how this engagement is expressed. In particular, we conducted video analysis of three types of socially significant events during student-robot interactions, aiming to answer the following research questions:

- Do children express social engagement with the robot as a reaction to socially significant events, and if so how?
- Are there different ways of expressing social engagement to different kinds of socially significant events?
- Do children's responses to different socially significant events change over time?

Regarding the first question, our analyses indicate that many children indeed show social engagement as a reaction to socially significant events initiated by the robot. The coding scheme developed based on observable behaviors of the children in our video material shows that those reactions cover several communicative channels. Since we have not been able to find a similar coding scheme using all these channels, we argue that our coding scheme in itself is a contribution for other researchers wishing to analyze children's behavioral reactions to social interactions with robots. However, we are aware that the coding scheme could be expanded by analyzing children's reactions to additional implicit probes, for example pointing behavior of the robot.

Regarding our second question, we can conclude that children use slightly different communicative channels to respond to different socially significant events. When it comes to greetings, facial expressions seem to constitute the primary channel for communication. If a robot could interpret facial expressions at the beginning of a task, it could potentially deduce whether students are comfortable or not with the task or with the robot and subsequently adapt its social behavior accordingly.

Interestingly, although the students in our study were informed that the robot would not be able to understand speech or nodding, they kept engaging in this mode of communication when asked whether they were ready to start the task. This resonates the findings of Sidner et al. [38] who concluded that adults naturally and automatically nod their heads in conversation with a robot, similar to human-to-human conversation. Moreover, it suggests that children treat robots as social actors similar to how adults have been shown to treat computers [36]. Developing robots capable of understanding verbal and gestural communication therefore seems promising for the future of CRI.

Regarding our third research question, it was found that all channels of communication reduced as time passed and the novelty effect wore off. As Salter Ainsworth [37] points out, when humans develop bonds to other human beings, this should be considered long-lasting. Perhaps then we might conclude that children do not develop bonds to robots in the human sense, but that they rather engage in some different sort of relationship with them. Without dismissing the possibility that children's responses to robots are what Nass and Moon [31] consider "mindless" expressions of overlearned social behaviors such as politeness, the nature of this child-robot relationship could perhaps be better understood in relation to how children might engage with robots in the presence of other children or teachers to which they might have some social bond already. Indeed, Kanda et al. [20] observed that children preferred interacting with robots in groups. We therefore consider that further work should entail studying CRI from the perspective of a triad in addition to the dyad which has been reported here. In future work we thus plan to study potential interaction differences between the individual map activity that was analyzed in this paper, and the collaborative sustainability game mentioned earlier.

ACKNOWLEDGMENTS

We thank all the students, teachers and staff at Leteboskolan (www.leteboskolan.se) for their participation, cooperation and support throughout the study. We would also like to thank our collaborating partners and technical team in the EMOTE-project for making this study possible. This work was partially supported by the European Commission (EC) and was funded by the EU FP7 ICT-317923 project EMOTE (www.emote-project.eu). 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. Anzalone, S.M., Boucenna, S., Ivaldi, S., and Chetouani, M. Evaluating the Engagement with Social Robots. *International Journal of Social Robotics 7* (2015), 465-478. DOI:10.1007/s12369-015-0298-7.
2. Argyle, M. and Dean, J. Eye-Contact, Distance and Affiliation. *Sociometry 28*, 3 (1965), 289-304.
3. Bainbridge, W.A., Hart, J.W., Kim, E.S., and Scassellati, B. The Benefits of Interactions with Physically Present Robots over Video-Displayed Agents. *International Journal of Social Robotics 3*, 1 (2011), 41-52. DOI:10.1007/s12369-010-0082-7.
4. Belpaeme, T., Baxter, P., De Greeff, J., Kennedy, J., Read, R., Looije, R., Neerincx, M., Baroni, I., and Zelati, M. Child-Robot Interaction: Perspectives and Challenges. In *Social Robotics*, Herrmann, G., Pearson, M., Lenz, A., Bremner, P., Spiers, A. and Leonards, U. eds. Springer International Publishing (2013), 452-459..
5. Belpaeme, T., Baxter, P., Read, R., Wood, R., Cuayahuitl, H., Kiefer, B., Racioppa, S., Kruijff-Korbayova, I., Athanasopoulos, G., Enescu, V., Looije, R., Neerincx, M., Demiris, Y., Ros Espinoza, R., Beck, A., Canamero, L., Hiolle, A., Lewis, M., Baroni, I., Nalin, M., Cosi, P., Paci, G., Tesser, F., Sommavilla, G., and Humbert, R. Multimodal Child-Robot Interaction: Building Social Bonds. *Journal of Human-Robot Interaction 1*, 2 (2012), 33-53. DOI:10.5898/JHRI.1.2.Belpaeme.
6. Beran, T. and Ramirez-Serrano, A. Can Children Have a Relationship with a Robot? In *Human-Robot Personal Relationships*, Lamers, M. and Verbeek, F. eds. Springer Berlin Heidelberg (2011), 49-56..
7. Bergin, C. and Bergin, D. Attachment in the Classroom. *Educational Psychology Review 21*, 2 (2009), 141-170. DOI:http://dx.doi.org/10.1007/s10648-009-9104-0.
8. Castellano, G., Leite, I., Pereira, A., Martinho, C., Paiva, A., and Mcowan, P.W. Affect Recognition for Interactive Companions: Challenges and design in real world scenarios. *Journal on Multimodal User Interfaces 3*, 1–2 (2010), 89–98.
9. Castellano, G., Paiva, A., Kappas, A., Aylett, R., Hastie, H., Barendregt, W., Nabais, F., and Bull, S. Towards Empathic Virtual and Robotic Tutors. In *Artificial Intelligence in Education*, Lane, H.C., Yacef, K., Mostow, J. and Pavlik, P. eds. Springer Berlin Heidelberg (2013), 733-736..
10. Castellano, G., Pereira, A., Leite, I., Paiva, A., and Mcowan, P.W. Detecting User Engagement with a Robot Companion Using Task and Social Interaction-based Features. In *Proc. ICMI-MLMI'09*, (2009), .
11. Corrigan, L.J., Basedow, C., Küster, D., Kappas, A., Peters, C., and Castellano, G. Mixing implicit and explicit probes: finding a ground truth for engagement

in social human-robot interactions. In *Proc. 2014 ACM/IEEE International Conference on Human-robot Interaction*, ACM (2014), 140-141. DOI:10.1145/2559636.2559815.

12. Datyner, A.C., Richmond, J.L., and Henry, J.D. The development of empathy in infancy: insights from the rapid facial mimicry response. *Frontiers in Human Neuroscience* (2013). DOI:10.3389/conf.fnhum.2013.212.00180.
13. Deshmukh, A., Castellano, G., Kappas, A., Barendregt, W., Nabais, F., Paiva, A., Ribeiro, T., Leite, I., and Aylett, R. Towards Empathic Artificial Tutors. In *Proc. 8th ACM/IEEE international conference on Human-robot interaction*, IEEE Press (2013), 113-114.
14. Duffy, B.R. Anthropomorphism and the social robot. *Robotics and Autonomous Systems 42*, 3–4 (2003), 177-190. DOI:http://dx.doi.org/10.1016/S0921-8890(02)00374-3.
15. Fogg, B.J. and Nass, C. Silicon sycophants: the effects of computers that flatter. *International Journal of Human-Computer Studies 46*, 5 (1997), 551-561. DOI:http://dx.doi.org/10.1006/ijhc.1996.0104.
16. Hall, L., Hume, C., Tazzyman, S., Deshmukh, A., Janarthanam, S., Hastie, H., Aylett, R., Castellano, G., Papadopoulos, F., Jones, A., Corrigan, L., Paiva, A., Alves-Oliveira, P., Ribeiro, T., Barendregt, W., Serholt, S., and Kappas, A. Map Reading with an Empathic Robot Tutor. In *Proc. HRI 2016: International Conference on Human-Robot Interaction (Extended abstracts)*, (2016), .
17. Han, J. Emerging Technologies: Robot Assisted Language Learning. *Language Learning & Technology 16*, 3 (2012), 1-9.
18. Johnson, L., Rickel, J., and Lester, J. Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments. *International Journal of Artificial Intelligence in Education 11* (2000), 47-78.
19. Jones, A., Küster, D., Basedow, C., Alves-Oliveira, P., Serholt, S., Hastie, H., Corrigan, L.J., Barendregt, W., Kappas, A., Paiva, A., and Castellano, G. Empathic Robotic Tutors for Personalised Learning: A Multidisciplinary Approach. In *Proc. International Conference on Social Robotics*, Springer International Publishing (2015), 285-295.
20. Kanda, T., Hirano, T., Eaton, D., and Ishiguro, H. Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial. *Human-Computer Interaction 19* (2004), 61-84.
21. Kanda, T., Sato, R., Saiwaki, N., and Ishiguro, H. A Two-Month Field Trial in an Elementary School for Long-Term Human-Robot Interaction. *IEEE Transactions on Robotics 23*, 5 (2007), 962-971. DOI:10.1109/TRO.2007.904904.
22. Kennedy, J., Baxter, P., and Belpaeme, T. The Robot Who Tried Too Hard: Social Behaviour of a Robot Tutor Can Negatively Affect Child Learning. In *Proc. 10th Annual ACM/IEEE International Conference on Human-Robot Interaction*, ACM (2015), 67-74. DOI:10.1145/2696454.2696457.
23. Kennedy, J., Baxter, P., Senft, E., and Belpaeme, T. Social Robot Tutoring for Child Second Language Learning. In *Proc. 11th Annual ACM/IEEE International Conference on Human Robot Interaction*, (2016), .
24. Kidd, C.D. 2003. Sociable Robots: The Role of Presence and Task in Human-Robot Interaction. In *Georgia Institute of Technology* Massachusetts Institute of Technology..
25. Leite, I., Castellano, G., Pereira, A., Martinho, C., and Paiva, A. Long-Term Interactions with Empathic Robots: Evaluating Perceived Support in Children. In *Social Robotics*, Ge, S., Khatib, O., Cabibihan, J.-J., Simmons, R. and Williams, M.-A. eds. Springer Berlin Heidelberg (2012), 298-307..
26. Leite, I., Martinho, C., and Paiva, A. Social Robots for Long-Term Interaction: A Survey. *International Journal of Social Robotics 5*, 2 (2013), 291-308. DOI:10.1007/s12369-013-0178-y.
27. Leyzberg, D., Spaulding, S., and Scassellati, B. Personalizing robot tutors to individuals' learning differences. In *Proc. 2014 ACM/IEEE international conference on Human-robot interaction*, ACM (2014), 423-430. DOI:10.1145/2559636.2559671.
28. Leyzberg, D., Spaulding, S., Toneva, M., and Scassellati, B. The Physical Presence of a Robot Tutor Increases Cognitive Learning Gains. In *Proc. 34th Annual Conference of the Cognitive Science Society*, (2012), .
29. Lusk, M.M. and Atkinson, R.K. Animated pedagogical agents: does their degree of embodiment impact learning from static or animated worked examples? *Applied Cognitive Psychology 21*, 6 (2007), 747-764. DOI:10.1002/acp.1347.
30. Mubin, O., Stevens, C.J., Shahid, S., Mahmud, A.A., and Dong, J.-J. A review of the applicability of robots in education. *Technology for Education and Learning* (2013), 1-7. DOI:10.2316/Journal.209.2013.1.209-0015.
31. Nass, C. and Moon, Y. Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues 56*, 1 (2000), 81-103.
32. Oh, K. and Kim, M. Social Attributes of Robotic Products: Observations of Child-Robot Interactions in a School Environment. *International Journal of Design 4*, 1 (2010), .

33. Okita, S.Y., Ng-Thow-Hing, V., and Sarvadevabhatla, R.K. Multimodal approach to affective human-robot interaction design with children. *ACM Trans. Interact. Intell. Syst. 1*, 1 (2011), 1-29. DOI:10.1145/2030365.2030370.
34. Pereira, A., Martinho, C., Leite, I., and Paiva, A. iCat, the chess player: the influence of embodiment in the enjoyment of a game. In *Proc. 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, (2008), .
35. Porges, S.W. Social Engagement and Attachment. *Annals of the New York Academy of Sciences 1008*, 1 (2003), 31-47. DOI:10.1196/annals.1301.004.
36. Reeves, B. and Nass, C. 1996. *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press, New York..
37. Salter Ainsworth, M. Attachments Beyond Infancy. *American Psychologist 44*, 4 (1989), 709-716.
38. Sidner, C.L., Lee, C., Kidd, C.D., Lesh, N., and Rich, C. Explorations in engagement for humans and robots. *Artificial Intelligence 166*, 1–2 (2005), 140-164. DOI:http://dx.doi.org/10.1016/j.artint.2005.03.005.
39. Sulčič, V. and Sulčič, A. Can Online Tutors Improve the Quality of E-Learning? *Issues in Informing Science and Information Technology 4* (2007), 201-210.
40. Takayama, L. Perspectives on Agency Interacting with and through Personal Robots. In *Human-Computer Interaction: The Agency Perspective*, Zacarias, M. and De Oliveira, J. eds. Springer Berlin Heidelberg (2012), 195-214..
41. Turkle, S. 2006. *A Nascent Robotics Culture: New Complicities for Companionship.* AAAI Technical Report Series..
42. Vacharkulksemsuk, T. and Fredrickson, B.L. Strangers in sync: Achieving embodied rapport through shared movements. *Journal of experimental social psychology 48*, 1 (2012), 399-402. DOI:10.1016/j.jesp.2011.07.015.
43. Vygotsky, L. 1978. *Mind in Society: The development of higher psychological processes*. Harvard University Press, Cambridge, MA..
44. Yılmaz, R. and Kılıç-Çakmak, E. Educational interface agents as social models to influence learner achievement, attitude and retention of learning. *Computers & Education 59*, 2 (2012), 828-838. DOI:http://dx.doi.org/10.1016/j.compedu.2012.03.020.