The Robot Who Tried Too Hard: Social Behaviour of a Robot Tutor Can Negatively Affect Child Learning

James Kennedy, Paul Baxter, Tony Belpaeme
Centre for Robotics and Neural Systems
Cognition Institute
Plymouth University, U.K.
{james.kennedy, paul.baxter, tony.belpaeme}@plymouth.ac.uk

ABSTRACT

Social robots are finding increasing application in the domain of education, particularly for children, to support and augment learning opportunities. With an implicit assumption that social and adaptive behaviour is desirable, it is therefore of interest to determine precisely how these aspects of behaviour may be exploited in robots to support children in their learning. In this paper, we explore this issue by evaluating the effect of a social robot tutoring strategy with children learning about prime numbers. It is shown that the tutoring strategy itself leads to improvement, but that the presence of a robot employing this strategy amplifies this effect, resulting in significant learning. However, it was also found that children interacting with a robot using social and adaptive behaviours in addition to the teaching strategy did not learn a significant amount. These results indicate that while the presence of a physical robot leads to improved learning, caution is required when applying social behaviour to a robot in a tutoring context.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems

Keywords

Robot Tutor; Social HRI; Child-Robot Interaction; Social Behaviour

1. INTRODUCTION

One-to-one tutoring has been shown to result in significantly higher knowledge gains than group education [5, 23]. Given the common school classroom arrangement, where one teacher is responsible for many children, it is not possible for teachers to offer as much one-to-one tutoring as would be desired. This presents an opportunity for social robotics. A robot tutor could be placed in a classroom to provide one-to-one support for children. However, it is currently unclear how a robot should behave in order to elicit the greatest learning gains from the children. Indeed, there is still a debate on how humans cause learning in tutoring [24].

We seek to explore how the social behaviour of robots can influence learning. This paper considers an experiment designed to explore the contribution of a robot and its social behaviour to dyadic educational interactions with primary school children. Educational interactions are conducted with and without a robot, where the robot's behaviour may be 'social', or 'asocial'. Particular attention is paid to the learning on the part of the children and their social responses to the robot tutor.

The rest of the paper is organised as follows. Background and motivations for this work will be discussed (Section 2) before the methodology for the study presented here is described (Section 3). The methodology will include details of the participants, conditions and robot behaviour. After this, results from both the task and video analysis will be presented and analysed (Section 4). The paper will be concluded with a discussion of the impact of robot social behaviour on child learning and what this may mean for future interactions of this nature (Sections 5 and 6).

HRI'15, March 2–5, 2015, Portland, Oregon, USA.

http://dx.doi.org/10.1145/2696454.2696457.

2. RELATED WORK

Educational robots have long been of interest within the field of HRI. Early exploratory efforts demonstrated the potential for robots in education. For example, a robot used to teach English in a school classroom resulted in an improvement in child learning over a 2 week period [7]. Robots have also been found to elicit advantages over web- and paper-based instruction in the home [6]. Large projects have now turned their attention towards child learning, for example the ALIZ-E project [3] in Europe, and an NSF funded Socially Assistive Robotics project in the USA [20].

These projects (and many others) have started to explore how the tutor behaviour can be manipulated in order to improve learning gains. Leyzberg *et al.* showed that the physical presence of a robot makes a difference to the knowledge gain of adults in an educational puzzle game [14], and that personalised tutoring strategies can lead to significant improvement in knowledge gain in the same puzzle task [13].

This paper seeks to further develop these advances by considering not just personalisation in tutor behaviour, but also how the social behaviour exhibited by the robot affects learning. Aspects of social behaviour, such as gestures, have been used by a robot to attract student attention when they lose focus, greatly improving their recall of information after the task [22]. Additionally, neutral and socially supportive robots have been compared, revealing that children's learning improved when the robot was socially supportive [18].

There are many studies which examine the impact that aspects of social behaviour can have on learning. For example, changing the language used to be personalised (e.g. changing 'the' to 'your') can lead to improved knowledge transfer [4, 15]. Human social behaviours are typically thought to increase a learners' interest, which is posited as the reason for greater learning gains in more social interactions [1].

Therefore it would appear to be desirable to make a robot tutor which is as close as possible to levels of human sociality in order to maximise the potential learning gains in interactions. This study seeks to explore how the social behaviour of a robot impacts upon the behaviour and the learning of children in dyadic interactions. By carefully controlling the manipulation of the social behaviour exhibited by the robot, this paper contributes to the field by comparing child learning when varying robot sociality. The novel learning content presented here also allows the confirmation of findings related to the presence of the robot on child learning in a different context.

3. METHODOLOGY

The methodology of the experiment was designed based on previous work with child learning in sorting tasks, as in [9], and with sub-tasks leading to a combination of knowledge in a larger primary learning objective, as in [13]. The following section will detail the participants involved, the interaction scenario, the task structure, the robot behaviour and the conditions used in the experiment.

3.1 Participants

A total of 53 children had permission to take part in the study. Due to technical issues, 8 of the children's data had to be excluded, leaving 45 children included in the study (23F, 22M). All participants were aged 7 or 8 and from the same year group at a primary school in the U.K. Participants were randomly distributed between conditions, whilst maintaining a balance of gender and mathematics ability (based on their teacher's assessment) between the groups. For the split between conditions, please see Section 3.3. Those in the robot conditions were requested for permission to film, which was granted in all but 2 of the cases. One video had to be excluded from analysis as it was not possible to see the child's eyes. Therefore, video analysis was conducted on 20 interactions.

3.2 Interaction Scenario

Interactions took place either in an unused classroom, or a relatively quiet public space in a primary school in the United Kingdom. The child was brought into the experiment area and would be sat facing a robot, an Aldebaran Nao, with a 27 inch touchscreen horizontally between them (Figure 1). A Microsoft Kinect was placed above and behind the robot to track the child's face. Two video cameras were also positioned around the setup: one to record the child's face and actions and another to record the robot's actions. The use of a touchscreen mediator [2] allows a consistent, constrained environment, so the robot's social behaviour can be manipulated without impacting on the nature of the task or the content of the learning [8].

The learning content for the interaction was devised with the help of primary school teachers from a different school to

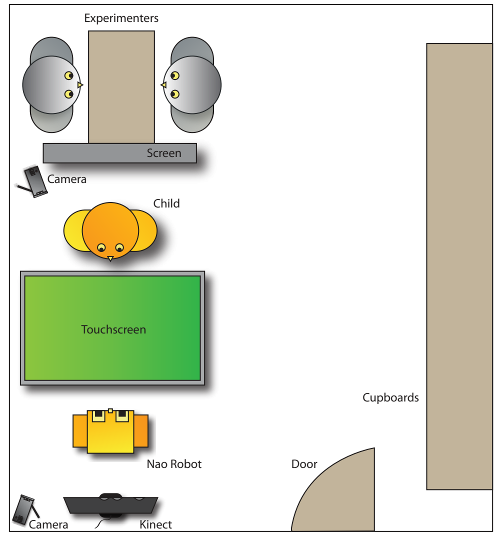

Figure 1: Schematic overview of the interactions under investigation in this paper. Two interactants (the child and the robot) face one another over the touchscreen. Two video cameras record the interactants during the studies. A Microsoft Kinect tracks the child's face. Two experimenters are in the room, but out of view of the child. Figure not to scale.

the one where the study took place. The aim was to select a topic with which children had no prior exposure, but could be learnt in a relatively short time. Prime numbers were determined to be an ideal solution. Calculation of whether a number is prime can be performed by using division (for more detail see Section 3.2.1). Children of the age used in the study are familiar with division, but have not been taught what a prime number is at this stage of their education.

The touchscreen presents different numbers for sorting. The child can touch the numbers to drag and drop them into categories. An example library here would display text labels at opposite sides of the screen, such as 'prime' and 'not prime', with some numbers in between (Figure 2). The child can touch these numbers and drag them to the label for categorisation. The touchscreen sends all state information to the robot so that the robot knows the child's moves, and the robot can make moves itself by synchronising movement with on-screen animation (the robot does not physically touch the screen) [2].

3.2.1 Task Structure

The structure for the task was created partly through necessity for measuring learning and partly through a logical method of calculating primes known as the Sieve of Eratosthenes [17]. The Sieve of Eratosthenes works through a group of numbers, eliminating non-prime numbers in a methodical manner to leave only the prime numbers. For the number range used in this study all composites can be eliminated by dividing by 2, 3, 5 and then 7.

The task was structured so that appropriate measures could be taken for both prime number learning and division learning. Additionally, the task structure allows the examination of the children's division skills prior to the prime

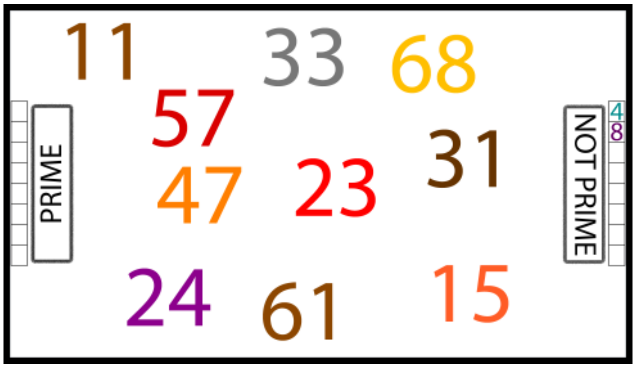

Figure 2: Example of the sorting task used. This is a screenshot of one of the tests used in the experiment. Children can touch a number, drag it over the 'prime' or 'not prime' label and release to make a categorisation. The number will then shrink and move into the boxes beside the category label.

number post-test, which is important as these division skills are necessary for the calculation of primes (Figure 3).

Pre- and post-tests each consisted of 12 numbers being presented on screen; 6 were prime and 6 were non-prime (Figure 2 shows an example test). Both tests avoided numbers from the prime lesson and had balanced distributions of numbers across the range being used (10-70).

Each practice library in the division 'pass one' consisted of 8 examples - 4 of which could be divided with no integer remainder by the number in question, and 4 which could not. This first pass was used to obtain a measurement for each child in how well they could divide by each of the divisors required for the main goal of calculating prime numbers. The number of examples in division 'pass two' totalled 24, but the distribution between each of the 4 divisors (2, 3, 5 and 7) was dependent upon the condition and performance in pass one (see Section 3.4).

3.2.2 Lesson Content

In the second division pass, a lesson was provided for each of the divisors: 2, 3, 5 and 7. This involved verbal instructions and categorisations on-screen. Each lesson consisted of a short verbal overview of the technique, followed by categorisation of 2 examples with verbal narration explaining the application of the technique to the examples. One example could be divided with no remainder, and the other could not. The lessons were not to teach the concept of the division, but often to provide a 'trick' whereby the division could be accomplished more easily. The lessons were explanations of the following concepts:

- Divisible by 2 - the number is even (ends in 0, 2, 4, 6 or 8)
- Divisible by 3 - sum the digits of the number and test if that divides by 3
- Divisible by 5 - the number ends in 0 or 5
- Divisible by 7 - no trick available; a reminder that a number in the 7 times table will be divisible by 7

The lesson about primes which took place after the second division pass used the information from the earlier division

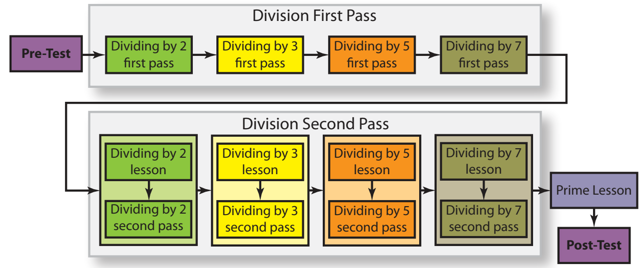

Figure 3: Structure of the task used in the interactions, showing robot lesson positions.

lessons to draw together the practice the child had with dividing by 2, 3, 5 and 7 into calculating whether numbers were prime. The concept of primes was explained (a number divisible, with no remainder, by only 1 and itself) before two worked examples were completed on-screen - one prime and one not prime. The Sieve of Eratosthenes was adapted to eliminate numbers one-by-one for categorisation. Children were instructed to consider each number to be categorised in turn, attempting to divide it by 2, 3, 5 and 7. If the number divided by any of these then it was not prime, otherwise it was prime.

3.3 Hypotheses and Conditions

Given the task described above, we seek to assess whether a social robot leads to increased learning. Our specific hypotheses to address this are as follows:

H1. The division lessons provided to the children in division pass two will result in significant improvement from division pass one. This serves as a check that the lessons provided do actually facilitate learning.

H2. The presence of a robot will result in greater learning gains than when a robot is not present given equal information content, as suggested by other studies ([6, 14]).

H3. A more social robot will result in greater learning gains than a less social robot. This hypothesis is based on observations of the positive effects of social behaviour in other HRI studies such as [13] and [18], and psychology studies such as [1] and [15].

In order to address the hypotheses, four conditions were devised:

1. **Division only** [n=11] - division pass one, followed by division pass two without any lessons. Conducted on the touchscreen only, with no robot present.
2. **Screen only** [n=11] - the full interaction as described in Section 3.2.1, but with no robot present. All feedback and lesson content is delivered by the speakers in the screen.
3. **Asocial non-personalised robot** [n=11] - identical script to the 'screen only' condition, but with the robot delivering the content. All verbal content and feedback is given by the robot; the screen now only displays the numbers for the task. Robot behaviour is designed to be non-social (see Section 3.4 for full details).

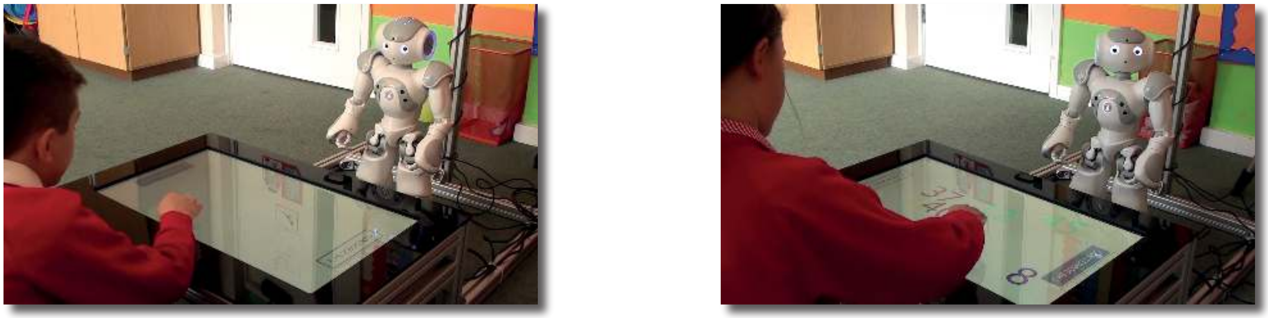

Figure 4: Snapshots taken from the video recordings of interactions. Both the social (left, looking at the child) and asocial robot (right, actively avoiding the gaze of the child) conditions are pictured to show the difference in gaze behaviour between them.

4. **Social personalised robot** [n=12] - a social version of the full interaction. All lesson content is kept the same as the asocial robot condition, but the non-lesson speech is adjusted to be more social. Robot non-verbal behaviour is also designed to be social.

3.4 Robot Behaviour

Human tutors are known to be effective, using social behaviour and adapting to the learning needs of the child. As such, the social robot behaviour was based on a human tutor's behaviour when taking five children through the task on the touchscreen. Section 3.4 outlines four observed behavioural dimensions that were implemented on the robot. Whilst maintaining balance between the conditions, the inverse for each dimension is used for the asocial robot behaviour in order to evaluate Hypothesis 3.

The phrases and actions used by the human were observed and implemented in the social robot model. It is posited that behaviour is perceived by the child as an integration of cues [26], meaning that each dimension must be considered in context of the others. Consequently, personalisation and social behaviour are considered inseparable in assessing Hypothesis 3 for this study, following the human model.

Both robot conditions adopted the following basic behaviour during the image categorisation portions of the task:

Move Suggestions - During each stage of the interaction, if the child was hesitant in making moves then the robot would move a number to the centre of the screen and suggest that the child work on that number next. The decision about when to move was probabilistic and cued by the child's behaviour. If the child did not make a categorisation for 6 seconds, then there was a 25% chance that the robot would move, with the decision repeated every 2 seconds until a move was made - the 6 second timer would then start again.

Categorisation Feedback - The robot would provide verbal feedback on the child's categorisations. Not every categorisation received feedback; there was a 25% chance of feedback on each categorisation - following the human tutor model.

Robot Condition Differences.

Verbal Content - The script for the social robot speech was taken from the human tutor; this was then modified for the asocial robot by removing any personalisation, i.e. "Johnny, we'll do dividing by 2 next" becomes "You'll do dividing by 2 next". We ensured that the total amount of speech was kept as close as possible between the conditions, and the lesson content was the same.

When providing speech alongside a suggestion, or when providing feedback, a number of phrases were available and selected at random. The asocial robot had only 2 options for each event (compared to the social robot's 8), thereby making it very repetitive.

Gestures - The social robot script used for the introduction and some of the lessons included iconic gestures. In the asocial condition, these were placed at inappropriate times, for example, the robot would wave its arm to greet the child half way through a sentence, rather than when it says hello at the start. The same gestures were used in both conditions, the only difference was their position in the script.

Personalisation - The social robot would use the child's name in greeting, just before the post-test and in the goodbye script. The asocial robot would not use the child's name at all. Personalisation of learning content was also provided by the social robot.

The performance of the child in the first division pass would dictate how many examples of each division library they would do in the second pass. A total of 24 numbers were always used in the second division pass. For the asocial condition, these were split equally between divisors, so 6 numbers for each of dividing by 2, 3, 5 and 7. In the social condition a minimum of 3 numbers were used per divisor, but the remaining 12 numbers were distributed between the divisors based on how many of each divisor the child got wrong in the first pass. Therefore, they had more practice on numbers that they were weaker at in the second pass.

In the second division pass, for each divisor library, there was also a reminder of the lesson available. In the asocial condition, this reminder would be delivered by the robot half way through the categorisations for that library (i.e. after the 3rd of the 6 categorisations to be made). In the social condition, the reminder was given after the first incorrectly categorised image.

Gaze - The social robot gaze was constrained so that it would generally be looking towards the touchscreen or in the direction of the child. Additionally, a Microsoft Kinect was used for tracking the child's head pose. If the child's head pose was directed towards the robot, then the robot would respond by looking back at the child. In the asocial condition, the robot was intentionally programmed to look up and to the side so that the gaze would avoid the child (Figure 4).

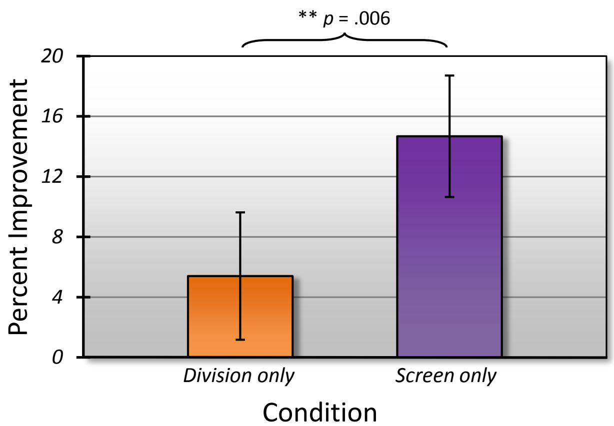

Figure 5: Improvement between division pass one and division pass two in percent for the division only and screen only conditions. Error bars show 95% Confidence Interval, ** indicates significance at the 0.01 level.

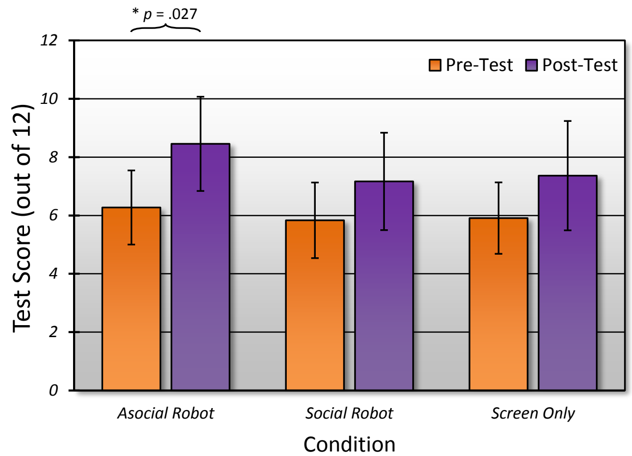

Figure 6: Pre-test and post-test scores for the asocial robot, social robot and screen only conditions. Error bars show 95% Confidence Interval, * indicates significance at the 0.05 level.

3.5 Procedure

One of the experimenters shown in Figure 1 controlled the start and end of the autonomous behaviour. This individual had three responsibilities: 1. to type in the name of the child for the social robot condition before the child arrived in the room, 2. to click a button once the child was sat down in front of the robot to denote the start of the interaction, and 3. to click an 'emergency' button if anything went wrong, where the robot would gracefully end the interaction. All other robot behaviour was fully autonomous.

4. RESULTS

This section will present the results from each of the conditions in relation to the hypotheses. Learning will be considered either between the pre-and post-test improvements, or for division, between the total percent correct in division pass one and division pass two. The behavioural analysis is derived from video coding of the child's gaze as previous work has highlighted gaze as the primary behaviour of interest in interactions of this nature [10]. The video coding was completed by one coder for all videos. Coding was verified by second-coding 20% of the videos, as in [16], with an average Cohen's Kappa of 0.80 signifying substantial agreement [12].

The conditions were split to have an equal balance of ability based on an estimate by the children's teacher (higher, middle and lower tiers). Comparing the approximate ability level of the children that was provided by their teacher against their performance in the first division pass (at which point they've had no lesson input), Pearson's r correlation is 0.638. This is a good correlation, which confirms that the teacher's estimate is reflected in the results of this study and therefore that the conditions are balanced for ability.

The mean average length of the interactions were: 974s (95% CI [750s,1199s]) in the asocial robot condition, 1011s (95% CI [786s,1236s]) in the social robot condition, and 873s (95% CI [680s,1066s]) in the screen only condition. The average length of the division only condition (M=452s, 95% CI [277s,629s]) was much shorter as the robot lessons, pre-test and post-test add a lot of time.

4.1 Learning from Lessons

A 2 tailed, unpaired t-test was conducted to compare the improvement between division pass one and division pass two in the division only (no lessons) and screen only (with lessons) conditions (with no robot present). There was a significant difference in the scores for division only (M=5.40, 95% CI [1.17,9.63]) and screen only (M=14.68, 95% CI [10.65,18.71]) conditions; $t(20)$=3.114, p = 0.006 (Figure 5). This shows that the improvement was significantly higher when the lessons were present, supporting Hypothesis 1. The result here is not surprising, but it is beneficial to show the effectiveness of the division lessons.

4.2 Robot Presence

To examine how the robot affects the learning of the child, we compared the improvement between pre-test and post-test scores between the screen only and (combined) robot conditions. All pre-test and post-test scores are out of 12. None of the children who took part in the study reported to know what a prime number was before the interaction. As a result, based on 2 options for each categorisation, it would be expected that the pre-test scores would be around chance (50% out of 12 correct).

In the screen only condition a 2 tailed, paired t test reveals no significant difference between the scores for the pre-test (M=5.91, 95% CI [4.68,7.13]) and post-test (M=7.36, 95% CI [5.49,9.24]); $t(10)$=1.027, p=0.329. However, when a robot is present there is a significant difference in scores between the pre-test (M=6.04, 95% CI [5.15,6.94]) and the post-test (M=7.78, 95% CI [6.61,8.95]); $t(22)$=2.997, p=0.007. This supports Hypothesis 2, that the presence of a robot will result in greater learning gains.

To further explore this result, we compared the screen only pre-test and post-test scores with those in the asocial robot condition. These two conditions are identical in the script that is used (the screen plays recorded clips of the robot voice) and the lack of personalisation. The previous paragraph showed that there is no significant difference in pre-test and post-test scores in the screen only condition. For the asocial robot, the difference is significant when the same

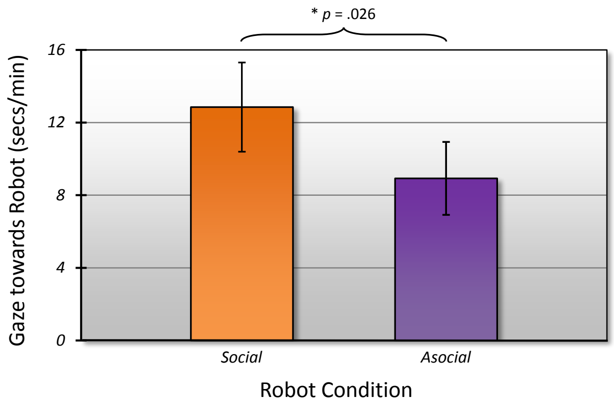

Figure 7: Child gaze towards the robot in seconds per minute, split by robot condition. Error bars show 95% Confidence Interval, * indicates significance at the 0.05 level.

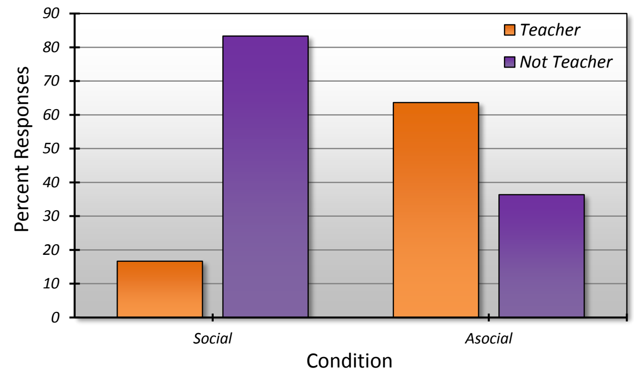

Figure 8: Post-questionnaire responses of the children when asked what they thought the robot was like. Eight options, including 'teacher' were available.

test is run between scores for the pre-test (M=6.27, 95% CI [5.00,7.54]) and post-test (M=8.45, 95% CI [6.84,10.07]); $t(10)$=2.597, p=0.027. This shows that children's learning gains are not significant in the screen only condition, but when the robot is added the learning gains become significant, providing further support for Hypothesis 2.

However, it should be noted that there was no significant difference in the improvement between the screen only (M=1.46, 95% CI [-1.32,4.23]) and asocial robot (M=2.18, 95% CI [0.54,3.83]) conditions; $t(20)$=0.442, p=0.664, when using a 2 tailed, unpaired t test. The fact that the learning becomes significant when the robot is added, despite careful control to match the conditions aside from the presence of the robot, indicates that the robot contributes to the learning that takes place. This result has been observed many times before in other contexts, for example [11] and [14].

4.3 Social Condition

As shown in the previous section, the learning gains for the asocial robot were significant. When conducting a 2 tailed, paired t-test for the social robot condition there is no significant difference between the pre-test (M=5.83, 95% CI [4.54,7.13]) and post-test (M=7.17, 95% CI [5.50,8.84]); $t(11)$=1.627, p=0.132. Whilst all conditions show improvement between the pre-test and the post-test, the only condition where the learning gain is significant is with the asocial robot; both the social robot and screen only conditions show non-significant improvement (Figure 6). This result contradicts Hypothesis 3, that a more social robot will result in greater learning gains than a less social robot.

To explore the impact that the learning personalisation may have had on the results, the lesson reminders and practice of numbers in the second division pass are considered. In the asocial condition a reminder of the lesson is given for each divisor, whereas in the social condition, reminders are only given when the child makes a mistake. This meant that in the asocial condition a total of 44 reminders were given (M=4.00 per interaction; no deviation), whereas in the social condition a total of 22 reminders are delivered (M=1.83, 95% CI [0.94,2.73] per interaction). This is not surprising, as most children can comfortably divide by 2 and 5 at this age; thereby eliminating the need for around half of the reminders. When correlating the number of reminders provided to children in the social robot condition with their improvement between pre- and post-test score, Pearson's correlation r=-0.418. This is a moderate negative correlation, which suggests that receiving fewer reminders does not reduce the child's performance.

Additionally, children in the social robot condition were given the opportunity to practice more of the numbers that they were weaker at (following the human model described in Section 3.4). There is a possibility that this could have been a de-motivator if they then performed poorly in this phase of the interaction. However, this seems unlikely as it is found that there is no significant difference between the performance in the second division pass between children in the social condition (M=82% correct, 95% CI [71%,93%]) and those in the asocial condition (M=85% correct, 95% CI [75%,95%]); $t(21)$=0.432, p=0.670.

In order to investigate the reasons behind why children's learning gains are not as great when the robot is social compared to when it is asocial, the children's behaviour and self-reported view of the robot were analysed. From video coding of the interactions, it was found that children look significantly more often at the social robot (M=12.9, 95% CI [10.4,15.3]) than at the asocial robot (M=8.9, 95% CI [6.9,10.9]); $t(18)$=2.425, p=0.026 (Figure 7). Values are provided in seconds of gaze at the robot per minute of interaction; this normalisation allows for comparison across interactions of different lengths.

The children completed a pre-questionnaire and a post-questionnaire before and after the interaction. These were very short, with just 4 questions in the pre-questionnaire and 2 questions for the post-questionnaire. The questionnaires were used to see what the children expected from the interactions, and subsequently how they viewed the robot afterwards. Despite being told by the experimenters several times before their interactions that they would be *taught* by a robot *teacher*, with the robot script emphasising this point too, the children in the social robot condition consistently reported that they thought the robot was a 'friend' after the interaction. The question asked "For me, I think the robot was like a -" with 8 options available: brother or sister, classmate, stranger, relative (e.g. cousin or aunt), friend, parent, teacher, and neighbour.

It was expected that the children would report the robot to be a teacher (as this is what they had been told), so their responses were grouped into either 'teacher' or 'not teacher'. In the social condition, 17% of the children reported the robot to have been like a teacher, compared to 64% in the asocial condition. Fisher's exact test shows that the responses differ significantly by condition, p=0.036 (Figure 8).

It is clear from the children's gaze and self-reported responses that the difference in robot behaviour between conditions has an effect on the children's behaviour and attitudes towards the robot. It is suggested that these differences could account for the difference in learning gains observed between the social and asocial robot conditions. Whilst the robot is providing the lesson about prime numbers, it demonstrates two examples on the screen by highlighting the numbers, discussing them and correctly categorising them. Therefore, during this period it is useful to look at the screen. During the prime lesson the average amount of gaze towards the social robot (M=26.9 secs/min, 95% CI [22.9,30.9]) is significantly higher than the gaze towards the asocial robot (M=17.0 secs/min, 95% CI [11.0,23.1]); $t(18)$=2.669, p=0.016. It is suggested that the additional attention directed towards the social robot's behaviour could distract the children from the content that it is delivering; this possibility is further discussed below (Section 5).

5. DISCUSSION

From the analysis of the results it is clear that the lessons for division have a positive effect on the children's performance, supporting Hypothesis 1. This validates part of the teaching behaviour and demonstrates that the children have the ability to understand the robot's voice and apply knowledge gained from the lessons in the task on-screen.

When the asocial robot is present, despite having the same content as the screen only condition, the improvement between pre-test and post-test becomes significant, providing partial support for Hypothesis 2. This is a demonstration of the social presence effect; the addition of an agent into the interaction leads to improvement in task performance, as observed before in other contexts, for example [11] and [14]. However, the improvement is lost when the robot behaviour is changed to become more social. This is a surprising result, which contradicts both Hypotheses 2 (that a robot will provide greater learning gains than the screen alone) and 3 (that a more social robot will result in improved learning gains).

This result is in contrast to existing studies in the literature that Hypothesis 3 was based on. As described in Section 3.4, the robot behaviour was derived directly from that of a human tutor. This necessitates a perspective that integrates behavioural dimensions [26] that emphasises sets of behavioural competencies (similar to the use by [18]). This differs from the more typical focus on individual social cues, as in [15] and [22]. With the interaction context (child-robot interactions in a school) and task content (learning mathematical concepts) also differentiating the work here from previous studies, this integrated cues perspective may merit further investigation in terms of the effects on the perceptions and performance of human interactants.

One possible explanation for the unexpected findings with respect to learning is that although the children looked at the social robot significantly more than the asocial robot during the lesson phase (which could be considered advantageous as the robot provides the lessons), they were paying attention to the social behaviour instead of the lesson content. An alternate explanation is that the social behaviour presented by the social robot places more cognitive load on the children, which may inhibit their capacity to process information related to the task [21]. It may be that in the long-term, as the novelty of the social behaviour wears off, the social robot would then elicit better learning, as indicated by [7]. However, further research is required to explore these ideas explicitly and in more detail.

5.1 Child Perception and Ability

In Section 4.3 it was shown that children in the asocial robot condition were more likely to report that they viewed the robot as a teacher than those in the social robot condition. The infrequency with which those in the social condition reported the robot to be like a teacher was surprising. The children were told several times before and during the interaction by both the robot and the experimenters that the robot was a teacher. It is suggested that there may be two reasons as to why this was the case. Firstly, it may be that the directness of the asocial robot conformed more to their expectations of what a robot teacher would be like than the social robot, which was less direct in its instructions. Secondly, the behaviour of the asocial robot may not have had enough character to change the children's perception of the robot as a teacher, whereas the social robot did. Interestingly, there was almost no correlation between the children's perception of the robot as a teacher and their performance; Pearson's r correlation = -0.11.

There was only a weak correlation (Pearson's r = 0.13) between the teacher-provided mathematics ability levels of the children and their subsequent improvement between pre-test and post-test. This is somewhat surprising, as one would expect the higher ability students to progress more given the same practice as those who were lower ability. This may highlight a limitation in the adaptiveness of the robot's behaviour used in this study. It is possible that a robot which is more adaptive could better respond to each individuals' needs and push them more effectively through the Zone of Proximal Development [25].

Due to the relatively small sample sizes used here, it only requires 2 or 3 subjects to perform particularly poorly or well to impact on the significance of the results. However, there is a trade-off between trying to carefully control the experiment and get greater subject numbers. Subjects were selected from the same school and year group so that they would have similar educational experiences and backgrounds. Due to limits on the sizes of school classes, it is likely that to get greater numbers would mean selecting subjects across multiple schools. This then introduces the risk of large variability between subjects' mathematical ability and the environment in which the experiment is conducted.

5.2 Gender Differences

One interesting aside that was noticed through additional exploratory analysis are differences between the genders. These results were not included in Section 4 as they were not part of the original hypothesis for this study. However, as an interesting observation they have been included here, with the suggestion that they may be worth further research. A significant difference is found between the improvement between pre-test and post-test of girls (M=2.77, 95% CI

[1.18,4.36]) and boys (M=0.40, 95% CI [-0.85,1.65]) when interacting with a robot present (both social and asocial conditions combined); $t(21)$=2.192, p=0.040. These results show that the boys barely improved with a robot, whilst the girls improved quite substantially.

Additionally, girls who interacted with a robot present (M=2.77, 95% CI [1.18,4.36]) improved more than those without a robot present (M=-0.40, 95% CI [-3.71,2.91]). Whilst this difference is not quite significant ($t(16)$=1.907, p=0.075), it seems as though there may be a possible trend. Gender differences due to social presence have been observed in other contexts in HRI, such as [19], where females saw a robot as more machine-like. This could support the argument that the robot social behaviour distracts from the lesson content that it is delivering; girls, who may perceive the robot as less social, therefore outperformed the boys. Whilst there is not enough evidence here to make firm conclusions about this point, the effect of gender certainly merits more research in the context of educational interactions.

6. CONCLUSION

As expected, the use of lessons improved the children's performance between the first division pass and the second division pass, as shown in Section 4.1. Partial evidence was found in support of the social presence effect. Section 4.2 showed that when a robot delivered the lessons to the child, the learning was significant, whereas when the same information was provided by just a screen, without a robot, it was not. By further breaking down the robot results into the two different behavioural conditions, it was found that the learning remains significant with the asocial robot, where the script is identical to the condition without the robot present (where the learning was not significant). However, these positive effects were not maintained when the robot was more social.

The results here have shown that a robot which is not appropriately social led to greater learning gains of children in a maths task than a robot with appropriate social behaviours. This result contradicts expectations and predications made based on other studies in the literature (for example [15] and [18]). It is hypothesised that the social behaviour of the socially appropriate robot may distract from the content it is delivering with regards to the learning task, whilst the asocial robot leads to disinterest, and therefore less distraction from the learning task. Gaze behaviour of the children throughout the interaction and specifically during the prime numbers lesson is used to provide evidence for this suggestion.

7. ACKNOWLEDGEMENTS

This work is partially funded by the EU FP7 projects ALIZ-E (grant 248116) and DREAM (grant 611391), and the School of Computing and Maths, Plymouth University, U.K. The authors would like to thank Okehampton Primary School, U.K. for taking part in the study.

8. REFERENCES

[1] R. K. Atkinson, R. E. Mayer, and M. M. Merrill. Fostering social agency in multimedia learning: Examining the impact of an animated agent's voice. *Contemporary Educational Psychology*, 30(1):117–139, 2005.

[2] P. Baxter, R. Wood, and T. Belpaeme. A touchscreen-based 'sandtray' to facilitate, mediate and contextualise human-robot social interaction. In *Proc. HRI'12*, pages 105–106, 2012.

[3] T. Belpaeme, P. Baxter, R. Read, R. Wood, H. Cuayáhuitl, B. Kiefer, et al. Multimodal child-robot interaction: Building social bonds. *Journal of Human-Robot Interaction*, 1(2):33–53, 2012.

[4] O. A. Blanson Henkemans, B. P. Bierman, J. Janssen, M. A. Neerincx, R. Looije, H. van der Bosch, et al. Using a robot to personalise health education for children with diabetes type 1: A pilot study. *Patient Education and Counseling*, 92(2):174–181, 2013.

[5] B. S. Bloom. The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational researcher*, pages 4–16, 1984.

[6] J. Han, M. Jo, S. Park, and S. Kim. The educational use of home robots for children. In *IEEE RO-MAN'05*, pages 378–383, 2005.

[7] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro. Interactive robots as social partners and peer tutors for children: A field trial. *Human-Computer Interaction*, 19(1):61–84, 2004.

[8] J. Kennedy, P. Baxter, and T. Belpaeme. Constraining content in mediated unstructured social interactions: Studies in the wild. In *Proc. AFFINE'13, at ACII'13*, pages 728–733, 2013.

[9] J. Kennedy, P. Baxter, and T. Belpaeme. Children comply with a robot's indirect requests. In *Proc. HRI'14*, pages 198–199, 2014.

[10] J. Kennedy, P. Baxter, and T. Belpaeme. Comparing robot embodiments in a guided discovery learning interaction with children. *International Journal of Social Robotics*, accepted.

[11] H. Kose-Bagci, E. Ferrari, K. Dautenhahn, D. S. Syrdal, and C. L. Nehaniv. Effects of embodiment and gestures on social interaction in drumming games with a humanoid robot. *Advanced Robotics*, 23(14):1951–1996, 2009.

[12] J. R. Landis and G. G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, 1977.

[13] D. Leyzberg, S. Spaulding, and B. Scassellati. Personalizing robot tutors to individual learning differences. In *Proc. HRI'14*, pages 423–430, 2014.

[14] D. Leyzberg, S. Spaulding, M. Toneva, and B. Scassellati. The physical presence of a robot tutor increases cognitive learning gains. In *Proc. CogSci'12*, pages 1882–1887, 2012.

[15] R. E. Mayer, S. Fennell, L. Farmer, and J. Campbell. A personalization effect in multimedia learning: Students learn better when words are in conversational style rather than formal style. *Journal of Educational Psychology*, 96(2):389, 2004.

[16] L. Moshkina, S. Trickett, and J. G. Trafton. Social engagement in public places: a tale of one robot. In *Proc. HRI'14*, pages 382–389, 2014.

[17] M. E. O'Neill. The genuine sieve of eratosthenes. *Journal of Functional Programming*, 19(01):95–106, 2009.

[18] M. Saerbeck, T. Schut, C. Bartneck, and M. D. Janse. Expressive robots in education: Varying the degree of social supportive behavior of a robotic tutor. In *Proc. CHI'10*, pages 1613–1622, 2010.

[19] P. Schermerhorn, M. Scheutz, and C. R. Crowell. Robot social presence and gender: Do females view robots differently than males? In *Proc. HRI'08*, pages 263–270, 2008.

[20] E. Short, K. Swift-Spong, J. Greczek, A. Ramachandran, A. Litoiu, E. C. Grigore, et al. How to train your dragonbot: Socially assistive robots for teaching children about nutrition through play. In *IEEE RO-MAN'14*, pages 924–929, 2014.

[21] J. Sweller. Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, 4(4):295–312, 1994.

[22] D. Szafir and B. Mutlu. Pay attention!: Designing adaptive agents that monitor and improve user engagement. In *Proc. CHI'12*, pages 11–20, 2012.

[23] K. VanLehn. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4):197–221, 2011.

[24] K. VanLehn, S. Siler, C. Murray, T. Yamauchi, and W. B. Baggett. Why do only some events cause learning during human tutoring? *Cognition and Instruction*, 21(3):209–249, 2003.

[25] L. S. Vygotsky. *Mind in society: The development of higher psychological processes*. Harvard university press, 1980.

[26] J. Zaki. Cue integration a common framework for social cognition and physical perception. *Perspectives on Psychological Science*, 8(3):296–312, 2013.