Heart vs Hard Drive: Children Learn More From a Human Tutor Than a Social Robot

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

***Abstract*—The field of Human-Robot Interaction (HRI) is increasingly exploring the use of social robots for educating children. Commonly, non-academic audiences will ask how robots compare to humans in terms of learning outcomes. This question is also interesting for social roboticists as humans are often assumed to be an upper benchmark for social behaviour, which influences learning. This paper presents a study in which learning gains of children are compared when taught the same mathematics material by a robot tutor and a non-expert human tutor. Significant learning occurs in both conditions, but the children improve more with the human tutor. This difference is not statistically significant, but the effect sizes fall in line with findings from other literature showing that humans outperform technology for tutoring. We discuss these findings in the context of applying social robots in child education.**

I. Introduction

An increasing quantity of research in HRI has considered the use of robot tutors, particularly for educating children [1], [2], [3]. It has been found that robot embodiment [1], [2], social behaviour [4], and teaching strategies [3], [5] can improve child learning. One question that often arises, particularly from non-academic audiences, is how robots compare to human tutors. The aim of such research is rarely to replace human teaching, but to supplement it, so such a comparison is not typically part of experimental hypotheses.

Given the link between robot social behaviour and learning [4], human behaviour is often used to derive behaviour for robots to provide an upper benchmark of social behaviour that robots can aim for in tutoring. The literature from other fields suggests that human tutoring also provides an upper benchmark in terms of learning gains [6], but this has not been verified in HRI. Serholt *et al.* [7] found no significant difference between the performance of children who had been tutored by a humanoid robot compared to a human, but the robot speech was controlled using a Wizard-of-Oz method, introducing additional variability between conditions. The present paper reports on a study in which the lesson content delivered by a human and an autonomous robot is kept consistent in order to explore the differences in child learning depending on the character (and their social behaviour) providing the content. The aim is to address the following hypothesis:

H1: Human tutoring will lead to more child learning when compared to robot tutoring.

II. Methodology

The study employs the same methodology as seen in [2] and [4]. Children aged 8 and 9 engage in a dyadic interaction in their school with a tutor who guides them through a method for prime number identification. The children's learning is measured through a pre-test and a post-test consisting of 12 numbers which need to be categorised as 'prime' or 'not prime' (6 per category). Prior to the interaction, children have not learnt about prime numbers, but the technique relies on their ability to divide by 2, 3, 5 and 7, so this is also tested. The tutor provides hints to help with the division, as well as a lesson about how to identify prime numbers using the Sieve of Eratosthenes technique. Two tests for prime number identification are used in a cross-testing strategy to control for exposure to the tests.

Two conditions were employed: *(1)* an autonomous 'high immediacy' robot tutor [4], and *(2)* a human tutor (Fig. 1). The robot tutor was designed to regularly gesture, look at the child, make small body movements to appear 'relaxed', and lean forwards. The human was given a word-by-word script to match the lesson content of the robot, but was not constrained in terms of social behaviour. Due to the script providing precise lesson content (and the study focus on social behaviour and embodiment differences) an expert tutor was not required. A total of 22 children took part: 11 in the robot condition and 11 in the human (age M=8.8, SD=0.4; 12F, 10M). Interactions lasted for M=14m05 (SD=3m16) in the robot condition, and M=13m10 (SD=3m39) in the human condition.

III. Results and Discussion

Children improve significantly in both conditions (Fig. 2). Paired t-tests show the post-test score (M=7.6, 95% CI [5.5,9.8]) is significantly higher than the pre-test score (M=5.2, 95% CI [3.7,6.7]) in the human condition; $t(10)$=2.425, p=.036. The post-test score (M=7.0, 95% CI [4.9,9.1]) is also significantly higher than the pre-test score (M=5.1, 95% CI [3.4,6.8]) in the robot condition; $t(10)$=3.057, p=.012. Although the children improve more between the pre-test and post-test in the human condition (M=2.5, 95% CI [0.2,4.7]) than in the robot condition (M=1.9, 95% CI [0.5,3.3]), this difference is not found to be statistically significant using an independent samples t-test; $t(20)$=0.459, p=.652.

The improvement from pre- to post-test score is not significantly different between the robot and human conditions, but

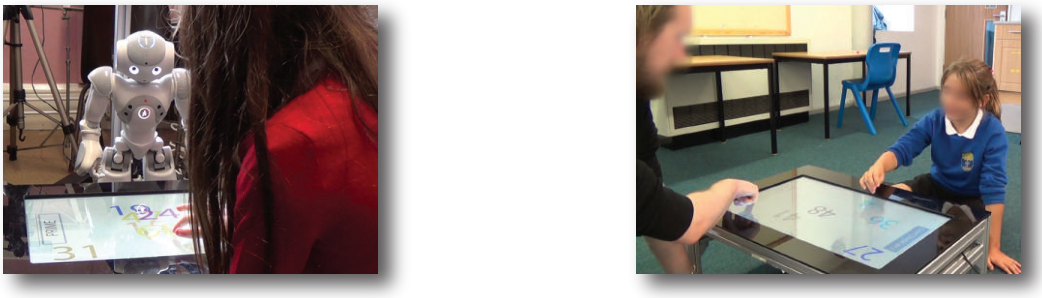

Fig. 1. Images of the interactions: (*left*) the robot condition, (*right*) the human condition. Interactions take place around a touchscreen which displays the learning material. Both the child and the tutor (whether human or robot) can move numbers on the screen. Feedback is provided by the tutor, and not on screen.

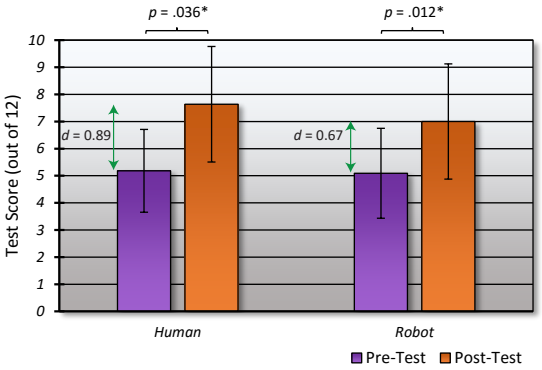

Fig. 2. Child pre-test and post-test scores for the robot and human conditions. The improvement is significant in both conditions, showing that the children learn. The difference between conditions is not significant, but the improvement effect size is larger in the human condition. *Error bars* show the 95% Confidence Interval.

this may be due to the relatively small sample size (although *t*-test assumptions are met). If the trends here were to continue, then this difference would become significant with more subjects. The effect size seen in each condition provides a clearer indication of the difference between them; Cohen's d=0.67 for the robot, but d=0.89 for the human. As such, this provides some support for H1: that child learning is greater when tutored by a human when compared to a robot. These effect sizes are similar to those found in other literature [6]. It should be noted that the effect sizes in [6] compare to a no tutoring control, which is not done here since the nature of the task makes learning unlikely without tutoring.

It is also worth noting that the human condition mean was lowered by one instance where the child had clearly learnt the technique, but confused the categories, and so scored 0 on the post-test (i.e. 100%, but incorrect). The child asked for clarification, but as this help would not have been available in the robot condition, it was not given by the human at the time.

The specific robot and human used in the tutoring task will have had a large impact on the results. One robot (and its behaviour) was compared to one human; these results are likely to vary depending on the robot and human used. The learning content was kept consistent between the conditions, but the social behaviour was not constrained in the case of the human. This means that the human can take advantage of some social cues that the robot could not, and could subsequently be more socially adaptive (for example, in mutual gaze) than the robot, which may account for some of the learning differences. A non-expert human was used due to the tightly specified learning content, but an expert tutor may have used different social behaviour, potentially leading to more learning. It remains to be seen if the robot could close the gap in learning outcomes with improved social sensitivity and behaviour.

Of course, the aim is not to replace human tutors; robots offer additional opportunities to supplement current human tutoring provision. Robots can assume a wider variety of roles, for example, to assist teachers [1], or to offer children a chance to teach a less-able peer [3], [5]. Alternatively, robots could provide personalised support which falls outside of typical lessons or the school environment, such as additional language support for non-native children, as discussed in [8].

IV. Acknowledgements

This work is partially funded by the EU FP7 DREAM project (grant 611391), the H2020 L2TOR project (grant 688014), and the SoCEM, Plymouth University, U.K.

References

[1] M. Alemi *et al.*, "Employing Humanoid Robots for Teaching English Language in Iranian Junior High-Schools," *Int. Journal of Humanoid Robotics*, vol. 11, no. 3, 2014.

[2] J. Kennedy *et al.*, "The Robot Who Tried Too Hard: Social Behaviour of a Robot Tutor Can Negatively Affect Child Learning," in *Proc. of the 10th ACM/IEEE Int. Conf. on HRI*. ACM, 2015, pp. 67–74.

[3] F. Tanaka and S. Matsuzoe, "Children Teach a Care-Receiving Robot to Promote Their Learning: Field Experiments in a Classroom for Vocabulary Learning," *Journal of Human-Robot Interaction*, vol. 1, no. 1, pp. 78–95, 2012.

[4] J. Kennedy *et al.*, "Higher Nonverbal Immediacy Leads to Greater Learning Gains in Child-Robot Tutoring Interactions," in *Proc. of the Int. Conf. on Social Robotics*. Springer, 2015, pp. 327–336.

[5] D. Hood *et al.*, "When Children Teach a Robot to Write: An Autonomous Teachable Humanoid Which Uses Simulated Handwriting," in *Proc. of the 10th ACM/IEEE Int. Conf. on HRI*. ACM, 2015, pp. 83–90.

[6] K. VanLehn, "The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems," *Educational Psychologist*, vol. 46, no. 4, pp. 197–221, 2011.

[7] S. Serholt *et al.*, "Comparing a humanoid tutor to a human tutor delivering an instructional task to children," in *Proc. of the 14th IEEE-RAS Int. Conf. on Humanoid Robots*. IEEE, 2014, pp. 1134–1141.

[8] T. Belpaeme *et al.*, "L2TOR - Second Language Learning Tutoring using Social Robots," in *Proc. of the First Int. Workshop on Educational Robots at ICSR'15*, 2015.