A Novel Paradigm for Children as Teachers to the Kaspar Robot Learner

A. Zaraki[1], M. Khamassi[2,3], L.J. Wood[1], G. Lakatos[1], C. Tzafestas[2], B. Robins[1] and K. Dautenhahn[1]

Abstract— **This paper presents a contribution to the active field of robotics research to support the development of social skills and capabilities in children with Autism Spectrum Disorders as well as Typically Developing children. We present preliminary results of a novel experiment where classical roles are reversed: children are here the teachers giving positive or negative reinforcement to the Kaspar robot to make it learn arbitrary associations between toys and locations where to tidy them. The goal is to help children change perspective, and understand that sometimes a learning agent needs several repetitions before correctly learning something. We developed a reinforcement learning algorithm enabling Kaspar to verbally convey its uncertainty along learning, so as to better inform the interacting child of the reasons behind successes and failures made by the robot. Overall, 30 children aged between 7 and 8 (19 girls, 11 boys) performed 16 sessions of the experiment in groups, and managed to teach Kaspar all associations in 2 to 7 trials. Kaspar only made a few unexpected associations, mostly due to exploratory choices, and eventually reached minimal uncertainty. All children expressed enthusiasm in the experiment.**

Keywords: Human-robot interaction, reinforcement learning, autism, autonomous robotics, children social skills, teaching.

I. INTRODUCTION

In this short paper, we present recent progresses in developing robot learning abilities for the Kaspar robot [1], [2] in order to make it learn from human feedback during social interaction. More precisely, we propose a novel experiment where children provide Kaspar with positive or negative reinforcement to make it learn arbitrary associations between 6 different toys and 3 possible locations where these toys could be placed. The main goal of this work is to contribute to the development of social skills and capabilities in children with Autism Spectrum Disorders.

A growing number of studies have tackled the complex challenge of applying reinforcement learning techniques for robot's behaviour adaptation during social interaction with humans (*e.g.* [3]). However, these studies have so far mostly involved adults interacting with robots, and to our knowledge, no one has yet addressed the question whether enabling children to reinforce a robot while it learns arbitrary associations could help develop social skills and perspective taking in these children [4], [5]. In contrast, robot learning studies in social interaction contexts have typically focused on extending reinforcement learning algorithm to make them cope with the high degrees of uncertainty, volatility and non-stationarity associated with non-verbal communication (*e.g.* [6]).

In this study, we adopt a much simpler reinforcement learning algorithm where the goal is to have the robot *making mistakes*, and the children providing the *right feedback* to make the robot correctly learn arbitrary associations. The task consists in showing a set of locations and asking the robot which toys among a small set should be placed there. Prior to the experiment, the children are informed of the correct associations. During the experiment, the children can press two buttons (green and red) for either positively or negatively rewarding the robot. To facilitate the understanding by the children of the reasons the robot makes mistakes, the strategy adopted here consists in having the robot verbally convey the degree of uncertainty it has for each choice it has to make. In the next section, we describe the task performed by 30 children. We then describe the reinforcement learning algorithm, the results of the experiments, and a short discussion.

II. EXPERIMENTAL PROTOCOL

We have set up a novel experiment (game)[1] where the Kaspar robot is interacting with two/three children (Fig. 1).

In this game, children work together to teach Kaspar how to recognize and point at different animal toys in different locations within the room. The children achieve this in a number of steps. Importantly, before starting the sessions Kaspar is already aware of having 6 potential toys with different names. However, Kaspar does not know which name is associated with which animal toy. The children take it in turns to teach the robot to recognize the toys.

- **Step 1:** The children place bagged mystery toys around the room with the help of the researcher making sure that they are given sufficient space.
- **Step 2:** When the bags have been positioned, each child teaches the robot to point at all bags successively. To do this, one of the children physically manipulates the arm of the robot to point at each of the bags, whilst another child indicates to Kaspar when its arm is in the right position with a button. Kaspar indicate when he has logged the position. Once Kaspar has been shown

[1]School of Computer Science, University of Hertfordshire, Hatfield, UK.
[2]School of Electrical and Computer Engineering, National Technical University of Athens, Greece.
[3]Sorbonne Université, CNRS, Institute of Intelligent Systems and Robotics, Paris, France.
Correspondence: a.zaraki@herts.ac.uk, mehdi.khamassi@upmc.fr, l.wood@herts.ac.uk, g.lakatos@herts.ac.uk, ktzaf@cs.ntua.gr, b.robins@herts.ac.uk, k.dautenhahn@herts.ac.uk

[1]This experiment has been approved by the ethical committee of **University of Hertfordshire** with the protocol number aCOM/SF/UH/03320(1). The children's parents provided written consents.

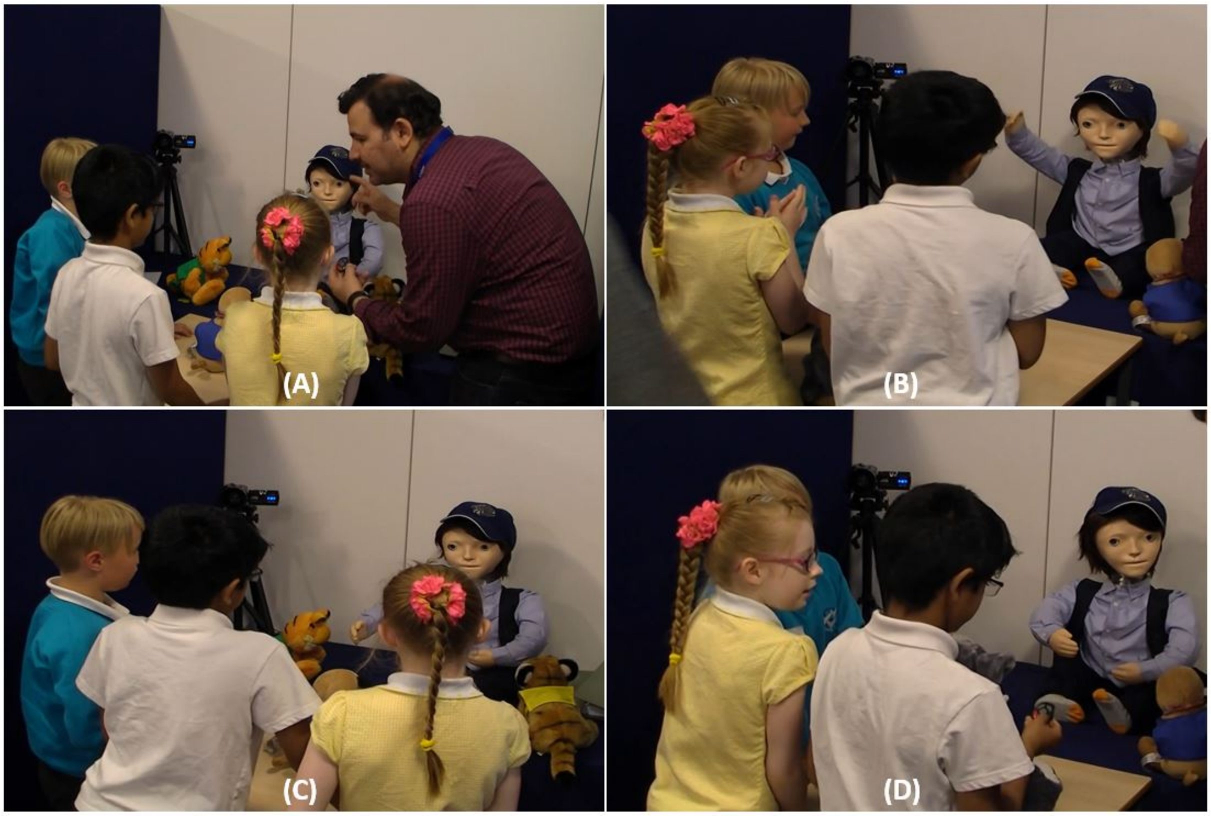

Fig. 1. (A) The researcher gives the instruction to students on how to give positive/negative reinforcement to the Kaspar robot, (B) Shows Kaspar behaviour once he has learned about the name of a toy, (C)(D) Kaspar is pointing at different toys and guessing the name of the toys.

how to point at all of the locations, the children switch roles and follow the same procedure again. This is mainly to ensure that all present children have an equal experience. Using this approach is similar in style to clicker training, a reinforcement learning method which is often used with dogs.

- **Step 3:** Once Kaspar has been shown how to point in the direction of the bags, the children next reveal the toys before moving onto the next part of the game.
- **Step 4:** Now all the toys are visible to Kaspar. The children must teach Kaspar the names of these animals. This is achieved by Kaspar by autonomously pointing at each of the toys and say the name of the animal that Kaspar thinks it is. Once again, the children answer by using the keyfob and pressing either the green or the red button dependent on Kaspar's actions (green = correct or red = incorrect). The children take this part in turns. Kaspar continues to guess the names of the animal toys until they can all be named correctly.
- **Step 5:** When all of animal toys can be correctly identified by Kaspar the game concludes with a thanks and farewell.

It should be noted that as Kaspar gets each animal correct this should be eliminated from the list of potential animals that are proposed for the ones that Kaspar is guessing after. In addition, Kaspar should work through the animal toys methodically and focus on an animal until it names it correctly. Since the focus of the current pilot study is to evaluate the performance of the proposed reinforcement learning model and to enable the robot to learn-on-the-fly in interaction with children, we have implemented and tested a simplified version of the experimental scenario in which Kaspar already knows how to point to different locations in an autonomous manner and the main goal is to make it learn arbitrary associations between the toys and locations.

III. Robot learning algorithm

The proposed algorithm is summarised in Algorithm 1. It is based on the reinforcement learning framework [7] where the set of discrete actions $A = \{a_1, a_2, ..., a_k\}$ represent the possible toys among which the robot can choose, and the set of discrete states $S = \{s_1, s_2, ..., s_j\}$ represent the possible locations where to put these toys. For the experiment described in Section II, we consider 6 toys and 3 locations. Learning the value of discrete action $a_t \in A$ selected at time step t in state $s_t \in S$ is done through Q-Learning [8], which is a parsimonious algorithm for reinforcement learning with discrete state and action spaces:

$$\Delta Q_t(s_t, a_t) = \alpha \left(r_t + \gamma \max_a (Q_t(s_{t+1}, a)) - Q_t(s_t, a_t) \right) \tag{1}$$

where α is a learning rate and γ is a discount factor. The probability of executing action a at timestep t is given by a

Algorithm 1 Interactive reinforcement learning algorithm

Initialize $Q_0(s, a)$
for $t = 0, 1, 2, ...$ **do**
  Observe the current state s_t (*i.e.,* location)
  Select discrete action a_t (Eq. 2)
  Communicate choice uncertainty H_t (Eq. 3)
  Observe reward r_t given by the children
  Update $Q_{t+1}(s_t, a_t)$ (Eq. 1)
end for

Boltzmann softmax equation:

$$P(a|s_t) = \frac{exp\left(\beta Q_t(s_t, a)\right)}{\sum_{a'} exp\left(\beta Q_t(s_t, a')\right)} \tag{2}$$

where β is the inverse temperature parameter which controls the exploration-exploitation trade-off. Finally, following [9], we measure the choice uncertainty as the entropy in the action probability distribution:

$$H_t = -\sum\nolimits_a \left(P(a|s_t)\log_2 P(a|s_t)\right) \tag{3}$$

This choice uncertainty is verbally expressed by the robot before each action execution, so as to help the children understand why the robot may hesitate, be sometimes sure of its answer, and sometimes not. Figure 2 in the results section illustrates the different types of phrases that the robot may use to express different levels of choice uncertainty.

IV. Results

Before testing this paradigm with children with autism, this paper presents the results of a first study with Typically Developing (TD) children. Kaspar performed 16 sessions of this experiment with 30 children aged between 7 and 8 (19 girls, 11 boys). Each session included a group of 3 children. Some groups asked to do one more session of the experiment since they really enjoyed it. For each group of children, Kaspar was successively confronted to 3 *problems*, each one consisting in learning which among 6 toys should be placed in a given location. The robot thus faced a total of 48 *problems* during 202 trials. On average, it took the robot 4.21±1.46 trials to learn each (location, toy) association, with a minimum of 2 trials and a maximum of 7. Figure 2 shows the trial-by-trial evolution of the robot's choice uncertainty averaged over the learning of all experienced problems. Uncertainty was measured as the entropy in probability distribution over the 6 toys. The maximal uncertainty thus starts at 2.5850, which is obtained for 6 equiprobable actions (*i.e.*,, $P = 1/6$) with Equation 3. This initial maximal uncertainty makes the robot verbally express that it initially has no clue about the correct toy associated to the current location. For an uncertainty around 1.5, the robot verbally expresses that it has an idea about the correct toy but is not certain. For an uncertainty below 0.5, the robot verbally expresses that it is sure of the answer.

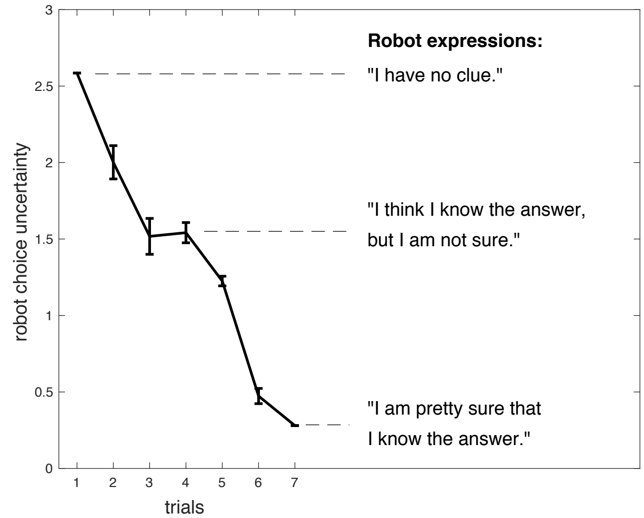

Fig. 2. Trial-by-trial evolution of the robot's choice uncertainty averaged over the learning of all experienced problems. Uncertainty was measured as the entropy in probability distribution over the 6 toys. For each level of entropy, the robot used a different verbal expression to convey its current uncertainty in selecting a toy for the considered location.

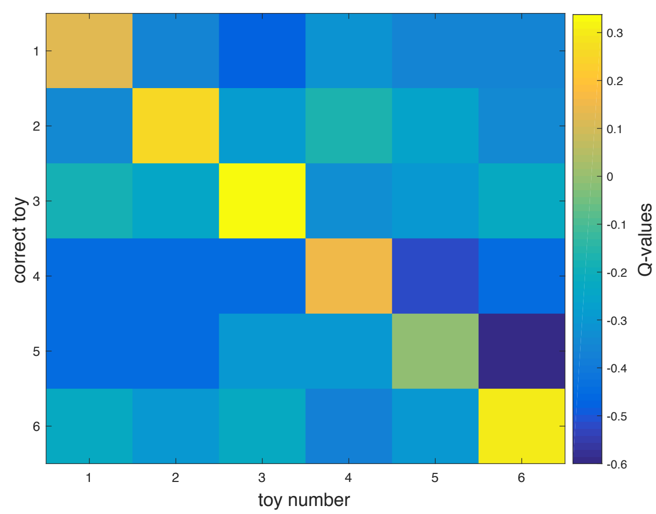

Fig. 3. Final Q-values obtained on average at the end of learning problems. Problems are here regrouped depending on which toy number was the correct answer to the problem (lines in the figure). Columns represent each possible toy number. Colors indicate the final Q-value of a given toy number averaged over all problems. Kaspar successfully learned the correct toys in each problem.

Overall, Figure 2 illustrates that the children successfully managed to make Kaspar progressively reduce its choice uncertainty along learning. The final entropy at the end of learning problems was on average 0.73±0.58. Moreover, Figure 3 shows that at the end of each learning problem, Kaspar found the correct toy: the represented matrix is diagonal, illustrating that the final Q-value of the correct toy in a given problem was always higher than the Q-values learned for the other toys.

Importantly, among the total of 202 trials performed by the robot in interaction with the 30 children, Kaspar made a single perseverative error (selecting two consecutive toys despite the negative feedback given by the children after the first selection), and a single win-shift (picking a different toy despite the positive feedback given by the children after the previous selection). These cases were due to exploratory choices occasionally made by Kaspar. These events were rare because we used a high learning rate ($\alpha = 0.6$) and a high inverse temperature ($\beta = 8$) for these experiments. Moreover, the robot picked the correct toy by chance at the first trial of a given problem only 7 times in total, maximum once per session of the experiment (thus once per group of children). The rarity of this event thus avoided children to be disturbed by the choice behaviour of the robot. Finally, and of particular importance, only 3 children gave an incorrect feedback to Kaspar, and this happened only once by each of these 3 children. Strikingly, these 3 cases where false positives where children rewarded the robot for a wrong choice, and no case of false negative occurred in our task.

V. Conclusions and future work

In this short paper, we presented recent progresses in developing robot learning abilities for the Kaspar robot [1], [2] in order to make it learn from human feedback during social interaction. The goal is to help develop children's social skills by putting them in the position of teachers having to assign feedback to the robot to make it learn. We used a simple reinforcement learning algorithm combined with verbal expression of the robot's choice uncertainty, in order to facilitate the understanding by the child of the reasons the robot makes mistakes.

The experiment yielded promising results, where all tested 30 TD children managed to make the robot progressively reduce its choice uncertainty and learn (toy, location) associations in less than 7 trials. Moreover, the robot displayed a coherent behaviour with very rare perseverative errors and win-shifts. These results, plus the positive impression expressed by children in the post-experiment questionnaire, suggest that the proposed paradigm is adequate and can now be transferred to children with autism.

In future work, we first plan to perform the experiment with children with autism and investigate the differences in teaching behaviour that they may adopt compared to TD children. Later, we also plan to extend the robot learning algorithm to different levels of adaptivity to make it look more or less smart in the eyes of interacting children. The goal would be to have each group of children interact with 2 different Kaspar robots, one being a fast learner and the other a slow learner, to help children with autism understand that different agents may need different numbers of repetitions, and different types of feedback to learn a given task. Besides, it would be interesting to compare learning performance of the present algorithm with a version where Kaspar does not verbally convey information about its choice uncertainty, to evaluate the impact of this communication on children's teaching behaviour during the task.

Acknowledgements

This research work has been partially supported by the EU-funded Project BabyRobot (H2020-ICT-24-2015, grant agreement no. 687831), and by the Centre National de la Recherche Scientifique (MI ROBAUTISTE & PICS 279521).

References

[1] L. J. Wood, A. Zaraki, M. L. Walters, O. Novanda, B. Robins, and K. Dautenhahn, "The iterative development of the humanoid robot kaspar: An assistive robot for children with autism," in *International Conference on Social Robotics*. Springer, 2017, pp. 53–63.

[2] A. Zaraki, K. Dautenhahn, L. Wood, O. Novanda, and B. Robins, "Toward autonomous child-robot interaction: Development of an interactive architecture for the humanoid kaspar robot," in *Proceedings 3rd Workshop on Child-Robot Interaction at HRI 2017*, vol. 3. https://childrobotinteraction.files.wordpress.com/2017/03/abolfazl-zaraki.pdf, 2017, pp. 1–4.

[3] M. Khamassi, S. Lallée, P. Enel, E. Procyk, and P. Dominey, "Robot cognitive control with a neurophysiologically inspired reinforcement learning model," *Frontiers in Neurorobotics*, vol. 5:1, 2011.

[4] L. Wood, K. Dautenhahn, B. Robins, and A. Zaraki, "Developing child-robot interaction scenarios with a humanoid robot to assist children with autism in developing visual respective taking skills," in *The 26th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2017*. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8172434, 2017.

[5] B. Robins, K. Dautenhahn, L. Wood, and A. Zaraki, "Developing interaction scenarios with a humanoid robot to encourage visual perspective taking skills in children with autism–preliminary proof of concept tests," in *International Conference on Social Robotics*. Springer, Cham, 2017, pp. 147–155.

[6] M. Khamassi, G. Velentzas, T. Tsitsimis, and C. Tzafestas, "Robot fast adaptation to changes in human engagement during simulated dynamic social interaction with active exploration in parameterized reinforcement learning," *IEEE Transactions on Cognitive and Developmental Systems*, 2018, in press.

[7] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press, 1998.

[8] C. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, no. 3-4, pp. 279–292, 1992.

[9] G. Viejo, M. Khamassi, A. Brovelli, and B. Girard, "Modeling choice and reaction time during arbitrary visuomotor learning through the coordination of adaptive working memory and reinforcement learning," *Frontiers in behavioral neuroscience*, vol. 9, 2015.