A Novel Reinforcement-Based Paradigm for Children to Teach the Humanoid Kaspar Robot

Abolfazl Zaraki, Mehdi Khamassi, Luke J Wood, Gabriella Lakatos, Costas S Tzafestas, Farshid Amirabdollahian, Ben Robins, Kerstin Dautenhahn

A Novel Reinforcement-Based Paradigm for Children to Teach the Humanoid Kaspar Robot

Abolfazl Zaraki[1] · Mehdi Khamassi[2] · Luke J. Wood[1] · Gabriella Lakatos[1] · Costas Tzafestas[3] · Farshid Amirabdollahian[1] · Ben Robins[1] · Kerstin Dautenhahn[1,4]

Abstract

This paper presents a contribution aiming at testing novel child–robot teaching schemes that could be used in future studies to support the development of social and collaborative skills of children with autism spectrum disorders (ASD). We present a novel experiment where the classical roles are reversed: in this scenario the children are the teachers providing positive or negative reinforcement to the Kaspar robot in order for it to learn arbitrary associations between different toy names and the locations where they are positioned. The objective is to stimulate interaction and collaboration between children while teaching the robot, and also provide them tangible examples to understand that sometimes learning requires several repetitions. To facilitate this game, we developed a reinforcement learning algorithm enabling Kaspar to verbally convey its level of uncertainty during the learning process, so as to better inform the children about the reasons behind its successes and failures. Overall, 30 typically developing (TD) children aged between 7 and 8 (19 girls, 11 boys) and 9 children with ASD performed 25 sessions (16 for TD; 9 for ASD) of the experiment in groups, and managed to teach Kaspar all associations in 2 to 7 trials. During the course of study Kaspar only made rare unexpected associations (2 perseverative errors and 2 win-shifts, within a total of 314 trials), primarily due to exploratory choices, and eventually reached minimal uncertainty. Thus, the robot's behaviour was clear and consistent for the children, who all expressed enthusiasm in the experiment.

Abolfazl Zaraki and Mehdi Khamassi have contributed equally to this work.

✉ Mehdi Khamassi
mehdi.khamassi@upmc.fr

Abolfazl Zaraki
a.zaraki@herts.ac.uk

Luke J. Wood
l.wood@herts.ac.uk

Gabriella Lakatos
g.lakatos@herts.ac.uk

Costas Tzafestas
ktzaf@cs.ntua.gr

Farshid Amirabdollahian
f.amirabdollahian2@herts.ac.uk

Ben Robins
b.robins@herts.ac.uk

Kerstin Dautenhahn
k.dautenhahn@herts.ac.uk

1 Introduction

In recent years, the field of social robotics has expanded widely from using robots to assist with learning language [29], to robotics pet for the elderly with dementia [15], to assisting children with autism [10]. An area particularly relevant to this paper is assistive robotics for children with autism which is often referred to as autism spectrum disorder (ASD). Children with ASD can have difficulty communi-

[1] School of Computer Science, University of Hertfordshire, Hatfield, UK

[2] Institute of Intelligent Systems and Robotics, CNRS, Sorbonne Université , Paris, France

[3] School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece

[4] Departments of Electrical and Computer Engineering/Systems Design Engineering, University of Waterloo, Waterloo, Canada

cating and forming social relationships and as a result of this they can become very isolated. The field of assistive robotics for children with ASD emerged during the late 1990s with K. Dautenhahn pioneering studies in this area [6]. Over the years, many different types of robots have been used by various researchers to engage autistic children in playful interactions, e.g., mobile robot IROMEC [8] artificial pets such as the baby seal robot Paro and the teddy bear Huggable [17,25], and creature/cartoon-like robots such as Keepon [14] and Probo [23] to mention just a few. Humanoid robots, e.g., Robota, Nao, Kaspar, Milo have been used with children with autism to help mediate interactions with peers and adults and also develop other skills, e.g., tactile interaction, emotion recognition [2,7,18,28]. It is thought that one of the reasons that children with ASD respond well to the humanoid robots is because they are predictable and can have a simplified appearance [19,20].

Unlike the conventional teaching scenarios where the robot functions as a teacher for children, some other works have followed the concept of "learning by teaching" [22] for education purposes and devised scenarios in which children and adults play role as the teacher to give instructions and teach the robot. Tanaka and Shizuko [27] applied this concept into a classroom of Japanese children (aged 3–6 years) and the tele-operated Nao robot as a WoZ mediator. Conducting a study they confirmed the feasibility of applying learning by teaching concept with an agent and children and demonstrated it can promote learning English verbs in children. In a later study, Hood et al. [9] applied the same concept to allow children (aged 6–8 years) to teach and develop the English handwriting of an autonomous teachable humanoid robot with the final aim of building an interaction, so as to stimulate meta-cognition, empathy and increased self-esteem in the children users. They showed that the children were successfully engaged in interaction with the robot where they did teach and improved the handwriting skill of the robot. There are other works in educational robotics field which also followed the learning by teaching concept with virtual screen-based computer agents rather than a physical robot, in different contexts e.g. [3,4]. Although the outcomes of using virtual agents were successful, comparing the screen-based agents, the robotic agents have been shown to be more effective in long-term relationships [13], increase users' compliance with tasks [1], and produce greater learning gains when acting as tutors [16].

In this paper, we present a novel experiment where children provide humanoid Kaspar robot [34] with positive or negative reinforcement to make it learn arbitrary associations between 6 different toys and 3 possible locations where these toys have been placed. The main goal of this work is to come up with a novel paradigm where children interact and collaborate to teach the robot something through trial-and-error, and where the robot verbally expresses its level of uncertainty to help the children understand its state of mind. To facilitate this game we developed robot learning abilities for the Kaspar robot which has a proven track record in working with children with autism [33,35]. The learning algorithm enabled Kaspar to learn about the environment from human feedback during the interactions. In a more long-term perspective, we further discuss at the end of the paper how this type of experiments could be used in future studies to assist in the development of social and collaborative skills of children with ASD via fun and friendly interactive games. By putting the children in control and in a position of power as teacher, this could help build the children's confidence which will further assist their development.

A growing number of studies have tackled the complex challenge of applying reinforcement learning (RL) techniques for robot's behaviour adaptation during social interaction with humans (e.g. Khamassi et al. [11]). Machine learning and in particular supervised learning algorithms can play a role in improving the human–robot interaction quality, by providing means to learn from interaction and ability to personalise to user's preferences. However, most 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 collaborative abilities in these children [21,32].

Our earlier study in this field explored using a two-layered reinforcement learning algorithm to enable a robot to preserve energy and adapt to the human companion's preferences regarding choice of play [5]. This study showed the feasibility of using machine learning in interaction design but was limited to a very simple learning scenario. In contrast, other robot learning studies in social interaction contexts have typically focused on extending reinforcement learning algorithms to make them cope with the high degrees of uncertainty, volatility and non-stationarity associated with non-verbal communication (e.g. Khamassi et al. [12]).

In this study, we adopt a much simpler reinforcement learning algorithm where the goal is to have the robot *making mistakes*, and the children provide the *correct feedback* in order to assist the robot correctly learn arbitrary associations. The basic principle of the game consists of the robot pointing to predefined locations and guessing which toy is in that location. During the game, the children can press two buttons (green and red) on a keyfob 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 TD children during a main experiment, and by 9 children with ASD during a pilot test serving as proof-of-concept that the paradigm can work similarly for children with ASD.

We then describe the reinforcement learning algorithm, the results of the experiments, and a discussion.

2 Development of the Game Protocol

We have set up a novel experiment (game)[1] where the Kaspar robot is interacting with two/three children (Fig. 1). Because the primary objective of this game was to encourage collaboration and social skills with other children, it was important that the game includes two children and that they have to collaborate. Based on our previous work in assisting children with ASD to learn about visual perspective taking (VPT) [32], we devised an interaction scenario where the children would work together to teach Kaspar point at specific locations and guess what animal toy was in that location. 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 turns to teach the robot to recognise the toys.

2.1 The initial Proof of Concept Tests

To test the concept of the reinforcement learning scenario, a pilot-study was conducted in the lab with 3 children (two boys and one girl), see Fig. 1. The children were aged between 8 and 12, and were accompanied by their mother at all times. Two of the children were typically developing whilst one child was diagnosed with ASD. Two iterations of the game were conducted, the child with ASD took part in both of the sessions whilst the other children took part in one session each. Generally the children responded well to the game and found teaching Kaspar very engaging. This pilot indicated that the game had the potential to work effectively with some children with ASD and led us to develop the procedure further. The main lesson learned from the first pilot was that the robot will need to have a device in its hand to point with as the children found it difficult to see where exactly Kaspar was pointing to. In the second iteration of the game we put a piece of card with the arrow shape in Kaspar's hand to make it clearer where it was pointing and this was very effective. In addition to this, to create an additional fun and mystery element to the game we changed the procedure so that the toys would be in bags and were therefore mystery toys whilst the robot learned the locations and the children would then reveal the toys before Kaspar tried to guess the names of the

animals. A second pilot study was run with the same two typically developing children and the changes implemented proved to be successful, which led to finalise the procedure into the following 6 steps.

- *Step 1* The children will place three bagged mystery toys around the room in three locations with the help of the researcher making sure that they are given sufficient space.
- *Step 2* When the bags have been positioned, each child will teach the robot to point to all of the bags. To do this one of the children will physically manipulate the arm of the robot to point at each of the bags, whilst the other child will indicate to Kaspar when its arm is in the right position by pressing the green button (for positive feedback) on a keyfob. Kaspar will indicate when he has logged the position. Once Kaspar has been shown how to point at all of the locations the children will switch roles and follow the same procedure again. This is mainly to ensure that each child has an equal experience.
- *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. Kaspar has several trials to guess the name of the toy at each of the 3 locations. This is achieved by Kaspar by autonomously pointing at each of the toys and saying the name of the animal that it 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). Kaspar makes several trials until finding the correct name. It then repeats a second correct trial to make the children sure that it has correctly learned. The children take this teaching part in turns. Kaspar continues to guess the names of the 3 animal toys until they can all be named correctly.
- *Step 5* Once Kaspar can correctly identify all of the animal toys the children will be asked to swap the toys locations.
- *Step 6* After the toys have all been set at their new place, Kaspar will again try to point and name all the animal toys following the same procedure as step 4. When all of animal toys can be correctly identified by Kaspar the game will conclude with a thanks and farewell.

In this game, Kaspar works through the animal toys methodically and focuses on an animal until it names it correctly. We did not include deduction mechanisms that would have allowed Kaspar to eliminate a toy name from the list of potential animals once this name has already been correctly assigned to a previous toy. This is because in this study we want to assess the effect of a simple learner verbally con-

---

[1] This research was approved by the University of Hertfordshire's ethics committee for studies involving human participants, protocol numbers: acCOM SF UH 02069 and aCOM/SF/UH/03320(1). Informed consent was obtained in writing from all parents of the children participating in the study.

Fig. 1 Children playing the reinforcement learning game with Kaspar—proof of concept tests. **a** Placing the toys on different chairs. **b**, **c** Manipulating the robot's arm to teach it the location of interest. **d** Inverting roles and repeating the steps described in **a–c** so that each child has experienced each role. **e**, **f**, Teaching the name of the animal toys

veying its uncertainty along learning to the teacher, without requiring the children to understand that the robot can also make deductions. Moreover, 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 in all three locations and their respective names.

3 The Implementation of RL Algorithm on Kaspar Robot

3.1 Robot Learning Algorithm

The proposed algorithm is summarised in Algorithm 1. It is based on the reinforcement learning framework [26] where the set of discrete actions $A = \{a_1, a_2, \ldots, a_k\}$ represent the possible toy names among which the robot can choose, and the set of discrete states $S = \{s_1, s_2, \ldots, s_j\}$ represent the possible locations where the toys can be encountered. For the experiment described in Sect. 2, 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 [31], 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 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 [30], we measure the choice uncertainty as the entropy in the action probability distribution:

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

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 5 in the results section illustrates the different types of phrases that the robot may use to express different levels of choice uncertainty.

3.2 Using the Interactive Deliberative–Reactive Robot Control Architecture

In order to implement the reinforcement learning based game and conduct our child–robot interaction studies, we have developed an interactive deliberative–reactive architecture

Fig. 2 The interactive semi-autonomous deliberative–reactive control architecture for child–robot interactions in different contexts. The system enables robot to learn about environment based on the reinforcing feedbacks given by children. It then generates dynamic behaviours for the robot based on the level of uncertainty. As it can be seen the interactive system closes the interaction loop between child and robot in an autonomous manner

---
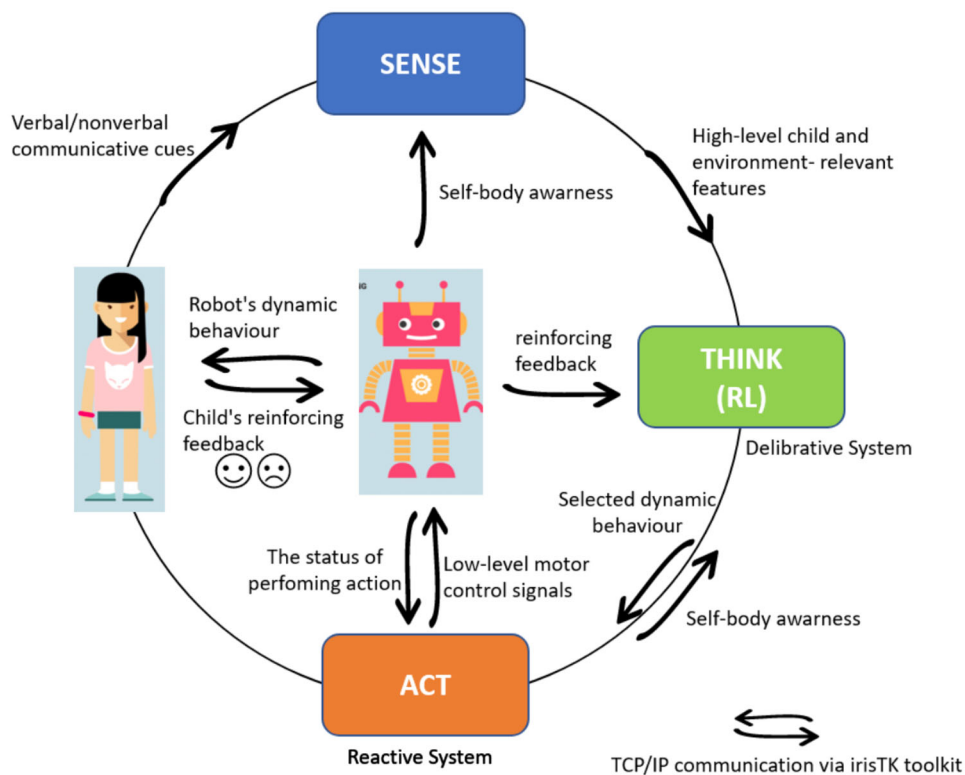
Algorithm 1 Interactive reinforcement learning algorithm

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

---

which controls the robot's dynamic behaviour base on the perceptual information shown by children while interacting with the robot. As shown in Fig. 2, the architecture has three main components called Sense, Think and Act which manage the interaction data flow from the low-level data acquisition and analysis, to the high-level decision making and robot control. In addition to the handling the interaction data flow, the system has a self-body awareness module which actively monitors the current and previous actions (i.e. body gesture and speech) of the robot and passes this information to the system which enables the Think layer to make a right decision for the next action of the robot. In addition to the system capability in detecting and tracking a wide range of children social features (e.g. body gesture, proximity, facial features, and spoken words), it receives the children reinforcing feedback during the interaction through two buttons (red and green) which are integrated directly into the robot's hardware. As shown in Fig. 2, the components of the system are

fully interconnected via a TCP/IP network which allow each component to send and receive the high-level data to other components as events as standard JSON data packets (please refer to [36] for the details of implementation).

The RL algorithm has been implemented as a part of the Think component (deliberative system) of the architecture. The deliberative system activity receives the child's reinforcing feedback along with the other high-level perceptual information and learns about the interaction by updating the level of uncertainty about the interaction which results in controlling the robot's behaviour in a dynamic and autonomous manner.

The deliberative system of the architecture as well as the communication modules have been developed using the IrisTK toolkit, which is a powerful Java-based framework for developing multi-modal interactive systems. Thanks to the IrisTK capability, the data flow at any level can be distributed to a Broker as JSON events and any system, machine, or robots can have a bidirectional communication by subscribing to this Broker (please refer to [24] for the detail of the IrisTK toolkit).

4 Testing the Game with Typically Developing (TD) Children

Once the core learning algorithm had been developed and integrated into the Kaspar architecture, we created a basic version of the game to test the system with 30 typically devel-

Fig. 3 Children playing the reinforcement learning game with Kaspar. **a** Researcher explaining the rules of the game to the children. **b** Kaspar introducing itself and asking the children to teach it the toy names. **c** Children looking at the toy in the direction in which Kaspar is pointing. **d** One of the children clicks on a button of a keyfob to send Kaspar a positive or negative feedback
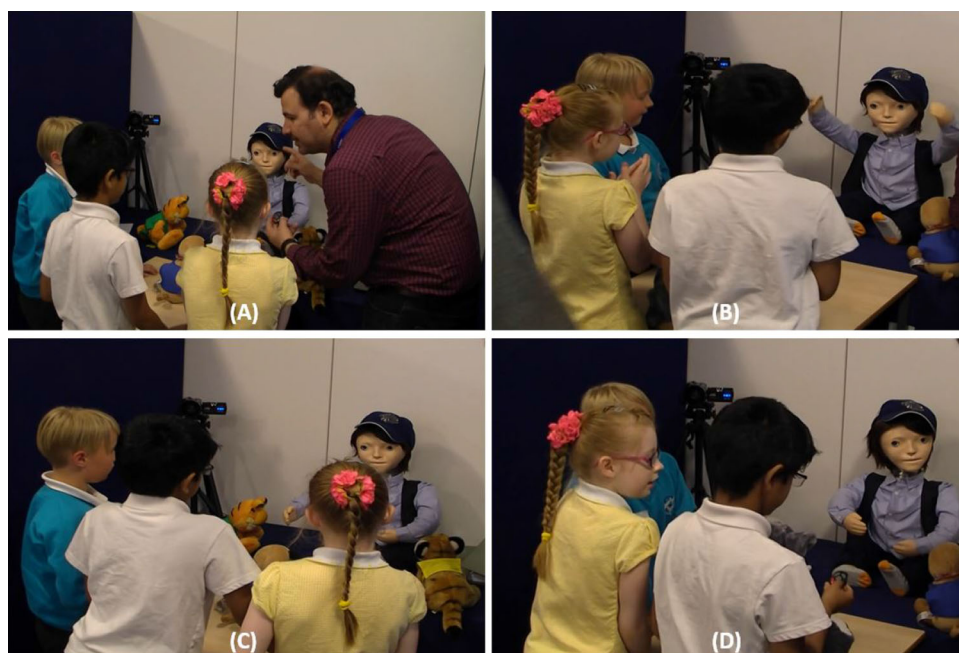
oping children prior to testing the system with children with ASD. In this basic version of the game there were 6 animal themed toys and 3 pre-defined locations marked on a table in front of Kaspar with an *X*. A group of 3 to 4 children would randomly place 3 of the 6 toys into the 3 locations, then Kaspar would try to guess which toy was in each location. The children would provide feedback via a keyfob with a red and green button, red indicated incorrect, green indicated correct. Each child got the opportunity to teach Kaspar where the 3 toys were. Figure 3 shows children playing the game with Kaspar. The primary goal here was to test the system with a more flexible user group before working with the primary target user group. During this testing, we established that the reinforcement learning algorithm was effective but there were some usability issues with the system that would need to be resolved prior to working with children with ASD. The primary issue was the small keyfob that was being used to provide feedback to the robot. This was far too small and the children would sometimes press the button by accident while they were holding it and would sometimes press the wrong button because it was so small. To remedy this we mounted 2 large buttons on a box for the children to press so they would not press it accidentally and would not mistakenly press the wrong button. Further to this we also added a beep noise so that the children would know that Kaspar had received their feedback and would not press the button multiple times.

## 5 Using the System with Children with ASD

Figure 4 shows two children with ASD playing the game with Kaspar. So far, three pairs and a trio of children with ASD played the game with Kaspar. All of the children were of different abilities. The first 2 pairs and the trio of children played the game with Kaspar well, whilst the third pair had some difficulty because one of the children was easily distracted and was not interested unless he was leading the interaction. This is very typical for some children with ASD and is to be expected.

6 Results

We present results with 30 TD children aged between 7 and 8 (19 girls, 11 boys), and 9 children with ASD. Kaspar performed a total of 25 sessions of this experiment. Each session included a group of children. TD children participated in the experiment in larger groups (3–4 children per session) while children with ASD worked mostly in pairs and once in a group of 3. 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 of learning which of the 6 possible animal names should be associated to a toy at a given location. The robot thus faced a total of 48 *problems* during the sessions with TD children, and 27 *problems* during the sessions with children with ASD. In total, 202 trials were performed during the 16 sessions with the TD children, and a total of 112 trials during the 9 sessions involving children with ASD. On average, it took the robot 4.19 ± 1.46 trials to learn each (location, toy) association, with a minimum of 2 trials and a maximum of 7. We found no difference in this number between TD children (4.21 ± 1.46 trials) and chil-

Fig. 4 Children with autism playing RL game with Kaspar. **a** Manipulating the robot's arm while teaching locations. **b** Teaching the animal name. The small key fob to reward the robot by TD children has been here replaced by 2 large buttons on a box on the Table
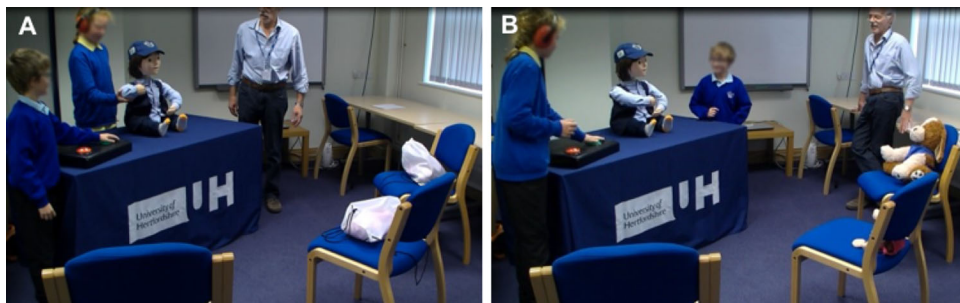
dren with ASD (4.15 ± 1.49 trials; Wilcoxon Mann–Whitney test, $z = -0.1633$, $p = 0.87$).

6.1 Model-Based Analyses

Figure 5 shows the trial-by-trial evolution of the robot's choice uncertainty averaged over the learning of all experienced problems with both TD children and children with ASD. Uncertainty was measured as the entropy in probability distribution over the 6 toys with Eq. 3. The maximal uncertainty thus starts at 2.5850, which is obtained for 6 equiprobable actions (i.e., $P = 1/6$). 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.

Overall, Fig. 5 illustrates that both groups of children successfully managed to make Kaspar progressively reduce its choice uncertainty along learning. In order to compare the two curves with the same number of sample measures, we focused on the 5 first trials for which there were at least 11 samples (i.e., 11 problems) for both groups (TD and ASD). We made a two-way ANOVA comparing choice uncertainty as variable and trial number $\times$ group (TD, ASD) as predictors, and found a significant effect of trial number ($F = 16$, $p = 3.7\mathrm{e}{-}10$), thus revealing a significant improvement trial after trial. Importantly, we found neither an effect of group ($p = 0.87$) nor an interaction between factors ($p = 0.07$). This finding suggests that the two groups of children did not differ in the reduction of the robot's choice uncertainty that resulted from their feedback. We also compared the final entropy at the end of learning problems. Note that the end of a problem could occur at different trial numbers, depending on how lucky the robot was in quickly finding the solution of a given problem. The average final entropy for TD children (0.73 ± 0.58) and children with ASD (0.68 ± 0.64) did not significantly differ (Wilcoxon Mann–Whitney test, $z = -0.89$, $p = 0.37$), suggesting that the final level of choice uncertainty reached by the robot was not different between the two groups.

In order to further analyse the effect of learning induced by children's feedback, we compared the final Q-values reached by the robot at the end of each learning problem within the two groups. Figure 6 illustrates for both TD and ASD groups the average final Q-values associated to each possible name for each of the 6 different encountered toys. The fact that for each group the represented matrix is diagonal illustrates that the final Q-value of the correct toy name in a given problem was always higher than the Q-values for the other toy names. This illustrates that Kaspar eventually found the correct toy name in each encountered problem. Importantly, in order to compare the quality of learning between the two groups, we shifted the final Q-values of each problem to make them all positive, we then normalised them and then measured the entropy with Eq. 3. This constitutes a way to measure the contrast between the correct Q-value and the other Q-values: a high entropy means little contrast; a low entropy means a large contrast, and thus reliable learning. The entropy of final Q-values for TD children (1.06 ± 0.88) and children with ASD (1.12 ± 0.89) were not significantly different (Wilcoxon Mann–Whitney test, $z = 0.41$, $p = 0.68$), thus suggesting no difference in the quality of learning between the two groups.

6.2 Analysis of the Robot's Behaviour During the Experiments

Overall, the ability of the robot to successfully learn from the children's feedback during this game is due to the very consistent behaviour that the robot displayed during each toy-guessing problem. During a pilot study with lab members as subjects, we tuned the two main parameters of the reinforcement learning algorithm, α and β, so as to find parameters that produce some behavioural variability, which is useful for learning, but nevertheless most of the time produce choices consistent with past feedback (exploitation). We ended up choosing a high learning rate ($\alpha = 0.6$) and a high inverse temperature ($\beta = 8$), which were then used for all the experiments with children. As a consequence, during the game with children, the robot systematically started a problem by exploring different possible toy names, almost never repeating an incorrect guess (apart from 2 exceptions which we analyse in the next paragraph). Then the robot almost sys-

Fig. 5 Trial-by-trial evolution of the robot's choice uncertainty averaged over the learning of all experienced problems with both TD children and children with ASD. 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
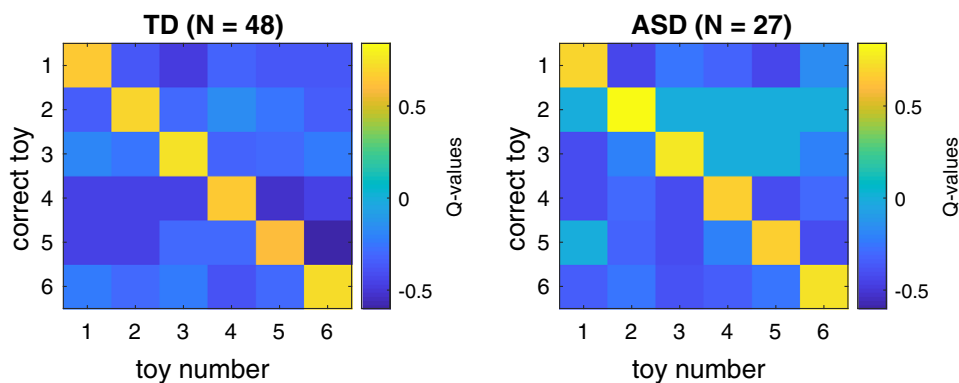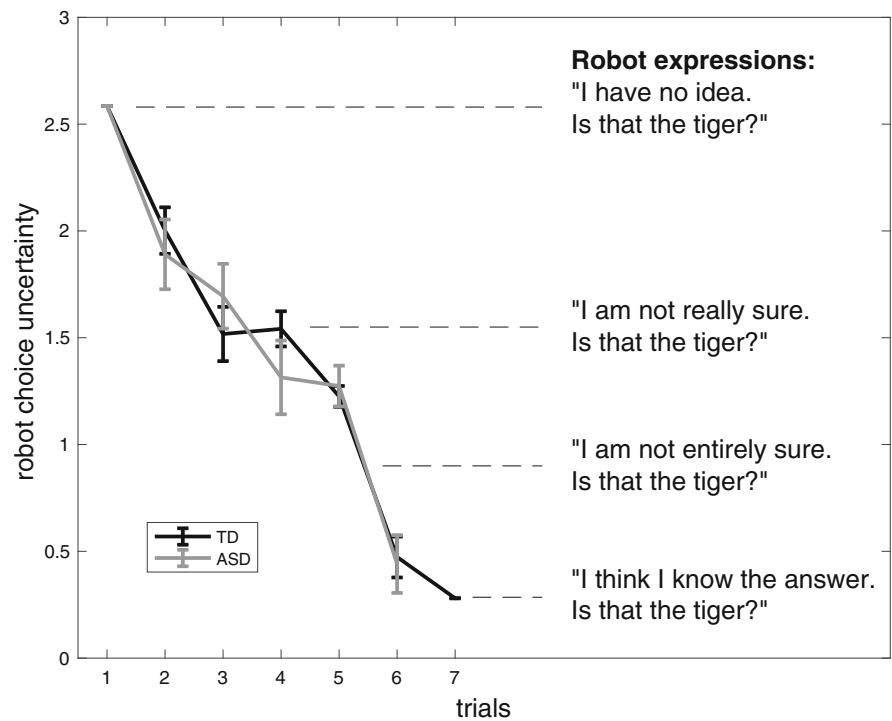
Fig. 6 Final Q-values obtained on average at the end of learning problems. Problems are here regrouped depending on which toy number was the correct answer to the problem (lines in the figure). Columns represent each possible toy number. The toy numbers correspond to the following toys used in the experiment: 1, monkey; 2, tiger; 3, cat; 4, elephant; 5, pig; 6, dog. Colors indicate the final Q-value of a given toy number averaged over all problems. The toy with the highest final Q-value is indicative of what the robot thinks is the correct answer at the end of the game. The diagonal shape of the two matrices indicates that Kaspar successfully learned the correct toys in each encountered problem with both groups of children: TD and ASD

tematically sticked to the correct toy name after a child's positive feedback (again apart from 2 exceptions which we discuss hereafter). This consistent robot behaviour, together with Kaspar's tendency to verbally express its choice uncertainty before each guess, made the interaction fluid and easy to understand by the children (as further illustrated with the post-experiment questionnaire analyses in the next section).

Importantly, among the total of 314 trials performed by the robot in interaction with the children (202 trials with TD children and 112 with children with ASD), Kaspar made only a few cases of unexpected behaviour which could have

disturbed the children, but did not, thanks to its ability to provide justifications through expression of choice uncertainty. These rares cases were the following: 2 perseverative errors and 2 win-shifts. Kaspar made only two perseverative errors (selecting the same toy twice consecutively despite the negative feedback given by the children after the first selection): one with a group of 3 TD children, and one with a group of 3 children with ASD. Moreover, Kaspar made only two win-shifts (picking a different toy despite the positive feedback given by the children after the previous selection): one with a TD group, one with a ASD group. These cases were due

to exploratory choices occasionally made by Kaspar. These events were rare, again, because we used a high learning rate and a high inverse temperature for these experiments, as explained in the previous paragraph.

Other robot behaviours are interesting to analyze to better understand the dynamics of interaction that took place during these experiments. For instance, the robot picked the correct toy by chance at the first trial of a given problem only 7 times in total with TD children, and 5 times in total with children with ASD. This mostly occurred at maximum once per session of the experiment (thus once per group of children), except with one group of TD children and one group of children with ASD for which Kaspar twice luckily guessed the correct name of a presented toy. The strength of the present approach is that Kaspar expressed its choice uncertainty at each trial before making a guess. As a consequence, it was clear for the children that the robot had no clue and found the correct answer by chance, thus avoiding the belief that Kaspar may have known the answer in advance, or that something magical may have happened.

Finally, and of particular importance, the system is resilient to occasional incorrect feedback given by the children. This happened for 3 TD children and 3 children with ASD. This happened only once for each of these 6 children. Strikingly, 5 out of these 6 cases where false positives where children rewarded the robot for a wrong choice. A single case of false negative occurred in our task, performed by a child with ASD. Table 1 illustrates a typical case of incorrect positive feedback given by a child to Kaspar. The below retranscription of the dialog recorded during these trials gives a better idea of the interaction dynamics. The child incorrectly rewarded the robot at the second trial, pushing the robot to repeat the same toy name at the next trial. Then after discussion with the other children and the adult experimenter, the child realised that they should have sent a negative feedback to the robot, which they did at the next trial. Then the problem finished normally, with the robot correctly guessing the name of the encountered toy.

Retranscription of the dialog for the example shown in Table 1:

- *Experimenter (explaining the instructions): Kaspar can learn from you. Kaspar wants to learn about the name of the animal toys that we have. So please choose three of them and put them on the table.*
- *(The children place three toys in front of Kaspar)*
- *Experimenter (continuing to explain the instructions): If Kaspar guesses right, you press the green button, and if Kaspar guesses wrong, you press the red button. Ok?*
- *Children: Yes!*
- *Experimenter: Then let's start.*
- *Kaspar (pointing at the first toy): I have no idea. Is that the pig?*
- *Experimenter: Is that the pig?*

Table 1 Illustration of a typical case of incorrect positive feedback given by a child (here TD) to the robot

Trial number	1	2	3	4	5	6
Robot's guess	Pig	Tiger	Tiger	Elephant	Dog	Cat
Child's feedback	-1	1	-1	-1	-1	1

Here the correct answer was: cat

- *Child 1 (in charge of pressing buttons for giving Kaspar feedback): No.*
- *(Child 1 presses the red button, giving a negative feedback to Kaspar)*
- *Kaspar: I have no idea. Is that the tiger?*
- *Experimenter: Tiger?*
- *Child 2 (seeing the stripes on the back of the cat): Yes!*
- *Child 1 (repeating): Yes.*
- *(Child 1 then presses the green button, incorrectly giving a correct feedback to Kaspar)*
- *Experimenter: Yes. (hesitating) No, it's no the tiger.*
- *Child 2: No, it's not! It's not! It's not!*
- *Kaspar (speaking in the background): I think I know the answer. Is that the tiger?*
- *Child 1 (looking at the experimenter with a guilty face): I pressed 'yes'.*
- *Experimenter: Ah! That's fine. Press 'no' again.*
- *(Child 1 presses the red button)*
- *Kaspar: I have no idea. Is that the elephant?*
- *Experimenter: The elephant?*
- *Child: No.*
- *Experimenter: No, it's not the elephant.*
- *(Child 1 presses the red button and the game continues)*
- *Kaspar (after having eliminated most options): I am not entirely sure. Is that the cat?*
- *(Children are hesitating)*
- *Child 3: Yes!*
- *Experimenter: Yes, it's the cat.*
- *(Child 1 presses the green button)*
- *Kaspar: Yes! I got it right.*
- *Experimenter: Bravo Kaspar.*
- *(Kaspar imitates the sound of the cat)*

6.3 Post-experiment Questionnaire Analyses

In order to learn about children's pre-interaction expectations of Kaspar as well as about their opinion once having interacted with Kaspar, TD children were asked to fill in two questionnaires, one before and one after their interaction with Kaspar. Using questionnaires with ASD children who took part in the studies was not feasible.

Results indicated that while TD children rated the question "Do you think Kaspar can recognise different animals?"

as a maybe (mean = 3.24; SD = 1.05) on a 5-point Likert scale before interaction, having interacted with Kaspar they rated this question significantly higher (mean = 3.88; SD = 0.95; Wilcoxon signed rank test for pre/post comparison: $Z = -2.15$; $p = 0.03$). We found the same result for the question "Do you think Kaspar is able to copy the sound of animals?". While children rated this question as a maybe (mean = 3.10; SD = 0.90) before interaction, after interaction their rating was significantly higher with the average answer being yes or definitely yes (mean = 4.42; SD = 0.70; Wilcoxon signed rank test for pre/post comparison: $Z = -3.53$; $p = 0.00$). These findings strengthen the results that the children perceived indeed the interaction as if Kaspar had successfully learned, through this game, the name of the animals through the children's feedback. In addition, TD children rated the interaction experience very high on a 5-point scale (mean = 4.50; SD = 0.94), where 1 was very boring and 5 was very fun. 76 % of the children even said that Kaspar is more like a friend than a toy having interacted with it, while 57.14% of the children said the same before the interaction, which was statistically no different (McNemar test: $p = 0.22$).

7 Discussion

In this paper, we presented recent progresses in developing robot learning abilities for the Kaspar robot in order to make it learn from human feedback during social interaction. The goal was 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 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 why the robot sometimes makes mistakes, sometimes guesses the correct response by chance, and sometimes is certain about the answer after receiving correct feedback.

The experiment yielded promising results, where all tested 30 TD children and 9 children with ASD managed to make the robot progressively reduce its choice uncertainty and learn (toy, location) associations in less than 7 trials. The experiments with such a small number of children with ASD so far are preliminary, and can only serve as proof of concept that this paradigm can be used with children with ASD. Interestingly, statistical analysis of the robot learning process reveals no significant differences between the two groups of children (TD and ASD), meaning that the proposed child–robot interactive learning paradigm and the game scenario designed for this study, seem to work equally well and to be intuitively used by both groups. Furthermore, the findings from analysing TD children responses in the post-experiment questionnaire show that the children in overall did perceive

the whole interactive game as if Kaspar has indeed been able to successfully learn the correct toy/name associations through the series of trials performed and the rewarding feedback provided by the children effectively assuming the role of the teacher. These results, together with the positive impression expressed by children in the post-experiment questionnaire, suggest that the proposed paradigm is adequate and can now be transferred for further studies with children with autism.

Getting post evaluation of such types of experiments by children with ASD is often tricky, as it is difficult to have them fill questionnaires. Nevertheless, we got feedback from the teachers who found that interactions with the Kaspar robot had been positive and beneficial for the children with ASD. A quantitative measure of how such child–robot interaction may have contributed to improve children's social skills is beyond the scope of the present work. This would require repeating the interaction over weeks and establishing proper measures for the evaluation. So far, the present work rather validates the proposed novel protocol for child–robot interaction, and shows that Kaspar can learn from children through a smooth interaction facilitated by choice uncertainty expression by the robot throughout the learning process.

From multiple pilots, this study highlighted the importance of choosing the correct input device for training the robot, for example choice of correct switch and separation between positive and negative keys. One of the limitations of this study was the imbalance in number of children involved in each group. However, as the robot was able to assimilate the correct names with similar Q-values and uncertainty rating between the two groups, the study still showed the feasibility of using RL in practice. In order to allow for a reasonable session length where all participants can try the same process to have an equal experience, and to have a clear separation between the location of the objects for the robot to point to, we had a limited set of locations, and subsequently a small set of toys. Having proven feasibility, future studies can explore impact of larger set of toys and locations, on enjoyment and also on how well the algorithms performs with larger sets.

In future work, we first plan to perform the experiment with more 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.

Compliance with Ethical Standards

References

1. Bainbridge WA, Hart JW, Kim ES, Scassellati B (2011) The benefits of interactions with physically present robots over video-displayed agents. Int J Soc Robot 3(1):41–52

2. Billard A, Robins B, Dautenhahn K, Nadel J (2006) Building robota, a mini-humanoid robot for the rehabilitation of children with autism. RESNA Assist Technol J 19:37–49

3. Biswas G, Leelawong K, Schwartz D, Vye N et al (2005) Learning by teaching: a new agent paradigm for educational software. Appl Artif Intell 19(3–4):363–392

4. Brophy S, Biswas G, Katzlberger T, Bransford J, Schwartz D (1999) Teachable agents: combining insights from learning theory and computer science. Artif Intell Educ 50:21–28

5. Castro-González Á, Amirabdollahian F, Polani D, Malfaz M, Salichs MA (2011) Robot self-preservation and adaptation to user preferences in game play, a preliminary study. In: 2011 IEEE international conference on robotics and biomimetics (ROBIO). IEEE, pp 2491–2498

6. Dautenhahn K (1999) Robots as social actors: aurora and the case of autism. In: Proceedings of CT99, the third international cognitive technology conference, August, San Francisco, vol 359, p 374

7. Dautenhahn K, Nehaniv CL, Walters ML, Robins B, Kose-Bagci H, Mirza NA, Blow M (2009) KASPAR—a minimally expressive humanoid robot for human–robot interaction research. Appl Bionics Biomech 6(3–4):369–397

8. Ferari E, Robins B, Dautenhahn K (2009) Robot as a social mediator—a play scenario implementation with children with autism. In: 8th international conference on interaction design and children. IDC, pp 243–252

9. Hood D, Lemaignan S, Dillenbourg P (2015) When children teach a robot to write: an autonomous teachable humanoid which uses simulated handwriting. In: Proceedings of the tenth annual ACM/IEEE international conference on human–robot interaction. ACM, pp 83–90

10. Ismail LI, Verhoeven T, Dambre J, Wyffels F (2018) Leveraging robotics research for children with autism: a review. Int J Soc Robot 11:1–22

11. Khamassi M, Lallée S, Enel P, Procyk E, Dominey P (2011) Robot cognitive control with a neurophysiologically inspired reinforcement learning model. Front Neurorobot 5:1

12. Khamassi M, Velentzas G, Tsitsimis T, Tzafestas C (2018) Robot fast adaptation to changes in human engagement during simulated dynamic social interaction with active exploration in parameterized reinforcement learning. IEEE Trans Cogn Dev Syst 10(4):881–893

13. Kidd CD, Breazeal C (2008) Robots at home: understanding long-term human-robot interaction. In: 2008 IEEE/RSJ international conference on intelligent robots and systems. IEEE, pp 3230–3235

14. Kozima H, Yasuda Y, Nakagawa C (2007) Social interaction facilitated by a minimally-designed robot: findings from longitudinal therapeutic practices for autistic children. In: The 16th IEEE international symposium on robot and human interactive communication, 2007. RO-MAN 2007. IEEE, pp 599–604

15. Leng M, Liu P, Zhang P, Hu M, Zhou H, Li G, Yin H, Chen L (2018) Pet robot intervention for people with dementia: a systematic review and meta-analysis of randomized controlled trials. Psychiatry Res 271:516–525

16. Leyzberg D, Spaulding S, Toneva M, Scassellati B (2012) The physical presence of a robot tutor increases cognitive learning gains. In: Proceedings of the annual meeting of the cognitive science society, vol 34

17. Marti P, Pollini A, Rullo A, Shibata T (2005) Engaging with artificial pets. In: Proceedings of the 2005 annual conference on European association of cognitive ergonomics. University of Athens, pp 99–106

18. Milo (2019). Retrieved November 19, 2019, from https://robots4autism.com/

19. Robins B, Dautenhahn K, Dubowski J (2004) Investigating autistic children's attitudes towards strangers with the theatrical robot—a new experimental paradigm in human–robot interaction studies. In: 13th IEEE international workshop on robot and human interactive communication, 2004. ROMAN 2004. IEEE, pp 557–562

20. Robins B, Dautenhahn K, Dubowski J (2006) Does appearance matter in the interaction of children with autism with a humanoid robot? Interact Stud 7(3):479–512

21. Robins B, Dautenhahn K, Wood L, Zaraki A (2017) 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, pp 147–155

22. Roscoe RD, Chi MT (2007) Understanding tutor learning: knowledge-building and knowledge-telling in peer tutors' explanations and questions. Rev Educ Res 77(4):534–574

23. Saldien J, Goris K, Yilmazyildiz S, Verhelst W, Lefeber D (2008) On the design of the huggable robot probo. J Phys Agents 2(2):3–11

24. Skantze G, Al Moubayed S (2012) Iristk: a statechart-based toolkit for multi-party face-to-face interaction. In: Proceedings of the 14th ACM international conference on Multimodal interaction. ACM, pp 69–76

25. Stiehl WD, Lee JK, Breazeal C, Nalin M, Morandi A, Sanna A (2009) The huggable: a platform for research in robotic companions for pediatric care. In: Proceedings of the 8th international conference on interaction design and children. ACM, pp 317–320

26. Sutton R, Barto A (1998) Reinforcement learning: an introduction. MIT Press, Cambridge

27. Tanaka F, Matsuzoe S (2012) Children teach a care-receiving robot to promote their learning: field experiments in a classroom for vocabulary learning. J Hum Robot Interact 1(1):78–95

28. Tapus A, Peca A, Aly A, Pop C, Jisa L, Pintea S, Rusu AS, David DO (2012) Children with autism social engagement in interaction with nao, an imitative robot: a series of single case experiments. Interact Stud 13(3):315–347

29. van den Berghe R, Verhagen J, Oudgenoeg-Paz O, van der Ven S, Leseman P (2018) Social robots for language learning: a review. Rev Educ Res 89:259–295

30. Viejo G, Khamassi M, Brovelli A, Girard B (2015) Modeling choice and reaction time during arbitrary visuomotor learning through the coordination of adaptive working memory and reinforcement learning. Front Behav Neurosci 9:225

31. Watkins C, Dayan P (1992) Q-learning. Mach Learn 8(3–4):279–292

32. Wood L, Dautenhahn K, Robins B, Zaraki A (2017a) 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. Retrieved November 19, 2019, from https://ieeexplore.ieee.org/document/8172434

33. Wood LJ, Zaraki A, Walters ML, Novanda O, Robins B, Dautenhahn K (2017b) The iterative development of the humanoid robot kaspar: an assistive robot for children with autism. In: International conference on social robotics. Springer, pp 53–63

34. Wood LJ, Zaraki A, Robins B, Dautenhahn K (2019) Developing kaspar: a humanoid robot for children with autism. Int J Soc Robot 1–18. https://doi.org/10.1007/s12369-019-00563-6

35. Zaraki A, Dautenhahn K, Wood L, Novanda O, Robins B (2017) 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, pp 1–4

36. Zaraki A, Wood L, Robins B, Dautenhahn K (2018) Development of a semi-autonomous robotic system to assist children with autism in developing visual perspective taking skills. In: 2018 27th IEEE international symposium on robot and human interactive communication (RO-MAN), IEEE, pp 969–976