When robot personalisation does not help: Insights from a robot-supported learning study

Yuan Gao, Wolmet Barendregt, Mohammad Obaid, Ginevra Castellano

Abstract— **In the domain of robotic tutors, personalised tutoring has started to receive scientists' attention, but is still relatively underexplored. Previous work using reinforcement learning (RL) has addressed personalised tutoring from the perspective of affective policy learning. However, little is known about the effects of robot behaviour personalisation on user's task performance. Moreover, it is also unclear if and when personalisation may be more beneficial than a robot that adapts to its users and the context of the interaction without personalising its behaviour. In this paper we build on previous work on affective policy learning that used RL to learn what robot's supportive behaviours are preferred by users in an educational scenario. We build a RL framework for personalisation that allows a robot to select verbal supportive behaviours to maximise the user's task progress and positive reactions in a learning scenario where a Pepper robot acts as a tutor and helps people to learn how to solve grid-based logic puzzles. A between-subjects design user study showed that participants were more efficient at solving logic puzzles and preferred a robot that exhibits more varied behaviours compared with a robot that personalises its behaviour by converging on a specific one over time. We discuss insights on negative effects of personalisation and report lessons learned together with design implications for personalised robots.**

I. Introduction

Robots are now used to support humans in new social roles, such as providing assistance for the elderly at home, serving as tutors, acting as therapeutic tools for children with autism, or as game companions for entertainment purposes [1]. However, human social skills remain unmatched in robots. To meet the demands of Europe's citizens in the 21st century, our prospective robotic companions need to learn to interact socially with humans [2] and adapt to their needs, preferences, interests, and emotions in order to become highly personalised to their users. Simulating the tremendous social adaptation abilities that characterise human interactions requires the establishment of bidirectional processes in which humans and robots synchronise and adapt to each other in real-time by means of an exchange of verbal and non-verbal behaviours (e.g., facial expressions, gestures, speech) in order to achieve mutual co-adaptation.

In recent years, technical advances in machine learning methods [3] have opened the door to new ways of building co-adaptive human-robot interactive systems. In the domain of robotic tutors [4], which are used to support user learning in instructional scenarios (e.g., in the classroom or in the factory), personalised tutoring has started to receive scientists' attention, but is still relatively underexplored, especially when it comes to building robot abilities enabling robots to interact and adapt to users over extended periods of time.

In the social human-robot interaction (HRI) literature, personalised tutoring has started to be addressed from the perspective of affective policy learning: affect and affect-related states such as engagement have been used to build reward signals in reinforcement learning (RL)-based frameworks to select motivational strategies [5] or supportive behaviours [6] personalised to each student. These works primarily address effects of personalisation on interaction quality and users' positive emotions. Other work has shown the positive effect of modelling learners' skills to deliver personalised lessons [7]. However, little is known about the effects of robot behaviour personalisation on users' task performance. Moreover, it is also unclear if and when personalisation may be more beneficial than a robot that adapts to its users and the context of the interaction without personalising its behaviour.

We build on previous work on affective policy learning that has used RL to learn what supportive behaviours of a robot are preferred by users in an educational scenario [6]. In this work we take a step forward: we develop an RL framework for personalisation that allows a robot to select verbal supportive behaviours to maximize user's task performance and positive reactions to the robot's interventions in a learning scenario where a Pepper [1] robot acting as a tutor helps people learning how to solve grid-based logic puzzles.

We address the following questions: (1) Does a robot's behaviour personalisation improve users' overall task performance? (2) Is a personalised robot more effective than a robot that adapts to its user without personalising?

We conducted a between-subjects experiment with two groups of participants, one interacting with a personalised robot and and one with a non-personalised one. We found that (1) the time taken to complete the logic puzzles significantly decreases over the tutoring sessions for both groups in a similar manner; and (2)

Yuan Gao (alex.yuan.gao@it.uu.se), Mohammad Obaid and Ginevra Castellano are with the Department of Information Technology, Uppsala University, Sweden. Wolmet Barendregt is with the Department of Applied IT, Gothenburg University, Sweden.

[1] https://www.ald.softbankrobotics.com/en/cool-robots/pepper

participants are more efficient at solving logic puzzles and prefer a robot that exhibits more varied robot behaviours compared to a robot that personalises its behaviour by converging on a specific one over time. We discuss insights on negative effects of personalisation and report lessons learnt together with design implications for personalised robots.

This work is relevant for the development of socially interactive embodied agents and social robots used to support users in instructional scenarios. In particular, this paper makes a contribution towards the development of algorithms for human-robot co-adaptation that enable robots and agents to select effective strategies to establish long-term relationships with human users.

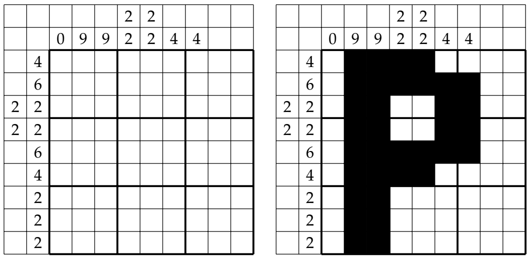
II. Related work

Research on robotic tutors has intensified over the last few years (see, for example, [4] [8]). Several works have investigated the qualities needed in a robotic tutor. Saerbeck et. al. [9], for example, analysed the influence of supportive behaviours in a robotic tutor on learning performance. They developed supportive behaviours in the iCat robot to help students in language learning, and compared the effects of a robot with such behaviours with one that did not provide any social support. They found that the introduction of social supportive behaviours increased students' learning performance. Kennedy et al. [10] likewise investigated the effects of adopting social behaviors. Their work suggests that a robot capable of tutoring strategies may lead to better learning. However, the authors also cautioned that social behaviours in robotic tutors can potentially distract children from the task at hand.

Recently, researchers have started to explore how robots can be used to support personalised learning [11]. Examples include studies exploring the effects of personalised teaching and timing strategies delivered by social robots on learning gains [7], [12] and the use of personalised robotic tutors to promote the development of students' meta-cognitive skills and self-regulated learning [13].

Other works have explored personalised tutoring from the perspective of affective policy learning: affect and affect-related states such as engagement have been used to build reward signals in reinforcement learning (RL)-based frameworks to select motivational strategies [5] or supportive behaviours [6] personalised to each student. RL-based approaches have also been proposed to decide how to employ different social behaviours to achieve interactional goals in task-oriented HRI [14]. Moreover, dynamic probabilistic models and Bayesian networks have been used in robotic tutors to model learners' skills and behaviours and their relationships with a robot's tutoring actions [15] and to assess learners' skills to deliver personalised lessons [7].

However, to our knowledge, previous research has neither explored effects of personalisation that maximises users' positive reactions to the robot and progress in the task, nor has it investigated whether there are any possible negative effects of personalisation.



Unsolved nonogram

Solved nonogram

Fig. 1. An unsolved nonogram puzzle (left) and its corresponding solution (right).

III. Scenario

We developed a scenario where a Pepper robot acting as a robotic tutor helps people solve grid-based logic puzzles called nonograms. These puzzles have previously been used to study robot personalisation to people's learning differences [7]. Nonograms have the advantage of not being well known to most people, thus ensuring that users interacting with the robot start the task-oriented interaction with the robot on a similar skill level.

In the task, the user is asked to solve several nonogram puzzles while a Pepper robot stands in front of the user observing their progress, learning a user model based on the interaction process, and generating verbal utterances in order to provide social support to the user during learning. Robot personalisation to individual users is achieved by combining a decision tree model with a Multi-Armed Bandit (MAB) algorithm called Exponential-Weight Algorithm for Exploration and Exploitation (Exp3) [16] to learn which robot's supportive behaviours (described in Section 3.2) maximise users' task performance and positive reactions to the robot's interventions in the puzzle-solving task.

A. Nonograms

Nonograms are puzzles where cells in a grid must be filled with black or left blank. In these puzzles, the numbers indicate how many black lines are needed to fill continuous lines for each row or column. Figure 1 shows an example of the unsolved Nonogram game and its corresponding solution.

IV. System

The system consists of different components: the nonogram interface, the user model, and the personalisation module, which consists of the Exp3 module and

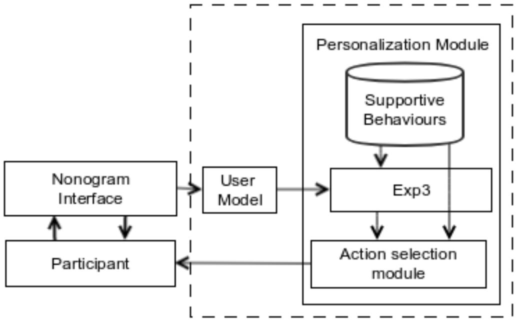


Fig. 2. System components. Arrows indicate the flow of information. The dotted line means that User Model and Personalisation Module are integrated together as one software system.

an action selection module. Figure 2 shows the relationships between the different components. The robot monitors the user's progress in the task through the nonogram interface and builds a user model extracting task indicators that convey information about whether the user is experiencing difficulties during the puzzle solving task. If that is the case, the personalisation module selects a category of supportive behaviour based on a policy learned by the Exp3 algorithm. The selected category will then be passed to an action selection module using a decision tree, which will choose the most relevant robot action for the current situation according to the selected category. In the following sections, we describe the different components of the system.

A. User model

The user model extracts a number of task indicators:

- **TimeLastMove** It measures the time taken to make the last action.
- **TimeLastSetOfMoves** It measures the time taken to make the last N actions, where N is a predefined arbitrary number.
- **CorrectMove** It measures whether the last action made by the user is correct or not.

These are used to define a set of rules that assess whether the user is experiencing difficulties in the puzzle solving task. The rules take the user's actions and their corresponding time into consideration. For example, if an action took the user more than T seconds to complete, where T is an arbitrary number, then it may indicate that the user experienced difficulty in the last decision.

The user model combines all the information it gathers from the task indicators to make a final decision whether the robot should generate an action or not. This decision is then passed to the personalisation module.

B. Personalisation module

The personalisation module uses an RL-based approach that learns which supportive behaviours delivered by the robot maximize the user's task performance and positive reactions to the robot's interventions. The former is defined as the time taken by the user to complete the nonogram puzzle. The latter is measured as positive verbal feedback to verbal supportive behaviours displayed by the robot. This is a problem of policy learning, which in an RL framework means optimising action selection policies to maximise a reward. The general idea here is for the robot to learn a policy of optimal supportive behaviours that maximises a user's task performance and positive reactions to the robot's interventions. Building on previous work on affect coadaptation mechanisms for a social robot [6], we model this problem as a Multi-Armed Bandit (MAB) problem and use an algorithm from the set of MAB learning algorithms – Exp3 [16]. We chose Exp3 based on the length of the interactive process in our study and relative efficiency of the learning algorithm.

1) Supportive behaviours: The robot behaviours in the form of verbal utterances are designed by adopting the categorisation proposed by Cutrona [17]. We select supportive behaviours belonging to four different categories, namely information support, tangible support, esteem support and emotional support. It has been shown that people differ in their preference for support [18]. In Table I, we give an example of each of the different categories of supportive behaviour.

Information support	”Do you need more information about the rules?”
Tangible support	”If you feel it is difficult, I can help you by completing the next one.”
Esteem support	”The game is hard this time.”
Emotional support	”But please don't worry, I am here for you.”

TABLE I
EXAMPLES OF SUPPORTIVE BEHAVIOURS IMPLEMENTED IN THE PEPPER ROBOT .

2) Exp3: We model the optimization of the supportive behaviours as a MAB problem. In our case, generating appropriate behaviours for different participants is the goal of the algorithm. In the following text, we explain how the Exp3 algorithm optimises the probability distribution over all categories of supportive behaviours.

We connect different categories of supportive behaviour with different actions in Exp3. Considering a process with K different actions, the Exp3 algorithm [16] functions as described in Algorithm 1, where γ is the exploration factor, and w_i is the weight of each action i. $p_i(t)$ is the probability of selecting action i at round t, while T means the total number of iterations. At the beginning, the algorithm initialises the exploration parameter γ. This parameter adjusts the possibility that the algorithm attempts to execute other actions while a certain action already has the highest probability. Next, the algorithm associates a weight with each action in order to give each action a probability to form a distribution over all possible actions.

After the exploration, the algorithm iterates the learning procedure T times, in order to learn from the environment and to generate a better probability distribution to receive more accumulative reward from the environment. In the learning procedure, the algorithm selects an action i based on the distribution $\mathcal{P}$, and then receives a reward $x_{i_t}(t)$ from the environment i.e. the reaction from the user. Thereafter, an estimated reward $\hat{x}_{i_t}(t)$ is calculated as $x_{i_t}(t)/p_{i_t}(t)$ to further include the influence of the probability on the reward. In the end, the algorithm updates the weight associated with the action, while the weights of other actions ($w_j, \forall j \neq i_t, j \in \{1, \ldots, K\}$) remain the same. After the algorithm converges, the eventual probability distribution over different actions is considered to be the best (and sometimes final) strategy of maximising the reward.

Algorithm 1 Exp3

```
procedure INITIALIZATION
    initialize $\gamma \in [0,1]$
    initialize $w_i(1) = 1, \forall i \in \{1, \ldots, K\}$
    for distribution $\mathcal{P}$,
        set $p_i(t) = (1-\gamma)\frac{w_i(t)}{\sum_{j=1}^{K} w_j(t)} + \frac{\gamma}{K}, \forall i \in \{1, \ldots, K\}$
end procedure
procedure ITERATION
    repeat
        draw $i_t$ according to $\mathcal{P}$
        observe reward $x_{i_t}(t)$
        define the estimated reward $\hat{x}_{i_t}(t)$ to be $x_{i_t}(t)/p_{i_t}(t)$
        set $w_{i_t}(t+1) = w_{i_t}(t)e^{\gamma\hat{x}_{i_t}(t)/K}$
        set $w_j(t+1) = w_j(t), \forall j \neq i_t \; and \; j \in \{1, \ldots, K\}$
        update $\mathcal{P}$:
            $p_i(t) = (1-\gamma)\frac{w_i(t)}{\sum_{j=1}^{K} w_j(t)} + \frac{\gamma}{K}, \forall i \in \{1, \ldots, K\}$
    until $T$ times
end procedure
```

To integrate Exp3 in our system, each action in this algorithm is associated with a possible category of supportive behaviours, which are described in Table I. In each iteration, the probability of selecting a certain action is adapted to the current environment. For instance, there are four actions ($K = 4$) in the learning procedure of algorithm by design, i.e., action 1, 2, 3, and 4. Respectively, actions 1, 2, 3 and 4 are mapped to four different categories of supportive behaviours: the robot can choose to select information support, tangible assistance, esteem support or emotional support.

That is, if the randomly sampled category of supportive behaviours i is 1, then the robot decides to use information support. After the algorithm receives the feedback, the weight of the corresponding action (i.e., action 1) is updated based on:

$$w_{1_t}(t+1) = w_{1_t}(t)e^{\gamma\hat{x}_{1_t}(t)/4}. \tag{1}$$

The weights of other actions (i.e., action 2 3, and 4) stay the same. In the final step, the distribution $\mathcal{P}$ is renewed to prepare for the next iteration round according to the following formula:

$$p_i(t) = (1-\gamma)\frac{w_i(t)}{\sum_{j=1}^{4} w_j(t)} + \frac{\gamma}{4}, \forall i \in \{1,2,3,4\}. \tag{2}$$

Until then, one learning iteration is done. The iteration continues in total T times. By design, the system's default T value is set to 200 $T = 200$, which means a fixed learning period for 200 iterations.

3) Action selection module: After a category of supportive behaviours is selected by the Exp3 algorithm, the action selection module, which consists of a decision tree, checks if the selected category is appropriate to the current situation. If the latter is not appropriate (for example, when the user has played several games and makes a mistake at the beginning of a new game, it is more likely that the user simply made a mistake and therefore the information support of explaining the rules of the game is not necessary), the robot will not perform any action, otherwise the robot will choose a specific robot action (i.e., supportive behaviour) from a pool of available actions within that category.

V. Methodology

A. Experimental set up

The experimental setting included a Pepper robot, a 27 inches IIYAMA touch screen placed on a table, an ubuntu 14.04 Linux server, a Logitech C920 1080p webcam and a laptop. The user was sitting on a chair in front of the robot, with the touch screen and the table placed between them, see Figure 3.

The Logitech webcam, positioned on a tripod on the side of the table, was connected to the laptop to record videos for offline video analysis. The software ran on a Linux server and consisted of two parts. One part contained the nonogram interface that interacted with the user and an algorithmic module that updated the parameters for the user model. The other part was responsible for controlling the robot and generation of verbal utterances.

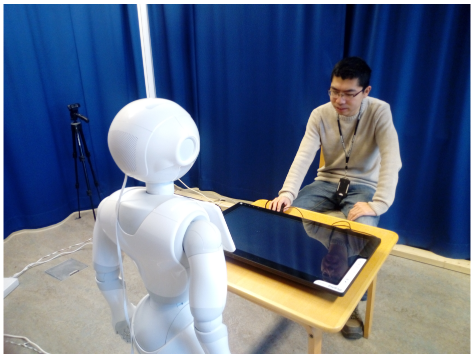


Fig. 3. Pepper interacting with a user.

B. Experimental Design and Participants

We performed an experiment with a between-subjects design, where participants were randomly assigned to two different groups (i.e., conditions), corresponding to two different parameterisations of the robot's behaviour: (1) *personalised adaptive* group, where the MAB-based personalisation module was used, vs *random adaptive* group, where the robot's supportive behaviours were randomly selected.

For the study, we recruited twenty-four (fourteen females and ten males) university students and researchers, primarily with a computer science background, took part in the experiment. We then randomly assigned twelve participants to each group. In the random adaptive group, four are males and eight are females. In the personalised random adaptive group, six are females and six are males.

C. Procedure

The experiment took place in a university laboratory environment. Before the experiment started, the participants received a document that described the experiment and tasks that they would need to solve and a consent form that they needed to sign. After signing the consent form, the participant was asked to enter the experiment's area.

During the experiment, each group performed three sessions, namely a pre-interaction session, a human-robot interaction session and a post-interaction session. In both the pre-interaction and post-interaction sessions, participants were asked to solve three nonogram puzzles of similar difficulty on their own. In the human-robot interaction session, the participants were asked to solve three nonogram puzzles with the assistance of a robot. We chose to include three nonograms in the human-robot interaction session because during preliminary tests of the system we found that normally three nonograms are necessary for the robot's learning process to converge while making sure the participant is not frustrated by the long interaction. In the human-robot interaction session, participants were informed that, after each supportive behaviour used by the robot, they had the possibility to give verbal feedback to the robot, in case they appreciated, or not, what the robot said. More specifically, participants were instructed to either ignore the robot (when they did not like what the robot said) or reply by saying "thank you" to the robot (when they liked what the robot said). This information is used, alongside with task performance, to calculate the reward function in the RL-based personalisation module.

After the post-interaction session the participant was asked to walk out of the experimental area and the researcher conducted a very short interview and asked the participants to fill in a set of questionnaires to collect information on the user experience and perception of the robot.

Normally, due to learning effects, the average time taken to solve nonogram puzzles in the post-interaction session is shorter than the average time taken to solve nonogram puzzles in the pre-interaction session. Here, we are interested in whether personalisation affects the time taken by people to solve the puzzle, i.e., if in the personalised adaptive group, the difference in the time taken by people to solve the puzzle between pre- and post- interaction session is more obvious.

D. Measures

Task performance. To measure task performance, we calculated the average amount of time taken by participants to complete the three puzzles in the pre-interaction session and the three puzzles in the post-interaction session. We computed the difference between these average values for both groups of participants for further analysis and comparison, as detailed in the next section.

Personalisation algorithm's output. We extracted the output of the personalisation algorithm to assess the probability with which each robot's supportive behaviour was selected when the personalisation process was completed.

Quality of interaction. We measured quality of interaction using a set of affective, friendship, and social presence dimensions that have been previously shown to be successful in measuring the influence of a robot's behaviour on the relationship between the robot itself and participants [19]. Participants were asked to rate these dimensions throughout the interaction using a 5-point Likert scale, where 1 meant "totally disagree" and 5 meant "totally agree". For each dimension presented below, we considered the average ratings of all the questionnaire items associated to it.

Social engagement: this metric has been extensively used to measure both human-human [20] and humans-robot [21] quality of interaction. The engagement questionnaire we used was based on the questions formulated by Sidner et al. [21].

Help: measures how the robot is perceived to have provided guidance and other forms of aid to the user. In particular, we refer to help as a friendship dimension that measures the degree to which a friend fullfils a support function existing in most friendship definitions, as suggested by Mendelson and Aboud in the McGill Friendship Questionnaire [22].

Self-validation: another friendship dimension taken from the McGill Friendship Questionnaire [22]. It has been used to measure the degree to which children perceived the robot as reassuring and encouraging, and helped the children to maintain a positive self-image.

Perceived affective interdependence: this dimension measures social presence, that is, "the degree to which a user feels access to the intelligence, intentions, and sensory impressions of another" [23]. Here, perceived affective interdependence measured the extent to which the user's emotional states were perceived to affect and be affected by the robot's behaviours.

User preferences. We collected participants' preferences for the robot's supportive behaviours using questionnaires. Participants were asked to indicate how much they liked each category of robot's behaviour on a scale from 1 to 5.

VI. Results

A. Task performance

In order to analyze the effect in the experiments, a mixed ANOVA was used to compare the difference between the time taken to complete the puzzles in the pre-interaction session and the time taken to complete them in the post-interaction session. Figure 4 shows the mean and standard deviation of average time difference between pre-interaction and post-interaction sessions. There is an overall time improvement for the time difference between pre-interaction session and post-interaction session ($F(1, 20) = 42, 524$, $p < 0.05$, partial $\eta^2 = 0.680$). Although in the random adaptive group the time difference ($M = 153.85$, $SD = 97.15$) is larger than in the personalised adaptive group ($M = 128.55, SD = 89.30$), the result was not significant ($F(1, 20) = 0.07$, $p > 0.05$, partial $\eta 2 = 0.004$).

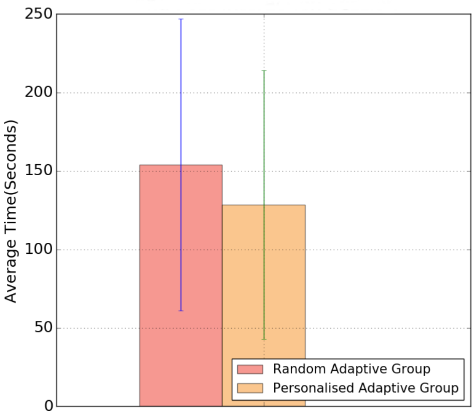


Fig. 4. The mean and standard deviation of average time difference between pre-interaction session and post-interaction session.

B. Learning algorithm's performance

In order to show the optimization process during the human-robot interaction session, we would like to present an example of the output of the algorithm. Figure 5 illustrates that the algorithm tries to personalise towards a specific participant in the session. In this case, the esteem support and emotional support are the most preferred categories. The most preferred categories are different for each participant.

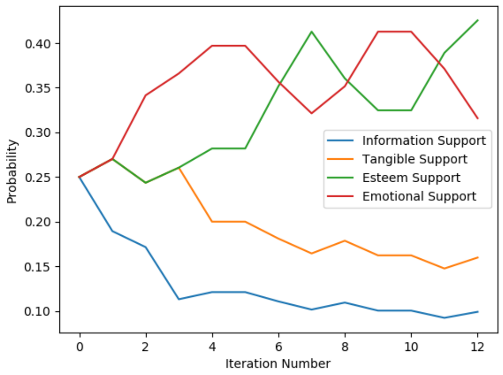


Fig. 5. An optimization process over a human-robot interaction session. For this participant, esteem support and emotional support are the most preferred categories.

C. Participants' preferences and quality of interaction

To compare the output of the personalisation algorithm with the participants' preferences of the robot's behaviour we define a similarity metric.

We converted the data on the participants' preferences from the questionnaires to a probability distribution using the following formula:

$$p_i = \frac{x_i - 1}{\sum_{i=1}^{4}(x_i - 1)} \tag{3}$$

Where x_i is the score of the preferred category i, given by a participant.

In order to compare the participants' preferences $\mathcal{P}$ with the output of the algorithm $\mathcal{Q}$, we use a metric called Hellinger distance [24] to quantitatively analyze how much do these two distributions differ from each other. The Hellinger distance is defined as follows:

$$H(P, Q) = \frac{1}{\sqrt{2}} \sqrt{\frac{1}{\sum_{i=1}^{n}(\sqrt{p_i} - \sqrt{q_i})^2}} \tag{4}$$

Where p_i and q_i are discrete numbers of two different distributions defined as $\mathcal{Q}$ and $\mathcal{P}$.

Figure 6 shows results of the calculated Hellinger distances for the two groups. We observe that the average Hellinger distance is smaller for the random adaptive group. We conducted a Mann-Whitney U test and found that the difference between the two groups is significant ($u(22) = -2.03, p < 0.05$).

We assess the quality of interaction with the robot using a set of social dimensions, as discussed in Section V. For each dimension we considered the average ratings of all the questionnaire items associated to it. A t-test was conducted to compare the two groups of participants. We did not find any significant differences between the two groups in terms of engagement ($t(94) = 1.36, p > 0.05$), help ($t(70) = 1.18, p > 0.05$), self-validation ($t(46) = 1.05, p > 0.05$) and perceived affective interdependence ($t(46) = 0.93, p > 0.05$)

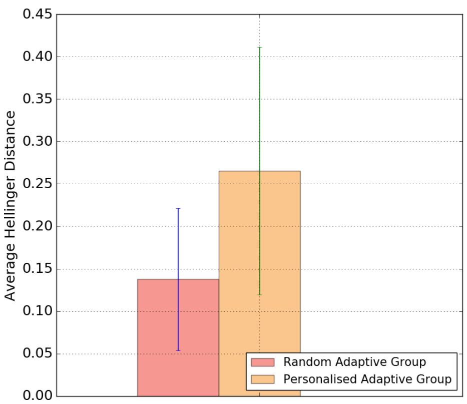


Fig. 6. Hellinger distances for the two study groups.

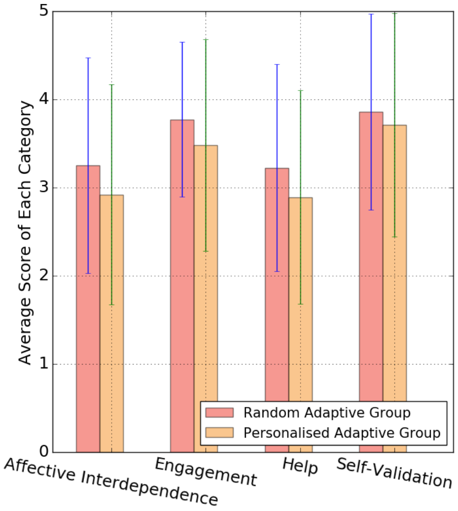


Fig. 7. This figure shows the average scores of the four dimensions of quality of interaction considered in our study.

Figure 7. However, when we analysed the individual questions for each dimension, we observed one significant result in one question in the engagement questionnaire: we found that participants felt more engaged with the robot in the random adaptive group ($t(22) = 2.07, p < 0.05$).

VII. Discussion

(1) *Does a robot's behaviour personalisation improve users' overall task performance?*

Our results showed that the time taken to complete the logic puzzles significantly decreased in both groups over the tutoring sessions. When we compare the difference between the average time taken to complete the three puzzles in the pre-interaction session and the average time taken to complete the three puzzles in the post-interaction session, we can observe that this difference is slightly higher for the participants in the random adaptive group. Although this result is not significant, it seems to suggest that participants are more efficient at solving logic puzzles when the robot does not attempt to personalise its supportive behaviour. This result is in line with previous work that showed that a robot that is too social and adaptive may not lead to increased learning gain in a tutoring scenario [10]. It must also be pointed out that people may not always need to hear feedback they like in order to learn better. Previous work in educational psychology, for example, discusses how giving praise may undermine learning motivation [25]. A teacher (or robot) that praises can be perceived very positively by children, but it does not necessarily mean it is good for them. Similarly, personalising to other kinds of supportive behaviours may have similar effects.

(2) *Is a personalised robot more effective than a robot that adapts to its user without personalising?*

The analysis on the similarity metric based on the Hellinger distance showed that participants' preferences for the robot's behaviours (i.e., collected via the Likert scale questions in the questionnaires) are more aligned with how the robot selects its supportive behaviour when it does not use personalisation. In other words, people might prefer to interact with a robot that exhibits more varied behaviours (i.e., a robot that selects different types of supportive behaviours with equal probability) compared with a robot that converges on selecting more often the behaviour that, for each participant, maximises the participant's positive reactions to it and task performance. Moreover, as far as measures of quality of interaction are concerned, there was no difference in how participants perceived the robot in terms of social engagement, help, self-validation and affective interdependence, except for one question concerning social engagement, which suggests that participants were more engaged with the robot that did not personalise its behaviour.

There are some considerations to make when interpreting these results. By selecting the supportive behaviours in a more balanced manner, the non-personalised robot may be perceived as less predictable than the personalised robot, thus possibly appearing as more interesting to interact with. We observed that the learning algorithm was successful at converging to a specific behaviour for each participant (meaning that such a behaviour is eventually selected more often than others), which seems to suggest that a longer learning session may not be necessary. There could be, however, a need for a more balanced trade-off between exploration and exploitation in the learning process. Moreover, participants may not express their real preferences by providing verbal feedback to the robot, which suggests that this aspect requires further investigation, both from an interaction design and an algorithmic perspective. However, it could also simply be that personalising a robot's behaviours to maximise a user's positive reactions to them does not help, as also suggested by studies in other domains [25] and discussed above. Therefore, a stronger focus on maximising task performance might be needed when designing a personalisation framework. Although not the focus of this paper, an analysis of the interview material that we collected might offer more insights towards answering these questions.

VIII. Conclusion

This work investigated the effects of a robot's behaviour personalisation on user's task performance and perception of the robot in a robot-supported learning scenario. We found that people are more efficient at solving logic puzzles and prefer a robot that exhibits more varied behaviours compared with a robot that personalises its behaviour by converging on a specific one over time. Our interpretation is that (1) the robot does learn which behaviours maximise users' positive reactions throughout the interaction, but (2) this does not necessarily mean that, after having experienced the whole interaction, users prefer a robot that personalises in this manner. Following in the steps of Kennedy and colleagues [10], we conclude that caution is needed when designing social and adaptive robot behaviours to support people's learning. While other studies showed a positive effect of robot personalisation on task performance and quality of interaction, this work confirms that little is known about the effects of a robot's social behaviour personalisation on user task performance. Therefore, we suggest that more work needs to be conducted to assess the real implications of personalisation in complex scenarios with both a social and a task component, such as robot-supported learning.

Acknowledgement

This work is partly supported by the Swedish Research Council (grant n. 2015-04378) and partly by the COIN project (RIT15-0133) funded by the Swedish Foundation for Strategic Research.

References

[1] C. L. Breazeal, *Designing sociable robots*. MIT press, 2004.

[2] K. Dautenhahn, "Socially intelligent robots: dimensions of human–robot interaction," *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 362, no. 1480, pp. 679–704, 2007.

[3] S. Levine, C. Finn, T. Darrell, and P. Abbeel, "End-to-end training of deep visuomotor policies," *Journal of Machine Learning Research*, vol. 17, no. 39, pp. 1–40, 2016.

[4] G. Castellano, A. Paiva, A. Kappas, R. Aylett, H. Hastie, W. Barendregt, F. Nabais, and S. Bull, "Towards empathic virtual and robotic tutors," in *International Conference on Artificial Intelligence in Education*. Springer, 2013, pp. 733–736.

[5] G. Gordon, S. Spaulding, J. K. Westlund, J. J. Lee, L. Plummer, M. Martinez, M. Das, and C. Breazeal, "Affective personalization of a social robot tutor for children's second language skills." in *AAAI*, 2016, pp. 3951–3957.

[6] I. Leite, G. Castellano, A. Pereira, C. Martinho, and A. Paiva, "Empathic robots for long-term interaction," *International Journal of Social Robotics*, vol. 6, no. 3, pp. 329–341, 2014.

[7] D. Leyzberg, S. Spaulding, and B. Scassellati, "Personalizing robot tutors to individuals' learning differences," in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. ACM, 2014, pp. 423–430.

[8] J. Kennedy, P. Baxter, E. Senft, and T. Belpaeme, "Social robot tutoring for child second language learning," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, ser. HRI '16. Piscataway, NJ, USA: IEEE Press, 2016, pp. 231–238. [Online]. Available: http://dl.acm.org/citation.cfm?id=2906831.2906873

[9] M. Saerbeck, T. Schut, C. Bartneck, and M. Janse, "Expressive robots in education - varying the degree of social supportive behavior of a robotic tutor," in *28th ACM Conference on Human Factors in Computing Systems (CHI2010)*. Atlanta: ACM, 2010, pp. 1613–1622.

[10] J. Kennedy, P. Baxter, and T. Belpaeme, "The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning," in *Proc. of the Tenth Annual ACM/IEEE Int. Conference on Human-Robot Interaction*. ACM, 2015, pp. 67–74.

[11] C. Clabaugh, G. Ragusa, F. Sha, and M. Matarić, "Designing a socially assistive robot for personalized number concepts learning in preschool children," in *The 2015 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*. IEEE, 2015, pp. 314–319.

[12] A. Ramachandran, C.-M. Huang, and B. Scassellati, "Give me a break!: Personalized timing strategies to promote learning in robot-child tutoring," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17. New York, NY, USA: ACM, 2017, pp. 146–155. [Online]. Available: http://doi.acm.org/10.1145/2909824.3020209

[13] A. Jones, S. Bull, and G. Castellano, "Personalising robot tutors' self-regulated learning scaffolding with an open learner model," in *Proceedings of WONDER (International Workshop on Educational Robots) workshop, International Conference on Social Robotics 2015 (ICSR15), Paris, France, October 2015*, 2015.

[14] J. Hemminghaus and S. Kopp, "Towards adaptive social behavior generation for assistive robots using reinforcement learning," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2017, pp. 332–340.

[15] T. Schodde, K. Bergmann, and S. Kopp, "Adaptive robot language tutoring based on bayesian knowledge tracing and predictive decision-making," *Proceedings of ACM/IEEE HRI 2017*, 2017.

[16] S. Bubeck, N. Cesa-Bianchi, *et al.*, "Regret analysis of stochastic and nonstochastic multi-armed bandit problems," *Foundations and Trends® in Machine Learning*, vol. 5, no. 1, pp. 1–122, 2012.

[17] C. E. Cutrona, J. A. Suhr, and R. MacFarlane, "Interpersonal transactions and the psychological sense of support," *Personal relationships and social support*, pp. 30–45, 1990.

[18] G. M. Reevy and C. Maslach, "Use of social support: Gender and personality differences," *Sex roles*, vol. 44, no. 7, pp. 437–459, 2001.

[19] I. Leite, "Long-term interactions with empathic social robots," Ph.D. dissertation, Universidade Técnica de Lisboa, Instituto Superior Técnico, 2013.

[20] I. Poggi, *Mind, hands, face and body. A goal and belief view of multimodal communication*. Weidler, Berlin, 2007.

[21] C. L. Sidner, C. D. Kidd, C. H. Lee, and N. B. Lesh, "Where to look: A study of human-robot engagement," in *IUI '04: Proceedings of the 9th international conference on Intelligent user interfaces*. Funchal, Madeira, Portugal: ACM, New York, NY, USA, 2004, pp. 78–84.

[22] M. J. Mendelson and F. E. Aboud, "Measuring friendship quality in late adolescents and young adults: Mcgill friendship questionnaires," *Canadian Journal of Behavioural Science*, vol. 31, no. 1, pp. 130–132, April 1999.

[23] F. Biocca, "The cyborgs dilemma: Embodiment in virtual environments," in *Cognitive Technology, 1997. 'Humanizing the Information Age'. Proceedings., Second International Conference on*. IEEE, 1997, pp. 12–26.

[24] E. Hellinger, "Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen." *Journal für die reine und angewandte Mathematik*, vol. 136, pp. 210–271, 1909.

[25] J. Henderlong and M. R. Lepper, "The effects of praise on children's intrinsic motivation: A review and synthesis," *Psychological Bulletin*, vol. 128, no. 5, pp. 774–795, 2002.