Proceedings of the 27th IEEE International Symposium
on Robot and Human Interactive Communication,
Nanjing, China, August 27-31, 2018

A reinforcement learning model for robots as teachers*

Sayanti Roy[1], Emily Kieson[2], Charles Abramson[2], and Christopher Crick[1]

***Abstract*— Robots are capable of training humans to achieve complex tasks, and their helpful feedback can lead to useful human-robot collaborations. In this research we present a reinforcement learning model influenced by human cognition which is repurposed to enhance human learning, investigate a robot's ability to encourage and motivate humans and improve their performance. During teaching the robot trades off between exploration and exploitation to understand the human perception and develop a successful motivational approach. We compare our learned reinforcement model with a baseline non-reinforcement approach and with a random reinforcer, and achieve more effective teaching in the learned reinforcement condition. In addition, we discovered an extremely strong relationship ($r = 0.88$) between the robot's regret, in a machine learning sense, and the performance of its human partner.**

I. Introduction

Teaching and learning involve complex cognitive models which are gradually constructed over a series of intrinsic and extrinsic social interaction episodes. This process is most effective if the facilitator can teach each individual while possessing some understanding of the subtleties of the student's motivations, regulating strategies accordingly. In this research, the humanoid robot Baxter motivates an individual extrinsically during the learning process using several positive reinforcers. During the interactions, the robot initially demonstrates several tasks to the participant, in ascending order of difficulty, involving the assembly of augmented-reality-tagged blocks into various patterns [1].

If a candidate is successful in every task, external motivation has negligible effect in changing anything, because people feel comfortable and they are in a familiar situation. Thus, people who are happier will sometimes be less motivated to push themselves toward action compared to someone in a negative mood, who will be more motivated to exert effort to change their unpleasant state. Hence a negatively-valenced mood can increase, and positively-valenced mood can reduce, perseverance with difficult tasks. This may be because people are less motivated to exert effort when they are already satisfied with their performance. Frustration, in turn, may increase perseverance as people see greater potential benefits of making an effort. Thus if people in a negative mood get some positive reinforcer to overcome their challenge, their learning rate is expected to increase [2].

*This work was supported by NSF award #1527828 (NRI: Collaborative Goal and Policy Learning from Human Operators of Construction Co-Robots).

[1]Robotic Cognition Laboratory, Computer Science Department, Oklahoma State University, Stillwater OK 74078 sayanti.roy@okstate.edu, chriscrick@cs.okstate.edu

[2]Department of Psychology, Oklahoma State University, Stillwater OK 74078 kieson@okstate.edu, charles.abramson@okstate.edu

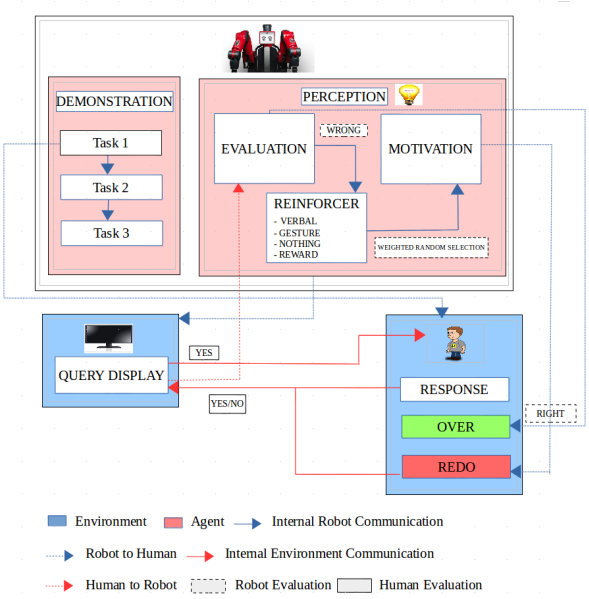

Fig. 1. Block diagram showing the human-robot interaction

In this research, the humanoid robot Baxter uses a reinforcement learning strategy to understand the effect of its reinforcement presentation on its human subjects, attempting to increase their performance over time. Here the subject pool is divided into sets of participants who receive no reinforcements, random reinforcements, or learned reinforcements respectively during their task performance. We compared the number of people commiting more than three mistakes in each group, because we expect our reinforcement strategies to be more effective for subjects who are performing somewhat poorly. We also look at the overall number of mistakes committed by each subject group. We discovered that participants in the learned group were more likely to perform well and committed comparatively fewer mistakes with respect to the other experimental conditions. We also learned that the robot's regret strongly correlates with the probability that a test subject makes more versus fewer mistakes.

II. Related Work

Many contemporary researchers are working in robotic feedback and implementing reinforcement learning strategies in a real-time environment. We discuss a few of them in this section.

A. Strategy determination

Thomaz [3], [4], [5] discussed novel policy shaping algorithms and how motivations and reward signals can be used as a channel to impact human-robot partnership in an HRI setting, simultaneously improving the future learning process of both humans and robots. Cakmak [6] demonstrated how social learning strategies vary with the particular environment when robots are allowed to explore and learn from their surroundings. Lopes [7] upgraded Multi-Arm Bandit techniques using different motivational resources in limited time to maximize skills and learing activities. The modification of traditional reinforcement learning algorithms using reward shaping produced important insights into how skill and accuracy can be improved for a particular task. In our work, we use exploration/exploitation strategies [8] to shape the reward, where the agent's feedback depends the performance of the participants.

B. Tutoring individuals

Scassellati [9], [10], [11], [12] and Park [13] have presented feedback-based human-robot interaction, demonstrating that if humans are guided by the robot at an interpersonal level, it increases the robot's perceived social reliability, making humans more eager to interact with them. This not only increases the performance accuracy of the human learner, but also connects with them emotionally [14] and provides social assistance throughout their learning process. A robot learning from human feedback tends develop a mental model [15], [16] of its own which positively influences human cognition.

C. Positive reinforcement and robots

Positive reinforcement plays a vital role in human learning. A few researchers have explored various areas where positive reinforcement from robots had a large impact on children. Boccanfuso [17] investigated the difference in responses between children with or without autism with an emotion-stimulating robot using positive reinforcement in an interactive environment. Nunez [18] described the use of positive reinforcers to overcome the underlying challenges in motivating a child to continue learning and to share the experience with others. Kim [9] addressed the unique positive effects and advantages a robot can have on autistic children, exploring areas where robots play an important role in the lives of specific individuals. Hence we wish to investigate how a robot can develop an understanding of the underlying motivations and cognitive traits of individual people, so that it can shape its teaching strategies appropriately and enhance the learning process.

III. Technical Description

The human mind is a cognitive structure that consists of memory, decision making, perception, thoughts, emotions, and so on. These features act differently when they are influenced by external factors like stress or negativity on a regular basis. Thus, understanding the needs of a human mind under particular circumstances can be difficult [2], especially when those needs are dynamic or specific to a certain individual. In this research, Baxter attempts to identify the reward orientation of the particular human with which it is interacting, depending on the task performance. In this section, we first describe the robot's evaluation procedure used to assess the learning performance of its partner. We then describe how the robot tries to identify the best course of action to improve its own teaching performance.

A. Motivation and reinforcers

The exploration/exploitation dilemma [8] is a common problem, where decisionmakers can either jump to a conclusion and make a decision on the basis of the partial knowledge they currently possess, or rather wait and invest more time and effort in accumulating further information, with the hope that a broader perspective will lead to a better decision in future. In our research, Baxter attempts to probe and understand a specific aspect of a human mind's cognitive orientation toward particular reinforcement strategies, on the basis of this exploration and exploitation trade-off, where human performance acts as the reward.

The robot employs random selection among potential reinforcement behaviors, weighted by its current mental model of its human partner's motivation. When Baxter is trying to motivate an individual, it provides a positive reinforcer. Four different kinds of positive reinforcement are used in this process: verbal, reward, gesture and none [1].

1) *Verbal reinforcer*: When using this reinforcer, the robot asserts that it is trying to encourage the subject with some positive feedback. For example, in our experiments, if the subject makes a mistake, the robot will verbalize something like, "Sorry dear, don't worry. You can do it!"
2) *Reward-based reinforcer*: This takes the form of a hint given to the participant during a task. The hint does not provide the correct answer but tries to influence the subject's thought process so that it increases the learning rate of the participant. For example, during the pattern making process, if a candidate places an incorrect marker, Baxter suggests flipping the marker box and trying the other side, before rejecting the block entirely. Thus people can track the blocks they have already tried to place in a particular spot. The reward-based reinforcers are different for each of the three experimental tasks.
3) *Natural reinforcer*: In this case, the robot only identifies the correct or the incorrect marker. It doesn't attempt to induce any kind of positivity or motivation in the participant.
4) *Gesture-based reinforcer*: In this case, the robot adds a consoling gesture to the positive verbal feedback referenced above.

B. Behavior selection

The robot initially assigns a uniform prior across its potential reinforcement behaviors, and they begin with equal weights. When a subject is given a particular reinforcement,

the robot reweights its reinforcement strategy immediately following each subtask by evaluating her performance.

$$S_t = \{\nu\phi s^+_{t-1}, \frac{\nu}{|S|-1}(1-\phi)s_{t-1}\forall s \in S_{s\neq s^+}\} \quad (1)$$

where S_t is the weight distribution over each reinforcement strategy at time t, ν (here equal to 0.03) is a learning rate parameter, ϕ is 0 if the participant is unable to complete its task after getting positive reinforcers, otherwise 1, and s^+_{t-1} is the particular reinforcement strategy selected at time $t-1$.

Hence after several interactions, the robot can identify those reinforcemenrs that are inducing the candidate to perform well; in other words, after receiving particular reinforcers, the candidate does not underperform. Fig. 1 shows the overall structure of the experiment used to answer this research question.

The robot demonstrates tasks in ascending order in the level of their difficulty. The task level increases with the number of markers or the complexity in the pattern. In the first task, the robot teaches the individuals about the markers and their corresponding labels. In the second task, the robot makes a pattern with seven markers, but they are not necessarily the same markers as in task 1; sometimes new markers are also introduced into the pattern. In task 3, the robot will make a complex pattern with 12 markers. If the robot manages to motivate the subject well, then the candidate will perform the difficult tasks with ease. To evaluate how well the robot can assess its own performance as a good teacher, we computed regret and correlated that with the total number of mistakes made by each student. If the regret increases linearly with mistakes made, Baxter's own self-assessment is reliable, and the reinforcement strategies selected are appropriate for the teaching task. Regret is defined as the difference between the reinforcer with maximum weight and the reinforcement strategy selected.

$$R = s_{max} - s^+ \quad (2)$$

Maximum weight is assigned to that reinforcer which the robot thinks is the most appropriate strategy for the human learner with which it is currently interacting. A regret of 0 means that the robot is exploiting this knowledge to increase the learning rate, while if it selects a different reinforcement strategy it is exploring to discover the orientation of the participant [1].

IV. Experiments

A. Positive reinforcer on success

Intially an experiment was carried out where Baxter uses positive reinforcers as appreciation if the person performs well in a task. $n = 19$ participants were invited to the laboratory where they were allowed to interact with Baxter. However, the different groups of people (none, random and learned) did not show any significant performance difference, judging from the number of mistakes they made in each condition. The median number of mistakes is the same across all conditions, although their overall range of mistakes varied somewhat. The subjects indicated in conversation with the researchers that when each participant is performing so well in all the tasks, they barely cared about the reinforcer from the robot because they are performing well anyway. Hence we redesigned the experiment in such a manner that the reinforcers can influence the subjects during their performance and the robot can help them better to accomplish the task.

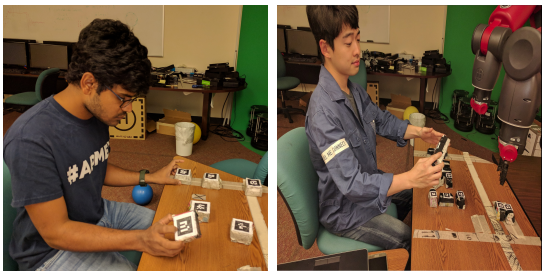

Fig. 2. Participants trying to recreate the patterns in tasks 2 and 3.

B. Positive reinforcer on failure

In this experiment, the positive reinforcements are provided by the robot if the candidate has unsatisfactory performance at any point. Section III-B explains how the reinforcement strategies are adjusted according to the human orientation uncovered by the robot's exploration and exploitation of effective strategies. Here, the candidates are divided into three categories, where they receive no reinforcement at all, a random reinforcement, or a reinforcement selected according to Baxter's understanding of what motivates the particular individual. In the no-reinforcement group, Baxter only demonstrates the task and declares "Right" or "Wrong" depending upon the performance of the participant. In the case of the random and learned model categories, the robot gives out positive reinforcers in the form of a reward, gesture, verbalization or just simply saying "You are right". Baxter also changes its facial expression [19] on the basis of the candidate's performance. Generally, Baxter puts up a neutral face while demonstrating the task, but if a candidate performs correctly, Baxter's face turns green with a smile, while it makes a sad face and turns red when wrong. The facial expressions are also applied as a form of reinforcer. The goal of this experiment was to determine if, when the robot has lower regret, whether the learner makes correspondingly fewer mistakes with time. Fig 2 shows participants performing task 2 and 3.

C. Procedure

$n = 110$ participants were recruited for the experiment (aged 18-20, 28 male, 82 female). The no-reinforcer group contains $n = 35$, the random reinforcer group contains $n = 22$ and the learned model group contains $n = 53$ participants. The blocks that are used in this experiment are two-faced having different markers on each side. During task 1, Baxter initially demonstrates seven markers, explaining their numbers and markings to the candidate. They are then

randomly asked to identify two markers out of seven. A participant proceeds to the next task only if the first is finished successfully. Accordingly, the markers the candidates study in task 1 may not be repeated in task 2, as they are all shuffled before each task. The robot is only responsible for placing the markers in their respective positions. Since markers are shuffled randomly, each participant is given different patterns.

In task 2, subjects are asked to solve some general math questions as a distraction while the robot constructs the pattern, in order to reduce the available observation time for the participant. When the robot is finished making the pattern, they are asked to turn back and observe the pattern for 30 seconds. The blocks are shuffled again and the students are asked to recreate the pattern in 50 seconds. After each task, the robot inquires if they require more time.

During task 3, the subjects are allowed to observe the pattern making process, but are not given any additional observation time. Baxter's pick-and-place manipulation is fairly slow, and it takes almost a minute for it to create the larger pattern.

Out of the three subject groups, substantially more participants in the learned reinforcement strategy group were able to advance to task 3.

D. Results

The results of the experiment is categorized into the following subsections.

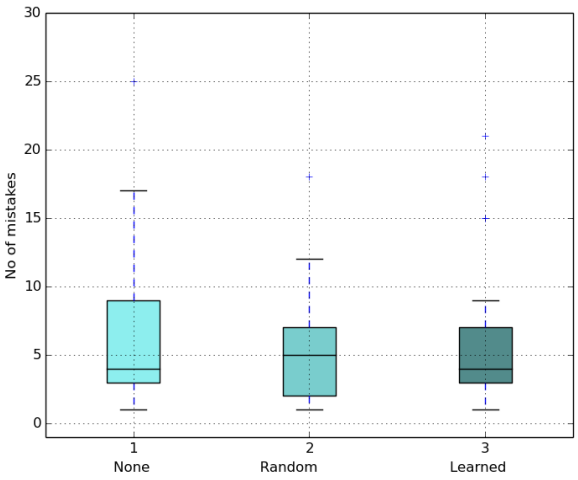

Fig. 3. Performance of participants in experiment II

1) Subjective Performance Evaluation: Fig. 3 shows the number of mistakes made by each participant, used as a metric to evaluate performance. Out of all the participants who performed the experiment, there are some ($n = 6$ for None, $n = 5$ for random and $n = 10$ for learned) within each experimental group who did not make any mistakes. Those participants did not receive any reinforcements regardless of which group they were assigned to, so are not considered as a part of the mistake data. Also there are cases where the participants responded to more than one reinforcer or made so many mistakes that the robot could not determine their reward orientation ($n = 20$). From within the learned model group, out of 53 participants 18.87 percent of people did not commit any mistakes, and the orientation of 37.73 percent of the participants could not be determined by the robot under the experimental conditions. This means that the robot successfully learned a good teaching strategy for slightly fewer than half of the participants. Hence Fig. 3 shows $n = 29$ for no reinforcers, $n = 17$ for random reinforcers and $n = 23$ for learned reinforcers.

Fig. 3 considers only those participants of the learned group whose orientation can be understood by the robot, and excludes the participants who performed so well as to receive no reinforcement at all. We can see that there is a suggestive difference between the different group of participants. Although the median performance is almost same in the none vs. the learned group, the range of the mistakes differ. From the data, we can see that more than a quarter of the participants in the no-reinforcement group made more mistakes than almost anyone in the learned group. Besides that, in spite of having a larger population, the overall range of mistakes of the learned group is smaller than any other group of participants. The group receiving no reinforcement has the largest magnitude of mistakes. To measure the standardized difference we calculated Cohen's $D = 1.93$ on these two populations, showing that there is only approximately 32% similarilty between them. The reinforcement strategy is considered to be working for a participant when the participant starts making fewer mistakes with same kind of reinforcer, and this also leads to a lower computed regret for the robot.

In the case of the random group, the range of mistakes is smaller than first group because some kind of positive motivation is given out, even if it isn't the most appropriate for the individual. Hence the number of mistakes are also smaller than the first group. From Fig. 3 we can tentatively conclude that the people in the learned model condition performed better than the people in the other groups.

Fig. 4 shows the fraction of people in all the three populations who made more than three mistakes. Since subjects who performed close to perfectly received little feedback regardless of their experimental category, we would not expect to see much of an effect among those subjects. In this figure, we restrict our attention to subjects who received significant feedback. In this case, a z-test performed between the no-reinforcement and the learned group shows a p-value of 0.03. Thus subjects who made mistakes in the learned group received helpful feedback and improved their performance significantly more than the others.

2) Reinforcer evaluation : To measure the effectiveness of reinforcers, we calculated the interactions elapse before Baxter realizes which reinforcer is working for a participant. In a few cases it discovered the best reinforcer in the first interaction, but this is not usually the case. Baxter initially starts with a random interaction, as they all have equal prior

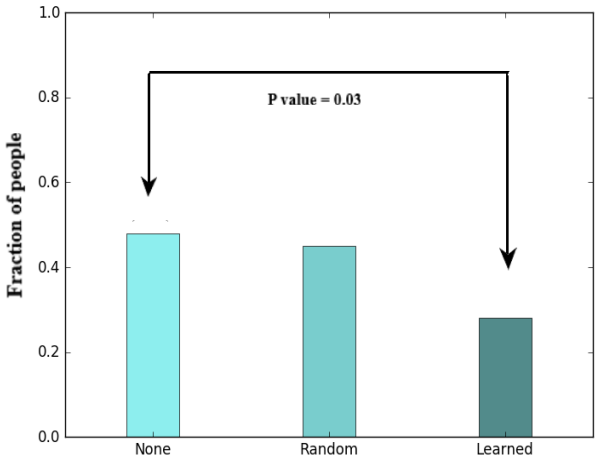

Fig. 4. Fraction of participants making more than three mistakes.

weights. But fairly quickly, one interaction tends to stand out among the others as the strategy which works best for the participant. We considered cases where Baxter realized in the very first interaction which reinforcer will be working for the participant. Here we also considered the case where subjects received only one reinforcement strategy, which induced them to perform very well through out the rest of the task without making any further mistakes. There are also cases where Baxter identified multiple reinforces which work equally well for a participant after several interactions.

The mean and the standard deviation for the various reinforcers are as follows: Gesture: $\mu = 3.0$, $\sigma = 0.76$, None: $\mu = 2.4$, $\sigma = 1.5$, Reward: $\mu = 3.67$, $\sigma = 3.01$, and Verbal: $\mu = 2.75$, $\sigma = 3.19$. We see that the least effective reinforcement strategy, or at least the one that took the longest to learn for the largest number of participants, was gesture-based. The experiment was performed on college-age subjects. Gestures are usually popular among young children; here it is assumed that the subjects lack emotional engagement, so gestures had less effect than other reinforcement strategies. In case of verbal reinforcement, more interactions were required by the robot to understand the orientation of human participants. This is because Baxter narrates the reinforcer in a machine voice, which is sometimes difficult to comprehend. Subjects encountered some difficulty in understanding and obtaining motivation from verbal interaction. Participants had similar responses to reward and verbal motivations.

3) Regret analysis: As mentioned in Section III-B, regret is calibrated on the basis of the decision making ability of the robot. It depends upon the subject's performance, which helps in characterizing the most appropriate reinforcement learning strategy. We correlated the number of mistakes made by the human participants and the total regret felt by the robot and found a linear relationship between the number of mistakes and the robot's regret. The value of the coefficient is $r = 0.88$; thus the robot's regret is strongly correlated and the reinforcement learning strategy used by the robot to understand human reponses and improve their performance is appropriately working. For the participants who had several interactions with the robot or made many mistakes, Baxter tried to explore different reinforcement strategies at different times, trying to increase their learning rate. Hence we can derive that Baxter can successfully train people to achieve complexs task using their preferred motivations.

4) Discussion: In the experiment, the results are not as strong as we might hope for several reasons. If the robot's grippers were closed, they occasionally hindered the camera, blocking the robot from identifying the markers, since we used the left hand camera for detection and evaluation purposes. The Baxter arm and gripper are not extremely dextrous; it is sometimes very unsophisticated in its attempts to pick up the blocks whenever they not lying perpendicularly to the camera. The markers, after several tasks and the degradation which resulted from repeated handling by both human and robot, became unclear and difficult for the marker tracking algorithm to recognize, which also contributed to system crashes. Again, many of the young adults who participated in the experiment failed to connect to the robot emotionally and lacked engagement. Some subjects paid very little attention to the robot's attempts to communicate a reinforcement strategy, to the point that the subjects attempted to interact with the researchers conducting the experiment rather than the robot. Some participants simply produced iteration after iteration of patterns until they happened upon the correct one, without paying attention to the robot gamely attempting to help. Rather, they simply tried each block at each position to figure out the right approach. Hence Baxter on its end was confused in providing the reinforcement strategy. For this reason, we see that Baxter was only able to identify a successful motivational strategy for half the participants in the learned group. Another potential point of alienation came from the fact that Baxter's voice did not issue from the robot itself, but rather a speaker off to the side (since the robot hardware lacks sound capability).

The students had to turn to their right side and interact with a computer console to give Baxter their feedback in form of yes or no, which is unsophisticated; verbal interaction would have been a better option. However, participants in the experiment came from different national backgrounds and language abilities, so it was very hard for the robot to understand their pronunciation. Hence we were forced to keep the human feedback in that format. Seven blocks with 14 markers can be placed in many, many ways, but no participants required nearly that many attempts to figure out the correct pattern. Thus, even when they did not directly engage with the robot's attempts to teach, it still had some impact on them. Although Baxter successfully developed a teaching strategy for half of the participants, it is enough to suggest that such robotic feedback does have impact on human behavior and learning. Furthermore, this approach allows the robot to assess its own success and learn to calibrate its own interactions in ways that lead to successful teaching.

V. Conclusion and Future Work

Feedback and positive reinforcement can be a powerful way to motivate people and influence their thought processes. Positive reinforcement has a salutory effect on the human mind, but each individual is different, and if a robot can learn which reinforcement strategy is most appropriate, it will be more successful at teaching complex tasks. Positive reinforcement can affect performance, interest level and motivation. A robot that can learn and experiment with the variety of potential reinforcement strategies available gains a powerful insight into human behavior. In our work, we developed a reinforcement learning strategy where Baxter behaves as a facilitator, but simultaneously also learns about human nature, goals and motivations. Humans and robots here both act as reinforcement learners sharing actions, policies and goals helping each other to accomplish a particular task.

In our future work we would like to determine the mental models arising out of this bidirectional learning policy, providing valuable information about how robots can train themselves over time to accomplish tasks and make necessary decisions with their human partners. Such models will give us a clearer idea about the heuristics responsible for the decision making process of both robots and humans. This research will have a number of real-world impacts, including enhancing a robot's ability to help train and transfer skills in an apprenticeship learning context. Among other research aims, we plan to apply this approach, specifically, to the problem of helping novice construction equipment operators acquire the skills to control and manipulate their machinery.

References

[1] S. Roy, E. Kieson, C. Abramson, and C. Crick, "Using human reinforcement learning models to improve robots as teachers," in *Proceedings of the 13th ACM/IEEE Conference on Human-Robot Interaction (HRI)*, 2018.

[2] M. Csikszentmihalyi and M. Wong, *Motivation and Academic Achievement: The Effects of Personality Traits and the Quality of Experience*. Springer Netherlands, 2014, pp. 437–465.

[3] S. Griffith, K. Subramanian, J. Scholz, C. L. Isbell, and A. Thomaz, "Policy shaping: Integrating human feedback with reinforcement learning," in *Proceedings of the International Conference on Neural Information Processing Systems (NIPS)*, 2013.

[4] T. Cederborg, I. Grover, C. L. Isbell, and A. L. Thomaz, "Policy shaping with human teachers," in *Proceedings of the 24th Annual International Joint Conference on Artificial Intelligence (IJCAI)*, 2015.

[5] A. L. Thomaz and C. Breazeal, "Teachable robots: Understanding human behavior to build more effective robot learners," *Artificial Intelligence*, vol. 172, pp. 716–737, 2008.

[6] M. Cakmak, N. DePalma, R. Arriaga, and A. L. Thomaz, "Computational benefits of social learning mechanisms: Stimulus enhancement and emulation," in *Proceedings of the 8th IEEE International Conference on Development and Learning (ICDL)*, 2009.

[7] B. Clement, D. Roy, P.-Y. Oudeyer, and M. Lopes, "Multi-armed bandits for intelligent tutoring systems," *Journal of Educational Data Mining*, vol. 7, no. 2, pp. 20–48, 2015.

[8] J.-Y. Audibert, R. Munos, and C. Szepesvari, "Exploration-exploitation tradeoff using variance estimates in multi-armed bandits," *Theoretical Computer Science*, vol. 410, pp. 1876–1902, 2009.

[9] E. S. Kim, L. D. Berkovits, E. P. Bernier, D. Leyzberg, F. Shic, R. Paul, and B. Scassellati, "Social robots as embedded reinforcers of social behavior in children with autism," *Journal of Autism and Developmental Disorders*, vol. 43, no. 5, pp. 1038–1049, 2013.

[10] A. Ramachandran and B. Scassellati, "Developing adaptive social robot tutors for children," in *Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI) Fall Symposium Series*, 2015.

[11] 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 ACM/IEEE Conference on Human-Robot Interaction (HRI)*, 2017.

[12] A. Litoiu and B. Scassellati, "Robotic coaching of complex physical skills," in *Proceedings of the 10th Annual ACM/IEEE Conference on Human-Robot Interaction (HRI)*, 2015.

[13] H. W. Park, R. Rosenberg-Kima, M. Rosenberg, G. Gordon, and C. Breazeal, "Growing growth mindset with a social robot peer," in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2017.

[14] I. Leite, G. Castellano, A. Pereira, C. Martinho, and A. Paiva, "Long-term interactions with empathic robots: Evaluating perceived support in children," in *Proceedings of the 4th International Conference on Social Robotics (ICSR)*, 2012.

[15] S.-L. Lee, S. Kiesler, I. Y. man Lau, and C.-Y. Chiu, "Human mental models of humanoid robots," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2005.

[16] B. Scassellati, "Foundations for a theory of mind for a humanoid robot," *Autonomous Robots*, vol. 12, pp. 13–24, 2002.

[17] L. Boccanfuso, E. Barney, C. Foster, Y. A. Ahn, K. Chawarska, B. Scassellati, and F. Shic, "Emotional robot to examine different play patterns and affect responses of children with and without asd," in *Proceedings of the 11the ACM/IEEE Conference on Human Robot Interaction (HRI)*, 2016.

[18] E. Nunez, S. Matsuda, M. Hirokawa, and K. Suzuki, "Humanoid robot assisted training for facial expressions recognition based on affective feedback," in *Proceedings of the 7th International Conference on Social Robotics (ICSR)*, 2015.

[19] N. T. Fitter and K. J. Kuchenbecker, "Designing and assessing expressive open-source faces for the Baxter robot," in *Proceedings of the International Conference on Social Robotics (ICSR)*, 2016.