Autonomous Storytelling for Social Robot with Human-Centered Reinforcement Learning

Lei Zhang[1], Chuanxiong Zheng[1], Hui Wang[1], Randy Gomez[2], Eric Nichols[2], Guangliang Li[1*]

***Abstract*— Social robots are gradually integrating into human's daily lives. Storytelling by social robots could bring a different experience to users through non-verbal and emotional capabilities compared to text-only one. However, as user needs and preferences over storytelling might change over time during long-term interaction with social robots, it is important for social robots to learn from social interactions with human users in real-time. In this paper, we propose to allow our social robot Haru to learn personalized storytelling styles for different human user's emotional states via human-centered reinforcement learning using the reward provided and delivered by directly interaction with the user explicitly. Results of our user study show that Haru can learn to adapt its storytelling style for detected human emotional states in a few number of interactions, and was perceived to have a better storytelling performance, experience and impact than a neutral one.**

I. INTRODUCTION

As a new type of human emotional companion, social robots are gradually integrating into human's daily lives, which can increase the literacy and creativity by following social norms in constant interaction with humans [1]. Social robots can bring a different experience to users through non-verbal and emotional capabilities [2]. With a figurative form, storytelling by social robots could be more interesting than text-only storytelling [3].

For example, the study of Striepe et al. [4] suggested that careful integration of emotions and non-verbal behaviors like gaze can enhance the human user's experience with robot storytelling. Robot storytelling enriched with gestures and emotions can also increase the user's empathy during the storytelling process [5]. To meet the human user's need and create a better experience, Wang et al. proposed an empathic and adaptive framework for robot storytelling, in which the robot can imitate the storytelling style and content of human teachers and provide users with a personalized storytelling experience [6]. Nichols et al. even proposed collaborative storytelling with social robot Haru, in which the robot can collaborate with a human user to create a unique, improvised story by using a large-scale neural language model to dynamically generate continuations to a story [7], [8], [9].

However, user needs and preferences over storytelling might change over time during long-term interaction with social robots. Therefore, it is important for social robots to learn from social interactions with human users in real-time [10], [11]. Reinforcement learning (RL) can facilitate agents to learn behavior policies by interacting with the environment via trial and error [12], [13], which shares the same key component — interaction — with social robotics. Reinforcement learning has been applied to a variety of scenarios and domains within social robotics with growing popularity [14]. For example, to improve the engagement in a joke-telling scenario, Weber et al. [15] allowed the Reeti robot to learn to adjust its sense of humor during joke-telling via reinforcement learning. The audience's voice and visual smiles which are also part of robot's state representation are used as implicit rewards to update the robot's learning policy.

In the storytelling task, Rudovic et al. [16] proposed a deep reinforcement learning architecture to learn a personalized policy to decide whether to estimate the child's engagement level or to query a human-expert for a video label in a child-robot storytelling interaction dataset. However, queried videos are labeled by a human expert in an offline manner to personalize the policy and engagement classifier to a target child over time. Glanz et al. [17] developed a robotic teddy bear — Robofriend — for telling stories, which can adapt its behavior to keep children's attention via reinforcement learning. However, the storytelling was performed by playing prerecorded video segments with a still image of one page in the printed book and a human reading the text on the page, and there is no text to speech in it. Moreover, the robot used pre-defined metrics assessing the children's engagement by detecting faces and gaze direction as rewards to learn optimal behaviors, which correspond to different types of feedback it can give to the children (e.g., asking questions, positive feedback like 'great job!'). Most related to our work, Ritschel et al. [18] proposed a reinforcement Q-learning approach to adapt the robot's linguistic style (i.e., the extent of extraversion) to keep the user engaged in the storytelling scenario using change in the engagement as a reward. The engagement of the user was estimated from the user's movement using a Dynamic Bayesian Network. Park et al. [19] also used reinforcement learning to facilitate a Tega robot to adapt its storytelling content (i.e., the lexical and syntactic complexity of a given sentence in a storybook) in an educational activity where a child and a robot tell stories to each other. In their work, rewards are extracted from a pre-defined weighted sum of engagement and lexical and syntax learning of the user.

The above work tried to maximize user's engagement and attention during storytelling using pre-defined metrics to implicitly extract rewards from estimated engagement. The

[1]College of Information Science and Engineering, Ocean University of China, {guangliangli}@ouc.edu.cn

[2]Honda Research Institute Japan Co., Ltd, Wako, Japan. {r.gomez, e.nichols}@jp.honda-ri.com

[*] Corresponding author

extracted rewards are also part representation of robot's state and used to update the robot's policy, which is similar to traditional reinforcement learning from pre-defined reward functions. Nichols et al. also used Haru for storytelling, but it does not learn to adapt to user's preference via RL.

In this paper, we propose to allow our social robot Haru to learn personalized storytelling styles for different human user's emotional states via human-centered reinforcement learning using. The rewards used in our work are provided and delivered by directly interaction with the user explicitly, which can directly reflect user's preference and be communicated via different interactive channels, e.g., speech. Results of our user study show that Haru can learn to adapt its storytelling style for detected human emotional states in a few number of interactions, and was perceived to have a better storytelling performance, experience and impact than a neutral one.

II. ROBOTIC PLATFORM AND METHODOLOGY

Our goal is to develop an adaptive emotional storytelling social robot that can tell stories with different styles according to detected emotions of a human user. An illustration of the learning mechanism with our method is shown in Fig. 1.

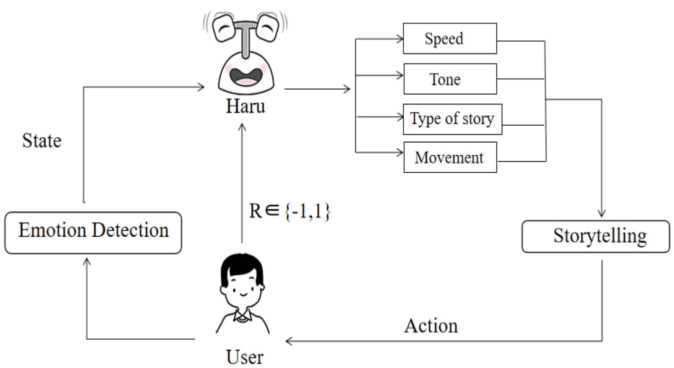

Fig. 1: Haru learns to adapt its style of storytelling and select the type of story for detected emotions from evaluative feedback provided by a human user.

A. Robotic Platform

Haru is an experimental tabletop robot which can be used to study human-robot emotional interaction research [20]. It has five degrees of freedom including base rotation, neck tilt, eye rotation, eye tilt, and eye stroke [21]. Haru's eyes are equipped with two three-inch TFT screen displays. In addition, Haru has stereo speakers and a microphone array, and its voice varies in intensity, subtlety, etc. Therefore, Haru can express various emotions through a combination of sounds, movements, and eyes to achieve reactive empathy with the user, as shown in Fig. 2. Haru is also equipped with RGB-D cameras, through which image and depth-sensing data are processed by designated software module for facial expression recognition.

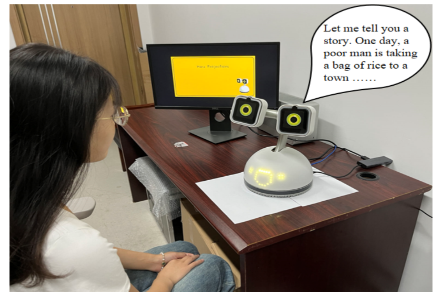

Fig. 2: Haru is telling a story to a user in our experiment.

B. Reinforcement Learning Module

A reinforcement learning agent learns optimal policies mapping from its environmental states to actions through interaction with the environment via trial and error [12]. Ideally, a deep reinforcement learning method would be used to learn the mapping from raw facial expressions to optimal storytelling behaviors. However, this will take a long time to learn facial features for representing the human user's emotional state before it can effectively learn the optimal behaviors. Therefore, we use the recognised results of the perceived modality from the perception suite of Haru as input to the learning system, which could hasten the optimal behavior learning. Haru can also identify different users with the perception system. The learning mechanism we used is human-centered reinforcement learning or human-in-the-loop reinforcement learning [10]. In this case, the rewards for learning are not provided by a pre-defined reward function as the traditional reinforcement learning, but delivered by a human user.

For simplicity, we used tabular Q-learning [22] as the learning algorithm in our system. At the beginning, the Q values for all actions of all states will be initialized to be 0. Then, when Haru detects the current emotional state of a human user interacting with it, it will randomly select an action from the action set with equal Q values. For example, at time step t, Haru detects the human emotional state s_t and selects an action randomly at the beginning. Then, the human user will evaluate the performed action by Haru and provide binary feedback to it. If the human user approves Haru's selected action, she will provide a positive reward (+1); if she disapproves Haru's selected action, a negative feedback (-1) will be provided. The received feedback from the human user will be perceived as reward R to update corresponding actions, as below:

$$Q(s_t,a_t)_{new} = Q(s_t,a_t)_{old} + \alpha(R + \gamma \max_a Q(s_{t+1},a) - Q(s_t,a_t)), \quad (1)$$

where s_t is the current human emotional state detected by Haru, a_t is the performed action by Haru, s_{t+1} is the next state, γ is the discount factor, α is the learning rate, and R is the received evaluative feedback from the human user. Then, at next time step $t+1$, Haru will detect a new human emotional state s_{t+1}, and selects the action with the largest Q value, as below:

$$\pi(s) : a \leftarrow \underset{a \in A}{argmax} Q(s_{t+1},a), \quad (2)$$

TABLE I: Haru's state and action space.

Emotional States		"Happy", "Sad", "Neutral", "Surprise"
Haru's Actions	Speed	Rate$\{slow, fast\}$
	Tone	Pitch $\{low, high\}$
	Type of story	$\{comedy, science fiction, sad\}$
	Movement	$\{smile, idle\}$

where s_{t+1} is the detected next emotional state of the human user and A is the action set of possible actions that can be performed by Haru.

Then, a new cycle of receiving human evaluative feedback, updating the corresponding Q-value, detecting new emotional state of the human user and selecting the action with the largest Q value starts. Haru will learn as long as it receives feedback from the human user until she is satisfied with Haru's behaviors for detected emotional states.

C. State and Action Space

Ideally, the more are the number of human emotional states, the better would it fit with the real application scenarios with our system. However, more human emotional states mean longer learning time for Haru in our study. Therefore, for simplicity and to keep the training time within the human physical endurance in our study, we considered four easily recognized emotional states for the human user: Happy, Sad, Neutral, Surprise, as Haru's possible states. The four emotional states are detected by Haru in a random order for each subject. The action of Haru for storytelling can be represented in four dimensions: speed, tone, type of story and movement. The speed of Haru storytelling can be slow or fast and the tone can be low or high. There are three types of stories can be selected: comedy, science fiction and sad. The movement of Haru during storytelling can be smile or idle. Therefore, the action space of Haru storytelling is 24-dimensional, as shown in Table I.

III. EXPERIMENTS

A. Experimental Conditions

Our goal is to facilitate Haru to learn simple autonomous storytelling behaviors via interacting with the human user and adapt to their preferences. To this end and as a first step, we set three experimental conditions:

- Human storyteller condition: the participant watches a human telling stories;
- Neutral Haru condition: the participant listens to Haru's storytelling with a fixed speech rate, intonation, and movement, the human user can provide feedback to shape the type of story told by Haru;
- Affective Haru condition: the participant listens to Haru's storytelling, the speech rate, intonation, movement and type of story can be shaped by a human user according to her preference.

In the human storyteller condition, human storytellers narrated stories in three genres: comedy, science fiction and sad, in a lively and vivid way. The human storyteller condition was set to test whether participants will perceive Haru's storytelling with our method in affective Haru condition is similar to that of human users. The difference between neutral Haru and affective Haru is that in neutral Haru condition, Haru can only learn to select the optimal type of story with the received feedback from the human user and tell the selected story with a fixed speech rate, intonation, and movement, while in the affective Haru condition, the human user can shape the speech rate, intonation, movement and type of story for Haru by providing feedback according to her preference. That is to say, the actions of neural Haru only include the type of story in Table I.

B. Experimental Setup

In our experiments, participants were asked to imagine which type of story and the way they would like Haru to tell stories in various emotional states, and then try to enact the four emotions and teach Haru appropriate storytelling behaviors according to their preferences. In the experiment, Haru first detects the human's emotional state, and selects and performs the storytelling behavior with an initialized policy. The participants listen to the way Haru telling the story and the content of the storytelling, and provide feedback to shape Haru's behavior. Haru will update its storytelling policy based on the received feedback. Then, in a new cycle, Haru will detect participants' new emotional states, perform storytelling behavior with updated policy and receive participants' feedback, until participants are satisfied with Haru's storytelling behavior for all emotional states. Fig. 2 shows Haru's storytelling to a human user in our experiment.

We recruited 20 subjects from one university campus for our study. Of them, 13 are male and 7 are female, aged from 20 to 26 years old. 3 are novice master students in robotics but know nothing about the system, 17 are bachelor students who are ignorant of machine learning and robotics. Each subject was invited to take part in all three experimental conditions (within-subject study) and received partial course credit for taking part in the study. The order in which each participant took part in the three conditions was randomly assigned. All participants filled out two questionnaires after finishing experiments in the three conditions. A third questionnaire was filled out by participants in the neural Haru and affective Haru conditions. The first two questionnaires are from [2], [8], [23] and used to evaluate participants' perception of storytelling performance and experience. All questions in the first two questionnaires are on a 5-point Likert scale. The third questionnaire with four questions is designed by ourselves to asses and compare the storytelling impact of neutral Haru and affective Haru. The purpose of the study was revealed at the end of the study. Due to the time constraint and the physical endurance of the human participants, we set the maximum number of interactions to 100 and the whole study lasted for about 1.5 hours.

IV. RESULTS AND DISCUSSION

A. Visualized Learning Process

To better understand Haru's learning for storytelling from human feedback, we visualized the learned Q model (i.e.,

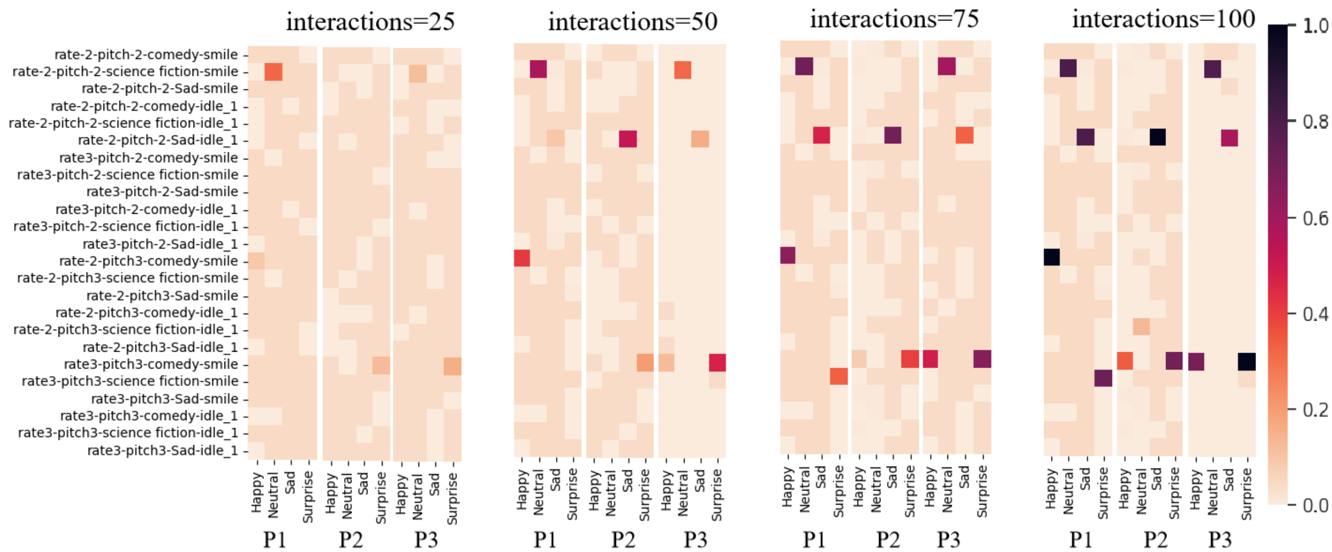

Fig. 3: Visualized heat map of Q values for each storytelling action in all emotional states during the training process. Haru's behavior was trained by three (P1 P2 P3) of the 20 participants in the affective Haru condition.

the Q value function for all storytelling behaviors) in all four detected human emotional states every 25 interactions in a heat map trained by three subjects, as shown in Fig. 3. The horizontal axis represents the four emotional states of three human users, and the vertical axis represents the 24 possible actions for storytelling in each state. Each block shows the Q value for an storytelling action in one emotional state. For easy comparison, all Q values are normalized to the same scale. The deeper is the block's color, the larger is the Q value. Haru will select and execute the storytelling action with the largest Q value for each emotional state. From Fig. 3 we can see that, as the number of interactions increases, Haru gradually learned different optimal storytelling behaviors for the four emotional states according to three subjects' preferences. For example,in the 'Happy' state, P1 preferred different speeds of storytelling from P3, and different type of story in the 'surprise' state (science fiction by P1 and comedy by P3.)

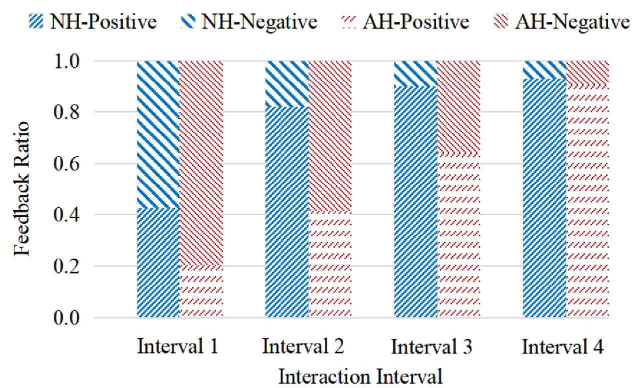

Fig. 4: The ratio of positive and negative feedback provided by all participants every 5 interactions during the training process in neutral Haru condition and every 25 interactions during the training process in affective Haru condition. Note that 'NH' represents the Neutral Haru Condition, 'AH' represents the Affective Haru condition.

B. Number of Feedback

We also analyzed the ratio of positive and negative feedback provided by all subjects in the neutral Haru and affective Haru conditions. The training process is divided into 4 intervals (every 5 interactions for neutral Haru since only selecting the type of story is needed to learn and it took only about 20 interactions to learn an optimal behavior, and every 25 interactions for affective Haru because of the much longer training time). As shown in Fig. 4, In the first interval, the participants provided more negative feedback than positive one, especially for the affective Haru condition. As Haru learns, the ratio of positive feedback becomes higher and higher. And during the last interval, almost all feedback provided by participants was positive one. This is consistent with the results in Fig. 3, as Haru already learned a stable policy for storytelling.

C. Personalization

We analyzed the final optimal storytelling behavior for all emotional states trained by all participants in affective Haru condition and found that all participants were trying to personalize Haru's behavior. As shown in Fig. 5, the horizontal axis represents the percentage of each optimal way of storytelling (low and high tone, slow and fast speech, smile and idle) trained by the 20 participants for all emotional states, and the vertical axis represents the four emotional states. From Fig. 5 we can see that, when participants' emotional state is 'happy', they prefer Haru to tell story in a fast speed with a high tone and a smile. When the detected emotional state is 'sad', participants prefer Haru to tell the story in a slow speed with a low tone and a smile. When the participants' emotional state is 'neutral', they prefer Haru to tell the story in a slow speed with a low tone and a smile.

For the emotional state of 'surprise', participants prefer Haru to tell the story at a fast speed with a high tone and a smile.

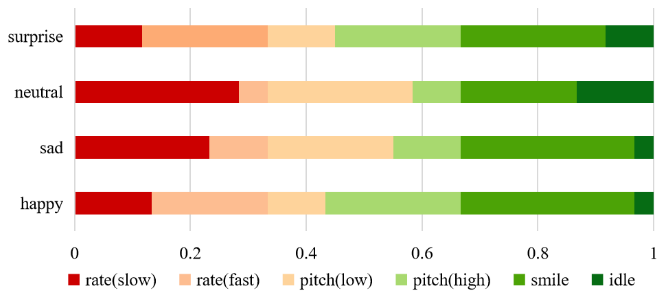

Fig. 5: The percentage of each optimal way of storytelling trained by the 20 participants for all emotional states in the affective Haru condition.

In Fig. 6, the horizontal axis represents the percentage of preferred types of story (comedy, sad, science fiction) trained by the 20 participants for all emotional states in the neutral Haru and affective Haru conditions, and the vertical axis represents the four emotional states. From Fig. 6 we can see that, participants prefer Haru to tell comedy stories when their emotional state is 'happy', and science fiction stories in the 'surprise' and 'neutral' emotional state. However, Haru was trained to prefer telling sad stories in the 'sad' emotional state, which is a bit surprising since people usually prefer to cheer up a sad friend, e.g., by telling comedy stories.

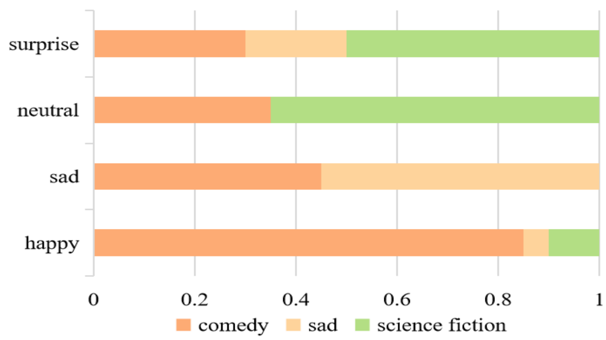

Fig. 6: The percentage of preferred types of story trained by the 20 participants for all emotional states in the neutral and affective Haru conditions.

D. Storytelling Performance and Experience: Human Storyteller vs. Neutral Haru vs. Affective Haru

We compared the performance of Haru's storytelling in neural Haru condition and affective Haru condition to human storyteller condition. Haru's storytelling performance was assessed by mean scores rated by the 20 participants in the questionnaire after the study in terms of six aspects: expressiveness, intelligence level, storytelling effectiveness, storytelling ability, overall evaluation and degree of popularity, as shown in Fig. 7. We did an n-way analysis of variance (ANOVA) as well as multiple comparisons tests (Significance level: $p < 0.01$) of storytelling performance in the three conditions. Our ANOVA analysis shows that there is a significant difference in terms of storytelling performance

TABLE II: The n-way analysis of variance (ANOVA) of differences in terms of "storytelling performance" between human storyteller, neutral Haru and affective Haru.

	F	PR(>F)
C(Condition)	118.74	<0.001
C(Performance)	0.49	0.79
C(Condition): C(Performance)	0.64	0.76

TABLE III: Multiple comparisons of "storytelling performance" between human storyteller, neutral Haru and affective Haru.

Group1	Group2	meandiff	p-adj
Human storyteller	Neutral Haru	-1.56	0.001
Human storyteller	Affective Haru	-0.25	0.05
Neutral Haru	Affective Haru	1.31	0.001

between conditions (p <0.001). Multiple comparisons tests show that the storytelling performances of human storyteller and affective Haru were significantly better than the neutral Haru ($p = 0.001$ and $p = 0.001$, respectively), while there was no significant difference in terms of storytelling performance between affective Haru and human storyteller.

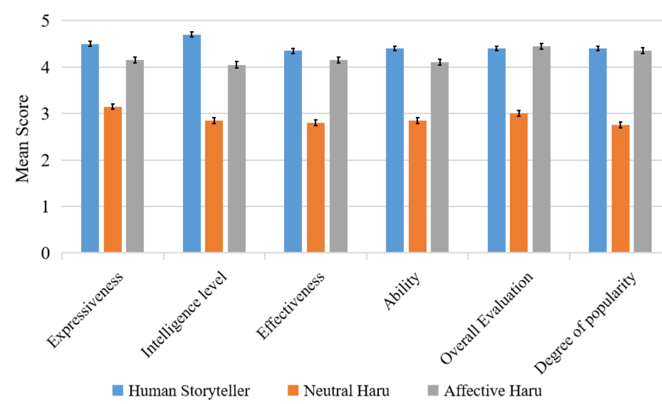

Fig. 7: Mean scores of affective Haru, neutral Haru and human "storytelling performance" rated by the 20 participants in the questionnaire after the study. Note: black bars represent standard deviation.

We also did an n-way analysis of variance (ANOVA) as well as multiple comparisons tests (Significance level: $p < 0.01$) over participants' storytelling experience in the three conditions, which was assessed by comparing mean scores rated by the 20 participants in the questionnaire after the study from four perspectives: interestingness, wonderfulness, pleasure level and enjoyment, as shown in Fig. 8. Our ANOVA analysis shows that there is a significant difference in terms of storytelling experience between conditions (p <0.001). Results of multiple comparisons tests show that the storytelling experience with human storyteller and affective

TABLE IV: The n-way analysis of variance (ANOVA) of differences in terms of "storytelling experience" between human storyteller, neutral Haru and affective Haru.

	F	PR(>F)
C(Condition)	73.27	<0.001
C(Experience)	0.88	0.45
C(Condition): C(Experience)	0.30	0.94

TABLE V: Multiple comparisons of "storytelling experience" between human storyteller, neutral Haru and affective Haru..

Group1	Group2	meandiff	p-adj
Human storyteller	Neutral Haru	-1.35	0.001
Human storyteller	Affective Haru	0.15	0.51
Neutral Haru	Affective Haru	1.5	0.001

Haru were significantly better than with the neutral Haru ($p = 0.001$ and $p = 0.001$, respectively), while participants perceived that there was no significant difference in terms of storytelling experience with affective Haru and human storyteller.

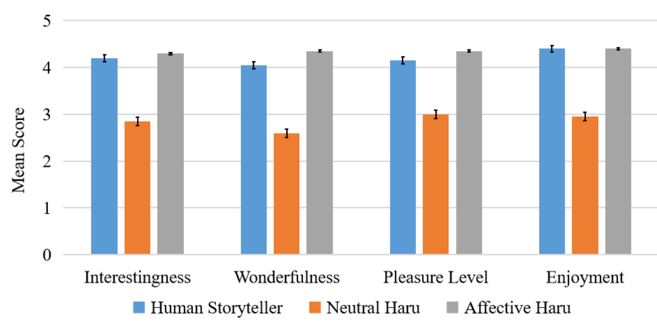

Fig. 8: Mean scores of affective Haru, neutral Haru and human "storytelling experience" rated by the 20 participants in the questionnaire after the study. Note: black bars represent standard deviation.

E. Storytelling Impact: Affective Haru vs. Neutral Haru

We further evaluated the storytelling impact of the neutral Haru and affective Haru with four questions in the questionnaire filled by the 20 participants. Fig. 9 shows the mean scores of the four questions in the questionnaire averaged over collected data. Student's t-test was performed to test the significance of differences between neutral Haru and affective Haru, as shown in Table VI. From Fig. 9 and Table VI we can see that, both our affective Haru and neural Haru can fully explain the story (no significance, $p = 0.22$). However, our affective Haru was perceived to have a significantly clear change in the mood, tone of voice and expression during storytelling ($p < 0.01$) and the perceived

TABLE VI: The significance of differences in terms of "storytelling impact" between neutral Haru and affective Haru. Significance level: p<0.01.

	Question	T-Test
Q1	During storytelling, do you feel a change in Haru's mood, tone of voice and expression?	t=8.21, p<0.01
Q2	Do you think Haru's changing emotions have a positive impact on storytelling?	t=4.42, p<0.01
Q3	Do you think Haru can fully explain the story?	t=1.24, p=0.22
Q4	Do you think Haru can learn to cater to your preferences	t=4.30, p<0.01

changing emotions of our affective Haru had significantly positive impact on storytelling ($p < 0.01$). Moreover, our affective Haru can learn to cater the participants' preferences better compared to neural Haru ($p < 0.01$).

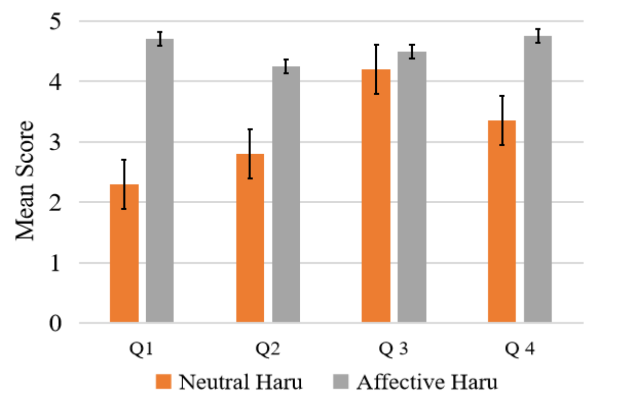

Fig. 9: Mean scores of the four questions in the questionnaire to measure the "storytelling impact" of neutral Haru and affective Haru averaged over data collected from the 20 participants. Note: black bars represent standard deviation.

V. CONCLUSION

This paper proposed a human-centered reinforcement learning approach to facilitate a social robot Haru to learn personalized storytelling behaviors in terms of speed, tone, movement and type of story, for detected emotional state of a human user. The study results show that with the proposed approach, Haru can learn different behaviors for all detected moods of human users according to their preferences. Further analysis of the user study reveals that the storytelling performance, experience and impact of Haru learning with the proposed method were perceived to be significantly better than a neutral Haru with fixed speed, tone, movement for storytelling. We believe our results could generalize to other social robots, since they can have the same sensors (e.g., stereo speakers, microphone array, camera etc.) to perform the personalized storytelling behaviors. In the future, multimodal data to perceive the emotional state of human user will be used, e.g., content and tone of speech, and we will try to increase the number of human emotional states and test our system in a more natural setting. In addition, large language models will be considered for generating the content of story based on the human user's feedback.

REFERENCES

[1] G.-D. Chen and C.-Y. Wang, "A survey on storytelling with robots," in *Proceedings of International Conference on Technologies for E-Learning and Digital Entertainment*. Springer, 2011, pp. 450–456.

[2] H. Striepe and B. Lugrin, "There once was a robot storyteller: measuring the effects of emotion and non-verbal behaviour," in *Proceedings of the 9th International Conference on Social Robotics (ICSR)*. Springer, 2017, pp. 126–136.

[3] J. Sjöbergh and K. Araki, "Robots make things funnier," in *New Frontiers in Artificial Intelligence*. Springer, 2009, pp. 306–313.

[4] H. Striepe, M. Donnermann, M. Lein, and B. Lugrin, "Modeling and evaluating emotion, contextual head movement and voices for a social robot storyteller," *International Journal of Social Robotics*, vol. 13, pp. 441–457, 2021.

[5] A. Augello, "Unveiling the reasoning processes of robots through introspective dialogues in a storytelling system: A study on the elicited empathy," *Cognitive Systems Research*, vol. 73, pp. 12–20, 2022.

[6] H. Wang, L. Zhang, C. Zheng, R. Gomez, K. Nakamura, and G. Li, "Personalized storytelling with social robot haru," in *Proceedings of the 14th International Conference on Social Robotics (ICSR)*. Springer, 2023, pp. 439–451.

[7] E. Nichols, L. Gao, and R. Gomez, "Collaborative storytelling with large-scale neural language models," in *Proceedings of the 13th ACM SIGGRAPH Conference on Motion, Interaction and Games*, 2020, pp. 1–10.

[8] E. Nichols, L. Gao, Y. Vasylkiv, and R. Gomez, "Collaborative storytelling with social robots," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 1903–1910.

[9] E. Nichols, D. Szapiro, Y. Vasylkiv, and R. Gomez, "I can't believe that happened!: Exploring expressivity in collaborative storytelling with the tabletop robot haru," in *Proceedings of the 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2022, pp. 59–59.

[10] G. Li, R. Gomez, K. Nakamura, and B. He, "Human-centered reinforcement learning: A survey," *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 4, pp. 337–349, 2019.

[11] J. Lin, Z. Ma, R. Gomez, K. Nakamura, B. He, and G. Li, "A review on interactive reinforcement learning from human social feedback," *IEEE Access*, vol. 8, pp. 120 757–120 765, 2020.

[12] R. Sutton and A. Barto, *Reinforcement learning: an introduction*. MIT Press, 1998.

[13] J. Kober, J. A. Bagnell, and J. Peters, "Reinforcement learning in robotics: A survey," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1238–1274, 2013.

[14] N. Akalin and A. Loutfi, "Reinforcement learning approaches in social robotics," *Sensors*, vol. 21, no. 4, p. 1292, 2021.

[15] K. Weber, H. Ritschel, I. Aslan, F. Lingenfelser, and E. André, "How to shape the humor of a robot-social behavior adaptation based on reinforcement learning," in *Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI)*, 2018, pp. 154–162.

[16] O. Rudovic, H. W. Park, J. Busche, B. Schuller, C. Breazeal, and R. W. Picard, "Personalized estimation of engagement from videos using active learning with deep reinforcement learning," in *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2019, pp. 217–226.

[17] I. Glanz, M. Weksler, E. Karpas, and T. Horowitz-Kraus, "Robofriend: An adpative storytelling robotic teddy bear–technical report," *arXiv preprint arXiv:2301.01576*, 2023.

[18] H. Ritschel, T. Baur, and E. André, "Adapting a robot's linguistic style based on socially-aware reinforcement learning," in *Proceedings of the 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2017, pp. 378–384.

[19] H. W. Park, I. Grover, S. Spaulding, L. Gomez, and C. Breazeal, "A model-free affective reinforcement learning approach to personalization of an autonomous social robot companion for early literacy education," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 687–694.

[20] R. Gomez, D. Szapiro, K. Galindo, and K. Nakamura, "Haru: Hardware design of an experimental tabletop robot assistant," in *Proceedings of ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2018, pp. 233–240.

[21] H. Brock, S. Sabanovic, K. Nakamura, and R. Gomez, "Robust realtime hand gestural recognition for non-verbal communication with tabletop robot haru," in *Proceedings of the 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2020, pp. 891–898.

[22] C. J. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, pp. 279–292, 1992.

[23] M. Appel, B. Lugrin, M. Kühle, and C. Heindl, "The emotional robotic storyteller: On the influence of affect congruency on narrative transportation, robot perception, and persuasion," *Computers in Human Behavior*, vol. 120, p. 106749, 2021.