Learning Socially Appropriate Robo-waiter Behaviours through Real-time User Feedback

Emily McQuillin, Nikhil Churamani, and Hatice Gunes
Department of Computer Science and Technology, University of Cambridge, Cambridge, UK
emm95@cantab.ac.uk, {nikhil.churamani, hatice.gunes}@cl.cam.ac.uk

***Abstract*—Current Humanoid Service Robot (HSR) behaviours mainly rely on static models that cannot adapt dynamically to meet individual customer attitudes and preferences. In this work, we focus on empowering HSRs with adaptive feedback mechanisms driven by either implicit reward, by estimating facial affect, or explicit reward, by incorporating verbal responses of the human 'customer'. To achieve this, we first create a custom dataset, annotated using crowd-sourced labels, to learn appropriate approach (*positioning* and *movement*) behaviours for a *Robo-waiter*. This dataset is used to pre-train a Reinforcement Learning (RL) agent to learn behaviours deemed socially appropriate for the robo-waiter. This model is later extended to include separate implicit and explicit reward mechanisms to allow for interactive learning and adaptation from user social feedback. We present a within-subjects Human-Robot Interaction (HRI) study with 21 participants implementing interactions between the robo-waiter and human customers implementing the above-mentioned model variations. Our results show that both explicit and implicit adaptation mechanisms enabled the adaptive robo-waiter to be rated as more *enjoyable* and *sociable*, and its *positioning* relative to the participants as more *appropriate* compared to using the pre-trained model or a randomised control implementation.**

***Index Terms*—Humanoid Robo-waiter, Reinforcement Learning, Explicit Feedback, Implicit Feedback and Facial Affect.**

I. Introduction

The use of Humanoid Service Robots (HSRs) operating specifically in service domains such as restaurants or hotels has increased in recent years [1]. Even though many HSRs do not interact with customers and only provide operational assistance, they increasingly are being deployed in more front-of-house roles, such as the *Robo-waiter* introduced in Pizza Hut restaurants across Asia [2]. Developing such HSRs requires careful consideration of Human-Robot Interaction (HRI) techniques to ensure the users do not find the experience uncomfortable or frustrating.

When humans engage in dialogue, they use socio-emotional signals to communicate different preferences [3]. Whilst humans are able to learn the preferences of those with whom they experience repeated interactions, existing HSRs rely on pre-defined behaviours and are not able to modify their behaviours with human feedback [4], [5]. This lack of emotional intelligence can hinder the performance of an HSR, which can be costly. For instance, in the hospitality domain, higher levels of emotional intelligence are seen to increase customer satisfaction and profit performance [6]. From the customers' perspective, employees in the service or hospitality industry who have better social skills are seen as more competent and approachable [7].

In this paper, we present a study that aims to improve on static models currently employed by HSRs by learning socially appropriate behaviours based on real-time feedback from the customer. We propose an adaptive HSR, that uses feedback from a human-in-the-loop to modulate behaviours in real-time. The performance of this adaptive HSR is evaluated through an HRI study to understand whether adaptive behaviours improved perceptions of the HSR across several dimensions, including *sociability* and *appropriateness*. As humans use both verbal (explicit) and non-verbal (implicit) signals to communicate, we implement and evaluate two separate adaptive agents; one that adapts using explicit feedback, and another using implicit feedback in the form of facial affect. A restaurant setting is realised for the *robo-waiter*, providing context for repeated interactions with the same customer, giving it the opportunity to learn and adapt for subsequent interactions.

II. Background and Related Work

A. Humanoid Service Robots

Robotic technologies have started influencing service industries. In 2014, Savioke tested a robot in Aloft Hotels [8] for deliveries to guests' rooms. The Pepper robot has also been employed by Pizza Hut restaurants to take customer orders and accept payments [2]. Enabling such applications for HSRs requires their extensive evaluation under various application domains. Lee *et al.* [9] evaluated a snack-delivery HSR for the workplace where, over a 4-month field study, they found employees extended the social roles of the robot beyond deliveries, attaching several different roles to it. This created a 'ripple effect' in the workplace, triggering new behaviours in employees. Herse *et al.* [10] investigated robot persuasion for food recommendations and found that the human-like features of an agent may contribute to boosting persuasion. Jie *et al.* [11] conducted a beverage-tasting study with a human *vs.* a humanoid social robot facilitator, with priming and non-priming instruction styles showing that the facilitation style and facilitator type did not have a significant influence on taste-liking. However, people were more willing to follow

E. McQuillin was supported by the 2020/21 DeepMind Cambridge Scholarship. N. Churamani is funded by the EPSRC grant EP/R513180/1 (ref. 2107412). H. Gunes is supported by the EPSRC project ARoEQ under grant ref. EP/R030782/1.

instructions, and felt more comfortable, with the humanoid robot facilitating with priming.

To effectively model behaviours for HSRs such as robo-waiters, Schmidt-Rohr *et al.* [12] applied Reinforcement Learning (RL) for modelling robot behaviours by understanding human activity through speech and posture. Yet, the focus of this work was to only model correct waiter behaviour, without any adaptation. More recently, Sawadwuthikul *et al.* [13] sought to improve on these static behaviour policies through dynamic adaptation of positioning and trajectory by learning from a human-in-the-loop. Human feedback was only used to assist the robots' ability to remember the location of a customer when several customers were present. In both of these studies, the focus was on task-based learning, in that the agent successfully learns to accomplish tasks associated with being a robo-waiter. They did not, however, focus on learning socially appropriate behaviours or personalising robot behaviour to specific customers as a result of their feedback.

B. Learning Socially Appropriate Behaviours in HRI

For robots to operate effectively in society, they need to understand various social norms and individual user-behaviour as inaccurate situational behaviour can decrease trust in social robots [14]. This is particularly challenging while learning appropriate behaviours during the first interactions with a user [15] as first impressions may impact competence ratings subsequently [16]. A key factor to consider while designing robots for human-centric environments is how they position themselves while interacting with humans. Understanding the basic principles of proxemics [17], [18], that is, approaching humans in an appropriate manner or the robot positioning itself at an appropriate distance to encourage a conversation, becomes essential for meaningful human-robot interactions.

To learn socially appropriate robot behaviours, most current approaches use static datasets, created using crowd-sourced labelling platforms, that provide common consensus annotations on what is considered socially appropriate. Tjomsland *et al.* [19] proposed the MANNERS-DB consisting of 3D scenes created in Unity, where the appropriateness of robot actions in each scene was labelled on a 5-point Likert scale, ranging from very inappropriate to very appropriate using a crowd-sourced labelling platform. Similarly, Gao *et al.* [20] trained an agent to learn socially appropriate approach behaviour using a 3D simulated environment in Unity. They formulated the task to be learnt as an RL problem, defining the reward function in terms of existing social theories [21], [22]. They demonstrated, in a within-subjects HRI study, that the agent trained on offline data was rated as more *polite*, *sociable* and *human-like* compared to an agent following a static policy [22]. With the help of offline training on large datasets, RL has been shown to be effective at learning socially appropriate behaviours [23] outperforming static models of social behaviour [20]. Yet, real-time adaptation in social robots is still relatively less explored. Recent works have shown that robots are able to learn socially appropriate behaviours through a combination of feedback mechanisms, which might include implicit, explicit or pre-trained behaviours [13], [20], [24]. Implicit signal processing in HRI, such as evaluating facial expressions [25] or body language, allows for more feedback to be collected from the participant, and also reduces 'feedback fatigue' [26]–[28]. Both implicit and explicit feedback have been used as input for Interactive Reinforcement Learning (IRL) models [13], [24]. However, their efficacy has not yet been adequately explored, and their impact on user perceptions is still to be investigated extensively.

C. Automatic Facial Reaction Analysis

As human communication relies on the successful exchange of verbal and non-verbal affective signals [29], embedding such an understanding of human behaviour in social robots plays a pivotal role in improving interactions with users and increasing trust in HRI [30]. Human affect is communicated in a myriad of ways, but the most frequently researched affective signals include speech, facial expressions and body gestures [31], [32]. These signals can be evaluated either in terms of user expressions categorised into emotion classes [33], or a dimensional approach such as the Circumplex Model [34] can be used to provide a more flexible and realistic representation of affect with *valence* depicting the positive or negative nature of the expression and *arousal* depicting its intensity.

Understanding user expressions can not only provide the necessary motivation for robots to adapt their behaviour but also act as feedback on how the interaction is going. A straightforward way to evaluate user behaviour during an interaction is to observe their facial expressions. Facial affect has been used to improve perceptions of robots in HRI [24] or to provide evaluations for robot behaviour [35], [36] but using facial affect to provide real-time implicit feedback, particularly in interactions with robo-waiters, is yet to be explored.

D. Reinforcement Learning in HRI

RL strategies such as Interactive Reinforcement Learning (IRL) [37] allow for training the agents through natural interactions with humans who provide feedback to the agent, shaping their behaviour in real-time [38]. This feedback can either be used to shape the action-policy, directly influencing the agent's actions [39], [40], or modulate the reward function [36], [41] guiding it to learn optimal behaviours. However, learning with the *human-in-the-loop* can be challenging as humans tend to provide more positive than negative feedback, at times ignoring the robots' mistakes. Additionally, as the interaction progresses, the frequency of providing feedback decreases. Modelling implicit feedback, such as using facial affect, can thus be more effective as the human 'teacher' will be less conscious of providing feedback and will be less likely to suffer from 'feedback fatigue' [27]. Existing approaches have explored such implicit feedback signals for robot learning [27], [42] such as Weber *et al.* [24] who focus on learning appropriate humorous behaviour using audiovisual data to detect laughs and smiles to reward the robot. They found that the robot was considered to be funnier when it adapted its behaviour in response to users' affective responses.

III. Research Questions and Contributions

This work investigates whether adapting robot behaviours improves human perceptions of robots in HSR settings. Prior work has used pre-annotated datasets to learn appropriate behaviours and found this to positively improve the robot evaluations [22]. Hence, we first investigate **(RQ1:)** whether pre-training robot behaviours with crowd-sourced data improves robo-waiter perceptions as compared to a 'control' condition with randomly sampled, static robot behaviours. Recent works have also highlighted the importance of using objective feedback in HRI evaluation, arguing that the time-consuming nature of subjective feedback (via questionnaires) makes it less likely to be effective in longitudinal learning settings such as with HSRs [43]. Thus, we employ an objective evaluation strategy, via explicit (using speech) and implicit (using facial expressions) feedback given by the participants during their interactions with the HSR. We first investigate **(RQ2:)** whether person-specific adaptation based on explicit feedback improves robo-waiter perceptions as compared to generalised learning with crowd-sourced data. In contrast, we also investigate **(RQ3:)** whether person-specific adaptation based on implicit feedback improves robo-waiter perceptions as compared to generalised learning with crowd-sourced data. Furthermore, to compare the two different feedback strategies, we explore if **(RQ4:)** there is a significant difference in robo-waiter perceptions when using explicit or implicit feedback.

To the best of our knowledge, this is the first study investigating different adaptation strategies in HSRs in the context of a robo-waiter. The contributions of this work are three-fold. Firstly, we create a dataset of varied HSR behaviours and undertake a web-based evaluation (see Section IV-A) to understand potential user perceptions and obtain labels for appropriateness via crowd-sourcing. We provide detailed statistical analyses for the appropriateness of positioning and speed labels for the robo-waiter. Secondly, we evaluate three popular RL methods using the collected data to train the robo-waiter agent and implement the best performing method (see Section IV-C) on the Pepper robot for real-time evaluations. Finally, we conduct a within-subjects HRI study with the Pepper Robot comparing 4 different strategies (see Section V-A4) for implementing robot behaviour. We compare the perceptions of the robot under these 4 strategies via statistical analyses and present insightful findings for future research.

IV. Web-based Survey and Model Training

A. Materials and Methods

For the robot to learn socially appropriate behaviours, it is important to understand how different individuals rate the appropriateness of HSR behaviours in restaurant settings. Thus, we conducted a crowd-sourced study where participants were asked to imagine themselves being served by a robo-waiter. They were presented with a survey, using Google Forms[1], where they were shown images and GIFs illustrating the robot positioning itself when serving a customer and were asked to provide appropriateness labels for these behaviours using a 5-point Likert scale (ranging from *very inappropriate* to *very appropriate*). Along with appropriateness ratings, participants also provided annotations based on their affective responses towards the robot's positioning in terms of *valence* and *arousal* labels using Self-assessment Manikin (SAM) [44] annotations.

1) Setup: The images and GIFs used in the web-based survey were created using a 9×9 grid, in which the middle 3×3 grid-locations were reserved for the placement of the table and the person sitting at the table. The first section of the survey consisted of 24 questions where the participants were shown 12 images depicting the robot's position with respect to the customer. For each image, they were asked to rate the appropriateness of the robot's positioning. Additionally, for every other position (6 out of the 12) they were also asked to annotate their affective responses (valence/arousal) towards the robot in that position. The second section of the survey consisted of further 24 questions, where participants were shown 12 different GIFs in which the robot moved one or two steps from its original position in any direction (*up, down, left or right*). For each GIF, the participants rated the appropriateness of both the movement (direction of approach) and the speed of the robot, considering it was approaching the customer. An example of the questions can be found in Fig. 1 of the supplementary material provided.

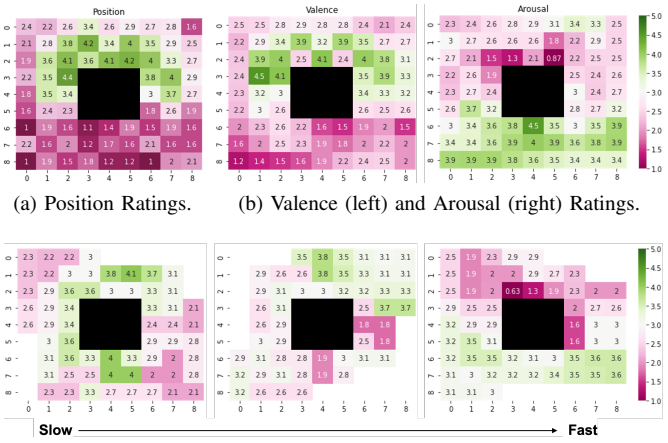
2) Participants and Assessment: After the study-design was approved by the Departmental Ethics Committee, the survey was distributed to the students at the university. All the generated images and GIFs were divided into 6 different survey forms with a total of 48 questions each. Each form had the same question structure and style, but with different images and GIFs depicting different positions on the 9×9 grid as well as different movements and speeds for the robot. Each survey form was annotated by $8 - 11$ unique participants.

B. Collected Data and Agreement Analysis

1) Robot Positioning: The average appropriateness ratings for each position can be seen in Fig. 1a. The areas directly in front, and to the left or right of the table are rated as the most appropriate when serving the customer. The further away the robot is positioned from the table, the less appropriate it is considered. Furthermore, the area behind the table is rated as less appropriate, suggesting the participants prefer the robot to keep a reasonable distance and not come too close. This is in line with the suggestions made by Mead *et al.* [18] who suggest that robots should maintain a reasonable 'social distance' from the participants in HRI settings.

2) Affective Responses: The participants only annotated every alternate adjacent position in terms of affective responses towards the robot in that position to reduce the time taken to complete the survey. These ratings were interpolated to provide a rating for each grid-position. As seen in Fig. 1b, *valence* ratings (left) closely reflected the appropriateness ratings with positions closer and in front of the table resulting in the highest ratings. For arousal (right), however, the positions with high valence and appropriateness ratings received lower overall

[1]https://forms.gle/mE5FEBBJS133Y36KA



(a) Position Ratings. (b) Valence (left) and Arousal (right) Ratings.

(c) Speed Ratings.

Fig. 1: Participants ratings for (a) Appropriateness of Robot Positioning, (b) Affective Responses to Robot Positioning and (c) Appropriateness of the Speed of Robot Movements.

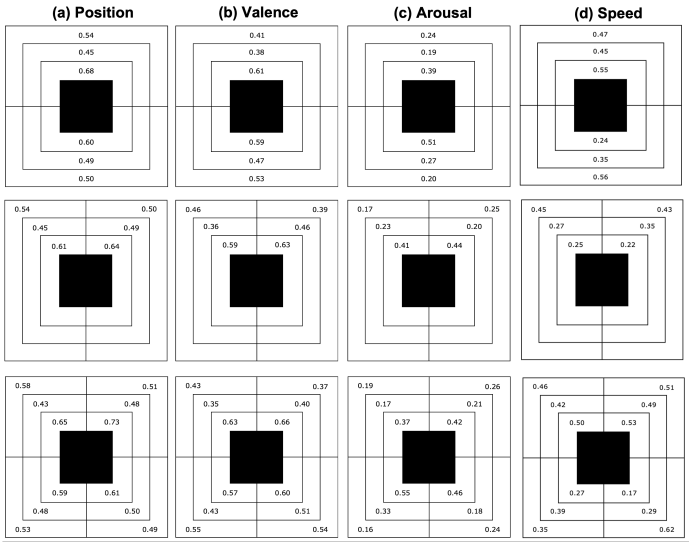


Fig. 2: Fleiss Kappa (κ) Scores for (a) Position, (b) Valence, (c) Arousal and (d) Speed Ratings. Scores are computed by splitting up the 9×9 grid horizontally (top), vertically (middle) and into quadrants (bottom).

scores which may be associated with feelings of calmness and safety [45] in those positions. In contrast, the participants reported higher arousal for positions immediately behind the table, hinting at feelings of stress or discomfort [45].

3) Speed of Movement: For collecting participant ratings for the appropriateness of the robot's speed of movement, multiple GIFs, with different frame-rates, were generated depicting the robot moving between two given positions. Fig. 1c presents the average appropriateness ratings provided for positions in the grid traversed by the robot following a speed ranging from *slow* to *fast*. As can be seen, the participants preferred slower speeds when the robot was closer to the table, and higher speeds when further away from the table.

4) Agreement Analysis: To evaluate the quality of the data collected and assess agreement levels in participants' ratings, we compute reliability scores for collected ratings using *Fleiss' Kappa (κ)* [46] values. Fleiss' κ measures how reliable ratings from two or more raters are by measuring the amount of agreement between the scores. For the position ratings, on average, we observe $\kappa = 0.55$, while for speed ratings, we get $\kappa = 0.46$. Similarly, for valence and arousal ratings we obtain κ values of 0.40 and 0.48, respectively. These values indicate a *moderate*-to-*substantial* agreement between the participants. For an in-depth analysis of these scores, we split the 9×9 grid representing the room settings into several sections and individually compute κ for these sections (see Fig. 2).

Robot Positioning: For position scores (see Fig. 2a), we observe a *moderate*-to-*substantial* agreement across the entire grid with a higher agreement for positions near the table (the black region), both front and behind (Fig. 2(a) top), and left and right (Fig. 2(a) middle) of the table, validating the scores witnessed in Fig. 1a. When split into quadrants (Fig. 2(a) bottom), we see the highest agreement scores for positions close to the top-right of the table.

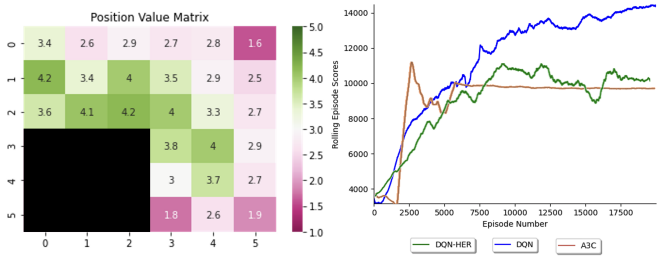
Affective Responses: For *valence* and *arousal* evaluations we see a similar trend where high agreement is witnessed for positions near the table (see Fig. 2(b,c)). However, even though a higher agreement is witnessed for *valence* scores in front of the table, *arousal* scores witness a higher agreement for positions behind the table. This is in line with the scores witnessed in Fig. 1b where positions behind the table invoked stronger responses in the participants, resulting in high arousal ratings. When split into quadrants, *valence* scores (Fig. 2b bottom) witness a high agreement for positions to the top-right of the table while *arousal* scores (Fig. 2c bottom) witness a high agreement for positions at the bottom-left of the table.

Speed of Movement: For *speed* evaluations, we do not observe high agreement scores in any sections of the grid (see Fig. 2(d)). While there is a moderate agreement for positions in front and close to the table, we also get such agreement scores for positions behind and far from the table.

Overall, despite a *moderate*-to-*substantial* agreement for most sections of the grid, participants differed in their evaluation of the agent's behaviour. This motivates us to investigate whether considering real-time user feedback to improve robot behaviour may improve their perceptions of the robo-waiter.

C. Model Training

To learn socially appropriate behaviours for the robo-waiter, we use the crowd-sourced data to train an RL agent. Filtering different grid sections based on the reliability scores received (see Fig. 2), as well as owing to experiment-room constraints for the in-person user-study (see Fig. 4), we model the environment for the RL agent using only the top-right quadrant of the grid. The action-space for the agent is consisted of $4\times3 = 12$ different actions to be selected as a combination of different speeds *(faster, same, slower)* and different movement options *(up, down, left, right)*. In each state, defined by the



(a) Position-wise scores. (b) Rolling Episode Scores.

Fig. 3: Training dynamics for *DQN, DQN+HER and A3C* trained on (a) Appropriateness ratings from the web-based survey. Models are compared on the basis of the (b) Rolling episode scores at the end of $20,000$ episodes.

grid-position and current speed, it chose from one of the 12 actions and was rewarded (see Eq. 1) using the average crowd-sourced appropriateness ratings (see Fig. 3a).

$$\mathcal{R}(s_t, a) = e^{M_p[s_t]} \tag{1}$$

where s_t is the current state, a is the action taken by the agent and M_p is the matrix of average appropriateness ratings.

We compared three popular RL methods, namely, Deep Q-Learning (DQN) [47], DQN with Hindsight Experience Replay [48] (DQN+HER) and Asynchronous Advantage Actor-Critic (A3C) [49] to learn socially appropriate robot behaviours. We implemented our grid-based environment as a custom OpenAI Gym [2] environment, realising the dynamics of our restaurant settings. The RL models were implemented adapting open-source PyTorch implementations[3].

Fig. 3b shows the rolling episode scores for $20,000$ episodes where the 3 strategies are trained on the same environment. DQN+HER requires a goal/destination to be selected and stored in the replay buffer along with the agent's experience. For this, we modify the environment to allow for the selection of random destinations during training for the DQN+HER agent. The episode terminates when the agent either reaches the maximum permissible step-count ($s_{count} = 250$) or if it reaches the destination. The DQN agent performs the best, accumulating the highest episode score at the end of $20,000$ episodes. DQN+HER training, however, showed relatively unstable reward dynamics that may be due to the early termination of episodes as the agent reached the randomly defined destination. As no reward is achieved for reaching this arbitrarily defined goal, the model struggled to learn the optimal policy in a short time. The A3C model was the quickest to converge, however, it repeatedly converged to a sub-optimal solution. A grid-search-based optimisation of hyper-parameters was performed, resulting only in shorter convergence times with no substantial improvements in final episode scores for the three models.

[2]https://gym.openai.com
[3]https://bit.ly/39tfMNU

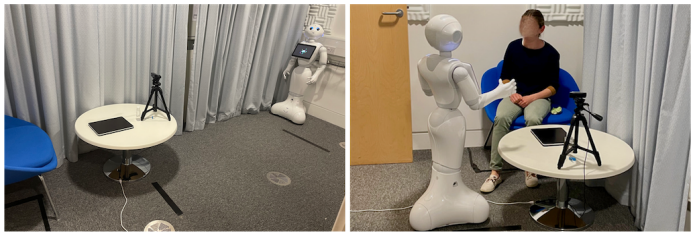


Fig. 4: *Left:* Room setup for the study. *Right:* Participant experiencing an interaction with the HSR.

V. HRI Study with Adaptive Learning

After training the DQN model on crowd-sourced data, we implemented it on the Pepper robot to evaluate the learnt behaviours using an in-person HRI study. As only *moderate* agreement was observed in participant ratings for robot behaviour (positioning and movement), we included adaptive feedback mechanisms (explicit or implicit) to extend the agents' learning and investigated whether using such mechanisms affect participants' perceptions of the robot's behaviour.

A. Materials and Methods

1) Setup: The experiment (see Fig. 4) was set-up with a low-table and chair in one corner of a well-lit room. Pepper was able to navigate the entire room as the participants remained seated on the chair at all times. A camera was placed on the low-table to record the participants' facial expressions and speech along with a tablet to be used by them to fill survey forms. Key elements of a usual restaurant setting (drinks or food) were missing due to Covid-19 restrictions in place.

2) Participants: The user study was conducted with a total of $N = 21$ participants (9 female, 12 male) recruited amongst students at the university. The majority of the participants ($N = 18$) were members of the Department of Computer Science, having some familiarity with robots and technology. All participants provided *informed consent* for their participation as well as on how the data collected during the experiments were to be used. The consent forms as well as the design and experiment protocol of the user-study were approved by the Departmental Ethics Committee. The participants were compensated in the form of Amazon vouchers.

3) Robotic Platform: We use the Pepper robot by SoftBank Robotics as it has been used in several HSR settings, making it an appropriate choice for our experiments. The humanoid features assist in making the robot seem welcoming and more human-like for such interactions [50]. Additionally, to make the interactions seem more naturalistic [51], the robot selected certain animations for its upper-body which played while the robot-waiter interacted with the participants. These animations were randomly pre-selected from a set of animations chosen for their similarity to expected movements during such an interaction, and kept consistent across all interactions. For example, Pepper gestured towards the participant to indicate that it was speaking to them, or waved when introducing itself.

4) Experiment Conditions: The user-study was conducted as a within-subjects study where each participant witnessed 4 different conditions (C1−C4) with the ordering of these conditions randomised for each participant. The participants were not informed of any differences in the conditions to avoid priming their responses and behaviours. Under each condition, the robot held a series of interactions with the participant designed to replicate interactions between a customer and a waiter in a restaurant. The 4 conditions were as follows:

C1−Control: Under the control condition, the robot followed a pre-selected random behaviour by moving (with an arbitrarily decided speed) and positioning itself randomly with respect to the participant without modifying its behaviour at all during the interactions. This allowed us to evaluate whether learning socially appropriate behaviours was actually desirable.

C2−Pre-trained: Under this condition, the robot was embedded with a behaviour (*positioning* and *speed* of movement) model pre-trained on the crowd-sourced data. The pre-trained DQN agent was used as it performed the best amongst the compared approaches (see Section IV-C). Unlike C1, the robot did not return to the same position each time but used the pre-trained RL model to determine its behaviour. Again, the robot did not modify its behaviour during the interactions and each time followed the pre-trained RL model.

C3−Explicit Feedback: Here, using the pre-trained DQN model as the starting point, the robot adapted its behaviour using the verbal feedback provided by participants as an evaluation of its behaviour. The robot utilised a simplistic 'keyword-spotting' method while listening to the participant.

C4−Implicit Feedback: Under this condition, the robot determined *valence* values from the facial expressions of the participants using the FaceChannel [52], [53], an *off-the-shelf* facial affect recognition model. Depending upon the positive or negative *valence* values, the robot was rewarded accordingly, allowing it to adapt its behaviour during the interactions.

5) Experimental Protocol: Before the experiment, the participants were explained how the robot will interact with them across 4 rounds over the course of ≈ 45 minutes. At the end of each round, they were asked to provide their evaluation, using the tablet placed in front of them, based on their experience interacting with the robot during that round. At the start of the experiment, the robot held an introduction round with the participants which was used to record baseline *arousal* and *valence* scores for the participant. Pepper, acting as the robo-waiter, interacted with the participants under 4 separate tasks, reflecting interactions a customer might experience with a waiter in restaurants:

i) Pepper welcoming the participant to 'Pepper's Diner'.
ii) Pepper taking a 'food' order from the participant.
iii) Pepper returning to the participant with their order.
iv) Pepper returning to collect the dishes.

6) Adapting to User Feedback: After each task, the robot asked the participants for feedback on its behaviour. This feedback, explicit for C3 by asking, *"How did you find the service just now?"* and implicit for C4 using the rolling average of the *valence* values estimated over the last 3 seconds, provided rewards for the robot under these conditions. To ensure the 4 conditions were identical, feedback was sought in C1 and C2 as well but not used to update the model.

Explicit Reward: Explicit reward was determined using 'keyword spotting' on the participants' responses. Based on the keywords, the robot was given a positive $= 10$ (for *Yes, Good, Great, Yeah, Perfect, Sure*), neutral $= 1$ (for *Sort of, Fine, Alright, OK, Adequate, Acceptable*) or negative $= -10$ (for *Nope, No, Not, Bad, Too Close, Too Far*) reward for its behaviour, updating the model for the subsequent interactions.

Implicit Reward: Implicit reward was computed using deviations in *valence* values ($V \in [-100, 100]$) determined from participants' facial expressions compared to the *introduction* round. These values were normalised to $V_n \in [-10, 10]$, using eq. 2 to make it comparable to explicit reward values.

$$V_n = 20.0 \times \frac{V - V_{\min}}{V_{\max} - V_{\min}} - 10.0 \tag{2}$$

where V represents the rolling average of *valence* values computed over the last 3 seconds ($3 \times 30 = 90$ frames), with $V_{\min}$ and $V_{\max}$ representing the minimum and maximum deviation detected from the baseline readings during this time.

Combined Reward: The reward function for C3 and C4 was defined by combining the reward computed for the *positioning* of the robot, following the position-wise scores computed using the crowd-sourced data (see Fig. 3), as well as the *implicit* or *explicit* reward using the following equation:

$$\mathcal{R}(s_t, a) = M_p[s_t] + e^{\mathcal{F}_t} \tag{3}$$

where s_t is the current state of the robot, a is the action taken, M_p is the matrix of average crowd-sourced appropriateness ratings and $\mathcal{F}_t$ is the explicit or implicit reward.

For the adaptive (C3 and C4) conditions, the robot started the interaction following the pre-trained model while subsequent interactions followed an ϵ-greedy ($\epsilon = 0.2$; determined following a grid-search) approach to learn optimal behaviour.

7) Implementation: For speech recognition, the SpeechRecognition [54] python library was used which uses the Google Cloud-based ASR API. This was found to be more accurate and reliable than the in-built speech recognition software of Pepper. For analysing facial affect, the FaceChannel [52], [53] was used, providing a lightweight off-the-shelf solution. Working with a dimensional model meant we were able to compare these values against baseline participant behaviour to compute the implicit reward for the model. We only use the valence values as they are sufficient to quantify the positive or negative experience of the customers [55]. Furthermore, as the crowd-sourced ratings for *arousal* received low agreement scores for the top-right quadrant (see Fig 2)), these scores were ignored.

8) Questionnaires: After each condition, that is after the robot completed all of the 4 tasks (see Section V-A5), the participants were asked to fill out a questionnaire evaluating their experience interacting with the robot. This questionnaire consisted of a subset of the Game Experience Questionnaire (GEQ) [56] measuring the *enjoyment* levels for the participants

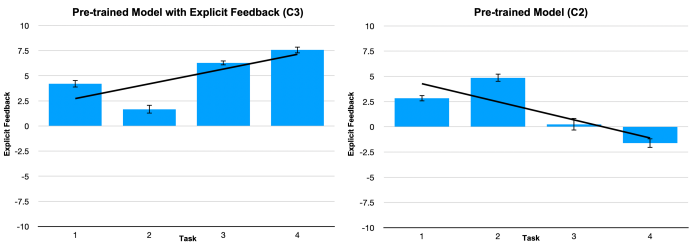


Fig. 5: Average explicit feedback values across the 4 tasks for C3 (left) and C2 (right).

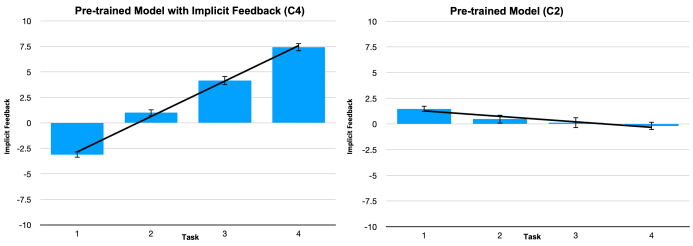


Fig. 6: Average implicit feedback values across the 4 tasks for C4 (left) and C2 (right).

due to their interaction with the robot. It also included questions measuring the participants' impressions on the *sociability* (evaluating politeness and human-like qualities), *adaptability* (sensitivity towards the participants' behaviour) and *appropriateness* (positioning, speed and approach) of robots' behaviour.

B. Results and Data Analysis

The results from the user-study are divided into *objective* and *subjective* feedback categories. The objective feedback constitutes measurements of *explicit* or *implicit* feedback provided by the participants evaluating the robot's behaviour. As one condition sought explicit user feedback while another focused on implicit feedback, in the interest of consistency, both forms of feedback were collected across all conditions, even when these were not used to adapt robot behaviour. Due to technical issues during the study, 2 participants were excluded while analysing the objective feedback. Subjective evaluations were collected using the different questionnaires filled by all the participants ($N = 21$) after each condition.

1) Quantitative Results:

a) Objective Feedback: Objective feedback was collected from the participants in the form of explicit verbal responses as well as implicit observation of how their facial affect (*valence*) changed during the interaction in response to the robot's behaviour. The positive or negative nature of these measurements provides an understanding of how appropriate and satisfactory they found the robot's behaviour.

Both implicit and explicit feedback was computed for each condition to provide insights into how the Explicit Feedback (C3) and Implicit Feedback (C4) conditions improved on the Pre-trained Model (C2), even though these values were not used to update the model under C2. Comparing C3 and C2, a one-tailed Mann-Whitney U Test [57] showed that the average explicit feedback (the reward value computed using keyword spotting) was significantly higher ($U = 88, p = 0.010$) for C3. Furthermore, as each condition consisted of the Pepper performing 4 tasks, we observe that these values increase for C3 (see Fig. 5) as the participants witnessed more interactions while decreasing for C2. Similarly, when comparing C4 and C2, a one-tailed Mann-Whitney U Test showed that the average implicit feedback (computed using eq. 2) was significantly higher ($U = 87.0, p = 0.015$) for C4. Similar to the explicit feedback, as the participant interacted with Pepper across the 4 tasks, these values steadily increase for C4 while decreasing for C2, as seen in Fig. 6.

In general, we see that a higher magnitude of explicit rewards were provided by the participants during C3 while implicit awards were dominant during C4 (see Section II of the supplementary material provided for the respective plots).

b) Subjective Feedback: To evaluate our research questions (see Section III), participants' responses to the questionnaire recording their impressions of robot behaviour in terms of *enjoyment*, *sociability*, *adaptability* and *appropriateness* were compared for the 4 conditions using a pairwise Wilcoxon Signed-Rank Test [58] (see Fig. 7).

RQ1: When comparing the robot following the RL model pre-trained on crowd-sourced data (C2) with a random static behaviour policy (C1), the robot under C2, on average, received improved ratings across all evaluations. In particular, it was rated significantly more impressive ($W = 36.0, p = 0.044$) and significantly higher in terms of the appropriateness of the speed of robot's movements ($W = 16.5, p = 0.005$).

RQ2: When evaluating whether receiving explicit feedback from the participants (C3) improved their perceptions of the robot compared to following the pre-trained model (C2), C3 was found to be rated higher across all dimensions. In particular, the robot under C3 was rated significantly less tiresome ($W = 37.0, p = 0.035$) and more sociable ($W = 21.5, p = 0.022$), better understood what the participant said ($W = 36.0, p = 0.013$) and adapted to what they said and did ($W = 14.0, p = 0.001$). C3 was also rated significantly higher for the positioning ($W = 22.5, p = 0.008$) of the robot.

RQ3: Similarly, on comparing adapting with implicit feedback (C4) to using the pre-trained model (C2), the robot under C4 improved upon the scores received for C2 across all dimensions. It was found to be rated significantly higher on its ability to adapt to what the participant said ($W = 27.5, p = 0.008$) as well as on the appropriateness of the positioning of the robot ($W = 33.5, p = 0.011$).

RQ4: Finally, comparing the explicit (C3) *vs.* implicit adaptation (C4) strategy, although no clear preference was discovered for the participants across most dimensions, C3 did receive significantly better results in a number of dimensions. It was rated significantly less tiresome ($W = 3.5, p = 0.016$) and more sociable ($W = 32.5, p = 0.017$), better understood what the participant said ($W = 105.0, p = 0.023$) and better adapted to what the participant said and did ($W = 81.0, p = 0.028$).

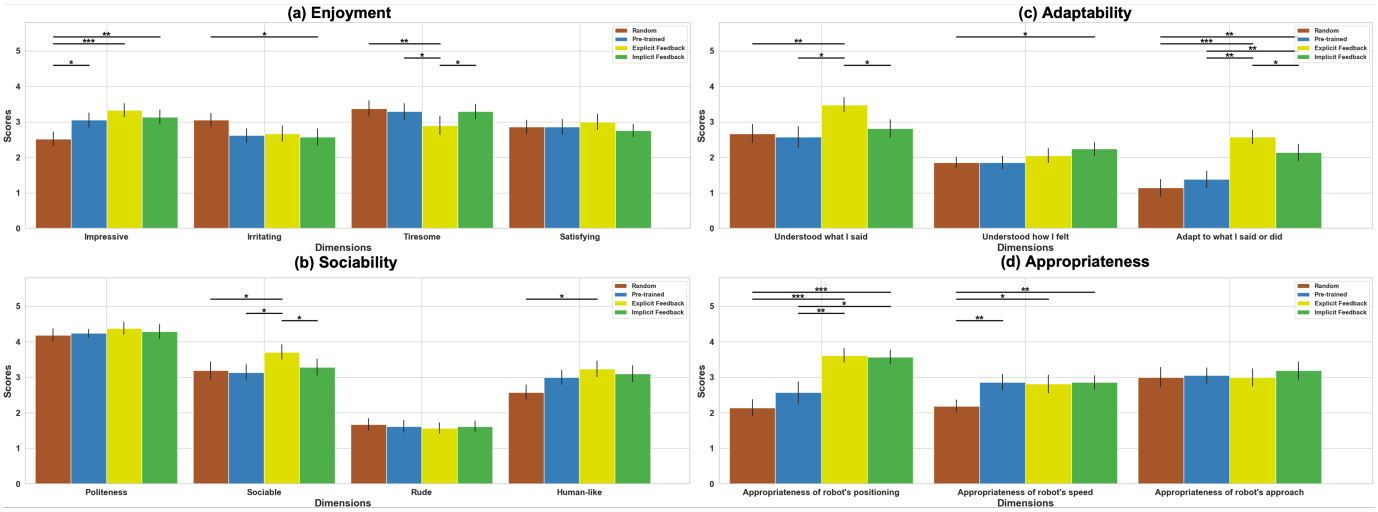


Fig. 7: Participants' evaluations of the robo-waiter across (a) *Enjoyment*, (b) *Sociability*, (c) *Adaptability* and (d) *Appropriateness* dimensions. * represents $p < 0.05$, ** represents $p < 0.01$, and *** represents $p < 0.001$.

2) Qualitative Results: To evaluate the qualitative experience of the participants interacting with the robo-waiter, semi-structured interviews were conducted at the end of the sessions. Despite having to repeatedly give feedback ($4 \times 4 = 16$ times), none of the participants found the interactions tedious. All of them said that they could imagine being served by a robo-waiter in the future. Several participants, however, highlighted aspects of the robot's behaviour that they did not appreciate. For instance, they found the robot approaching them "face-on" to be *frightening*, especially when it moved straight towards them. This is in line with the *compensation* model of proxemic behaviour [17] with the participants expressing discomfort over such robot behaviour. One participant mentioned their discomfort when "it came up to them with its eyes glowing green" while another compared its movement to a "doll in a horror movie". However, the same participants enjoyed how the robot moved its upper-body and hands during the interactions describing it as "very human-like". Three participants mentioned that the robot did not seem to listen to them, some other ($N = 6$) found the entire interaction boring or repetitive while a couple others suggested that inclusion of real food might have improved their interaction experience.

VI. Summary and Conclusions

To the best of our knowledge, this is the first work investigating different adaptation strategies for learning appropriate approach and positioning behaviour for HSRs in real-time HRI settings. The main outcomes of this work are summarised as:

1) Using an RL model pre-trained on crowd-sourced data improves *sociability*, *enjoyment*, *appropriateness* of robot's behaviour compared to following a random policy.
2) Adaptation using explicit or implicit user-feedback improves *appropriateness* and *sociability* ratings compared to using a model pre-trained only on crowd-sourced data.
3) Modelling adaptation using an explicit feedback strategy results in improved *enjoyment*, *sociability* and *adaptability* ratings, compared to using implicit feedback.

These findings highlight the need to expand real-time social interaction capabilities for HSRs, enabling them to adapt their behaviour in response to user-feedback. Pre-training the behaviour model using crowd-sourced data improves the participants' interaction experience as learning from such a *generalised* understanding of user preferences improves how the robot interacts with the participants. Following an explicit feedback strategy results in an improved performance evaluation compared to using implicit feedback. This may be due to explicit feedback being more targeted and direct. Implicit feedback may get convoluted with other interaction factors such as the participants' politeness or their interest in the robot.

Our results also show that an adaptive agent is preferred, in line with other works [59] where the robot was pre-trained based on responses to a survey completed by the participants before the interaction but did not adapt its behaviour in real-time. Our work demonstrates that not only do participants find the adaptive robot to be more *enjoyable*, *sociable*, *adaptable* and *appropriate*, but they also interact more positively with the robot, providing it with positive implicit and explicit feedback.

A. Limitations and Future Work

The implicit feedback strategy explores only *valence* estimations from facial expressions, which may not be sufficient to exactly capture the participants' feedback as such evaluations may get convoluted with other interaction factors. Thus, future work should explore a multi-modal analysis of user behaviour in terms of both extrinsic (*audio-visual signals*) and intrinsic (*biosignals*) factors to better capture participant reactions and feedback to robot actions [29]. Furthermore, due to COVID-19 restrictions, only department members were allowed to participate in the study. This limits the diversity amongst the participants. Future works should explore if the findings in this study hold within a more realistic restaurant environment, and with a wider and more diverse user demographics.

REFERENCES

[1] M. Mende, M. L. Scott, J. van Doorn, D. Grewal, and I. Shanks, "Service robots rising: How humanoid robots influence service experiences and elicit compensatory consumer responses," *Journal of Marketing Research*, vol. 56, no. 4, pp. 535–556, 2019.

[2] S. Curtis. (2016) Pizza hut hires robot waiters to take orders and process payments at its fastfood restaurants. mirror. [Online]. Available: https://www.mirror.co.uk/tech/pizza-hut-hires-robot-waiters-8045172

[3] W. Scholl, "The socio-emotional basis of human interaction and communication: How we construct our social world," *Social Science Information*, vol. 52, no. 1, pp. 3–33, 2013.

[4] R. Kirby, J. Forlizzi, and R. Simmons, "Affective social robots," *Robotics and Autonomous Systems*, vol. 58, no. 3, pp. 322–332, 2010.

[5] G. R. Collins, "Improving human–robot interactions in hospitality settings," *International Hospitality Review*, 2020.

[6] S. Langhorn, "How emotional intelligence can improve management performance," *International Journal of Contemporary Hospitality Management*, 2004.

[7] D. Nickson, C. Warhurst, and E. Dutton, "The importance of attitude and appearance in the service encounter in retail and hospitality," *Managing Service Quality: An International Journal*, 2005.

[8] J. Markoff. (2014) 'beep,' says the bellhop. the new york times. [Online]. Available: https://www.nytimes.com/2014/08/12/technology/hotel-to-begin-testing-botlr-a-robotic-bellhop.html

[9] M. K. Lee, S. Kiesler, J. Forlizzi, and P. Rybski, "Ripple effects of an embedded social agent: a field study of a social robot in the workplace," in *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, 2012, pp. 695–704.

[10] S. Herse, J. Vitale, D. Ebrahimian, M. Tonkin, S. Ojha, S. Sidra, B. Johnston, S. Phillips, S. L. K. C. Gudi, J. Clark *et al.*, "Bon appetit! robot persuasion for food recommendation," in *Companion of the ACM/IEEE Int'l Conf. on Human-Robot Interaction*, 2018, pp. 125–126.

[11] Z. Jie and H. Gunes, "Investigating taste-liking with a humanoid robot facilitator," in *Proceedings of IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2020, pp. 1–6.

[12] S. R. Schmidt-Rohr, M. Losch, and R. Dillmann, "Human and robot behavior modeling for probabilistic cognition of an autonomous service robot," in *RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2008, pp. 635–640.

[13] G. Sawadwuthikul, T. Tothong, T. Lodkaew, P. Soisudarat, S. Nutanong, P. Manoonpong, and N. Dilokthanakul, "Visual goal human-robot communication framework with few-shot learning: a case study in robot waiter system," *IEEE Transactions on Industrial Informatics*, 2021.

[14] H. Cramer, J. Goddijn, B. Wielinga, and V. Evers, "Effects of (in) accurate empathy and situational valence on attitudes towards robots," in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2010, pp. 141–142.

[15] J. Avelino, L. Garcia-Marques, R. Ventura, and A. Bernardino, "Break the ice: a survey on socially aware engagement for human–robot first encounters," *International Journal of Social Robotics*, pp. 1–27, 2021.

[16] M. Paetzel, G. Perugia, and G. Castellano, "The persistence of first impressions: The effect of repeated interactions on the perception of a social robot," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020, pp. 73–82.

[17] J. Mumm and B. Mutlu, "Human-robot proxemics: Physical and psychological distancing in human-robot interaction," in *Proceedings of the 6th International Conference on Human-Robot Interaction (HRI)*. Lausanne, Switzerland: ACM, 2011, p. 331–338.

[18] R. Mead and M. J. Matarić, "Autonomous human–robot proxemics: socially aware navigation based on interaction potential," *Autonomous Robots*, vol. 41, no. 5, pp. 1189–1201, Jun. 2016.

[19] J. Tjomsland, S. Kalkan, and H. Gunes, "Mind your manners! a dataset and a continual learning approach for assessing social appropriateness of robot actions," *arXiv preprint arXiv:2007.12506*, 2020.

[20] Y. Gao, F. Yang, M. Frisk, D. Hemandez, C. Peters, and G. Castellano, "Learning socially appropriate robot approaching behavior toward groups using deep reinforcement learning," in *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2019, pp. 1–8.

[21] D. Jan and D. R. Traum, "Dynamic movement and positioning of embodied agents in multiparty conversations," in *Proceedings of the 6th International Joint conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2007, pp. 1–3.

[22] C. Pedica and H. Vilhjálmsson, "Social perception and steering for online avatars," in *International Workshop on Intelligent Virtual Agents*. Springer, 2008, pp. 104–116.

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

[24] 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*, 2018, pp. 154–162.

[25] J. M. Garcia-Haro, E. D. Oña, J. Hernandez-Vicen, S. Martinez, and C. Balaguer, "Service robots in catering applications: A review and future challenges," *Electronics*, vol. 10, no. 1, p. 47, 2021.

[26] C. S. Barathi, "Lie detection based on facial micro expression body language and speech analysis," *International Journal of Engineering Research & Technology*, 2016.

[27] G. Li, H. Dibeklioğlu, S. Whiteson, and H. Hung, "Facial feedback for reinforcement learning: a case study and offline analysis using the tamer framework," *Autonomous Agents and Multi-Agent Systems*, vol. 34, no. 1, pp. 1–29, 2020.

[28] L. J. Corrigan, C. Basedow, D. Kïster, A. Kappas, C. Peters, and G. Castellano, "Mixing implicit and explicit probes: finding a ground truth for engagement in social human-robot interactions," in *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2014, pp. 140–141.

[29] H. Gunes, B. Schuller, M. Pantic, and R. Cowie, "Emotion representation, analysis and synthesis in continuous space: A survey," in *Face and Gesture 2011*. IEEE, 2011, pp. 827–834.

[30] L. Tian and S. Oviatt, "A taxonomy of social errors in human-robot interaction," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 10, no. 2, pp. 1–32, 2021.

[31] E. Cambria, D. Das, S. Bandyopadhyay, and A. Feraco, "Affective computing and sentiment analysis," in *A practical guide to sentiment analysis*. Springer, 2017, pp. 1–10.

[32] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, "A review of affective computing: From unimodal analysis to multimodal fusion," *Information Fusion*, vol. 37, pp. 98–125, 2017.

[33] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *JPSP*, vol. 17 (2), p. 124, 1971.

[34] J. A. Russell, "A circumplex model of affect." *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

[35] N. Churamani, F. Cruz, S. Griffiths, and P. Barros, "iCub: Learning Emotion Expressions using Human Reward," in *Workshop on Bio-inspired Social Robot Learning in Home Scenarios, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016.

[36] N. Churamani, M. Kerzel, E. Strahl, P. Barros, and S. Wermter, "Teaching emotion expressions to a human companion robot using deep neural architectures," in *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, May 2017, pp. 627–634.

[37] C. Arzate Cruz and T. Igarashi, *A Survey on Interactive Reinforcement Learning: Design Principles and Open Challenges*. New York, NY, USA: Association for Computing Machinery, 2020, p. 1195–1209.

[38] F. Cruz, S. Magg, C. Weber, and S. Wermter, "Training agents with interactive reinforcement learning and contextual affordances," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 8, no. 4, pp. 271–284, 2016.

[39] S. Griffith, K. Subramanian, J. Scholz, C. L. Isbell, and A. L. Thomaz, "Policy Shaping: Integrating Human Feedback with Reinforcement Learning," in *Advances in Neural Information Processing Systems*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, Eds., vol. 26. Curran Associates, Inc., 2013.

[40] N. Churamani, P. Barros, H. Gunes, and S. Wermter, "Affect-driven modelling of robot personality for collaborative human-robot interactions," *arXiv:2010.07221*, 2020.

[41] A. L. Thomaz, G. Hoffman, and C. Breazeal, "Real-time interactive reinforcement learning for robots," in *AAAI 2005 workshop on human comprehensible machine learning*, 2005.

[42] S. K. Kim, E. A. Kirchner, A. Stefes, and F. Kirchner, "Intrinsic interactive reinforcement learning–using error-related potentials for real world human-robot interaction," *Scientific reports*, vol. 7, no. 1, pp. 1–16, 2017.

[43] Y. Mizuchi and T. Inamura, "Optimization of criterion for objective evaluation of hri performance that approximates subjective evaluation: a case study in robot competition," *Advanced Robotics*, vol. 34, no. 3-4, pp. 142–156, 2020.

[44] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of behavior therapy and experimental psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[45] M. D. McManus, J. T. Siegel, and J. Nakamura, "The predictive power of low-arousal positive affect," *Motivation and Emotion*, vol. 43, no. 1, pp. 130–144, aug 2018.

[46] J. L. Fleiss, "Measuring nominal scale agreement among many raters." *Psychological Bulletin*, vol. 76, no. 5, pp. 378–382, 1971.

[47] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.

[48] M. Andrychowicz, F. Wolski, A. Ray, J. Schneider, R. Fong, P. Welinder, B. McGrew, J. Tobin, P. Abbeel, and W. Zaremba, "Hindsight experience replay," *arXiv preprint arXiv:1707.01495*, 2017.

[49] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *International conference on machine learning*. PMLR, 2016, pp. 1928–1937.

[50] S. J. Stroessner and J. Benitez, "The social perception of humanoid and non-humanoid robots: Effects of gendered and machinelike features," *International Journal of Social Robotics*, vol. 11, no. 2, pp. 305–315, 2019.

[51] K. S. Lohan, H. Lehmann, C. Dondrup, F. Broz, and H. Kose, *Enriching the Human-Robot Interaction Loop with Natural, Semantic, and Symbolic Gestures*. Dordrecht: Springer Netherlands, 2016, pp. 1–21.

[52] P. Barros, N. Churamani, and A. Sciutti, "The FaceChannel: A fast and furious deep neural network for facial expression recognition," *SN Computer Science*, vol. 1, no. 6, Oct. 2020.

[53] P. Barros, N. Churamani, and A. Sciutti, "The facechannel: A lightweight deep neural network for facial expression recognition," in *Proceedings of the 15th International Conference on Automatic Face and Gesture Recognition (FG)*, 2020, pp. 652–656.

[54] A. Zhang. (2014) Speech recognition (version 3.8) [software]. [Online]. Available: https://github.com/Uberi/speech_recognition

[55] S. Shapiro and D. J. MacInnis, "Understanding program-induced mood effects: Decoupling arousal from valence," *Journal of Advertising*, vol. 31, no. 4, pp. 15–26, 2002.

[56] W. A. IJsselsteijn, Y. A. de Kort, and K. Poels, "The game experience questionnaire," *Eindhoven: Technische Universiteit Eindhoven*, vol. 46, no. 1, 2013.

[57] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other." *The Annals of Mathematical Statistics*, vol. 18, pp. 50–60, 1947.

[58] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bulletin*, vol. 1, no. 6, p. 80, Dec. 1945.

[59] A. Sekmen and P. Challa, "Assessment of adaptive human–robot interactions," *Knowledge-Based Systems*, vol. 42, pp. 49–59, apr 2013.