ARTICLE TEMPLATE

Emotion and Memory Model for Social Robots: A Reinforcement Learning based on Behavior Selection

Muneeb Imtiaz Ahmad[a], Yuan Gao[b], Omar Mubin[c], Suleman Shahid[d] and Fady Alnajjar[e]

[a]Department of Computer Science, Swansea University & The MARCS Institute, Western Sydney University, Australia; [b]Department of Information Technology, Uppsala University, Sweden; [c] School of Computing, Western Sydney University, Australia; [d]Department of Computer Science, Lahore University of Management Science, Pakistan; [e]College of Information Technology, UAE University, UAE.

ARTICLE HISTORY
Compiled October 12, 2021

ABSTRACT

In this paper, we propose a reinforcement learning (RL) mechanism for social robots to select an action based on users' learning performance and social engagement. We applied this behavior selection mechanism to extend the emotion and memory model, which allows a robot to create a memory account of the user's emotional events and adapt its behavior based on the developed memory. We evaluated the model in a vocabulary-learning task at a school during a children's game involving robot interaction to see if the model results in maintaining engagement and improving vocabulary learning across the four different interaction sessions. Generally, we observed positive findings based on child vocabulary learning and sustaining social engagement during all sessions. Compared to the trends of a previous study, we observed a higher level of social engagement across sessions in terms of the duration of the user gaze toward the robot. For vocabulary retention, we saw similar trends in general but also showing high vocabulary retention across some sessions. The findings indicate the benefits of applying RL techniques that have a reward system based on multi-modal user signals or cues.

KEYWORDS
Reinforcement learning; Social Robots; Educational Robots; Repeated Child Robot Interaction; Personalization; Children Engagement.

1. Introduction

"Reinforcement learning is the problem faced by an agent that must learn behavior through trial-and-error interactions with a dynamic environment" (Kaelbling, Littman, and Moore 1996). In essence, an agent (a robot) learns by interacting within an environment after receiving rewards for its actions.

Reinforcement learning (RL) has been applied to several fields, including robotics, personalized recommendations, and advertising. RL has also been applied in the human-computer interaction domain, especially in education technology to design and implement intelligent tutoring systems. For example, Sottilare et al. (2012) presented a *Generalized Intelligent Framework for Tutoring* (GIFT) that enables educational technologies to utilize RL to learn appropriate strategies for each user based on the user state information to maximize learning gains. The GIFT framework has also been applied in the design of different educational technologies (Kelsey et al. 2015). In addition, other studies have promoted RL in the design of educational technology. For example, an adaptive educational system that adapts teaching style by modeling the user's learning style has been proposed previously (Dorça et al. 2015).

The application of RL in human-robot interaction (HRI) and social robotics to enable robots to learn

CONTACT Muneeb Imtiaz Ahmad. Email: muneeb.ahmad@westernsydney.edu.au

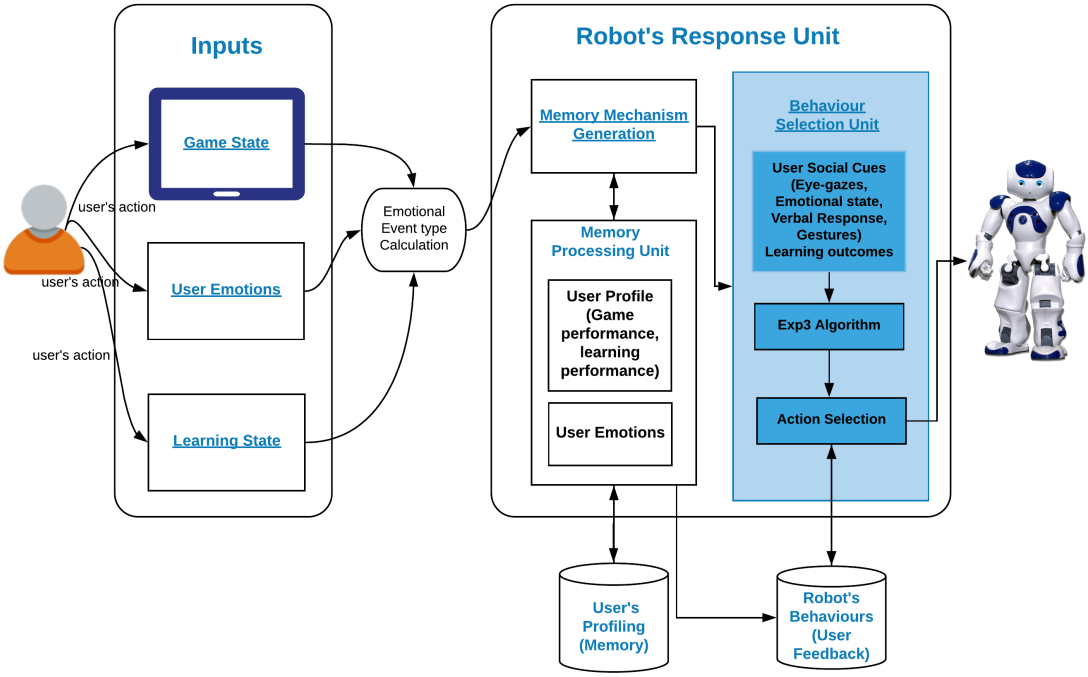


Figure 1.: Emotion and Memory model. The behavior selection unit (BSU), marked in blue, was updated by applying a reinforcement learning algorithm (Exp3). This enabled the system to select appropriate actions based on the reward computed by estimating social engagement measured by the data collected from the multi-model channels.

about the selection of appropriate actions is a growing trend. In particular, the personalization or adaptation implemented by using simple algorithms has greatly improved the effectiveness of robot tutoring compared to methods that do not consider personalization (Leyzberg, Spaulding, and Scassellati 2014). We conjecture that modeling the robot's feedback based on user characteristics, primarily the user's estimated amount of social engagement and/or learning, may promote learning performance in tutoring interactions. The rationale for this assumption is based on the findings highlighted in the literature (Haider, Sinha, and Chaudhary 2010), which state that the process of user modeling to inform teaching strategies improves learning (Liu, Graf et al. 2009). Further, it is acknowledged that the implementation of the RL-based algorithm for behavior selection in HRI would also result in other benefits, e.g., autonomy, lifelike, scalability, and more practical applications (Breazeal et al. 2003). RL enables agents to learn and adapt according to user characteristics, a dynamic approach which may not always be permissible in rule-based systems.

The work described in this paper extends the previous work; (Ahmad et al. 2019), where the robot provided positive, negative, and neutral feedback depending on the type of the emotional event (positive or negative or neutral) via a rule-based mechanism. In the mentioned prior work, the robot was programmed such that when the system detected a positive, negative, or neutral event, the robot would simply reciprocate the gesture by providing positive, negative, or neutral feedback. Note that, the previous work implemented a rule-based method and did not implement a RL-based learning mechanism to select behaviors. Therefore, in this paper, our *contributions* are the following:

- We present and demonstrate an extended version of the emotion and memory model by updating the behaviour selection unit, as shown in Figure 1, in which we introduce the RL-based learning

mechanism to learn the best strategy to select a behaviour for the given user according to their current level of social engagement and/or learning outcome.

- The RL-based learning algorithm implements a method to compute a reward function by taking inputs from the multi-modal channels. We used external cameras to capture the facial expressions, gesture, and eye gazes. Further, we used microphones to record the verbal responses. This reward function represented the current level of social engagement and was used by the Exp3 algorithm to learn and select behaviours for each user.
- This paper also presents the results of a new long-term user study performed to evaluate the updated version of the emotion and memory model, where we reused the stimulus from our previous study. The evaluation of the model showed that social engagement was sustained or increased across the four different interaction.
- We compared the trends of the results of the current study to the previous one. We noticed a higher amount of engagement in terms of eye gaze towards the robot. Further, we observed a more consistent amount of verbal response (an average increase of 11.8%) from the first to the last session for the current work.

The model was implemented on the NAO robot, which played a modified variant of the snakes and ladders game. We modified the game such that understanding RObot Interaction LAnguage (ROILA) vocabulary (Mubin et al. 2011) was an essential game component and children could study ROILA playfully and interactively. We programmed the robot to teach vocabulary because it is an integral part of language learning that helps improve language learning skills. In addition, efficient vocabulary may result in faster recognition and knowledge of grammar (Goldfield and Reznick 1990), and ROILA helps mitigate the confounding determinant of participants or learners in various linguistic settings (Mubin et al. 2011). The rationale for selecting games was based on learning theories that emphasize the value of play in learning processes (Vygotsky 1980; Rieber 1996).

We have organized the manuscript in the following manner. Section 2 briefly presents the background on the applications of RL in HRI and the role of emotions in human memory. Section 3 describes the Emotion and Memory Model. Sections 4 to 5 describe the research method used to evaluate the model and how to present the results of the evaluation. Section 6 discusses our findings.

2. Background

2.1. *Reinforcement Learning and Applications in HRI*

RL is a process that uses a reward signal to shape the behavior of an agent. It considers an agent taking an action in an environment to maximize its accumulated reward emitted by the environment. To model this process, RL utilizes the Markov Decision Process to learn the probabilities between states and actions. As one of the mathematical models capturing the essence of human-human interactive processes, RL has also been studied to help the development of systems for HRI. In such studies, researchers explored different aspects of HRI, including physical, affective, and long-term HRI problems (Ghadirzadeh et al. 2016; Leite et al. 2014; Irfan et al. 2019). Generally, RL was used to optimize a robot's behavior under various conditions. By combining social HRI and reinforcement learning, Leite et al. (2014) proposed a non-intrusive RL approach in which a robot can learn and adapt to the user based on different user preferences in real time. They applied this technique to enable a robot to choose appropriate behavior from a set of supportive behaviors during playful interactions. In addition, Castro-González, Malfaz, and Salichs (2011) applied an RL algorithm based on q-learning to enable the Maggie robot to select a correct action for each state to maximize user motivation. Moreover, Ritschel and André (2017) proposed the use of an RL algorithm to inform robot personality adaptation in real-time based on the user's social signals or cues. Their proposed method relied on the use of the social signals interpretation framework (Wagner et al. 2013) to estimate the users engagement and used it as an indicator to decide the use of the personality of the robot for each user.

Different RL algorithms have been applied to implement adaptive behavior selection to inform personalization in HRI across various domains, e.g., education. For example, Gordon et al. (2016) used an affective RL algorithm based on q-learning to determine a social robot's verbal and non-verbal behaviors in a language learning game to promote effective personalization. In addition, Ramachandran and Scassellati (2015) applied a contextual bandit algorithm that can adaptively control the pace of interactions based on user performances and effective feedback. Ramachandran and Scassellati (2014) also implemented personalization in robots using RL based on each individual's learning difficulty level, and Gao et al. (2018) proposed a RL framework that enabled a robot to select supportive behavior to maximize task performance in a game-based learning scenario. Furthermore, we observe an increasing trend of applying different learning-based mechanisms in the HRI field (Jones, Bull, and Castellano 2017; Jones and Castellano 2018). Similarly, RL models are being applied to inform better robotic tutors (Roy et al. 2018). The findings from these studies have resulted in influencing positive attitudes in individuals and have promoted task performance or improved learning.

Based on these findings, we have applied an RL algorithm to automate the decision-making process when selecting a robot's social behavior depending on the emotion and memory model (Ahmad et al. 2019). Furthermore, our previous study (Ahmad et al. 2017) highlighted the benefits of positive and negative feedback during child-robot tutoring interactions in a long-term configuration. We apply the multi-armed bandit (MAB) problem, where we omit the state differences meaning $s_t = s_{t+1}$. To the best of our knowledge, algorithms to solve MAB were first discussed in the social robotics field (Leite et al. 2014) to optimize robot behavior in a chess-teaching scenario. Then, additional studies utilized them for instructional scenarios from different perspectives such as sequence learning task where users are asked to "arrange language, thoughts, information, and actions in an effective order" (Tsiakas et al. 2018; Tozadore et al. 2018). However, to the best of our knowledge, such methods have not been applied to emotion and memory models together in the social robot context.

2.2. *The Role of Emotions in Human Memory*

Existing literature suggests that emotions aid memory (Canli et al. 2000; Cahill et al. 1996). We observe that human memory is influenced by emotions, and previous studies have claimed that central information is enhanced by emotions at the expense of external specifications, where central information refers to the information fundamental to the experience or event that evokes the emotion (Levine and Pizarro 2004). In particular, previous findings have shown that events inducing positive or negative emotions are remembered better compared to neutral events (Canli et al. 2000; Cahill et al. 1996). It was found that positive and negative emotions affect human memory in different ways (Levine and Burgess 1997). In addition, research has shown that pleasant or positive emotions are remembered better than negative or unpleasant emotions (D'Argembeau, Comblain, and Van der Linden 2003; Comblain, D'Argembeau, and Van der Linden 2005). Moreover, these effects have been reported in people with Alzheimer's (Kazui et al. 2000). Furthermore, previous studies have confirmed that emotions aid both short- and long-term memory (Bradley et al. 1992; Dolcos, LaBar, and Cabeza 2005).

From a theoretical perspective, the emotional modulation of memory theory also suggests that emotions positively impact the human memory system (Baumeister et al. 2007), which has been confirmed by various studies using positron emission tomography and functional magnetic resonance imaging. These studies have also shown that emotionally-charged events enhance human memory (Cahill et al. 1996; Dolcos, LaBar, and Cabeza 2005).

Our research also attempts to map the process in which humans' store or retrieve the memory of different emotional events on to the robot by following the taxonomy given by Levine and Pizarro (2004). Note that previous findings have shown that information attributed to an emotional state or event is remembered better compared to a neutral one (Christianson and Loftus 1991). Considering this finding, we expect that when a robot recalls the memory of a user, mainly through the use of the *central information* stored under various emotion states, in its behavior or dialogue, this may have a positive impact on the user's engagement and learning (i.e., retention of the taught words). Hence, we try to

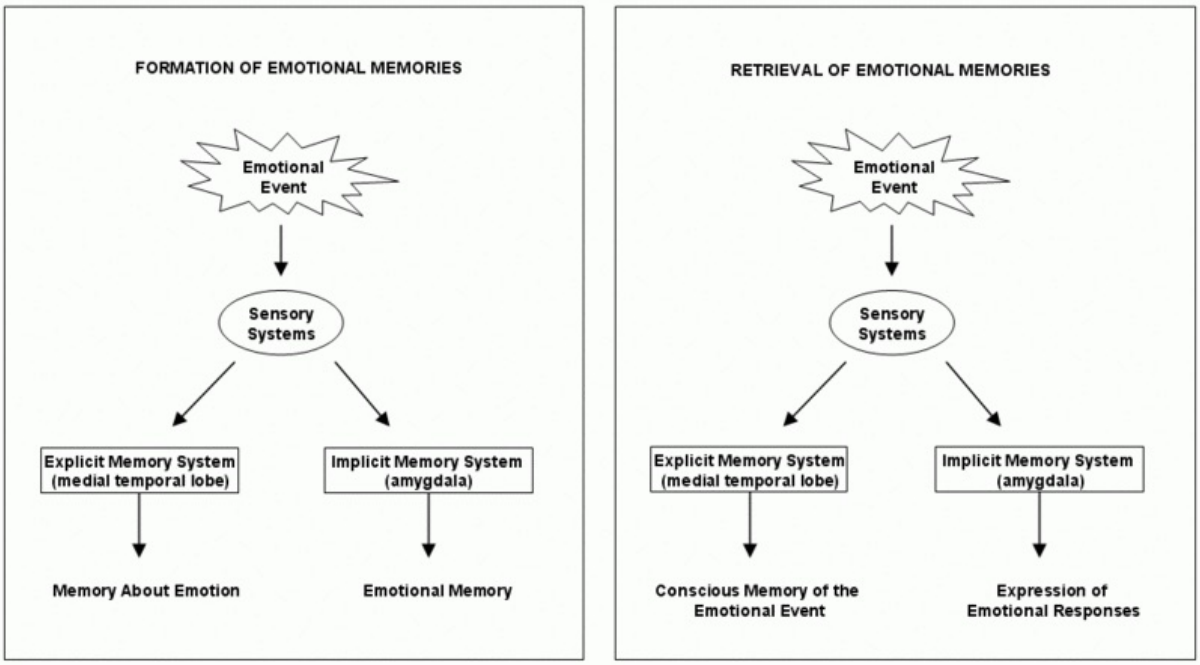


Figure 2.: Building emotional memories; (LeDoux, 2007)

apply this process in our emotion and memory model and evaluate it in terms of social engagement and vocabulary learning.

3. Emotion and Memory Model

We designed the model in reference to the formation and retrieval functionality of emotional memories (LeDoux 2007), as shown in Figure 2. The human memory system is broadly classified into short- and long-term memories. Here, we consider long-term memory because information about emotional experiences is associated with long-term memory. Long-term memories are broadly split into two parts: conscious memory (declarative memory) and unconscious memory (procedural memory). Conscious memory refers to explicit memory, and unconscious memory refers to the implicit memory system. Humans encode different emotional experiences in these memory systems. Emotional memory is a special category of memory that involves the implicit learning and storage of information about the emotional significance of events (LeDoux 1993). In contrast, memory of an emotional situation (memories about emotions) is part of conscious (i.e., explicit) memory. The creation and retrieval process of emotional memories is summarized as follow. The human sensory system processes an emotional experience or event and then transfers it to the temporal lobe to create explicit or implicit memory. Then, when an appropriate cue occurs, the memories of these emotional events can be retrieved. The retrieval of both memories occurs as the cue is processed by the sensory system. Our model for humanoid robots employs conscious, declarative, or explicit memory of both emotional situations of the frontal user and further uses this memory during repeated interactions with the same user. The model is broadly adopted from research into memory conception and retrieval. The justification for selecting the emotional events to be stored and retrieved by the robot was to make the robot more familiar, natural, and therefore a more effective and dynamic tool (Dautenhahn 1998).

To build a proper emotional events memory for the user, it is essential to distinguish positive and negative emotional events. Here, events in which goals are achieved with no critical predicaments or after minor or major difficulties are deemed positive emotional events, and events that are tagged as obstacles and loss relative to a plan or goal are considered negative emotional events. Note that events not tagged

Discrete Emotions	Motivational State	Central Information
Happiness	maintain current state; attain new goal.	broad range of information from general knowledge and the environment
Fear	avoid or escape threat of goal failure.	sources of threat; means of avoiding threat
Anger	remove obstacle to goal attainment.	goal; agents obstructing goal attainment.
Sadness	adjust to irrevocable goal failure.	outcomes and consequences of goal failure.

Table 1.: Types of information stored under one positive (happiness) and three negative emotional states (Fear, Anger, Sadness) (Levine and Pizarro 2004).

as positive or negative are considered neutral events (Bower 1992). Humans remember various types of central information under several emotions (Levine and Pizarro 2004), and any emotional state is directly mapped to its associated emotional situation (Table 1). As shown in Table 1, during happiness, humans store a range of information from the environment and general knowledge. However, in a negative emotional state, e.g., sadness, fear, and anger, different types of information are stored. For example, anger leads to storing information about agents that obstruct goal fulfillment, sadness leads to recalling the outcomes and consequences of goal failure, and fear can lead to recalling a primary threat source and ways to avoid it. Note that humans tend to show a range of emotions during the learning process such as boredom and frustration (DMello and Graesser 2012). However, currently, we only consider the emotions highlighted in Table 1 due to technical limitations to measure boredom and frustration. Hence, this is beyond the scope of our contribution. Nonetheless, both emotions may reflect negative events and in principle, can be added in the model for future applications. We further highlight that the choice of emotions (Happy, Sad, Angry, Fear) for the model is grounded in the body of work where we observed that children show these emotions and this resulted in the formation of the model (Ahmad, Mubin, and Orlando 2016a, 2017; Ahmad et al. 2019).

We designed the model (Figure 1) based on the understanding of the information stored by humans under different emotional states induced upon the occurrence of emotional events as proposed by Levine et al (2004). The first step in the creation of the emotional memory was the identification of the emotional event. Thus, to ensure that the identified event was represented correctly, we considered the following two aspects: 1) To mark an event as a positive or negative event, we checked the following parameters. First, classification was according to the definition of positive and negative events. Second, it relied on the detection of the user's emotional state at the occurrence of the event; 2) When the user did not show positive or negative emotions, we identified the event as a neutral event. We did this because typically a user does not always need to depict positive or negative emotions when trying to reach the goal. It may depend on the task because the task can be too easy to accomplish. Once identified, we used Levine's taxonomy (see Table 1) to create the memory and allowed the social robot to use this created memory in its dialog and behave accordingly. Finally, the created memory was selected according to the afore-described retrieval process for an emotional event during the robot's future interaction (LeDoux 2007).

The goal is to give the robot the ability to rather reflect a user's emotional memory and enable the robot to use this memory to motivate children learning or engagement in various educational setups. For instance, we had applied the model to promote mathematics learning, mainly learning how to compute the area and perimeter of regular and irregular shapes (Ahmad and Mubin 2018). We believe that the ability to encode emotional states will enhance personalized learning and interactions with the robot when dealing with children (Kennedy et al. 2016; Baxter et al. 2017). Besides, in various qualitative studies after exposure of interaction with the robot, teachers and children observed that feedback con-

taining a memory of past mistakes may motivate the individual to improve their learning performance (Ahmad, Mubin, and Orlando 2016b,a; Eriksson, Björklund Boistrup, and Thornberg 2017). Moreover, the model generates unique behaviors in terms of the robot's feedback, and this may maintain their engagement with the robotic partner in the longitudinal context lasting for months as discussed in the finding of past long-term studies (Kanda et al. 2004; Komatsubara et al. 2014; Leite 2015). Note that the model follows a process to form user emotional memory grounded in literature (LeDoux 1993) and applies it in an HRI context to investigate its impact on maintaining engagement and improving vocabulary learning.

Algorithm 1 The abstract representation for the implementation of Emotion & Memory Model. Note that, "STORE" broad range of information in the game environment, refers to storing the elements impacting the game outcome such as a player's position in the game (see section 4.1.2 for details).

Input: GameEvents, UserEmotions, LearningState;
GameEvents, UserEmotions, LearningState $\in [Positive, Negative, Neutral]$
Where: Positive UserEmotions $\in [Happy]$;
Negative UserEmotions $\in [Sad, Angry, Fear]$;
Neutral UserEmotions $\in [Neutral]$;
if *((GameEvents is (Positive) OR LearningState is Positive) AND UserEmotions is Positive)* **then**
  MARK Emotional Event as **POSITIVE** STORE "broad range of information from general knowledge and the environment."
end
if *((GameEvents is Negative OR LearningState is Negative) AND UserEmotions is Negative)* **then**
  MARK Emotional Event as **NEGATIVE**
  if *UserEmotion is SAD* **then**
    STORE "outcomes and consequences of goal failure."
  end
  if *UserEmotion is ANGRY* **then**
    STORE "goal; agents obstructing goal attainment."
  end
  if *UserEmotion is FEAR* **then**
    STORE "sources of threat; means of avoiding threat"
  end
end
if *(((GameEvents is Negative OR LearningState is Negative) AND UserEmotions is Positive) OR ((GameEvents is Positive OR LearningState is Positive) AND UserEmotions is Negative)))* **then**
  MARK Emotional Event as **POSITIVE** or **NEGATIVE** based on the context in the game
  STORE Based on the type of Emotion
end
if *((GameEvents is (Neutral)) AND UserEmotions is Positive or Negative)* **then**
  MARK Emotional Event as **POSITIVE** or **NEGATIVE** based on the emotional state
  STORE Based on the type of Emotion
end
if *((GameEvents is (Neutral)) AND UserEmotions is Neutral)* **then**
  MARK Emotional Event as **Neutral** No information is stored in this case
end
if *Event Occurred* **then**
  Use The Stored Information AND Generate Three Classes Of Robot Response
  Apply Exp3 To Learn And Decide The Action For The Robot To Display from the Selected Class.
end

The model has four sub-modules (see Figure 1): 1) inputs, 2) emotional event calculation (EEC), 3) memory mechanism generation (MMG), and 4) a behavior selection unit (BSU). The first three

modules are similar in functionality (Ahmad et al. 2019); whereas, the BSU was refined by applying the RL mechanism. The developed model has three inputs, i.e., game events, the user's emotional state, and the learning state. These inputs were estimated to create an emotional event during the interaction. We computed the emotional event of the user in the EEC module depending on the learning outcome **OR** the game event type, i.e., positive, negative, or neutral, **AND** the user's emotional state, i.e., happy, sad, angry, etc. The emotional event type estimated in the EEC module was then transmitted to the MMG module to provide data about the type of information stored for various emotional situations. This information was then sent to the memory-processing unit. For example, for a happy event, an event that takes the player close to the end goal in the game, we stored information linked with the happy event (e.g., the player's position, near or far from the end goal, in the game). Further, on the recurrence of the same happy event under the same circumstances in the future game interactions, we update the central information linked with the newly or most recently occurred event in the database. This updated information is sent to the MMG module. The MMG sends it to the BSU. Consequently, the BSU is responsible for selecting an appropriate behavior/response.

We show the abstract representation of the model in the shape of pseudo code in algorithm 1. The algorithm presents all the possible scenarios based on the inputs and describes the action performed during each case. It considers the following. 1) When all three inputs are positive, it marks the emotional event as positive and stores the broad range of information about the environment (e.g. player position); 2) When all three inputs are negative, it marks the emotional event negative, and stores the information based on the emotion detected (sad (e.g. robot ladder close to 100, history of snake near 100 causing loss of the game), angry (e.g. number of times a snakes appearing close to 100th mark), fear (e.g. robot catching ladder/missing snake close to the 100th mark, robot position)); 3) When a game event or learning state is negative and user emotion is positive, or vice versa, it marks the emotional event positive or negative based on the context (e.g. player ahead of the robot/robot ahead of the player) in the game and stores based on the type of emotion detected; 4) When a game event or learning state is neutral and user emotion is positive or negative, it marks the emotional event positive or negative based on the emotional state of the user and stores based on the type of emotion detected; 5) When a game event or learning state is neutral and user emotion is neutral, it does not store any information. The algorithm considers a special case where regardless of the game event (positive or negative or neutral), we recognize the emotional state and game context as the deciding factor to identify the event type. It is for instance due to the understanding that the event can be negative in general, but it did not have an impact on the game. For example, it happened at the beginning of the game.

The BSU employs a unique adaptive strategy for the interactive scenario. The core part of this strategy is modeled using the MAB RL framework (Mahajan and Teneketzis 2008). In MAB, the goal is to ask an agent to receive as much accumulated reward as possible in a fixed number of iterations T. For each iteration t, the agent has K actions, and the agent must determine what action to take. After each action, the agent receives an immediate reward r for the action. In this study, there are three classes of actions, and the robot must determine the best action. Therefore, we model it as a MAB problem. The Exponential Weight algorithm for exploration and exploitation (Exp3) is described below in detail (Bubeck, Cesa-Bianchi et al. 2012).

Here, we consider a process with K different actions. At each time step, the algorithm must take one of K actions and receive reward r from the environment. This reward represents feedback on the action generated by the robot operating in the given environment. The goal of the robot is to maximize its accumulative reward over time to ensure that it meets the goal of maintaining engagement. Note that the reward signal is critical to ensure that the agent learns an informed choice of an action and it should represent the current state of the situation or environment. Therefore, the reward was based on the social cues (eye-gaze direction, verbal responses, facial expressions, and gestures, see section 4.1.2) representing the social signals shown during a highly engaged social interaction. To solve this problem, the algorithm links each action with weight w_i to highlight the significance of each action. The weights are used to generate probabilities for each action and are adjusted based on the rewards. When the process initiates, the algorithm first determines whether it wants to exploit or explore based

on exploration rate γ, which determines the possibility of executing a random action. If exploration is selected, the algorithm randomly samples a single action from all possible actions. If exploitation is chosen, the algorithm selects action i based on the previous distribution P. After executing action i, the algorithm receives reward signal $x_{i_t}(t)$) from the environment. Then, the algorithm generates an estimated reward $\hat{x}_{i_t}(t)$ by considering the influence of the probability of each action $p_{i_t}(t)$. The estimated reward is defined as $\hat{x}_{i_t}(t) = x_{i_t}(t)/p_{i_t}(t)$. The algorithm then updates the selected action's weight w_i while maintaining other actions' weights. In total, the algorithm attempts to iterate the learning phase T times to optimize the policy. When the algorithm converges, the probabilistic distribution of all actions is considered the best distribution relative to maximizing the reward. The pseudocode for this algorithm is given in Algorithm 2.

The appropriate behavior is returned by the algorithm as a part of the BSU, and the BSU passes this information to the database and returns the behavior for the given event. At the end of this process, the behavior is displayed by the robot.

Algorithm 2 Exp3 Algorithm to select behaviors

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

4. Research Method

We focused on evaluating the updated behavior selection process in the emotion and memory model and studied its effects on sustaining social engagement and vocabulary retention during repeated HRI. We investigated the following questions (RQs).

RQ1: Will the social robot demonstrating the updated version of the model result in sustained social engagement quantified in relation to the duration (total time) of users gazing at the robot, showing positive emotions (e.g., smiles), generating oral responses, and displaying gestures during repeated HRI?

RQ2: Will the social robot adapting and selecting actions depending on the updated version of the model result help maintain immediate **(RQ2a)** and delayed vocabulary retention **(RQ2b)** during repeated HRI?

Our hypotheses are given as follow.

H1: The robot using our emotion and memory model will sustain social engagement estimated in terms of eye gaze, verbal responses, facial expressions, and gestures across all sessions in human-robot-

Figure 3.: Snakes and Ladders Game.

game interaction.

H2: The robot using our emotion and memory model will maintain immediate retention of vocabulary in each session (**H2a**) and delayed retention of vocabulary across sessions **(H2b)**.

Note that delayed vocabulary retention refers to retention of vocabulary, that was exposed in the first session, and retained in the second, third, and fourth sessions. In addition, as discussed in the introduction, we refined the emotion and memory model and introduced a learning mechanism to learn and select actions during child-robot interactions. We investigated whether the updated model maintains engagement and improves learning over time and, as highlighted by (Leite 2013a; Leite, Martinho, and Paiva 2013), for repeated HRI the session becomes a significant variable for comparison. Therefore, we used the session as an independent variable to study the impact of the updated model to see significant changes in the data. Nonetheless, it was important to investigate the value of the updated version of the model by comparing it with a control version. We did not use a control group in the current study due to logistical issues. The evaluation was conducted at a primary school. We were limited in terms of the time constraints due to the school schedule. Therefore, we decided to reflect on and compare the trends of the findings of the current study with the previous study (Ahmad et al. 2019). In essence, we compared the significant trends in the results of the current study with the results presented in the previous study, where we had similar experimental setup and the robot's behavior was predefined and dependent on event type during the game.

4.1. *System Description*

4.1.1. Snakes and Ladders Game

In the snakes and ladders game, the game board has 10 rows and 10 columns, and the snakes and ladders were modified (Ahmad, Mubin, and Escudero 2015) as shown in Figure 3. The rules of the game are simple. A player rolls dice and attempts to reach the 100th mark to win the game. The player begins at the 0th position on the board and moves forward equal to the number shown on the dice. The player that reaches the 100th mark first wins. During gameplay, a player is faced with stars, snakes, and ladders on the game board. On each snake, the NAO robot spoke about one of the vocabulary words from the ROILA language (Figure 4). For ladders, the player climbs a few steps closer to the 100th mark. On each star, a player moves one to six steps away from or closer to the 100th mark depending on the indicated positive or negative nature of the star.

4.1.2. Applying Emotion and Memory Model during the Game

We implemented the refined version of the model (updated behavior selection unit - Figure 1) in the snakes and ladders game task. The first step was to identify the significant events that happen during the

Figure 4.: ROILA words appearing in the game on the appearance of the snake on the board.

game that have an impact on goal achievement or goal failure to describe the process of the application of the model. We used similar numbers and types of positive and negative game events as identified in our previous studies (Ahmad, Mubin, and Orlando 2017; Ahmad et al. 2019, 2017). These game events include (but are not limited to) a snake or ladder appearing on the board distant or near to 100 (winning mark), occurrence of either positive or negative star near or distant from the winning mark, six appearing twice or more on the dice, and a positive or negative game outcome. We computed the emotional state of the players during the game through facial scans. We applied an online API (Indico 2016) to compute the emotional states of the players. This API returns six key-value pairs of six emotions (angry, surprised, fear, sad, neutral, and happy). The key-value pairs have the values of the probability of each emotion detected on the user's face. The threshold for each emotion was set to 0.05, and each emotion less than the threshold value was discarded. Here, six different emotion values were stored every 10 s during the interaction. Note that the API returns six emotions. However, we considered happy, sad, angry, fear, and neutral values. We calculated the emotional state in the following steps. First, we calculate the simple average of the last six values for each emotion (happy, sad, angry, fear, and neutral) stored in the database for each significant game event. This suggests that on the occurrence of the event, we computed the sum of the last six values for each emotion-value and divided it with 6. Second, the max function in *Python* language was used to select the maximum value from the list of mean emotion-value. Finally, the max value was considered as the current emotional state. More specifically, Happy, Sad, Angry, Fear, Neutral emotions were computed at times (t) $-$ ($t_1, t_2, .., t_6$), and later the value of each emotion was averaged, and the maximum value among all emotion-value pairs was selected: $MAX((Happy_1 + Happy_2 + .. + Happy_6)/6, (Sad_1 + Sad_2 + .. + Sad_6)/6, (Angry_1 + Angry_2 + .. + Angry_6)/6, (Fear_1 + Fear_2 + .. + Fear_6)/6, (Neutral_1 + Neutral_2 + .. + Neutral_6)/6))$. This further means that, for example, if the Happy emotion-value was high in the first two iterations and later Sad emotion value was high in the next four iterations, then the emotion-value of Sad will be set as the maximum value. The user's emotional state was computed by considering the current and recent history of the user's emotions values reflected on the user's face.

Note that, we followed a process proposed by Levine and Pizarro (2004) (see Table 1) on what central information is stored under the four considered emotions to store the memory during the game. Specifically, we followed a brute force practice and reflected on all the possible combinations of game-events and emotions. Particularly, we considered all the four emotions for all the significant game events (e.g., positive or negative star near or distant from the winning mark) during the snakes and ladders

game. We then followed the taxonomy to generate a response that uses the central information stored under various emotional states. We describe a few examples to show how we applied the model during the snakes and ladders game-based task. The examples consist of the selected emotional events during the game. The examples also provide specifications for the information remembered for different events and how the robot's memory was created. For example, an occurrence of the snake close to the 100th mark was referred to as a negative event because it impedes the player's chance to win. It further prevents the player from obtaining the end goal. As stated, for a negative event, we store information depending on the users' emotional state. Therefore, for the sad emotional state, the information regarding the previous outcome of the game after a snake appeared near 100 was stored with the user's emotional state. For the angry emotional state, information regarding the number of times a snake was faced close to the 100th mark was also saved. We considered fear to represent fear of losing the game; thus, information about the competitor, i.e., the robot's position on the board, was also stored. When a ladder was available near 100, we considered this a positive event because it facilitated reaching the final goal. For positive circumstances, we stored a wide spectrum of information regarding the situation; thus, we stored data about the number of times a ladder was near or distant from 100 or their respective former outcomes. When a happy emotional state was detected for negative events, e.g., the user facing a snake on the game board, we considered the position of the robot to determine whether the robot was leading or following the user. In summary, the robot's behavior was created by incorporating the memory (the central information fundamental to the event that evokes the emotion) to ensure generation of independently or simultaneously context-aware verbal and non-verbal responses. In Table 2, we enlist an example of the selected emotional event and a complete list can found on this link. Further, the pattern of dice was fixed for all the game interactions for all the participants. Therefore, we created and stored the three classifications (positive, negative, neutral) of robot response for all the game events and applied the Exp3 to learn the likable choice (class) of behavior for each participant.

The model was also demonstrated in the vocabulary testing phase. In this phase, the robot tested the knowledge of words exposed or learned during the game. We considered the emotional state and learning outcome of the user as the inputs to the model. For the learning state, we took the correct and incorrect answers of the user for the word during the interaction. For example, if a happy emotion was detected, and the learning outcome is correct answer, we follow the taxonomy (see Table 1) and stored information about the correctly answered "word" along with the current session number of the interaction. When the learning outcome is an incorrect answer and the emotional state is Happy, we stored the information about the detected emotion (Happy in this case) along with the "word" and the session number in progress. For the sad emotion, we collected data about the number of user attempts to memorize the word. For anger and fear emotions, the session in which user correctly answered the word was stored because the child could potentially get angry when forgetting a word or they could feel uncomfortable experiencing similar points on the test.

In the previous study (Ahmad et al. 2019), the BSU enabled the robot to react positively, negatively, and neutrally based on the event type during the game. However, we updated the BSU and used an MAB algorithm to enable the robot to select one of three behaviors (positive, negative, or neutral). We created three categories of robot behavior during the game. For each game event, the robot selected one of these behaviors based on the child's social engagement. The rationale for classification of behaviors was mainly due to the findings of the previous work, where we found that the robot's emotional feedback affects the child's vocabulary retention and social engagement in terms of eye gaze when applied during the same snakes and ladders game (Ahmad et al. 2017, 2019). Further, the choice was partly motivated by the findings suggesting an impact of the social role of the robot (competitive/cooperative) on the user's task performance and engagement (Zaga et al. 2015). Hence we speculated that children may have a different preference for robots in an educational environment. In addition, Mutlu et al. (2006) performed an exploratory study into the perception of a humanoid robot possessing cooperative and competitive characteristics. Their results revealed that people perceived robots as significantly more desirable in cooperative roles. However, they stated that the preference concerning social characteristics may vary according to the given task. We further find positive effects relative to informing a teaching

Inputs	Information	NAO's behavior
Game Event: Snake far 100 **Emotional State:** Happy, Sad, Fear, Angry	**Happy**: robot's position **Sad**: game's outcome, number of steps away from the goal. **Fear**: Position of the robot. **Anger**: No of times the snake has occurred.	*First Session* **Happy**: I am glad to learn you are looking **happy** and the snake is not worrying you and you are **ahead** of me . lets learn a new word. **Sad**: you are looking **sad**, I feel sad for you but you can still win the game as you are **ahead** of me. Lets learn a new word. **Fear**: you are looking **scared**; I am **hopeful** you can win the game as you are also **ahead** of me. Lets learn a new word. **Angry**: you are looking **angry**, I am **hopeful** you can still win the game as you are **ahead** of me. Lets learn a new word. *Other Sessions* **Happy**: I am glad to learn you are looking **happy**, in the **last session** you had **three** snakes in the beginning but you still **won** the game. Lets learn a new word. **Sad**: It's a Snake, don't be sad, you also had **three snakes** in the beginning last time but you did win the game. Lets learn a new word. **Fear**: Don't be scared, I was behind you when you **had snakes** in the last game but you did **win** at the end. **Angry**: you are looking **angry**, you had **three** snakes in the beginning last time and you **won** the game. Don't worry you are ahead of me.

Table 2.: The table shows an example of the selected emotional event (Snake far 100) and shows how the Emotion and Memory model was applied, mainly for the Happy, Sad, Fear and Angry emotions. Note that, the table only shows the positive/encouraging class of robot behaviors and do not enlist the negative/competitive/critical or neutral feedback behavior. For complete list, please click on the link.

strategy based on user modeling (Graf, Liu, and Kinshuk 2008; Bajraktarevic, Hall, and Fullick 2003). Due to these reasons, we did not choose to collect data on the children preferences in this work. Nonetheless, the feedback was positive, negative, and neutral and also consisted of providing encouragement, and being critical or competitive depending on the task.

The classification of behaviors along with examples is summarized as follows.

- **Positive (Emphatic/Supportive) Feedback:** The robot behaved positively in game events by encouraging the child and/or informing the child about its emotional feelings about game events. For example, when the snake was distant from the 100 mark[1], in the first session, the robot said, "I am glad to learn you are looking happy and the snake is not worrying you and you are ahead of me. Lets learn a new word." During the other sessions, the robot said, "I am glad to learn you are looking happy, in the last session you had three snakes in the beginning but you still won the game. Lets learn a new word."
- **Negative (Competitive/Critical Feedback):** The robot behaved competitively in game events by challenging the child and/or informing the child about its emotional feelings about game events. For example, when the snake was near 100, in the first session, the robot said, "A snake near 100, I feel happy you might lose today and you are also behind me, lets learn a word.". During the other sessions, the robot said, "A snake near 100, you also had a snake on 99 in the last game Although you won the game, but you can't be lucky every time, lets learn a new word."
- **Neutral Feedback:** The robot reacted neutrally to game events and reacted neutrally in terms of emotional feelings. For example, in case of six on the dice near 100, during the first session, the robot said, "You have a six near 100, it's ok, keep playing." During the other sessions, the robot said, "You have a six near 100, it's ok, I remember you had a snake after the six in the last game."

We applied the updated version of the emotion and memory model in a similar manner as in our previous evaluation (Ahmad et al. 2019) during the game and post-test. Here, we used the same method to store information during different emotional states and applied it to the robot's behavior after the first interaction. In the previous examples, the use of memory varied for each child based on the child's emotional state during different game events.

4.1.3. Adaptive Action Selection Strategy

We used the Exp3 algorithm (Bubeck, Cesa-Bianchi et al. 2012) to allow the social robot to select one of the three behaviors for a game event game phase. Note that game events in which the robot uttered a response were similar to those used in the previous study, e.g., "Snake near 100" and "Ladder far from 100." To compute reward $x_{i_t}(t)$, we measured the social engagement of each individual for each event during the game phase. The user's social engagement was measured based on eye gaze, verbal responses, facial expressions, and gestures. These four variables were considered based on our understanding of the social engagement measurement in our previous study (Ahmad, Mubin, and Orlando 2017). For each event, we computed all the four variables. When a robot generated a response to an event, if the child's *gaze* was facing the robot's face when the robot was displaying a response, we marked the reward for gaze value as 0.25 (otherwise 0). Similarly, we estimated the *user's emotional* state at the utterance of a robot's response for each event because one of the goals of applying the algorithm was to enhance social engagement. Here, happiness and smiling were considered positive indicators of social engagement. Therefore, if the child's emotional state was happy (or smiling) at the occurrence of the robot's response to an event, we marked the facial expressions reward as 0.25 (otherwise 0). For a *verbal response*, if the child reacted to an event verbally after the robot's response to an event, we marked the verbal response as 0.25. Finally, if the child displayed a *gesture* in the form of a fist or hand waving after the robot uttered a response to an event, we marked the gestures as 0.25 (otherwise 0). We were limited to the number of gestures due to a technical limitation. However, the choice of gestures was based on observations that fist and wave gestures were the most commonly displayed gestures in the

[1]Snake Far 100 refers to the case, where the position of the user is at least 50 steps away from100 (winning position).

previous studies. We computed reward $x_{i_t}(t)$ for each event after the robot has completed its response on the event by adding all values of the previously mentioned factors. The mathematical formula for the reward function is $x_{i_t}(t)$ = $Gaze$ + $userEmotion$ + $verbalResponse$ + $gestures$. Therefore, the maximum reward for an event was $x_{i_t}(t) = 1$. This suggests a state where the participant looks at the robot, displays a smile, gives a verbal response, and shows a gesture (fist or wave).

To measure eye gaze facing the robot, we used the Haar cascade frontal face and eye classifier (Kasinski and Schmidt 2010) to detect the user's eye gazes. We detected eye gazes in an image captured by an external camera. If gazes were found/detected in the image, we considered it as the user's eye gazes facing the robot or user was looking at the robot. To ensure that the gazes were directed at the robot, we used the camera installed on the head of the NAO robot. As the NAO robot performed an action, we determined whether the user's eye gaze faced the robot for the duration of the robot's response. We logged each occurrence of an eye gaze facing the robot in the database. When the robot's response was completed, we counted the occurrences of gazes according to the frames per second. Here, if the averaged frequency of gazes per second was greater than a given threshold (50%), we calculated the reward based on the eye-gaze as 0.25 (otherwise 0). This was done to ensure that, if the child did not look at the robot initially, we would not calculate the reward value as 0. We programmed this mechanism according to the logistics of our setup. The child could potentially look down to the tablet or could move their head away from the robot. In such occurrences, the eyes were not viewed by the camera. Note that, the Haar cascade frontal face and eye classifier is a commonly used classifier to identify eyes in a frame and has a reliability of over 95% (Padilla, Costa Filho, and Costa 2012).

To measure facial expressions, we detected the user's emotional state based on the human face. Here, we used the Indico API to measure the emotional state on the child's face. We computed the average of all the six value pairs returned by the Indico API throughout the duration of the robot's response. If the detected emotional state remained positive (happy), we assigned a reward of 0.25 (otherwise 0) based on the emotional state. Note that, the indico API is a reliable tool and has been used by other researchers to detect children emotions (Spitale et al. 2019) and has shown a relatively good accuracy to detect emotions (Balakrishnan et al. 2019).

We used the Google speech recognition API to evaluate for a verbal response. For each event, at the end of the robot's response, we recorded the child's verbal response. If the speech API detected a verbal response, we considered it as the child's verbal response to the robot and assigned a reward of 0.25 (otherwise 0).

Finally, we detected gestures using OpenCV (Bradski and Kaehler 2000). We captured the frame using an external camera and calculated the number of pointing fingers. To calculate the number of fingers, we initially created a bounding rectangular frame around the hand. In the bounding rectangular frame, we checked for fingertips and finger webbing. We then calculated the number of fingertips pointed toward the camera, where a fist was considered zero fingertips, and a waving hand was considered five fingertips. The gestures were evaluated throughout the robot's response and were simultaneously logged in the database. At the end of the robot's response, we checked for zero or five values throughout the duration of the response. If such values were found, we assigned a gesture-specific reward of 0.25 (otherwise 0). The gesture detection was tested by the author with various users and it was found that it detects the fist and wave gesture with an over 70% accuracy. Note that, detecting gestures is a challenging problem and this manuscripts presents a proof of concept.

We also used the Exp3 algorithm during the post-test (testing) phase to select positive, negative, or neutral feedback. To compute the reward function $x_{i_t}(t)$ during the post-test, we considered the learning outcome of the individual in the post-test. If the child remembered the word, we gave reward $x_{i_t}(t) = 1$; otherwise, $x_{i_t}(t) = 0$. Here, the goal was to enhance learning outcomes (vocabulary retention). So, we based our reward on the learning outcome. The justification for this choice of reward function was dependent on the findings of the previous study. We found that both positive and negative feedback was useful in improving the learning of the participants. Exp3 learns the best action based on the maximization of the reward. For example, when a child answers correctly in the second session, it may have been due to the positive feedback (encouragement) given in the first session. Assuming that the positive

feedback was selected by Exp3 and it received a +1 reward. Consequently, this enabled Exp3 to learn that positive feedback could be the reason for the better learning outcome. In other words, Exp3 will exploit and select the positive feedback 90% of the time and only explore the other two options 10% of the time.

4.2. *Study*

4.2.1. Ethics Approval

The study was approved by the ethics review board at the host Institution. This included pre-approvals via the management of the primary school which were then included as part of the ethics application with the institution.

4.2.2. Interaction Scenario

The interaction took place in four phases, i.e., the introduction, vocabulary pre-test, snakes and ladders gameplay, and the vocabulary post-test phases.

In the introduction phase, the NAO robot was programmed to introduce itself and conduct high-level dialog with the participant (child), including talking about their day and activities they have undertaken on the given day. In the second, third, and fourth phases, the robot personalized its dialog by calling children by their names and used personalized information, e.g., previous game outcomes.

In the vocabulary pre-test, the NAO robot tested the vocabulary knowledge introduced during the game. In the first session, the word appeared on the tablet, and the NAO robot asked if the child knew about the word. Here the robot asked: "Do you know the word Biwasa?", and the child answered "No". The NAO robot then responded with: "We will learn about the <WORD NAME> shortly." In the following sessions (second to fourth), the NAO robot either said this is correct or corrected a mistake by stating the correct answer.

In the game phase, the children played the snakes and ladders game with the NAO robot. In each game, the NAO robot was programmed to teach six different words. A fixed/predetermined pattern of turns was maintained by the robot during the game for each session for all participants. This suggests that the dice rolls were controlled. The fixed pattern of turns results in each participant winning the game in first and third session and losing in the second and fourth session. We applied the emotion and memory model during the game. We used the memory from the second session.

In the vocabulary post-test, the words taught during the game were asked by NAO. The emotion and memory model was demonstrated to generate positive, negative, and neutral feedback behaviors as portrayed by the NAO robot. Animated feedback was provided about each child's learning outcomes. Examples of the feedback can be found in the literature (Ahmad et al. 2019).

4.2.3. Participants

We conducted a within-subject study with 24 children (12 males and 12 females) aged 10-12 at a school (mean: 10.69, +/- 0.47). The participants were bilingual, and none had interacted with a robot before the sessions. All participants' characteristics were similar to those of our previous studies (Ahmad et al. 2019, 2017).

4.2.4. Procedure

The repeated within-subject evaluation lasted four school weeks. Each child individually interacted with the NAO robot and played the snakes and ladders game four times on four different days (one session per day) over four school weeks with a gap of six days between sessions (for a total of 96 sessions, i.e. 24 children times four sessions). We conducted the study for four consecutive weeks, and eight children participated on their assigned days (Wednesday, Thursday, or Friday of each week for four weeks).

No.	ROILA	ENGLISH	Session Used in
1	jabami	hi	1
2	make bama	good bye	1
3	kanek	go	1
4	botama	turn	1
5	babalu	stop	1
6	koloke	forward	1
7	webufo	left	2
8	besati	right	2
9	nole	back	2
10	jinolu	ball	2
11	lakowo	cat	2
12	fipuko	dog	2
13	belutu	boy	3
14	batuno	girl	3
15	piwaja	flower	3
16	wikute	fruit	3
17	fupuma	music	3
18	lobo	robot	3
19	lujusi	box	4
20	bubas	house	4
21	bokubo	school	4
22	panope	garden	4
23	kepete	shop	4
24	wapisi	bucket	4

Table 3.: Words from ROILA Language (Omar Mubin 2015).

Figure 5.: This figure demonstrates a child playing the snakes and ladders game with the NAO robot. The tablet on the table was used to show the game and also show the vocabulary to the child. The two cameras were used to collect the data on facial expressions and gestures.

Each interaction was completed in approximately 24 minutes and had five steps, similar to the procedure presented in the previous study (Ahmad et al. 2019; Ahmad, Mubin, and Orlando 2017): 1) two-minute introduction, 2) four-minute pre-test, 3) ten-minute game playing session, 4) four-minute break, and 5) four-minute post-test. The vocabulary post-test was similar to the pre-test to maintain consistency among the test results. We logged all test results, including mistakes.

4.2.5. Setup and Materials

The setup and materials were similar to those used in our previous studies (Ahmad et al. 2019, 2017), as shown in Figure 5. Here, the difference was that the NAO robot sat in a crouching posture. In our previous study, the robot was in a sitting posture. The choice of robot posture differed due to the differences in heights of the tables against the chairs in both studies. Height was an important issue because we needed to record eye gaze facing the robot using the robot's facial camera. We used a quiet room, which was divided into two parts. One researcher sat on one side to initially run the program on the robot. On the other side, the child interacted with the NAO robot placed on a table with a Samsung tablet. In addition, participants interacted once a week for four weeks in the current study, where as, in the previous study, they had interacted twice a week.

The NAO robot is a humanoid robot measuring 58 cm in height with 25 degrees of freedom (Softbank Robotics). We also used the same 24 vocabulary words (Table 3) from the ROILA, similar to our previous study. The 24 words were taken from the first two chapters of a book about learning ROILA (Mubin et al. 2011). The words were presented in each session in the order they appear in the book to follow the learning pattern presented in the book.

Gaze	Facial Expression	Verbal Response	Gesture
Facing robot face	Smiles, timid Smiles, Surprised, Flushed	"Hello" "thank you" "Okay" ,"Yes"	Wave Fist Nod

Table 4.: Coding scheme used to measure social engagement (Ahmad, Mubin, and Orlando 2017). The table presents examples for spotted gestures and also for the verbal responses.

4.2.6. Measurements

We measured the impact of the model in terms of supporting ROILA vocabulary development and maintaining social engagement during repeated HRI by considering the following dependent variables (DV).

(1) Total number of words used in the game remembered in the post-test.
(2) Total time a child looked at the robot, showed positive expressions (e.g., smiles), uttered oral responses, and displayed animated gestures during the interaction was used to quantify the promotion of social engagement.
To measure the impact of feedback generated during the post-test on vocabulary retention, we measured the following DV.
(3) The total number of old words (used in the post-test of the last session) remembered in the pre-test in the next sessions. For example, the number of ROILA words taught in the first session remembered in the second session. Further, the total number of ROILA words used in the first and second sessions tested in the pre-test of the third session. Finally, the total number of ROILA words used in the first, second, and third sessions, tested in the pre-test of the fourth session.

Note that the above mentioned DVs are the commonly used metrics to access the social engagement in the HRI literature (Serholt and Barendregt 2016). To measure social engagement, we performed video analysis to code the following DVs identified in the previous study (Table 4). Here, we manually coded videos for the following DVs: gaze facing robot, facial expressions, verbal responses, and gestures. We coded for the total time (duration) of the types of smiles, gestures as listed in the table except for the case of verbal responses. Table 4 shows examples of dialog, where as and we coded any type of verbal response provided by the child to the robot during the interaction. We used ELAN, a professional video annotation tool, to code the videos. Note that two researchers were involved in the video coding process, and one researcher did 100% of the video and was also involved in the coding process of the previous study (Ahmad, Mubin, and Orlando 2017). The other researcher who is one of the authors performed 10% the video coding. 10% of the videos were selected randomly. We conducted an Intraclass Correlation Coo-efficient (ICC) test to check the reliability of the video coding. The analysis showed excellent reliability with an average measure of the ICC for eye-gaze was .947 with a 95% confidence interval from .735 to .989 (F(7,7)= 18.82, p¡.001). The average measure of the ICC for facial expressions was .985 with a 95% confidence interval from .925 to .997 (F(7,7)= 66.73, p¡.001). The average measure of the ICC for verbal responses was .922 with a 95% confidence interval from .608 to .984 (F(7,7)= 12.75, p¡.001). The average measure of the ICC for gestures was .962 with a 95% confidence interval from .810 to .992 (F(7,7)= 26.24, p¡.001). Regular discussion sessions were conducted during the coding process to avoid any bias (Serholt 2018). Further, we chose to code the videos manually due to technical limitations. First, we considered a gesture other than a fist and waving. Second, the automatic system did not distinguish between the speaking-robot and the child as we take the verbal signals after the robot had stopped speaking. Lastly, for the facial expressions, we considered various types of smiles (timid, flushed) that may have been difficult to recognize automatically.

5. Results

5.1. *Social Engagement During Game Phase*

To investigate **RQ1** and test **H1**, a repeated measure analysis of variance (ANOVA) was conducted. The session was the within-subjects factor with four levels on the game play phase for the following DVs: total amount of time (duration) a child looked at the robot; showed positive expressions, such as smiles; uttered oral responses, and displayed animated gestures. We did not observe a difference in the amount of time for the game interaction phase for any of the participants; therefore, we did not normalize the data. All time measurements were recorded in ms.

We found that there was a significant effect of session during the game phase for the amount of time a child looked at the robot (F(3, 69) = 14.58, $p < .001$, $\eta_p^2 = .388$), showed positive expressions, such as smiles (F(3, 69) = 6.95, $p < .001$, $\eta_p^2 = .232$), and uttered verbal responses (F(3, 69) = 4.39, $p < .007$, $\eta_p^2 = .160$). In addition, there was no effect of session on the amount of time children displayed gestures (F(3, 69) = 1.34, $p = 0.266$, $\eta_p^2 = .055$).

The Bonferroni test was conducted to find the difference observed among sessions for the four DVs. No significant differences were observed for the amount of time a child looked at the robot in the first three sessions. However, a significant increase in the amount of time a child looked at the robot was observed in the fourth session. Thus, , the values for the amount of time a child looked at the robot in the fourth (last) session were statistically significant with the first ($p < .004$), second ($p < .002$), and third ($p < .001$) session. The mean values for the duration of children's gaze facing the robot were as follows: Session 1 (M: 171.80, SD: 75.18), Session 2 (M: 171.34, SD: 74.64), Session 3 (M: 173.40, SD: 77.00), and Session 4 (M: 258.01, SD: 135.72).

For the amount of time a child showed positive expressions, such as smiles, the third session was statistically significant compared to the second ($p < .02$), fourth ($p < .002$) session. We did not witness a statistically significant decline in the duration of children's facial expressions from the first to the second and fourth sessions. The mean values for the duration of children's facial expressions were as follows: Session 1 (M: 39.96, SD: 25.61), Session 2 (M: 30.41, SD: 26.13), Session 3 (M: 53.22, SD: 40.31), and Session 4 (M: 30.98, SD: 26.18).

For the amount of time a child uttered verbal response, we found that the fourth session was significant compared to the first ($p < .05$), second ($p < .05$) sessions. We witnessed an upward trend in the duration of children's verbal responses from the first to the fourth sessions. The mean values for the duration of children's verbal responses were as follows: Session 1 (M: 13.93, SD: 9.95), Session 2 (M: 15.94, SD: 8.07), Session 3 (M: 17.33, SD: 8.23), and Session 4 (M: 20.30, SD: 5.94).

Lastly, for the amount of time a child displayed gestures, we did not observe a significance difference between sessions. We witnessed consistent duration of gestures from the first to the fourth session. The mean values for the duration of children's gestures were as follows: Session 1 (M: 6.70, SD: 5.65), Session 2 (M: 5.46, SD: 7.20), Session 3 (M: 8.10, SD: 12.53), and Session 4 (M: 4.63, SD: 4.38).

5.2. *Vocabulary Learning During Game Phase*

To investigate **RQ2a** and to test **H2a**, we considered the total number of words used in the game that each participant remembered in the post-test. A repeated measure ANOVA was performed with the session as the within-subjects factor with four levels using the total number of words used in the game remembered in the post-test per session as the DV. A significant effect was found for the session (F(3, 69) = 13.46, p<.001, η_p^2 = .369) on children's vocabulary-learning performance.

The Bonferroni post hoc was done to further check the effect of the total number of words used in the game remembered in the post-test for each participant across sessions. We found that significantly more words were remembered during the third session than the first ($p < .001$) and fourth sessions ($p < .001$). Similarly, the number of words remembered during the second session were significantly more than the number remembered during the first ($p < .004$) and the fourth ($p < .001$) sessions. The

mean values of the learning outcome for all sessions were as follows: Session 1 (M: 4.62, SD: 0.96), Session 2 (M: 5.58, SD: 0.77), Session 3 (M: 5.70, SD: 0.55), and Session 4 (M: 4.62, SD: 0.87).

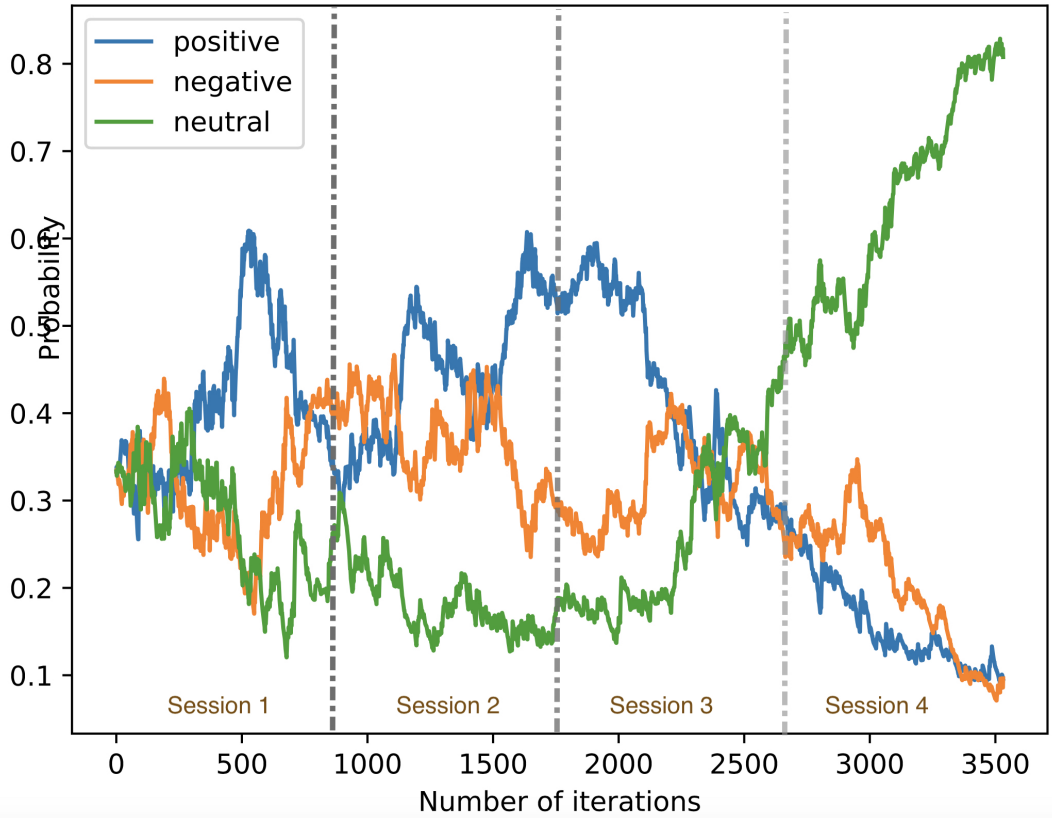


Figure 6.: A run of Exp3 against the reward based on eye gaze, verbal response, facial expressions, and gesture. The green line (neutral feedback) is the weight of the best action learnt in the fourth session. The blue and orange lines are the weights of the second and third best actions, respectively. The chart shows that neutral feedback was the learnt behavior choice of the robot behavior for children. The values on the x-axis show that each session had between 870-880 iterations. We applied the policy in a way that placed all interactions with all the children together. We did it due to the following reasons. First, we argue that the algorithm adapts to a pattern of behavior that is most likely to happen among all the participants during our interaction. Besides, the game behavior was pre-defined or fixed. In addition, it is evident from the Figure that it did not apply the same action to all of the participants until the fourth session. The variation of action selection was observed during the first three sessions and the learning of the likeable action reflected only in the fourth session.

5.3. *Vocabulary Learning During Post-test*

To investigate RQ2b and to test H2b, we checked for delayed retention, i.e., total number of old words remembered in the pre-test in the next sessions for each participant. A repeated measure ANOVA was conducted with the session as the within-subjects factor with $three, two$ levels using the total number of old words remembered in the pre-test in the next sessions as the DV. We did not find a significant effect of interaction (session) (F(2,46) = .870; $p = .426$, $\eta_p^2 = .036$) on the total number of words learned during the first session across the second, third, and fourth sessions. The mean values of the retention of Session 1 words across session was as follows: Session 2 (M: 4.50, SD:1.06), Session 3 (M: 4.70, SD: 1.26), and Session 4 (M: 4.83, SD: 1.16).

We found a significant effect (F(1,23)= 9.12; $p < 0.006$, $\eta_p^2 = .284$) of session on the delayed retention of words learned during the second session across the third and fourth sessions. The mean

values of the retention of Session 2 words across these sessions was as follows: Session 3 (M: 5.20, SD:0.58), and Session 4 (M: 4.58, SD: 1.17). Lastly, the mean values of the retention of Session 3 words during the fourth session were as follows: Session 4 (M: 5.29, SD: 0.69).

5.4. *Comparing Rule-based and RL across two studies*

Here, we report the formal analysis to compare the reported study with a prior study as mentioned earlier, which relied on rule-based learning. However, we will limit ourselves to making strong inferences based on this analysis. It is considering the highlighted difference of the settings of the two studies (as highlighted in the setup section). We believe that the trends (how much a particular measurement increased from the first to the last session) of both studies are more indicative in contrast with the formal analysis.

To investigate the differences between the current and previous study, a mixed analysis of variance (ANOVA) was conducted. The session was the within-subjects factor with four levels on the game play phase for the following DVs: total amount of time (duration) a child looked at the robot; showed positive expressions, such as smiles; uttered oral responses, and displayed animated gestures and type of study (current v.s. previous) as between-subject factor.

We found that there was a significant effect of the type of study during the game phase for the amount of time a child looked at the robot (F(1, 46) = 71.61, $p < .001$) uttered oral responses (F(1, 46) = 18.36, $p < .001$), and displayed animated gestures(F(1, 46) = 6.41, $p < .02$). This suggests that the amount of time a child looked at the robot was higher in the current study whereas the amount of uttered oral responses, and displayed animated gestures were higher in the previous study. We did not find a significant effect of the type of study for the positive expressions, such as smiles (F(1, 46) = 3.86, $p > .05$).

6. Discussion

We see from the results that children's social engagement measured in terms of duration of their eye gaze facing the robot during the interaction sustained from the first to the third sessions. However, a significant increase in the duration was observed during the fourth session. We understand that there could several reasons for this finding. Specifically, we can associate the sudden increase in the fourth interaction with the situation that children were aware that this is their last interaction with the robot. It may have resulted in a keener and more engaged behavior during the interaction. This inference is dependent on the past finding that tried to analyze verbal and non-verbal behaviors during farewells or last interactions (Knapp et al. 1973). Their findings note the significant difference between verbal and non-verbal interactions during good-byes. Hence, we speculate that children may have recognized that it is their last interaction with the robot, and this could have resulted in a sharp gaze increase during the fourth session. Nonetheless, we also believe that the adaptive strategy was learning to select either positive, negative, or neutral behavior according to child's preference, primarily social cues, and this may have enabled the robot to select the likable behavior according to each child and could have played the part to maintain the duration of the eye-gaze. As can be seen in Figure 6, the neutral feedback was the best action based on the reward function. It appeared to have converged in the fourth session. Further to this, RL is a trial and error process. In this case, the robot tried to explore and exploit several actions in the snakes and ladders game environment. It may have been that the actions (neutral) exploited in the last session may have had the most impact as the reward was generated based on the engagement level of the participant. In addition, the proposed model enabled the robot to employ memory in its behavior and thereby generate novel and personalized behaviors in terms of the use of the central information fundamental to the event that evoked the emotion during each session. Our findings strengthen the observation that personalization and adaptation to users in different forms (feedback, using the memory-related events inside the dialogue) during long-term HRI is an effective way to maintain engagement

(Irfan et al. 2019).

We also see that the duration of facial expressions, primarily smiles, declined from the third to the fourth sessions. In addition, there was an increase from the second to the third session. We conjecture that there can be several reasons for these findings. First, during the first session, not all participants had interacted with the NAO robot, which may have created a novelty factor that may have generated an excessive number of smiles during the first interaction (Kanda et al. 2004; Komatsubara et al. 2014; Leite 2013b; Serholt and Barendregt 2016). Second, the *Exp3* algorithm learns based on the reward function in an incremental manner (Kaelbling, Littman, and Moore 1996). Therefore, as can be seen in Figure 6, in the second session the algorithm appeared to have generated relatively more negative feedback behavior that may have been less likable to a few children. Third, during the third session, the game resulted in a winning outcome for the participants. Hence, the reaction of the robot may have been positive and may have been more benefiting than the other sessions. It has also been observed that positive game events mostly elicit positive feelings in children (Ravaja et al. 2006). However, this may not be generalised to all sessions as we did not see a significant decline of smiles from first to second and first to the fourth session as children also won in the first session. Lastly, we cannot discount that each child possesses a different personality and has different traits (Cole et al. 1996). This could have also been a reason for a varied reaction in the case of a favorable or unfavorable outcome. An increase in the third session can be attributed to better performance in terms of the algorithm in behavior selection. As it can be seen in Figure 6, from the $1600th$ iteration referring to the third, positive feedback was selected more often than the other two behaviors. This shows that the algorithm may have generated more likeable personalized behavior for each user. A decline in the fourth session also reflects the normalization of the duration of the facial expressions as it is notable that the duration is coherent with the second and first sessions. The same trend can also be seen in Figure 7 where the mean-value for the detection of happy emotion increased from the second to the third session and decreased from the third to the fourth session. Another conjecture can be the effect of the setting that participants were on the losing end during the second and fourth sessions in the game. This also explains that there was an increase in the mean-value of sad emotion from the third to the fourth session. Trends similar to those for happy and sad emotions were seen for the fear and angry emotions, respectively.

We also see that there was an increase in the duration of verbal responses across session, and the duration in the fourth session was significantly high compared to the first and second sessions. These findings can be attributed to the following reasons. First, the children may have become familiar with the robot and the overall system, and this may have increased social interaction (Kim et al. 2013). Second, the increase in the duration during the fourth session can further be attributed to the neutral feedback behavior during the fourth session. A recent exploratory study showed that likable or preferred feedback enhances child-robot tutoring interaction in regards to their verbal utterances (de Haas, Vogt, and Krahmer 2016). We understand that this finding indicates that more research should be conducted to investigate the effect of feedback on the duration or frequency of the children's verbal responses during HRI. Third, the robot reacted positively by providing encouragement more frequently during the third session (Figure 6), and eventually learned the neutral feedback as the best choice, which may have resulted in increased verbal responses. These positive findings of emphatic dialog are consistent with the recent findings identified by (Alves-Oliveira et al. 2019).

In summary, we understand that the overall social engagement measured in terms of eye-gaze, facial expressions, verbal responses, and gestures, either remained consistent, sustained, or enhanced across all sessions. Thus, **H1** was accepted.

We refer to the section 5.4, Table 5, Figures 8 and 9, to compare the results for all of the DVs used to measure social engagement in the current (in blue) and the previous study (in black). We observed difference between the two studies. In the current study, we observed a consistent duration of eye gaze towards the robot in the first, second and third session, where we did not find a statistical significance. But, later on, remarkably, a sharp increase of 150% on average was seen in the fourth session in the duration of eye gaze in comparison to the first session. On the contrary, in the previous study, a consistent eye gaze towards the robot was observed from first to the fourth session. We observed a very minor

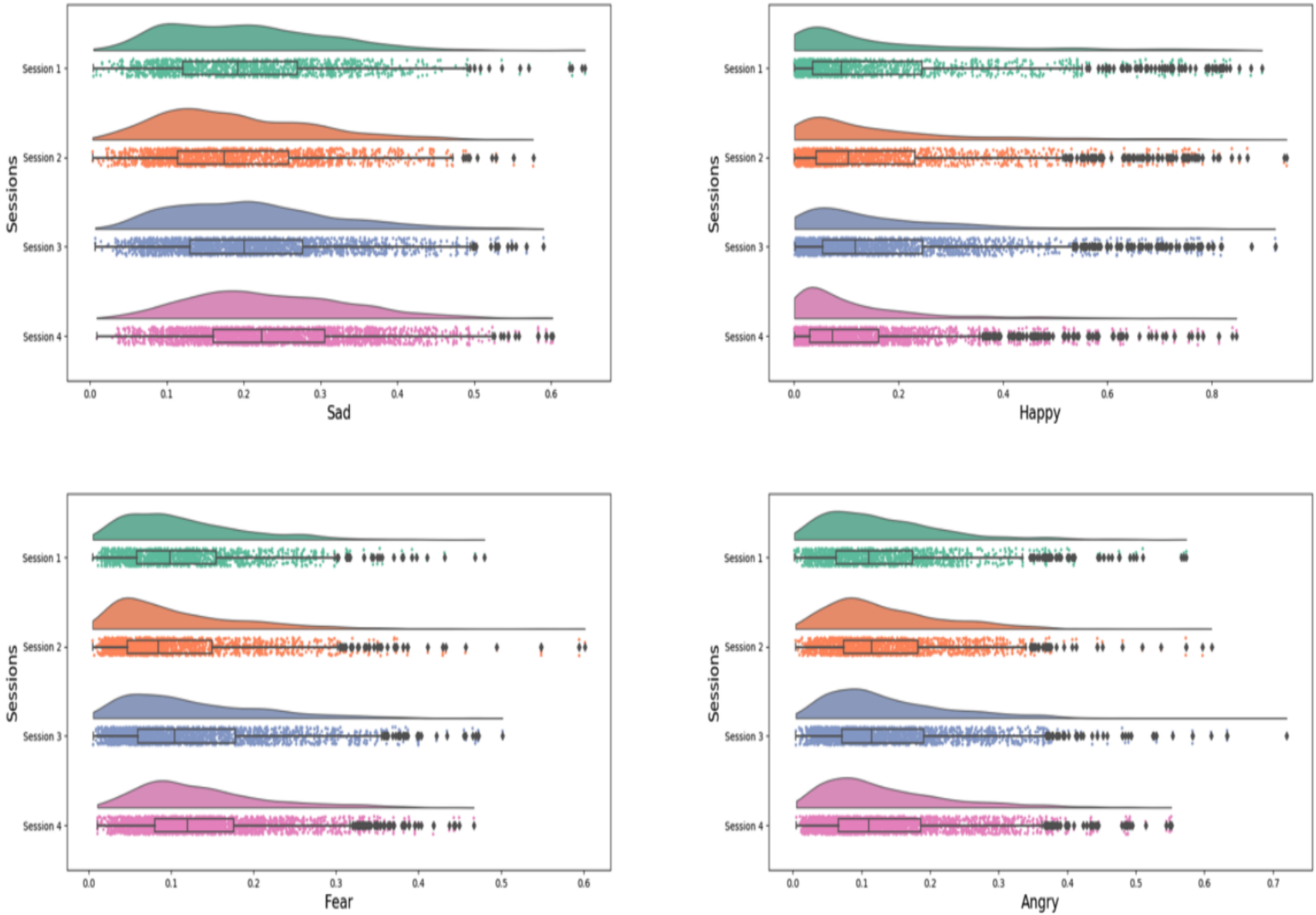


Figure 7.: The Figure shows the overall mean value of the emotions (Happy, Sad, Fear, and Angry emotions) displayed by the participants and detected by the automatic system across first, second, third, and fourth sessions. The figure provides an idea of thee distribution of the data points for the four emotions across sessions, and makes it easy to determine differences at a glance during different sessions.

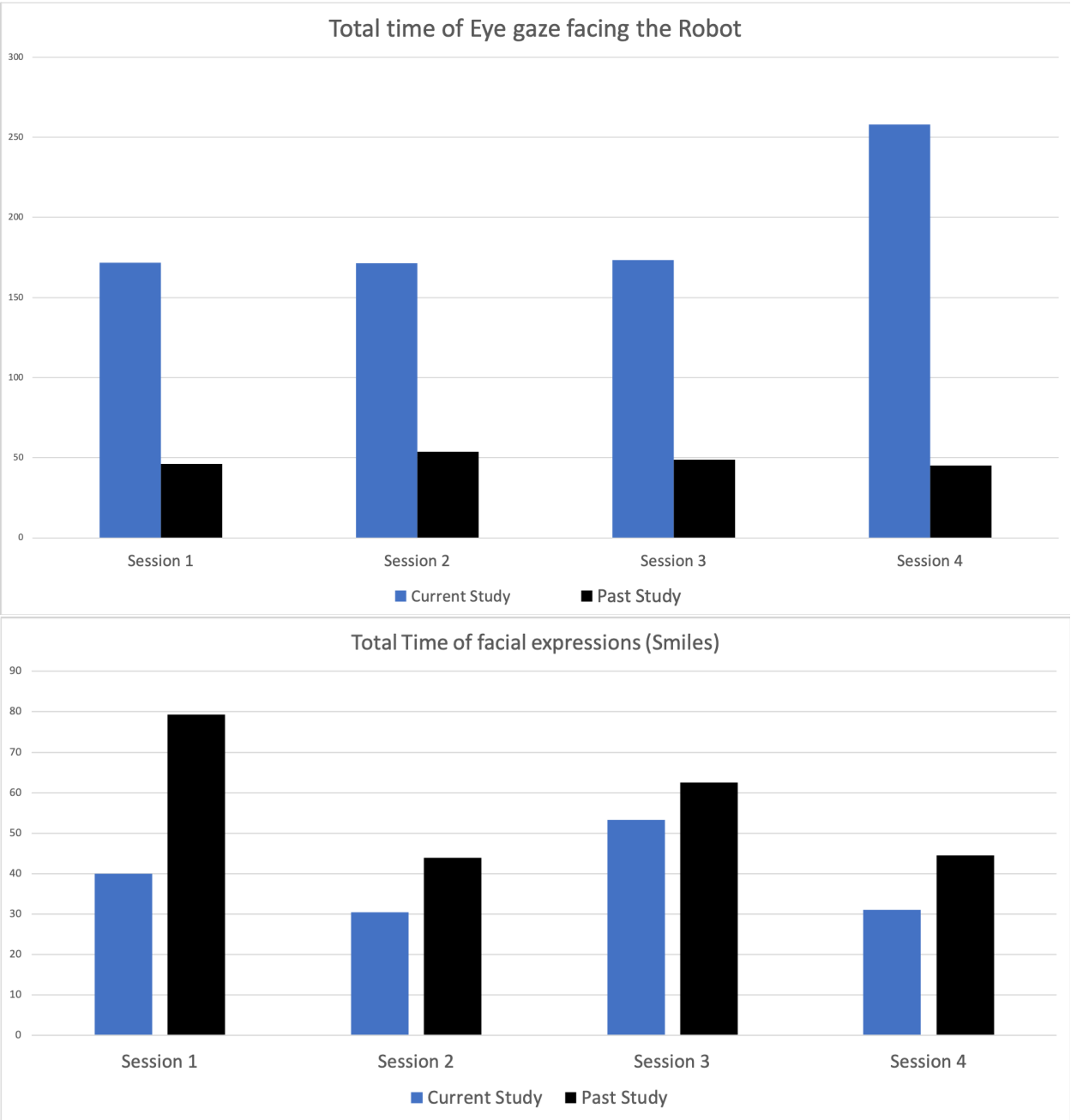
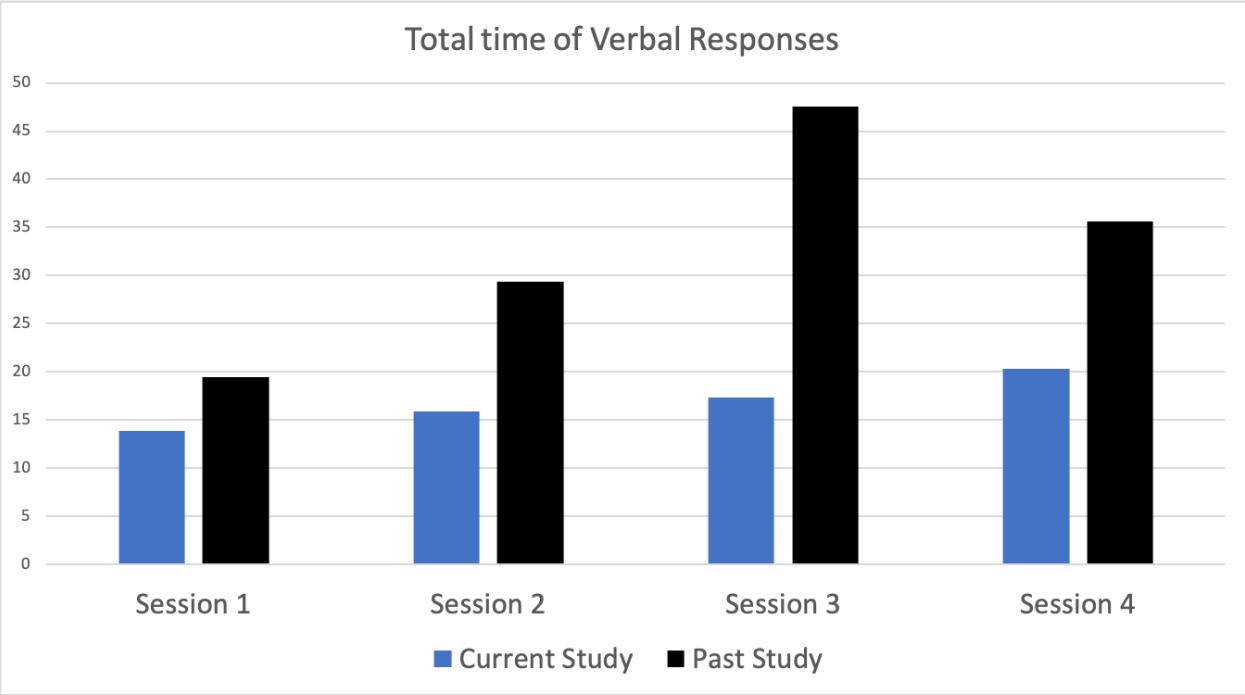


Figure 8.: Mean (M) for the total time of Eye-gaze and facial expressions measured in seconds.milliseconds format during the current and previous (Ahmad et al. 2017) studies.



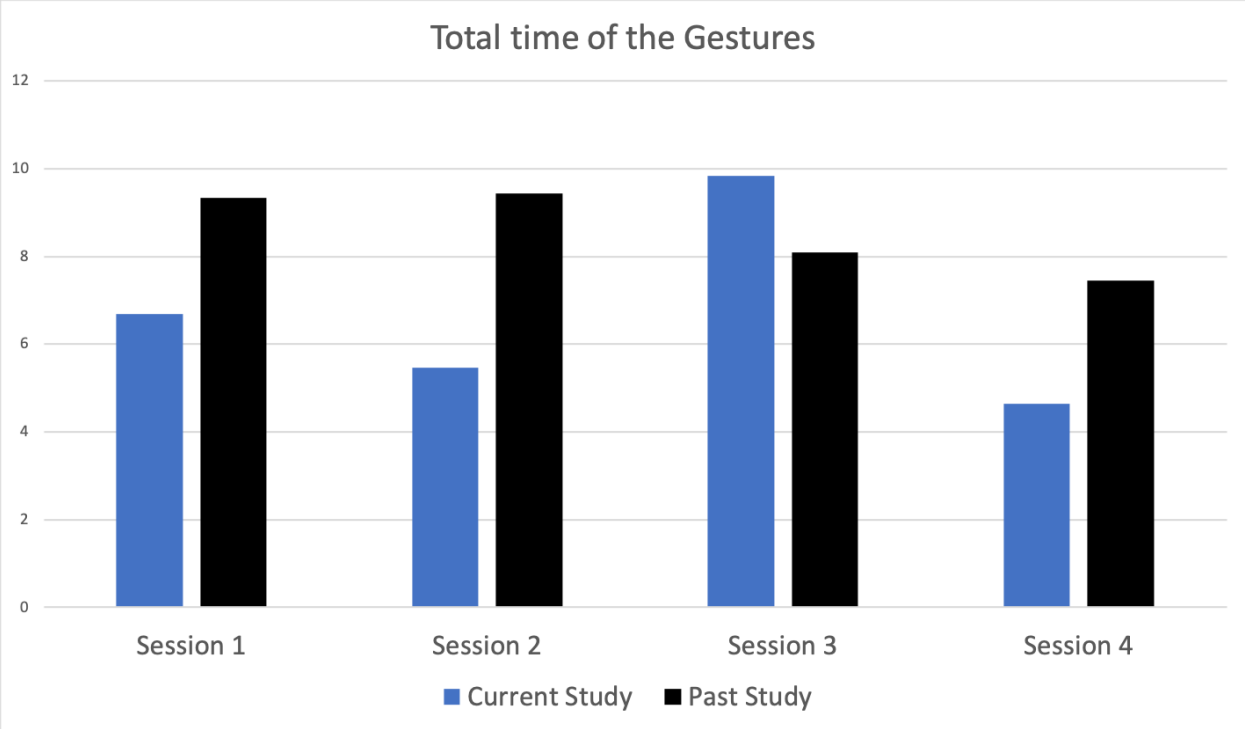


Figure 9.: Mean (M) for the total time of verbal responses and gestures measured in seconds.milliseconds format during the current and previous (Ahmad et al. 2017) studies.

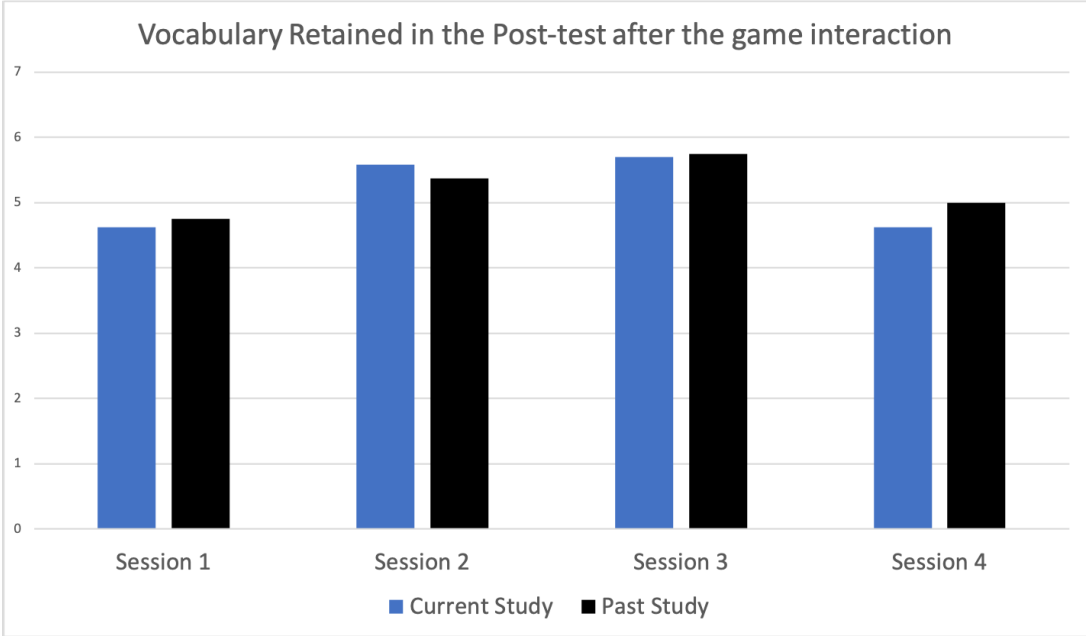


Figure 10.: Mean (M) for the vocabulary retained in the post-test conducted after the game interaction.

	Current Study V.S. Previous Study (Ahmad et al. 2017)			
Dependent Variables	**Session 1**	**Session 2**	**Session 3**	**Session 4**
Eye-Gaze across sessions	M: 171.80, SD: 75.18 M: 46.08, SD: 13.80	M: 171.34, SD: 74.64 M: 53.87, SD: 23.20	M: 173.40, SD: 77.00 M: 48.70, SD: 20.80	M: 258.01, SD: 135.72 M: 45.25, SD: 24.18
Facial Expressions across sessions	M: 39.96, SD: 25.61 M: 79.24, SD: 69.31	M: 30.41, SD: 26.13 M: 43.94, SD: 45.22	M: 53.22, SD: 40.31 M: 62.47, SD: 54.18	M: 30.98, SD: 26.18 M: 44.48, SD: 34.36
Verbal Response across sessions	M: 13.93, SD: 9.95 M: 19.42, SD: 21.00	M: 15.94, SD: 8.07 M: 29.43, SD: 24.34	M: 17.33, SD: 8.23 M: 47.57, SD: 34.53	M: 20.30, SD: 5.94 M: 35.67, SD: 18.35
Gestures across sessions	M: 6.70, SD: 5.65 M: 9.35, SD: 22.61	M: 5.46, SD: 7.20 M: 9.44, SD: 12.18	M:9.85, SD: 15.82 M: 8.10, SD: 12.53	M: 4.63, SD: 4.38 M: 7.45, SD: 13.22
Vocabulary retention across sessions	M: 4.62, SD: 0.96 M: 4.75, SD: 0.98	M: 5.58, SD: 0.77 M: 5.37, SD: 0.76	M: 5.70, SD: 0.55 M: 5.75, SD: 0.60	M: 4.62, SD: 0.87 M: 5, SD: 0.65

Table 5.: Mean (M) and Standard deviation (SD) for Eye-gaze, facial expressions, verbal responses and gestures, measured in seconds.milliseconds format, during the current and previous (Ahmad et al. 2017) studies. The values in **blue** represents the results of the current study and the values in **black** represents the results of the previous study.

decrease of 1% on average in the duration of eye-gaze in the fourth session in comparison to the first session. Further, there was no statistically significant difference observed between sessions. The sharp increase in the fourth session of the current study can be partly associated with the learning mechanism introduced in the BSU module of the emotion and memory model along with other afore-described conjectures. We further see the difference between the trends of two studies for the verbal response. The duration of verbal response increased significantly from the first to, and second to the fourth session referring to an on average increase of 11.8% across the sessions. In the previous study, the verbal responses declined significantly from the third session to the fourth session. This referred to a decline of nearly 25% in the duration of verbal responses in the fourth session in comparison to the third session. This finding further affirms the value of the current work. We also witnessed differences in the trends for the duration of the facial expressions. In the previous study, the duration of facial expressions declined significantly from the first to the second (nearly 45%) and the fourth (nearly 44%) session. On the contrary, in the current study, the duration of expressions did not decline significantly from the first to the second (nearly 25%) and to the fourth (nearly 25%) session. The conjectured reasons for the significant difference of the third session with second and fourth session have been afore-described. However, it does not discount the finding that the values were more consistent as compared to the previous study when compared from first to the fourth session. Lastly, we witnessed similar trends for the duration of gestures.

The statistical analysis comparing our current and prior study further showed that current study observed higher engagement in terms of eye gaze. It further showed that previous study observed higher engagement in terms of verbal response and gestures. However, we remain mindful of the differences in the absolute and hence mean values of the eye-gaze, verbal responses and gestures between two studies. This could be due to the following reasons. First, we understand that there was indeed variations in the experimental setup. In the current study, children interacted once a week for four weeks, where as, previously, they had interacted twice a week. Moreover, the sitting posture of the robot also varied because of logistical settings. Second, the study was conducted at two different schools, and potentially there can be demographic and behavioral differences among the participants. Likely, a group of children may not behave similarly to the other one. Nonetheless, we cannot discount the factor that the introduction of the learning method to select the behaviour did play a role in the *overall* increase in the duration of eye-gaze from first to the fourth session and more consistent rise of the verbal responses from first to the fourth session in comparison to the previous study.

It may have been due to a similar division of expressiveness and un-expressiveness among participants. It is due to the observation that waving or showing fist are commonly associated with expressing oneself. Note that, we assume that any of these effects would have been balanced out due to the equal distribution of such traits among participants. In summary, the finding showing a significant increase in the duration of the eye-gaze in the last session and showing higher values of eye-gaze in the current study. In addition, we also see a more consistent and significant increase in the duration of verbal response in the current study over sessions. We understand that all this asserts the value of the extension of the BSU of the model.

We see that the session had a significant effect in terms of immediate vocabulary retention. In particular, the immediate vocabulary retained during the second and third sessions was significant to the first and fourth session. We believe that the following reasons may account for these findings. First, it can be seen that the engagement measured in terms of facial expressions and verbal response during the third session is significantly high compared to the first session. Therefore, keeping the engagement and learning outcome in mind (Carini, Kuh, and Klein 2006) and also understanding that improved engagement results in superior performance in the human adaptive agent interaction scenario (Szafir and Mutlu 2012), we believe this explains our finding. Compared to our previous study (Table 5, Figure 10), in the current study, we witnessed a similar vocabulary-learning trend, except for the higher vocabulary retention for words used in the second session. We believe that the reason for a decline in vocabulary retention during the last session could be because the children were notified that this was the last interaction session. We suspect that the children might not have participated in the post-test with similar

enthusiasm or interest as they knew that they would not be tested for the similar words again. Consequently, the children may have done the post-test quickly. Similar observations have been reported by other researchers (Rosenthal-von der Pütten, Straßmann, and Krämer 2016). In summary, the children's immediate retention of words was encouraging as they retained 5 out of 6 words in all the sessions. However, a slight decline was seen in the last session. Hence, our **H2a** was partially accepted.

In terms of delayed retention of vocabulary for the words tested across sessions during the post-test, we understand that the updated behavior selection unit applying personalised policy enabled the robot to learn the feedback based on the children's learning outcomes, and it reflects on their learning performance across sessions. Especially, the combination of emotional feedback enabling the robot to express both positive and negative emotions as the part of its feedback behavior may have impacted the vocabulary retention as highlighted in the literature that the information of the topics associated with positive emotional content was remembered better than the neutral neutral ones (Christianson and Loftus 1991). A decline in vocabulary retention during the fourth session may have been due to a lack of interest in the task as described earlier children might not have participated with similar enthusiasm or interest like others (Rosenthal-von der Pütten, Straßmann, and Krämer 2016). As mentioned earlier, children may have done it quickly due because it was the last interaction session. We further understand that emotional feedback may have a positive impact on the delayed retention of vocabulary. It is a common finding that humans retained information for a longer period under positive emotional states (Baumeister et al. 2007). Positive (encouraging) feedback may have resulted in creating this effect. We also believe that our finding are preliminary and require further exploration. In summary, children's delayed retention of words was also encouraging and our **H2b** was partially accepted.

The delayed retention of the words learned in the first, second and third session also highlight the practical value of the presented child-robot-game interaction setup, of reinforcing over sessions. Children on average were able to retain or recall 5 out of 6 words for each session in the subsequent sessions. This shows that similar methods can be used with children to teach them vocabulary and other content. For example, we applied a similar setup in a mathematics learning task (Ahmad and Mubin 2018). In this study we showed that the learning performance based on calculating area and perimeter of regular and irregular shapes improved as children interacted with the robot on three different yet separate occasions.

7. Limitations

Our results are based on a low number of participants, and, therefore, are of exploratory nature. We did not run statistical analysis to compare the results based on the effects of the updated model on the social engagement with the previous study for two reasons. First, the robot's posture differed during both studies. Second, this study was performed for four weeks rather than two weeks. We were limited due to logistical issues. However, we did reflect on the trends in the of both studies in our discussion of results. The trends show differences of significant nature suggesting that a consistent increase in the total time of gaze facing robot and verbal response given to the robot. We believe these trends reflect on the value of the updated model in comparison to the previous version (Ahmad et al. 2019). Note that, this study builds on previous works, we also collected data with a non-adaptive version of the setup in (Ahmad, Mubin, and Orlando 2017) and we found that the total time of the gaze, smiles, verbal responses, and gestures decline significantly after the second session.

We were limited in our approach to detect emotional state every 10 seconds during the game. The choice was dependeny on the pace of the game in terms of user and robot turns. We acknowlege the loss of data because the emotion may have changed during the 10 second interval. Nonetheless, it is important to note that detecting human emotion in real-time is a challenge and hence we understand the limitation of the method. For instance, a recent review suggests that the reliability of current automatic emotion detection systems based on the Facial Action Coding to measure facial activity may not match the humans (Calvo and D'Mello 2010). We hope to improve our system in the future evaluation of our

model. It is important to note that the gesture recognition was dependent on specific lighting and was although robust but was also limited in nature.

The policy was not applied individually to each participant due to the limited iterations per participant in the game. We instead placed all interactions with all the children together. It can be claimed that the robot may not have personalised its policy to each participant in this case, and the robot may have applied the same actions to all the children. We argue that the algorithm adapts to a pattern of behavior that most likely happens among all the participants during our interaction as the game behavior was pre-defined or fixed. This reflects the rationale of enabling the algorithm to adapt to all the participants as a whole but not to each individual. In addition, the current research tries to present a proof of concept and it did not apply same actions to all of the participants and this was evident from the actions selected during first three sessions and the learning of the likeable action was reflected in the fourth session.

We did not choose to collect data on the children's preferences regarding their likability for encouraging or competitive feedback. The robot's feedback in principle is positive, negative, and neutral and has partial or limited elements in terms of giving encouragement and being critical or competitive depending on the task.

Finally, the vocabulary learning results were based on the recognition. We did not evaluate the model to investigate its impact on the pronunciation or other sections of language learning.

8. Conclusion

In this paper, we attempted to contribute the following knowledge in the field of Human-Robot Interaction. First, we demonstrated an extended version of the emotion and memory model by updating the behaviour selection unit. In this model we applied the MAB algorithm Exp3 to learn the best strategy to select a behaviour for the given user according to their current level of social engagement and/or learning outcome. Second, we implemented a method to compute a reward function by taking inputs from the multi-modal channels. We used external cameras to capture the facial expressions, gesture, and eye gazes. Further, we used microphone to record verbal responses. This reward function represented the current level of social engagement. Third, we evaluated the model to see whether children's engagement is sustained in a vocabulary-learning task during a children-game-robot interaction, where children interacted with the robot on four different occasions. Fourth, we found that children social engagement maintained during all the sessions. We also compared the findings of this study with the findings of our previous work, where we did not use reinforcement learning (rather rule-based) to select behaviours in the behaviour selection unit of the emotion and memory model. We noticed a higher amount of engagement in terms of eye gaze towards the robot. Further, we observed a more increasing amount of verbal response from the first to the last session for the current work, with an average increase of 11.8%.

We understand and acknowledge the limitations of our reported study which can be addressed as future work. First, we can carefully appoint a control condition to validate the model in other settings. It can be testing the model with or without reinforcement learning. Further, it can be using different reinforcement learning algorithms to select the behaviours such as Exp3 v.s. policy gradient based solution for Exp3 problem, together with meta-learning (see Gao et al. (2019)). Secondly, we can improve the method to compute the emotional state of the users by using a state of the art emotion recognition API and also rethinking the method to compute emotion on the given state in the game. Thirdly, we can create an experimental task where the robot interaction policy can be applied individually for each participant.

References

Ahmad, Muneeb, Omar Mubin, and Paola Escudero. 2015. "Using adaptive mobile agents in games based scenarios to facilitate foreign language word learning." In *Proceedings of the 3rd International Conference on Human-Agent Interaction*, 255–257. ACM.

Ahmad, Muneeb, Omar Mubin, and Joanne Orlando. 2016a. "Children views' on social robot's adaptations in education." In *Proceedings of the 28th Australian Conference on Computer-Human Interaction*, 145–149. ACM.

Ahmad, Muneeb, Omar Mubin, and Joanne Orlando. 2016b. "Understanding behaviours and roles for social and adaptive robots in education: teacher's perspective." In *Proceedings of the Fourth International Conference on Human Agent Interaction*, 297–304. ACM.

Ahmad, Muneeb, Omar Mubin, and Joanne Orlando. 2017. "Adaptive Social Robot for Sustaining Social Engagement during Long-Term Children–Robot Interaction." *International Journal of Human–Computer Interaction* 1–20.

Ahmad, Muneeb, Omar Mubin, Suleman Shahid, and Joanne Orlando. 2017. "Emotion and memory model for a robotic tutor in a learning environment." In *Proceedings of the Seventh ISCA workshop on Speech and Language Technology in Education 2017, August 25-26, 2017, Djur, Stockholm, Sweden*, 26–32.

Ahmad, Muneeb Imtiaz, and Omar Mubin. 2018. "Emotion and memory model to promote mathematics learning - an exploratory long-term study." In *Proceedings of the 6th International Conference on Human-Agent Interaction*, 214–221.

Ahmad, Muneeb Imtiaz, Omar Mubin, Suleman Shahid, and Joanne Orlando. 2019. "Robots adaptive emotional feedback sustains childrens social engagement and promotes their vocabulary learning: a long-term childrobot interaction study." *Adaptive Behavior* 0 (0): 1059712319844182. https://doi.org/10.1177/1059712319844182.

Alves-Oliveira, Patrícia, Pedro Sequeira, Francisco S Melo, Ginevra Castellano, and Ana Paiva. 2019. "Empathic Robot for Group Learning: A Field Study." *ACM Transactions on Human-Robot Interaction (THRI)* 8 (1): 3.

Bajraktarevic, Namira, Wendy Hall, and Patrick Fullick. 2003. "Incorporating learning styles in hypermedia environment: Empirical evaluation." In *Proceedings of the workshop on adaptive hypermedia and adaptive web-based systems*, 41–52.

Balakrishnan, Vimala, Marian Cynthia Martin, Wandeep Kaur, and Amir Javed. 2019. "A comparative analysis of detection mechanisms for emotion detection." In *Journal of Physics: Conference Series*, Vol. 1339, 012016. IOP Publishing.

Baumeister, Roy F, Kathleen D Vohs, C Nathan DeWall, and Liqing Zhang. 2007. "How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation." *Personality and social psychology review* 11 (2): 167–203.

Baxter, Paul, Emily Ashurst, Robin Read, James Kennedy, and Tony Belpaeme. 2017. "Robot education peers in a situated primary school study: Personalisation promotes child learning." *PLoS One* 12 (5): e0178126.

Bower, Gordon H. 1992. "How might emotions affect learning." *The handbook of emotion and memory: Research and theory* 3: 31.

Bradley, Margaret M, Mark K Greenwald, Margaret C Petry, and Peter J Lang. 1992. "Remembering pictures: pleasure and arousal in memory." *Journal of experimental psychology: Learning, Memory, and Cognition* 18 (2): 379.

Bradski, Gary, and Adrian Kaehler. 2000. "OpenCV." *Dr. Dobbs journal of software tools* 3.

Breazeal et al., Cynthia. 2003. "Toward sociable robots." *Robotics and autonomous systems* 42 (3-4): 167–175.

Bubeck, Sébastien, Nicolo Cesa-Bianchi, et al. 2012. "Regret analysis of stochastic and nonstochastic multi-armed bandit problems." *Foundations and Trends® in Machine Learning* 5 (1): 1–122.

Cahill, Larry, Richard J Haier, James Fallon, Michael T Alkire, Cheuk Tang, David Keator, Joseph Wu, and James L Mcgaugh. 1996. "Amygdala activity at encoding correlated with long-term, free recall of emotional information." *Proceedings of the National Academy of Sciences* 93 (15): 8016–8021.

Calvo, Rafael A, and Sidney D'Mello. 2010. "Affect detection: An interdisciplinary review of models, methods, and their applications." *IEEE Transactions on affective computing* 1 (1): 18–37.

Canli, Turhan, Zuo Zhao, James Brewer, John DE Gabrieli, and Larry Cahill. 2000. "Event-related activation in the human amygdala associates with later memory for individual emotional experience." *Journal of Neuroscience* 20 (19): RC99–RC99.

Carini, Robert M, George D Kuh, and Stephen P Klein. 2006. "Student engagement and student learning: Testing the linkages." *Research in higher education* 47 (1): 1–32.

Castro-González, Álvaro, María Malfaz, and Miguel A Salichs. 2011. "Learning the selection of actions for an autonomous social robot by reinforcement learning based on motivations." *International Journal of Social Robotics* 3 (4): 427–441.

Christianson, Sven-Åke, and Elizabeth F Loftus. 1991. "Remembering emotional events: The fate of detailed information." *Cognition & Emotion* 5 (2): 81–108.

Cole, Pamela M, Carolyn Zahn-Waxler, Nathan A Fox, Barbara A Usher, and Jean D Welsh. 1996. "Individual differences in emotion regulation and behavior problems in preschool children." *Journal of Abnormal Psychology* 105 (4): 518.

Comblain, Christine, Arnaud D'Argembeau, and Martial Van der Linden. 2005. "Phenomenal characteristics of autobiographical memories for emotional and neutral events in older and younger adults." *Experimental Aging Research* 31 (2): 173–189.

D'Argembeau, Arnaud, Christine Comblain, and Martial Van der Linden. 2003. "Phenomenal characteristics of autobiographical memories for positive, negative, and neutral events." *Applied Cognitive Psychology* 17 (3): 281–294.

Dautenhahn, Kerstin. 1998. "The art of designing socially intelligent agents: Science, fiction, and the human in the loop." *Applied artificial intelligence* 12 (7-8): 573–617.

de Haas, Mirjam, Paul Vogt, and EJ Krahmer. 2016. "Enhancing child-robot tutoring interactions with appropriate feedback." In *Proceedings of First Workshop on Long-Term Child-Robot Interaction. IEEE Ro-Man,(New York, NY)*, .

Dolcos, Florin, Kevin S LaBar, and Roberto Cabeza. 2005. "Remembering one year later: role of the amygdala and the medial temporal lobe memory system in retrieving emotional memories." *Proceedings of the National Academy of Sciences* 102 (7): 2626–2631.

Dorça et al., Fabiano. 2015. "Implementation and use of Simulated Students for Test and Validation of new Adaptive Educational Systems: a Practical Insight." *International Journal of Artificial Intelligence in Education* 25 (3): 319–345. `https://doi.org/10.1007/s40593-015-0037-0`.

DMello, Sidney, and Art Graesser. 2012. "Dynamics of affective states during complex learning." *Learning and Instruction* 22 (2): 145–157.

Eriksson, Elisabeth, Lisa Björklund Boistrup, and Robert Thornberg. 2017. "A categorisation of teacher feedback in the classroom: a field study on feedback based on routine classroom assessment in primary school." *Research Papers in Education* 32 (3): 316–332.

Gao, Yuan, Wolmet Barendregt, Mohammad Obaid, and Ginevra Castellano. 2018. "When robot personalisation does not help: Insights from a robot-supported learning study." In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 705–712. IEEE.

Gao, Yuan, Elena Sibirtseva, Ginevra Castellano, and Danica Kragic. 2019. "Fast adaptation with meta-reinforcement learning for trust modelling in human-robot interaction." In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 305–312. IEEE.

Ghadirzadeh, Ali, Judith Bütepage, Atsuto Maki, Danica Kragic, and Mårten Björkman. 2016. "A sensorimotor reinforcement learning framework for physical human-robot interaction." In *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*, 2682–2688. IEEE.

Goldfield, Beverly A, and J Steven Reznick. 1990. "Early lexical acquisition: Rate, content, and the vocabulary spurt." *Journal of child language* 17 (01): 171–183.

Gordon, Goren, Samuel Spaulding, Jacqueline Kory Westlund, Jin Joo Lee, Luke Plummer, Marayna Martinez, Madhurima Das, and Cynthia Breazeal. 2016. "Affective Personalization of a Social Robot Tutor for Children's Second Language Skills." In *AAAI*, 3951–3957.

Graf, Sabine, Tzu-Chien Liu, and K Kinshuk. 2008. "Interactions between students learning styles, achievement and behaviour in mismatched courses." In *Proceedings of the international conference on cognition and exploratory learning in digital age (CELDA 2008). IADIS International Conference*, 223–230. Citeseer.

Haider, M, A Sinha, and B Chaudhary. 2010. "An Investigation of relationship between learning styles and performance of learners." *International Journal of Engineering Science and Technology* 2 (7): 2813–2819.

Indico. 2016. "Facial Emotion Recognition." `https://indico.io/docs`. Accessed: 2016-06-08.

Irfan, Bahar, Aditi Ramachandran, Samuel Spaulding, Dylan F Glas, Iolanda Leite, and Kheng Lee Koay. 2019. "Personalization in Long-Term Human-Robot Interaction." In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 685–686. IEEE.

Jones, Aidan, Susan Bull, and Ginevra Castellano. 2017. "I Know That Now, Im Going to Learn This Next Promoting Self-regulated Learning with a Robotic Tutor." *International Journal of Social Robotics* 1–16.

Jones, Aidan, and Ginevra Castellano. 2018. "Adaptive Robotic Tutors that Support Self-Regulated Learning: A Longer-Term Investigation with Primary School Children." *International Journal of Social Robotics* 1–14.

Kaelbling, Leslie Pack, Michael L Littman, and Andrew W Moore. 1996. "Reinforcement learning: A survey." *Journal of artificial intelligence research* 4: 237–285.

Kanda, Takayuki, Takayuki Hirano, Daniel Eaton, and Hiroshi Ishiguro. 2004. "Interactive robots as social partners

and peer tutors for children: A field trial." *Human-computer interaction* 19 (1): 61–84.

Kasinski, Andrzej, and Adam Schmidt. 2010. "The architecture and performance of the face and eyes detection system based on the Haar cascade classifiers." *Pattern Analysis and Applications* 13 (2): 197–211.

Kazui, Hiroaki, Etsuro Mori, Mamoru Hashimoto, Nobutsugu Hirono, Toru Imamura, Satoshi Tanimukai, Tokiji Hanihara, and Larry Cahill. 2000. "Impact of emotion on memory: Controlled study of the influence of emotionally charged material on declarative memory in Alzheimer's disease." *The british journal of psychiatry* 177 (4): 343–347.

Kelsey, Elaine, Fritz Ray, Debbie Brown, and Robby Robson. 2015. "Design of a domain-independent, interactive, dialogue-based tutor for use within the GIFT framework." In *Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym3)*, 161.

Kennedy, James, Paul Baxter, Emmanuel Senft, and Tony Belpaeme. 2016. "Social robot tutoring for child second language learning." In *Human-Robot Interaction (HRI), 2016 11th ACM/IEEE International Conference on*, 231–238. IEEE.

Kim, Aelee, Jooyun Han, Younbo Jung, and Kwanmin Lee. 2013. "The effects of familiarity and robot gesture on user acceptance of information." In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 159–160. IEEE.

Knapp, Mark L, Roderick P Hart, Gustav W Friedrich, and Gary M Shulman. 1973. "The rhetoric of goodbye: Verbal and nonverbal correlates of human leave-taking." *Communications Monographs* 40 (3): 182–198.

Komatsubara, Tsuyoshi, Masahiro Shiomi, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2014. "Can a social robot help children's understanding of science in classrooms?" In *Proceedings of the second international conference on Human-agent interaction*, 83–90. ACM.

LeDoux, J. 2007. "Emotional memory. Scholarpedia 2 (7): 1806." .

LeDoux, Joseph E. 1993. "Emotional memory systems in the brain." *Behavioural brain research* 58 (1-2): 69–79.

Leite, Iolanda. 2013a. "Long-term interactions with empathic social robots." PhD diss., Ph.D. UNIVERSIDADE TECNICA DE LISBOA INSTITUTO SUPERIOR TECNICO.

Leite, Iolanda. 2013b. "Long-term Interactions with Empathic Social Robots." PhD diss., Instituto Superior Tcnico, Universidade de Lisboa.

Leite, Iolanda. 2015. "Long-term interactions with empathic social robots." *AI Matters* 1 (3): 13–15.

Leite, Iolanda, Ginevra Castellano, André Pereira, Carlos Martinho, and Ana Paiva. 2014. "Empathic robots for long-term interaction." *International Journal of Social Robotics* 6 (3): 329–341.

Leite, Iolanda, Carlos Martinho, and Ana Paiva. 2013. "Social robots for long-term interaction: a survey." *International Journal of Social Robotics* 5 (2): 291–308.

Levine, Linda J, and Stewart L Burgess. 1997. "Beyond general arousal: Effects of specific emotions on memory." *Social Cognition* 15 (3): 157.

Levine, Linda J, and David A Pizarro. 2004. "Emotion and memory research: A grumpy overview." *Social cognition* 22 (5: Special issue): 530–554.

Leyzberg, Daniel, Samuel Spaulding, and Brian Scassellati. 2014. "Personalizing robot tutors to individuals' learning differences." In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, 423–430. ACM.

Liu, Tzu-Chien, Sabine Graf, et al. 2009. "Coping with mismatched courses: students behaviour and performance in courses mismatched to their learning styles." *Educational Technology Research and Development* 57 (6): 739.

Mahajan, Aditya, and Demosthenis Teneketzis. 2008. "Multi-armed bandit problems." In *Foundations and Applications of Sensor Management*, 121–151. Springer.

Mubin et al., Omar. 2011. "ROILA: RObot Interaction LAnguage." PhD diss., Technische Universiteit Eindhoven.

Mutlu, Bilge, Steven Osman, Jodi Forlizzi, Jessica Hodgins, and Sara Kiesler. 2006. "Perceptions of ASIMO: an exploration on co-operation and competition with humans and humanoid robots." In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, 351–352. ACM.

Omar Mubin, Christoph Baternek. 2015. "Vocabulary." `http://roila.org/language-guide/vocabulary/`. Accessed 23-September-2015.

Padilla, R, CFF Costa Filho, and MGF Costa. 2012. "Evaluation of haar cascade classifiers designed for face detection." *World Academy of Science, Engineering and Technology* 64: 362–365.

Ramachandran, Aditi, and Brian Scassellati. 2014. "Adapting difficulty levels in personalized robot-child tutoring interactions." In *Proceedings of the Workshops at the 28th AAAI Conference on Artificial Intelligence*, .

Ramachandran, Aditi, and Brian Scassellati. 2015. "Fostering learning gains through personalized robot-child

tutoring interactions." In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, 193–194. ACM.

Ravaja, Niklas, Timo Saari, Mikko Salminen, Jari Laarni, and Kari Kallinen. 2006. "Phasic emotional reactions to video game events: A psychophysiological investigation." *Media psychology* 8 (4): 343–367.

Rieber, Lloyd P. 1996. "Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games." *Educational technology research and development* 44 (2): 43–58.

Ritschel, Hannes, and Elisabeth André. 2017. "Real-Time Robot Personality Adaptation based on Reinforcement Learning and Social Signals." In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 265–266. ACM.

Rosenthal-von der Pütten, Astrid M, Carolin Straßmann, and Nicole C Krämer. 2016. "Robots or agents–neither helps you more or less during second language acquisition." In *International Conference on Intelligent Virtual Agents*, 256–268. Springer.

Roy, Sayanti, Emily Kieson, Charles Abramson, and Christopher Crick. 2018. "Using human reinforcement learning models to improve robots as teachers." In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 225–226. ACM.

Serholt, Sofia. 2018. "Breakdowns in children's interactions with a robotic tutor: A longitudinal study." *Computers in Human Behavior* 81: 250–264.

Serholt, Sofia, and Wolmet Barendregt. 2016. "Robots tutoring children: Longitudinal evaluation of social engagement in child-robot interaction." In *Proceedings of the 9th nordic conference on human-computer interaction*, 1–10.

Sottilare, Robert A, Keith W Brawner, Benjamin S Goldberg, and Heather K Holden. 2012. "The generalized intelligent framework for tutoring (GIFT)." .

Spitale, Micol, Fabio Catania, Giulia Cosentino, Mirko Gelsomini, and Franca Garzotto. 2019. "WIYE: building a corpus of children's audio and video recordings with a story-based app." In *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion*, 33–34.

Szafir, Daniel, and Bilge Mutlu. 2012. "Pay attention!: designing adaptive agents that monitor and improve user engagement." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 11–20. ACM.

Tozadore, Daniel Carnieto, João Pedro Hannauer Valentini, Victor Henrique de Souza Rodrigues, Fernando Moura Leite Vendrameto, Rodrigo Geurgas Zavarizz, and Roseli Aparecida Francelin Romero. 2018. "Towards Adaptation and Personalization in Task Based on Human-Robot Interaction." In *2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE)*, 383–389. IEEE.

Tsiakas, Konstantinos, et al. 2018. "Interactive Learning and Adaptation for Personalized Robot-Assisted Training." PhD diss.

Vygotsky, Lev Semenovich. 1980. *Mind in society: The development of higher psychological processes*. Harvard university press.

Wagner, Johannes, Florian Lingenfelser, Tobias Baur, Ionut Damian, Felix Kistler, and Elisabeth André. 2013. "The social signal interpretation (SSI) framework: multimodal signal processing and recognition in real-time." In *Proceedings of the 21st ACM international conference on Multimedia*, 831–834.

Zaga, Chritina, Manja Lohse, Khiet P. Truong, and Vanessa Evers. 2015. "The Effect of a robot's social character on children task engagement: peer vs tutor." In *Proceedings of the ICSR2015 Conference*, Springer.