Exploring task and social engagement in companion social robots: a comparative analysis of feedback types

Advanced Robotics

ISSN: 0169-1864 (Print) 1568-5535 (Online) Journal homepage: www.tandfonline.com/journals/tadr20

Exploring task and social engagement in companion social robots: a comparative analysis of feedback types

Bahram Salamat Ravandi, Imran Khan, Alva Markelius, Martin Bergström, Pierre Gander, Engin Erzin & Robert Lowe

To cite this article: Bahram Salamat Ravandi, Imran Khan, Alva Markelius, Martin Bergström, Pierre Gander, Engin Erzin & Robert Lowe (2025) Exploring task and social engagement in companion social robots: a comparative analysis of feedback types, Advanced Robotics, 39:14, 884-899, DOI: 10.1080/01691864.2025.2526668

To link to this article: https://doi.org/10.1080/01691864.2025.2526668

ADVANCED ROBOTICS
2025, VOL. 39, NO. 14, 884–899
https://doi.org/10.1080/01691864.2025.2526668

FULL PAPER

OPEN ACCESS Check for updates

Exploring task and social engagement in companion social robots: a comparative analysis of feedback types

Bahram Salamat Ravandi[a], Imran Khan[a], Alva Markelius[a,b], Martin Bergström[a], Pierre Gander[a], Engin Erzin[c] and Robert Lowe[a,d]

[a]Department of Applied IT, University of Gothenburg, Gothenburg, Sweden; [b]Leverhulme Centre for the Future of Intelligence (CFI), University of Cambridge, Cambridge, UK; [c]Departments of Electrical and Electronics Engineering and Computer Engineering, Koç University, Istanbul, Turkiye; [d]Unit of Human-Centred AI, Research Institutes of Sweden – RISE, Gothenburg, Sweden

ABSTRACT
In recent years, the integration of social robots into various domains has received significant attention due to their potential to engage users in meaningful ways, offering companionship, support, and assistance in tasks, particularly in healthcare. This study investigates the impact of different types of feedback provided by the social robot Furhat on user engagement during a digital visuospatial memory training task. Using a $3 \times 2 \times 2$ mixed design ($N = 58$), we investigated three types of feedback: performance-based, affective-based, and a combination of both, across two levels of challenge (Easy and Medium) between subjects, incorporating a within-subject baseline control block. The results indicate that affective-based feedback leads to significantly higher social engagement, as evidenced by higher eye contact with the robot. However, this higher social engagement is associated with lower task performance in the affective-based feedback condition. Additionally, participants perceived the social robot as more user-friendly in the combined feedback condition and as more distracting within the Medium challenge level. This research provides insights into the ways in which social robots can be used to facilitate human performance and engagement in tasks where both positive attitudes towards the task and high performance are essential for long-term involvement.

ARTICLE HISTORY
Received 30 September 2024
Revised 19 March 2025
Accepted 14 June 2025

KEYWORDS
Socially assistive robots; companion social robots; computerized cognitive training; social engagement; task engagement

1. Introduction

The integration of social robots into various domains, particularly healthcare, has garnered significant attention in recent years. As the global population ages, with people over 60 now accounting for almost 16% of the world's population and set to nearly double in the next 25 years [1], demographic shifts will impact the cost and accessibility of healthcare services [2]. Therefore, there is a need for innovative solutions to support and enhance the quality of care. Social robots, being employed as 'companions', offer a scalable approach to buffering these rising demands for healthcare, having previously been used in cognitive therapy [3,4], reducing depression and loneliness among elderly individuals in long-term care facilities [5–8], assisting children with Autism Spectrum Disorder (ASD) in acquiring skills related to distrust and deception [9–11]. This is particularly important in long-term cognitive training interventions, where sustained engagement is critical for success, e.g. adherence to a full course of clinical intervention [12,13]. These social robots can adjust their interactions with users through various verbal or non-verbal feedback and help users feel more socially engaged and maintain their interest [14,15].

Measuring engagement in HRI settings, particularly with social robots, is vital for understanding user experiences and tailoring interactions to meet user needs. Engagement in HRI is a complex concept that has been defined and explored in various ways across numerous studies [16]. This diversity in interpretations has led to the adoption of various metrics and features for assessing user engagement in the HRI literature [17]. In this paper, *engagement* is defined as:

> a quality of user experiences with technology that is characterized by challenge, aesthetic and sensory appeal, feedback, novelty, interactivity, perceived control and time, awareness, motivation, interest, and affect ([18], p. 949).

Engagement interpretations can be divided into different components, such as cognitive, behavioral, and emotional engagement [16]. Each of these components can

CONTACT Bahram Salamat Ravandi bahramsalamat@ait.gu.se

Supplemental data for this article can be accessed online at http://dx.doi.org/10.1080/01691864.2025.2526668.

have various measures based on the HRI setup. For instance, attention is often seen as a cognitive measure of engagement [16], while emotional engagement can be measured through affect states, such as facial expressions [19], physiological responses [20], or subjective self-reports [21].

There are various studies that have explored the experience and perceptions of individuals interacting with social robots [22–24], focusing on specific aspects of human-robot interaction and highlighting the ways in which social robots can impact user experience. The dimensions of 'warmth' and 'competence' are two fundamental dimensions in human perceptions that users associate with robots [25,26]. Warmth refers to traits such as being friendly and trustworthy, while competence relates to abilities and effectiveness. Robots with these attributes can significantly affect users' willingness to accept and engage with social robots [27,28]. Feedback mechanisms in social robots can further enhance the perception of human-like qualities, such as warmth and competence, and increase anthropomorphism. By offering timely and relevant feedback that shows empathy and understanding, robots can create a sense of connection similar to human interactions.

Mollahosseini et al. focused on investigating empathy in human-robot interaction, using the Ryan Companionbot to assess the participants' perception of the robot's social likability and empathy [29]. The study found that automated empathic responses improved subjects' perception of empathy and likability. In a study focused on interactive feedback and performance, [30] investigated the Nao robot's impact on children's engagement and performance in understanding 3D geometric figures. The findings indicated that high interactivity, with behaviors such as high-fives and personalized feedback, led to increased engagement and improved performance on the Post-Experiment Exam (PEE) assessing the taught mathematical concept. Similarly, [31] explored cognitive task performance and social presence using the Nao robot during a Stroop task. The study demonstrated that the social presence of an anthropomorphized robot significantly enhanced participants' attentional control mechanisms, with the 'social robot' condition showing the most substantial performance improvements.

While these studies suggest that more interactive scenarios and task-related feedback foster stronger social connections with robots and enhance task performance, excessive feedback from social robots could potentially lead to cognitive overload, diminishing both user engagement and motivation [32]. The effectiveness of specified feedback may also depend on factors such as the specific HRI setup (e.g. concerning the extent and timing of robot feedback), the nature of the task (fast-paced or slower), and its challenge level [33]. However, there is limited research comparing user responses to different types of feedback, such as feedback corresponding to task performance or to a user's emotional state. This study aims to address this gap by investigating the impact of performance-based and affective-based feedback from a social robot, Furhat, on both task and social engagement during a fast-paced visuospatial memory training task. The term 'fast-paced task' refers to tasks that require continuous user engagement with the task with minimal idle time. These tasks are specifically designed to challenge working memory (e.g. [34]). This research has implications for applications in education, healthcare, and other fields.

In healthcare, social robots engage in interpersonal interactions to provide adaptive responses tailored to individual user needs, which is crucial for applications like mental health support. Their non-judgmental presence encourages open engagement, facilitating better user experiences in sensitive contexts. Additionally, these robots offer monitoring and guidance to promote healthy behavior [35]. Common characteristics of robot feedback in healthcare include emotion and expression recognition, allowing robots to connect empathetically with users; natural behavior to establish trust; and multimodal communication, incorporating verbal responses, facial expressions, and gestures for better engagement [36].

In educational settings, feedback from social robots is characterized by a blend of affective and cognitive outcomes, aiming to foster both empathy and learning gains, through personalization and adaptive responses that adjust based on students' actions and emotional states [37]. Such feedback often incorporates social engagement techniques, utilizing supportive behaviors like addressing students by name and offering encouragement to enhance motivation. Additionally, it can provide cognitive and affective support, addressing both understanding of material and emotional well-being while employing engagement-sustaining techniques to maintain interest and focus. Moreover, feedback incorporates social cues like gestures and facial expressions to create a natural interaction environment and is delivered clearly and constructively to guide learners effectively toward their educational goals [38].

Performance-based feedback is widely used in HRI to enhance users' intrinsic motivation and self-esteem and consequently increase performance, particularly when the feedback is positive, as it reinforces feelings of competence and enjoyment in tasks [28]. Meanwhile, affective-based feedback has been studied for its role in fostering social engagement and emotional connection between users and robots [39–41]. These feedback mechanisms align with the fundamental dimensions of warmth and competence in human perception. Given that warmth and competence influence users' acceptance and perception of social robots [25,26], it is essential to investigate how these types of feedback impact user engagement. By studying these methods, we aim to deepen the understanding of how different types of feedback shape user engagement and performance.

The present study investigates a number of metrics, such as eye contact with the social robot [42,43], distance from screen [44], performance [45], affect states [29,46], and blink rate [47,48] to evaluate different aspects of engagement within an established HRI setup. These measurements guide the design and development of more effective and enjoyable interactions between users and robots [49]. While task engagement and social engagement are interconnected, they represent distinct dimensions of the interaction that can influence user outcomes in different ways. By analyzing these two types of engagement independently, researchers can identify specific factors that enhance or hinder each dimension. This separation allows for a more targeted approach to designing social robots that target the unique needs of users in various contexts, e.g. in tasks where maintaining focus and a flow state [50] is essential. It is also important to ensure that social engagement does not interfere with task engagement.

To achieve the above, a general HRI framework is proposed, which outlines the interaction loop between the social robot, the human user, and a gamified task. The following section details this HRI framework, highlighting its key components and their roles in fostering effective task-based and social interactions.

2. HRI framework

Figure 1 depicts a proposed general HRI framework featuring several modules designed for customized user interaction. Within this framework, different gamification elements are implemented [51], e.g. users can engage with the task by receiving rewards and audiovisual feedback from both the robot and the task itself. The interaction loop, as presented in Figure 1, encompasses six components:

(1) *Challenge Modulation:* Adjusting the challenge level of the task based on the user's expressed state of engagement or disengagement.
(2) *Task State:* This involves providing information on human performance considering the current state of the game.
(3) *Action Selection:* This involves the use of touchscreen-based, verbal, or mouse inputs for selecting actions.
(4) *Reward Feedback*: The task provides direct feedback on user performance and specific actions.
(5) *Social Feedback:* This includes the robot's verbal/nonverbal feedback.
(6) *Engagement State:* Possible inputs that help determine the user's engagement state.

Integrating gamification elements in HRI frameworks can significantly boost user engagement, enhance the effectiveness of interactions, and increase users' motivation to interact with robots for longer durations [52]. These gamification features can be applied at two key levels: the interaction design [53] and task design [54]. At the interaction level, gamification aims to make the overall interaction between humans and robots more engaging and enjoyable. This can include elements like verbal or non-verbal feedback that encourage users to participate actively. Within task design, gamification is embedded into the specific tasks users undertake with the robot to enhance user experience and motivation.

Previous research by Markelius et al. [55] examined the impact of minimal feedback from Furhat on task performance, finding that its presence, compared to no feedback, did not negatively affect outcomes. Building on this foundation, the present study expands the scope by investigating how different types of feedback influence both engagement and performance. While prior work primarily focused on the absence or minimal presence of feedback, this study explores the effects of various feedback types, including performance-based feedback, affective-based feedback, and a combination of both. Affective-based feedback involves responses that convey emotions, acknowledge enjoyment, and inquire about the user's emotional state, whereas performance-based feedback provides information on accuracy, progress, and overall task performance. By examining these dimensions, this study seeks to provide deeper insights into how feedback strategies shape user interaction and outcomes.

2.1. Research questions and hypotheses

The research question of this study is:

- *Question:* How do different types of feedback from the social robot Furhat influence users' engagement during a digital visuospatial memory training task?

Based on the literature reviewed above, we set up two hypotheses:

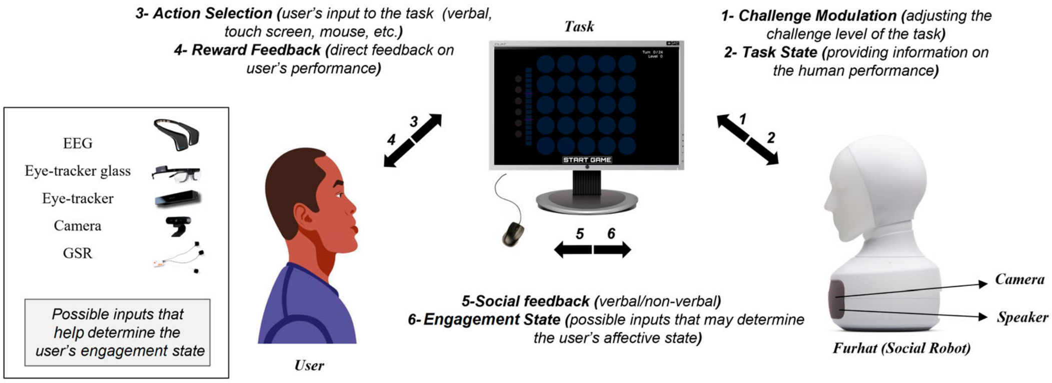

Figure 1. This diagram illustrates a general HRI framework, which consists of six interconnected components that facilitate how a social robot and a human user interact within the framework to complete a task required by the user.

- *Hypothesis 1:* Participants' social and affective engagement will be higher in conditions with affective-based feedback compared to conditions without such feedback.
- *Hypothesis 2:* Participants' task performance will be higher in conditions with performance-based feedback compared to conditions without such feedback.

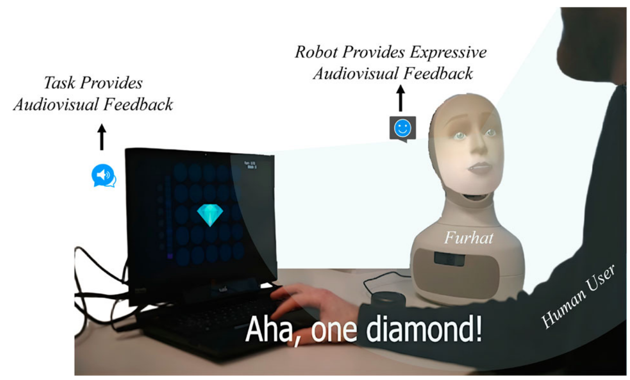

Figure 2. Experimental setup: The robot delivered audiovisual feedback through facial expressions, head movements (such as nodding or shaking its head), and verbal phrases, all related to the task.

3. Methodology

The experimental design follows a $3 \times 2 \times 2$ mixed design with three independent variables: types of feedback (IV1) and challenge levels (IV2) as between-subjects factors and a within-subject baseline control block (IV3). IV1 encompasses three types of feedback: (i) performance-based (PB), (ii) affective-based (AB), and (iii) combined (COM). IV2 comprises two challenge levels: (i) Easy challenge level and (ii) Medium challenge level. The baseline control block is not influenced by the feedback manipulation, as block feedback is provided only after the block is completed. By incorporating challenge levels as a factor, we account for its influence on users' working memory and cognitive states, such as mental fatigue and cognitive overload, which are linked to cognitive load [56] and how it affects the users' perception of appropriate feedback from a robot. In this study, the difficulty is determined by the number of stimulus sequences users must remember in a visuospatial memory training task. A more difficult task involves a greater number of sequences, leading to increased working memory taxation and cognitive load. By manipulating the challenge level, we aim to examine how users' engagement and perception of the robot change across different challenge levels. This allows us to assess our hypotheses and determine whether they hold true under varying challenge levels. Figure 2 illustrates the configuration layout of the experimental setup. In the illustration shown, the robot was positioned next to the human participant, providing a view of both the task and the participant from an angled perspective. This arrangement aimed to enhance interactions between the human and the robot.

For this study specifically, we utilized a memory training task, which involves a sequence of cognitive exercises designed to challenge and improve visuospatial memory. The task requires users to process and recall information under time constraints, making it both cognitively demanding and time-sensitive. The task is a modified (gamified) version of a visuospatial working memory task with Differential Outcome Training (DOT) developed by Vivas et al. [57]. In previous research, this task was used with different numbers of trials and challenge levels [55], investigating how different types of interactions (simulated versus physical robots) could influence participants' performance and affective responses during the task. Our current focus is solely on physical robots, and the task features two distinct challenge levels, each comprising three blocks of eight trials.

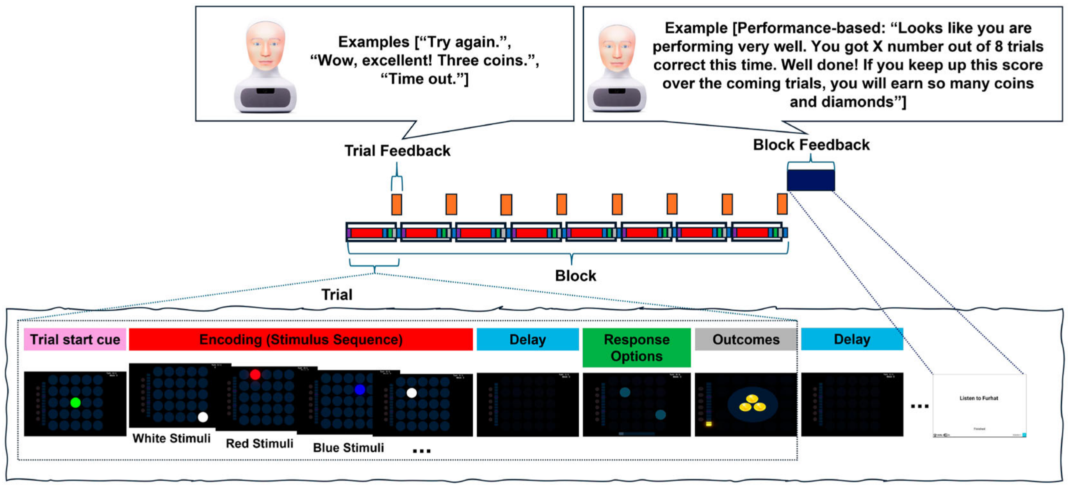

Figure 3. Sequence of steps within a single block of eight trials, outlining the phases from start cue to feedback delivery by the robot after each trial and also at the end of each block: Each trial begins with a start cue stimulus, followed by an encoding phase where participants memorize the locations of the white cells. After a delay phase, participants select a location corresponding to one of the previously presented sample stimuli (response phase). At the end of the trial, feedback is provided, indicating whether the response was correct (reward) or incorrect. Concurrently with the presentation of the trial outcome by the game, the robot provides trial feedback. Upon completion of all eight trials within the block, the robot delivers block feedback.

As illustrated in Figure 3, the process diagram outlines the sequence of steps within a single block comprising eight trials. During each trial, players must remember the locations of white cells that appear on a 5 × 5 grid and choose from the provided options at the end of each trial. Each trial begins with a start cue stimulus, followed by an encoding phase where participants are instructed to memorize the locations of white cells while also being presented with red and blue distractor cells. The distractor task was used to increase the cognitive load of working memory and make the task more challenging. Participants are instructed to press the spacebar when blue cells appear. A miss is recorded if a participant fails to respond to blue cells within two seconds. The encoding phase is followed by a delay phase, leading into the response phase, where participants select a location corresponding to a previously presented stimulus. Trial feedback is then provided to indicate whether the response was correct or incorrect, where participants receive audiovisual feedback from both the social robot Furhat and the task. The trial feedback provided by the robot is consistent with the audiovisual feedback that the task provided for reward types. Table 1 illustrates 6 different outcomes for each trial and the corresponding trial feedback provided by the task and the robot. Furthermore, while delivering this feedback, the robot occasionally turns its head toward the participant – this occurs randomly with a higher likelihood during the earlier trials – to enhance engagement.

A pause follows each trial, and upon completing all eight trials in the block, Furhat provides block feedback, which can be categorized as performance-based (PB), affective-based (AB), or a combination of both (COM). Performance-based feedback offers details on the number of correct responses, progress, and overall performance, such as 'Looks like you are performing very well. You got X out of eight trials correct this time. Well done! If you keep up this score over the coming trials, you will earn so many coins and diamonds.', whereas affective-based feedback involves responses that convey emotions, recognize enjoyment, and inquire about the individual's feelings, like 'I hope you still feel happy playing the game! How do you feel about your results? [wait for a response]. I am sure you will keep improving over the coming levels.' Finally, combined feedback states a combination of these two types of feedback, for example, 'Looks like you are performing very well. You got X out of eight trials correct this time. Looks like you are enjoying this game. Do you feel happy about your good results? [wait for a response]'.

The robot, Furhat, was introduced as a companion to the task, using a scripted scenario before the game. Positioned at a fixed angle, it greeted participants and explained the memory training task rules, which remained the same across conditions.

3.1. Participants

Recruitment was advertised and involved healthy adults between the ages of 18 and 45 to participate in the research study, conducted at Koç University, Turkiye. Participants were informed through email lists for students

Table 1. Trial feedback from the task and the robot: At the end of each trial, feedback is first provided by the task based on the trial's outcome, followed by additional feedback from the robot.

Outcome	Task Feedback	Robot Feedback
Outcome 1		Smile + random 'Yes, one coin', 'Huh, you got a coin', 'Hmm, good work'[Medium pitch]
Outcome 2		Big smile + random 'Wow, excellent! Three coins', 'Whoa, you got three coins!', 'Oh yes, jackpot!'[High pitch]
Outcome 3		Nod + random 'Aha, one diamond!', 'Okay! you got a diamond', 'Yes, well done!'[Medium pitch]
Outcome 4		Double nod + random 'Wow, you got three diamonds', 'Outstanding! So many diamonds', 'Yes! Brilliant!'[High pitch]
Incorrect		Shake head + random 'Incorrect, keep fighting!', 'Try to stay focused', 'No, try again!', 'Come on, you can do it!'[Low pitch]
Timeout		Shake head + random 'Uh, time out!', 'Time out, be faster!', 'Oh no, pick up speed!', 'Huh, catch up'[Low pitch]

Table 2. Distribution of participants across conditions.

Condition	Easy Challenge	Medium Challenge	Total
PB Feedback	8	9	17
AB Feedback	9	10	19
COM Feedback	11	11	22
Total	28	30	58

at the Departments of Electrical and Electronics Engineering and Computer Engineering. Participants were required to provide informed consent and could withdraw from the experiment at any time without consequences. The participants were divided evenly into the available time slots under different conditions. However, some participants did not attend, and some participants were under 18 years old. Fifty-eight engineering students (35 males, 23 females) from Koç University's Electrical and Electronics Engineering and Computer Engineering departments, aged 18 to 24 ($M = 20$, $SD = 1.87$), participated in the experiment. The distribution of participants across conditions can be seen in Table 2.

3.2. Data collection

Data collection included the task performance metrics, participants' video recordings, and eye-tracking data, which were gathered using the non-invasive Tobii Pro X2-30 screen-based eye tracker device. Additionally, questionnaires and structured interviews were conducted to obtain subjective feedback from the participants. Upon completing the task, participants were instructed to fill out two questionnaires: the Self-Assessment Manikin (SAM) scale [58], which measures emotional responses, and the Human-Robot Interaction Evaluation Scale (HRIES) [59], which assesses anthropomorphism in human-robot interactions. Following the completion of these questionnaires, participants also took part in a structured interview to provide further insights into their experiences and perceptions.

The iMotions software was used to collect real-time data from eye-tracking devices and facial expression recognition using Affectiva's Affdex toolkit. Affectiva's Affdex is a software development kit (SDK) designed for real-time emotion recognition [60]. It involves analyzing facial expressions to determine an individual's emotional state by capturing data such as facial movements and expressions.

3.3. Engagement and performance assessment metrics

Engagement in this study is categorized into two primary types: task engagement and social engagement.

- *Social Engagement*: Involves the interaction between participants and the robot. This is assessed through the frequency of eye contact with the robot, as well as affect data captured via facial expressions. Eye contact typically reflects social connection, indicating a participant's interest and emotional responsiveness towards the robot [61]. In human interactions, eye contact plays a crucial role in signaling attention, connection, and communication [62]. When participants were engaging with a robot, the frequency of eye contact can indicate the level of attentiveness or interest in the interaction. In this context, eye contact serves as a non-verbal cue, signaling how connected or engaged a user feels with the robot. The more frequently users make eye contact with the robot, the more likely they are to be socially engaged in the robot's feedback.
- *Task Engagement*: Refers to the participant's engagement with the task itself. This is measured through affect data captured via facial expressions, blinks, and distance from the screen captured through eye tracker data. Blink rate serves as an indicator of cognitive load [63], where higher rates during lost trials suggest decreased engagement or cognitive fatigue, while lower rates indicate heightened focus. Distance from

the task can reveal comfort levels and willingness to engage, with closer proximity suggesting positive interaction and greater distance indicating discomfort [64]. Task performance is intrinsically linked to task engagement, which is typically assessed using objective metrics such as accuracy, response time, or completion rate. The primary objective of task engagement is to enhance user outcomes [16]. Jain et al. [45] employed task performance as a key indicator in engagement modeling.

This study utilizes a combination of objective and subjective assessment metrics to evaluate engagement in the proposed HRI setup. The following objective and subjective metrics are employed:

- *Performance Metrics*:
  - *Accuracy*: Accuracy per block is assessed through the percentage of correct responses per block. Participants completed three blocks of eight trials each, with accuracy measured both per block and across all blocks. Since the accuracy is influenced by block feedback only after the first block (see Figure 3), the first block was considered as a baseline control block. The accuracy per block 2 and block 3 serves as a metric to compare the impact of different types of feedback on performance.
  - *Distractor Task*: Participants are instructed to press the spacebar when blue cells appear. The rate of successful responses to these blue cells (hit rate) is used as an additional measure of user performance. In each trial, two to four cells are displayed, depending on the challenge level and the stimulus sequences.
- *Behavioral data*:
  - *Eye Contact*: Eye contact with the robot during each trial is recorded as a binary measure, where a value of 1 indicates that the participant looked at the social robot at least once, and 0 indicates no eye contact. This measure was used to assess attention to the robot, particularly during feedback delivery and reward presentation (see Figure 3). Eye contact instances were manually labeled using the video recordings to determine whether the user made eye contact with the robot during each trial or not.
  - *Blinks*: The number of blinks per trial is extracted during the encoding phase of each trial (see Figure 3), using Affectiva's Affdex SDK within the iMotions software (version 9.4.0).
  - *Distance Measure*: Eye-tracking data is used to assess each participant's distance from the screen. The distance measure is calculated per trial to assess users' behavior. The distance measure for participant i during trial j is calculated as:

$$\Delta d_{i,j} = \overline{d}_{\mathrm{outcome}_{i,j}} - \overline{d}_{\mathrm{trial}_{i,j}}$$

    where $\Delta d_{i,j}$ is the shift in distance from the screen for participant i during trial j, $\overline{d}_{\mathrm{outcome}_{i,j}}$ is the average distance from the screen measured after the outcome was presented, for 5 s which also includes the pause between trials, for participant i during trial j, and $\overline{d}_{\mathrm{trial}_{i,j}}$ is the average distance from the screen measured during the trial period before the outcome is revealed for participant i during trial j.
- *Affect Measure*: The affect measure per trial is used to measure affective engagement, calculated as:

$$\Delta v_{i,j} = \overline{v}_{\mathrm{outcome}_{i,j}} - \overline{v}_{\mathrm{trial}_{i,j}}$$

  where $\Delta v_{i,j}$ is the shift in valence for participant i during trial j, $\overline{v}_{\mathrm{outcome}_{i,j}}$ is the average valence measured after the outcome is presented, for 5 s which also includes the pause between trials, for participant i during trial j, and $\overline{v}_{\mathrm{trial}_{i,j}}$ is the average valence measured during the trial period before the outcome is revealed for participant i during trial j. Valence is measured by Affectiva's Affdex SDK within the iMotions software through facial expression recognition [60]. It analyzed facial expressions to infer an individual's emotional state by capturing facial movements and expressions.
- *Self-Assessment Manikin (SAM)*: The study utilized the SAM questionnaire [58] to evaluate participants' emotional responses and engagement levels, incorporating three distinct dimensions: valence, arousal, and dominance. *Valence* measures the happiness or unhappiness of the emotional experience, *arousal* measures the intensity of that experience from calm to stressed, and *dominance* measures the sense of control or influence participants feel over their emotions, as illustrated in Figure 4. The SAM scale was specifically used to assess users' affective states, providing support for detecting any differences in emotional responses that might be observed through the iMotions Affdex SDK.
- *Human-Robot Interaction Evaluation Scale (HRIES)*: HRIES provides additional insights into participants' perception of anthropomorphism in human-robot interactions [59]. The HRIES Questionnaire encompassed 16 attributes, prompting participants to assess their perceptions of social robots using a 7-point Likert scale. The 16 attributes of the HRIES questionnaire were divided into four distinct categories: Sociability (including Warm, Likable, Trustworthy, and

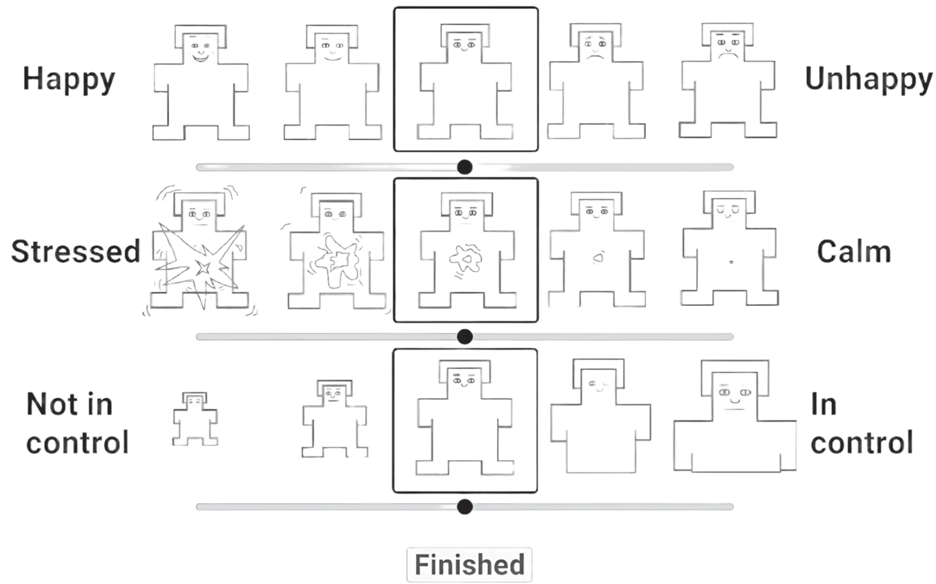

Figure 4. Example of the SAM Scale.

Friendly), Agency (comprising Self-Reliant, Rational, Intentional, and Intelligent), Animacy (including Alive, Natural, Real, and Human-Like), and Disturbance (comprising Creepy, Scary, Uncanny, and Weird). The HRIES scale was employed to measure participants' perceptions of the robot, which varied based on the type of feedback received.

4. Results

This section presents the findings of the study, focusing on how different types of feedback from the social robot Furhat influence user engagement and task performance. The results are divided into objective and subjective engagement metrics, each providing insights into the effectiveness of feedback types during the visuospatial memory training task.

4.1. Objective metrics

4.1.1. Eye contact

Table 3 shows the mean number of trials with eye contact with the robot across blocks, *challenge levels*, and *types of feedback*. The eye contact frequency with the robot in the first block was slightly higher than other blocks. However, the first block was considered a baseline control block because participants received the block feedback only after it. A two-way ANOVA was conducted to compare the main effect of *types of feedback* and *challenge levels* as well as their interaction effects on users' eye contact frequency with the robot during block 2 and block 3. The main effect of *challenge levels* was not significant, $F(1,45) = 1.09$, $p = 0.30$, $\eta_p^2 = 0.02$. The main effect of *types of feedback* was significant, $F(2,45) = 4.03$, $p = 0.02$, $\eta_p^2 = 0.15$). The interaction effect was not significant, $F(2,45) = 0.91$, $p = 0.40$, indicating that there was no combined effect of *challenge levels* and *types of feedback* on eye contact with the robot.

The Bonferroni post-hoc test results for the dependent variable *eye contact* with the robot show that there are no significant differences between COM and PB ($p = 0.80$) or COM and AB ($p = 0.12$). However, significant differences were found between PB and AB ($p = 0.01$), with AB scoring higher by 2.35 points on average on a scale from 0 to 8.

A between-subjects ANOVA analysis was conducted to examine the effects of feedback type and level on eye contact with the robot in the baseline control block. The main effect of feedback type was not significant, $F(2,52) = 1.45$, $p = 0.24$, $\eta^2 = 0.05$, nor was the main effect of level, $F(1,52) = 0.10$, $p = 0.75$, $\eta^2 = 0.00$. Additionally, the interaction between feedback type and level was not significant, $F(2,52) = 1.44$, $p = 0.24$, $\eta^2 = 0.05$.

4.1.2. Accuracy

Table 4 presents the mean accuracy and standard deviation across various *challenge levels* and *types of feedback* within each block and in total. The first block was considered as a baseline control block because participants received the block feedback only after the first block. A two-way ANOVA was conducted to compare the main effect of *types of feedback* and *challenge levels* as well as their interaction effects on accuracy (block 2 and block 3). The main effect of *challenge levels* was

Table 3. The mean number of trials with eye contact with the robot across blocks, *challenge levels*, and *types of feedback*.

Type of Feedback	Challenge Level	Block 1	Block 2	Block 3	Total
PB	Easy	1.38 (1.50)	0.63 (0.74)	0.38 (0.51)	0.79 (1.06)
AB	Easy	2.22 (2.43)	2.78 (2.83)	1.56 (1.74)	2.19 (3.35)
COM	Easy	2.80 (2.74)	1.50 (1.65)	0.80 (0.63)	1.70 (2.00)
PB	Medium	0.50 (0.54)	1.00 (1.09)	0.67 (0.81)	0.72 (0.82)
AB	Medium	3.71 (2.43)	1.57 (1.51)	1.29 (2.98)	2.19 (2.52)
COM	Medium	2.36 (2.37)	0.36 (0.92)	0.55 (0.68)	1.09 (1.73)

Table 4. Mean accuracy and standard deviation (SD) per block and in total across different *challenge levels* and *types of feedback* (unit: percentage).

Type of Feedback	Challenge Level	Block 1	Block 2	Block 3	Total
PB	Easy	84.37 (17.35)	82.81 (16.28)	96.87 (5.78)	88.02 (14.96)
AB	Easy	77.77 (19.54)	58.75 (27.66)	77.77 (21.44)	76.58 (19.82)
COM	Easy	73.86 (23.35)	81.81 (17.99)	79.54 (25.16)	78.40 (21.93)
PB	Medium	52.77(18.51)	70.83 (19.76)	63.88 (19.20)	62.50 (19.91)
AB	Medium	60.00 (9.86)	58.75 (27.66)	41.25 (17.72)	53.33 (21.00)
COM	Medium	68.18 (27.02)	71.59 (22.42)	69.31 (16.16)	69.69 (21.65)

significant, $F(1,52) = 19.58$, $p < 0.001$, $\eta_p^2 = 0.27$. The main effect of *types of feedback* was also significant, $F(2,52) = 3.94$, $p = 0.02$, $\eta_p^2 = 0.13$. The interaction effect was not significant $F(2,52) = 1.08, p = 0.34$, indicating that there was no combined effect of *challenge levels* and *types of feedback* on accuracy.

The Bonferroni post-hoc test results show that there is a statistically significant difference in accuracy between PB and AB ($p = 0.02$), with PB scoring higher by 14.71 percent on average. However, no significant differences are observed between COM and PB ($p = 1.00$) or COM and AB ($p = 0.09$). These results suggest that PB feedback leads to significantly better accuracy than AB, while COM does not significantly differ from either PB or AB.

A two-way ANOVA was conducted to compare the main effect of *types of feedback* and *challenge levels* as well as their interaction effects on accuracy over the baseline control block. The main effect of *type of feedback* is not significant, $F(2,52) = 0.08, p = 0.91$, suggesting that accuracy does not differ statistically among types of feedback in the control block. However, the main effect of *challenge level* is significant, $F(1,52) = 11.63, p = 0.00$, indicating that performance varies significantly based on challenge level. The interaction effect between *type of feedback* and *challenge level* is not significant, $F(2,52) = 1.94, p = 0.15$, suggesting that the effect of the types of feedback does not differ significantly across challenge levels.

The mean accuracy under the PB condition increased from 67.65 in block 1 (control) to 77.21 in block 2 and block 3, with improvements in both Easy (from 84.38 to 87.50) and Medium (from 52.78 to 68.06) conditions. This suggests that PB led to overall accuracy gains, particularly in the Medium challenge level, indicating better adaptation over time. Mean accuracy under the AB condition declined from 68.42 in block 1 (control) to 62.50 in block 2 and block 3, with a more significant drop in the Medium challenge level (from 60.00 to 50.00), while the Easy challenge level showed only a slight decrease (from 77.78 to 76.39). This suggests that participants in the AB condition struggled more in later blocks, particularly under the Medium challenge level. The mean accuracy under the COM condition slightly improved from 71.02 in block 1 (control) to 73.86 in block 2 and block 3, with a noticeable increase in the Easy challenge level (from 73.86 to 79.55), while the Medium challenge level remained stable at 68.18. Figure 5 illustrates a box plot of accuracy across types of feedback and challenge levels in the baseline control block and experimental blocks (block 2 and block 3). This visualization highlights the differences between the AB and PB conditions in the experimental blocks, while no such statistical distinction is observed in the baseline control block.

4.1.3. Distractor task

Table 5 presents the average hit rate for blue cells per block and in total across challenge levels and types of feedback. A two-way ANOVA was performed to examine the main effects of *types of feedback* and *challenge levels*, as well as their interaction on hit rate. The analysis revealed no significant differences in the average hit rate across the study conditions. These results suggest that the types of feedback provided and the challenge levels did not significantly influence the users' performance on the distractor task.

4.1.4. Blinks

Table 6 presents the average number of blinks per block and in total across different challenge levels and types of feedback. A two-way ANOVA was conducted to examine the main effects of *types of feedback* and *challenge levels*, as well as their interaction on the number of blinks. Neither

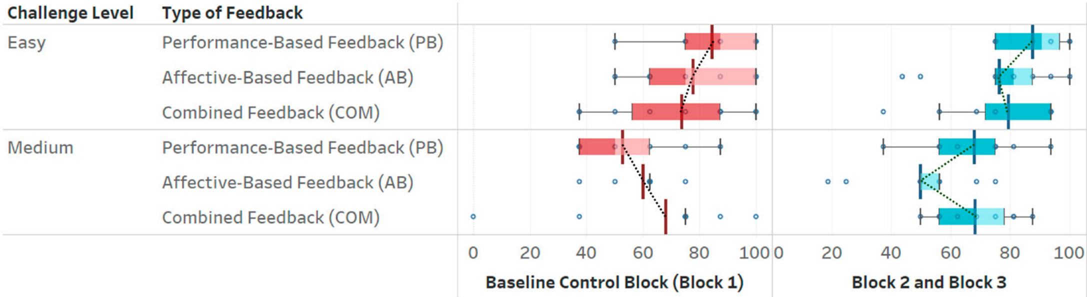

Figure 5. Box plot illustrating accuracy across *types of feedback* (PB: performance-based, AB: affective-based, COM: combined) and *challenge levels* (Easy, Medium). A dashed line connects mean accuracy values.

Table 5. The average hit rate for blue cells and the standard deviation (SD) per block and in total across different *challenge levels* and *types of feedback*.

Type of Feedback	Challenge Level	Block 1	Block 2	Block 3	Total
PB	Easy	0.85 (0.34)	0.91 (0.25)	0.98 (0.02)	0.98 (0.02)
AB	Easy	0.82 (0.31)	0.86 (0.32)	0.88 (0.33)	0.86 (0.31)
COM	Easy	0.86 (0.29)	0.89 (0.29)	0.89 (0.29)	0.88 (0.28)
PB	Medium	0.94 (0.03)	0.98 (0.02)	0.97 (0.03)	0.97 (0.03)
AB	Medium	0.86 (0.30)	0.88 (0.31)	0.82 (0.34)	0.86 (0.31)
COM	Medium	0.86 (0.29)	0.88 (0.29)	0.90 (0.29)	0.88 (0.28)

the main effect nor interaction effects of *challenge levels* and *types of feedback* were significant.

4.1.5. Distance measure

Table 7 displays the mean of the distance measure per block and in total across different challenge levels and types of feedback. A two-way ANOVA showed no significant differences between the study conditions.

4.1.6. Affect measure

Table 8 presents the mean of the affect measure and the standard deviation per block and in total across challenge levels and types of feedback. The results of a two-way ANOVA revealed no significant differences between the study conditions.

4.2. Subjective engagement metrics

The result of a two-way ANOVA for each category of Sociability, Agency, Animacy, and Disturbance indicates no significant difference between conditions. However, looking at individual items in these 16 attributes using the ANOVA indicated that within the Medium *challenge level*, participants perceived Furhat as more distracting, $F(1, 57) = 5.21, p = 0.02$. Also, participants deemed Furhat to be less rational in PB conditions, $F(2, 27) = 3.57, p = 0.02$, compared to COM ($p = 0.06$) and AB ($p = 0.00$), as well as between AB and COM ($p = 0.40$). Additionally, participants found COM to be more user-friendly, $F(2, 57) = 3.12, p = 0.05$, with $p = 0.01$ between AB and COM and $p = 0.21$ between AB and PB, as well as $p = 0.28$ between PB and COM.

The SAM questionnaire was also evaluated to indicate participants' emotional responses across challenge levels and types of feedback. Results indicated that participants reported significantly higher arousal during the Medium challenge level ($M = 3.33, \mathrm{SD} = 0.92$) compared to the Easy challenge level $F(1, 52) = 24.84, p < 0.001$ ($M = 2.21, \mathrm{SD} = 0.83$). However, no significant differences were found among the types of feedback.

4.3. Interaction of trial outcomes/feedback and objective engagement metrics

To further explore the interaction between trial outcomes/feedback and engagement metrics, the trials were categorized into two groups: wins and losses. A 'win' trial indicates that participants correctly identified the location of the white stimulus, while a 'loss' trial indicates an incorrect response. As the examination of the eye contact, participants demonstrated more eye contact with Furhat during lost trials, with the mean proportion of trials where participants made eye contact with the robot being in lost trials ($N = 332, M = 0.26, \mathrm{SD} = 0.43$) compared to won trials ($N = 892, M = 0.16, \mathrm{SD} = 0.63$). In examining the distance measure, the results suggest a tendency for participants to move backward more frequently after making incorrect responses, indicating potential disengagement. Descriptive statistics for the distance measure indicated that the lost trials ($N = 384$) had a mean of $M = 2.82$ ($\mathrm{SD} = 23.34$), while the won trials ($N = 943$) had a mean of $M = -0.23$ ($\mathrm{SD} = 21.61$). Regarding blinks, the results indicate that incorrect responses were associated with higher blinking. Descriptive statistics for

Table 6. Average number of blinks with standard deviation (SD) per block and in total across different *challenge levels* and *types of feedback*.

Type of Feedback	Challenge Level	Block 1	Block 2	Block 3	Total
PB	Easy	5.38 (7.36)	6.75 (7.10)	3.25 (3.88)	5.12 (6.21)
AB	Easy	9.67 (8.97)	12.56 (12.31)	11.44 (12.96)	11.22 (11.16)
COM	Easy	8.50 (11.84)	9.50 (13.68)	8.40 (15.57)	8.80 (13.31)
PB	Medium	14.67 (7.86)	18.00 (11.47)	20.17 (14.73)	17.61 (11.23)
AB	Medium	8.29 (9.39)	14.14 (12.28)	15.86 (15.79)	12.76 (12.55)
COM	Medium	9.82 (10.51)	11.45 (14.13)	9.18 (12.36)	10.15 (12.07)

Table 7. Mean distance measure and the standard deviation (SD) per block and in total across *challenge levels* and *types of feedback*.

Types of Feedback	Challenge Level	Block 1	Block 2	Block 3	Total
PB	Easy	−1.10 (4.48)	−1.70 (3.47)	−1.62 (7.58)	−1.47 (5.23)
AB	Easy	0.01 (5.26)	1.45 (2.69)	−1.23 (14.46)	0.07 (8.73)
COM	Easy	1.75 (7.24)	0.61 (2.21)	0.11 (5.31)	0.82 (5.181)
PB	Medium	3.88 (7.57)	5.61 (8.34)	−0.06 (9.15)	3.14 (8.40)
AB	Medium	3.58 (10.75)	1.51 (11.02)	2.01 (8.47)	2.36 (9.83)
COM	Medium	−1.18 (8.95)	−1.28 (8.83)	−1.39 (6.48)	−1.28 (7.91)

blink rate per trial indicated that the lost trials ($N = 332$) had a mean number of blinks of $M = 1.90$ (SD = 2.31), while the won trials ($N = 892$) had a mean number of blinks of $M = 1.10$ (SD = 1.66). Lastly, the affect measure showed that more positive facial expressions were exhibited when participants made incorrect responses compared to correct ones. Descriptive statistics for the affect measure indicated that the lost trials ($N = 328$) had a mean of $M = 6.38$ (SD = 24.13), while the won trials ($N = 886$) had a mean of $M = 0.34$ (SD = 15.60).

5. Discussion

This study examined how different types of feedback – performance-based (PB), affective-based (AB), and a combination of both (COM) – delivered by a social robot, Furhat, influence user task engagement and social engagement during a visuospatial memory training task. To potentially improve interaction, participants received one of these types of feedback at the end of each block of trials. The results offer valuable insights into the distinct impacts of these types of feedback on participants' social and task engagement, highlighting the complexities of HRI in a gamified setting.

The results provide partial support for Hypothesis 1, which is that the AB condition did result in higher social engagement, as indicated by higher eye contact with the robot compared to performance-based feedback. However, it did not support the aspect of Hypothesis 1 predicting higher affective engagement in the AB condition as measured by valence values. There were no significant differences in affective engagement between groups. Further analysis revealed that the combined feedback condition, which also includes affective-based feedback, failed to support higher social engagement. A deeper examination showed that participants made more eye contact with the robot during trials with negative outcomes (lost trials), suggesting that eye contact may be more closely related to trial outcomes or trial feedback rather than block feedback, where participants were given summarizing feedback over eight trials. Given the lower accuracy observed in the affective-based condition and the significant association between trial outcome and eye contact, the increased eye contact appears to be more closely related to the trial outcome rather than the block feedback itself. This suggests that while the block feedback may influence overall performance, the immediate trial outcomes are more directly linked to participants' eye contact with the robot.

The results support Hypothesis 2, demonstrating that conditions involving performance-based feedback lead to higher task performance, as evidenced by greater accuracy compared to conditions without performance-based feedback. The consistent findings in the combined

Table 8. Mean of the affect measure and standard deviation (SD) per block and in total across *challenge levels* and *types of feedback*.

Type of feedback	Challenge Level	Block 1	Block 2	Block 3	Total
PB	Easy	0.81 (6.16)	1.44 (5.38)	0.06 (0.32)	0.77 (4.55)
AB	Easy	0.72 (8.54)	2.40 (4.80)	−0.01 (2.53)	1.04 (5.71)
COM	Easy	5.36 (10.74)	2.77 (7.06)	−0.35 (4.58)	2.59 (7.96)
PB	Medium	0.44 (5.48)	−0.83 (1.28)	−1.02 (3.149)	−0.47 (3.56)
AB	Medium	10.53 (12.60)	1.22 (7.43)	2.75 (8.70)	4.83 (10.21)
COM	Medium	3.29 (5.66)	3.31 (12.06)	2.19 (4.85)	2.93 (8.20)

(COM) condition indicate that performance-based feedback, whether delivered alone or in combination with affective-based feedback, is more effective at maintaining or slightly enhancing accuracy.

Moreover, the study found no significant interaction effects between types of feedback and challenge levels on social engagement or task performance. This lack of interaction effects may indicate the consistent influence of feedback across different challenge levels.

The results further suggest that participants were more likely to disengage from the screen after losing a trial and receiving negative feedback, compared to winning a trial and receiving positive feedback. Moreover, participants tended to display facial expressions with more positive valence (as picked up by the Affectiva module of iMotions software) after losing a trial compared to after winning a trial. While interpreting 'positive valence happy' as the simplest inference, the literature provides examples of how humans smile, e.g. to cover an embarrassment or even in tragic situations [65,66]. Another noteworthy result showed that the blink rate was higher during lost trials compared to won trials, which indicates that the subjects were concentrating less during loss trials [67]. These findings highlight how engagement levels are tied to task events and feedback from both the robot and the game.

Furthermore, the questionnaire data provided additional insights. We found that participants in the combined feedback condition generally found the robot more friendly than affective-based feedback or performance-based feedback conditions, alongside higher performance in the COM and PB feedback conditions. This suggests that combined feedback is more desirable and yields better results. One participant from the combined feedback group noted, 'She's so beautiful and pretty, actually, and it seems really friendly. It's not scary; I like her.'. Another participant from the performance-based condition shared, 'When he said, 'Oh, you did eight out of eight,' I felt very happy and successful.'

Participants also reported the robot as more distracting during the Medium challenge level compared to the Easy challenge level, indicating that the appropriateness of feedback is linked to task difficulty and participants' cognitive capacity [68]. One participant engaged in the Medium challenge remarked,

> All right, now if you are trying to motivate me, just let me focus on the task. I think the sound feedback was more important- like when you do it wrong, you get a beep, and when you do it right, you get a sound.

The results have practical implications for designing social robots in educational and therapeutic environments. Incorporating performance-based feedback may benefit tasks where performance is the primary goal, such as cognitive training or educational exercises. Conversely, in contexts where building a social connection is crucial- such as companionship or social skills training- prioritizing affective-based feedback could enhance user engagement with the robot. Additionally, in long-term interventions or educational programs, it's crucial for participants to not only achieve strong performance but also maintain positive attitudes and ongoing engagement with the tasks. A social agent can effectively support this dual emphasis, fostering both performance outcomes and user engagement. This approach is vital for the success of long-term interventions or tasks, ultimately leading to better learning outcomes.

Although this study provides significant insights, it is not without limitations. The sample size, while sufficient for detecting main effects, may limit the generalizability of the findings. Despite the random assignment of participants to different conditions, individual variations in personality traits could influence the engagement metrics, rendering them more reflective of personal characteristics rather than universal indicators of engagement. Additionally, focusing exclusively on a specific task type – visuospatial memory training task – may not fully capture the potential variability in responses to different types of feedback across other tasks or contexts. While the study found that affective-based feedback might impair performance, these findings underscore the need for further research into long-term engagement and the complexities of managing multiple factors within the proposed HRI setting.

A direction for future research is to investigate the long-term effects of different types of feedback on user engagement and performance, particularly in settings like education, healthcare, and home environments. Metrics such as blink rate, gaze, distance, and valence could be examined to determine if they consistently reflect engagement over time, as they have shown a relationship with short-term performance. Additionally, it is possible that the novelty of interacting with a robot may wear off, which could impact user engagement and performance. Moreover, future research should consider providing systematically generated feedback to evaluate its effects. Understanding the types of feedback that promote sustained engagement will be essential for improving human-robot interaction over the long term, particularly in vulnerable user groups such as the elderly or young individuals with special needs. However, more research is needed to ensure that the nature and appropriateness of the engagement strategies are tailored to the unique cognitive and emotional needs of these populations, thereby maximizing the long-term benefits of social robot applications in contexts like healthcare and education.

6. Conclusion

This study examines how different types of feedback from a social robot influence user performance and engagement. The findings reveal that affective-based feedback negatively impacts performance but enhances social engagement. The observation that affective-based feedback led to poorer performance, compared to other types of feedback, raises interesting questions about the relationship between emotional responses and cognitive task performance, warranting further investigation.

Additionally, the study explored participants' perceptions of the robot, revealing that during more challenging tasks, feedback was perceived as more distracting than during easier tasks. This highlights the importance of designing feedback that accounts for participants' cognitive capacity. Moreover, participants reported higher arousal levels during the Medium challenge level compared to the Easy challenge level, suggesting that increased task difficulty may promote greater engagement.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Bahram Salamat Ravandi is a Ph.D. student in Cognitive Science. His research focuses on Human-Robot Interaction, Affective Computing, and studying engagement and human behavior analysis. He develops experimental setups with social robots for emotionally intelligent interactions and is also interested in human-autonomous vehicle interaction and explainable AI.

Imran Khan is a Marie-Sködowska Curie Actions EUTOPIA-SIF Postdoctoral Fellow at the DICE Lab (University of Gothenburg). His research focuses on areas of artificial intelligence and artificial life, affective computing, and human-computer interaction with a focus on socially assistive robots. http://www.imytk.co.uk

Alva Markelius is an MSt student in AI Ethics & Society at the Leverhulme Centre for the Future of Intelligence (CFI), University of Cambridge, and a research engineer at the DICE lab, University of Gothenburg. Her research focuses on the ethics of social AI and robotics, embodied cognition, and the integration of large language models in driver monitoring systems and social robots.

Martin Bergström is a research assistant at the DICE lab, University of Gothenburg, with a background in Cognitive Science and a focus on Interaction Design. His research interests include biometric analysis, gamification, and app development.

Pierre Gander is an Associate Professor, senior lecturer, and head of the bachelor's program in cognitive science at the University of Gothenburg. He also leads the UNREAL international network for research on cognition and fictionality. He conducts research on cognition, imagination, episodic memory, and the reality-fiction distinction in memory.

Engin Erzin received his Ph.D. degree, M.Sc. degree, and B.Sc. degree from Bilkent University, Ankara, Turkiye, in 1995, 1992, and 1990, respectively, all in Electrical & Electronics Engineering. Since January 2001, he has been with the Computer Engineering and Electrical & Electronics Engineering Departments of Koç University, Istanbul, Turkey. His research interests include speech signal processing, audio-visual signal processing, human-computer interaction, and pattern recognition.

Robert Lowe is an Associate Professor at the University of Gothenburg and Senior Researcher at RISE AB. He conducts research in Human-Technology Interaction with a focus on digitized cognitive training, Human-Robot Interaction, and Driver support systems in relation to the safe use of AI and monitoring cognitive-affective states. He also leads DICE Lab https://dice-r-lab.com/robert

Ethical approval

This project has received ethical approval from the Swedish Ethical Review Authority under reference number 2022-07017-01 to collect and analyze data in Sweden and Turkiye. We obtained a separate application to collect data in Turkiye. The data collected from experiments is digitally archived on password-protected servers at the University of Gothenburg.

References

[1] World Health Organization. Ageing and health, 2024. [accessed 2025 March 7]. Available from: https://www.who.int/news-room/factsheets/detail/ageing-and-health

[2] Prince M, Comas-Herrera A, Knapp M, et al. World Alzheimer report 2016: improving healthcare for people living with dementia: coverage, quality and costs now and in the future, Research Report, Alzheimer's Disease International, 2016.

[3] Cruz-Sandoval D, Morales-Tellez A, Sandoval EB, et al. A social robot as therapy facilitator in interventions to deal with dementia-related behavioral symptoms, In: 2020 15th ACM/IEEE International Conference on Human-Robot Interaction (HRI); Cambridge, UK. 2020. p. 161–169.

[4] Triantafyllidis A, Alexiadis A, Votis K, et al. Social robot interventions for child healthcare: a systematic review of the literature. Comput. Methods Programs Biomed. Update. 2023;3:100108. doi: 10.1016/j.cmpbup.2023.100108

[5] Martinez-Martin E, Escalona F, Cazorla M. Socially assistive robots for older adults and people with autism: an overview. Electronics. 2020;9(2):367. doi: 10.3390/electronics9020367

[6] Scoglio AA, Reilly ED, Gorman JA, et al. Use of social robots in mental health and well-being research: systematic review. J Med Internet Res. 2019;21:e13322. doi: 10.2196/13322

[7] Chen S-C, Jones C, Moyle W. Social robots for depression in older adults: a systematic review. J Nurs Scholarsh. 2018;50:612–622. doi: 10.1111/jnu.2018.50.issue-6

[8] Chen S-C, Moyle W, Jones C, et al. A social robot intervention on depression, loneliness, and quality of life for Taiwanese older adults in long-term care. Int Psychogeriatr. 2020;32:981–991. doi: 10.1017/S1041610220000459

[9] Zhang Y, Song W, Tan Z, et al. Could social robots facilitate children with autism spectrum disorders in learning distrust and deception? Comput Human Behav. 2019;98:140–149. doi: 10.1016/j.chb.2019.04.008

[10] Yu C, Sommerlad A, Sakure L, et al. Socially assistive robots for people with dementia: systematic review and meta-analysis of feasibility, acceptability and the effect on cognition, neuropsychiatric symptoms and quality of life. Ageing Res Rev. 2022;78:101633. doi: 10.1016/j.arr.2022.101633

[11] Chen Z, Zheng J, Gao Y, et al. Do children with autism spectrum disorders show selective trust in social robots? J Autism Dev Disord. 2024.

[12] Forbes A, Keleher MR, Venditto M, et al. Assessing patient adherence to and engagement with digital interventions for depression in clinical trials: systematic literature review. J Med Internet Res. 2023;25:e43727. doi: 10.2196/43727

[13] Winter N, Russell L, Ugalde A, et al. Engagement strategies to improve adherence and retention in web-based mindfulness programs: systematic review. J Med Internet Res. 2022;24:e30026. doi: 10.2196/30026

[14] Chan J, Nejat G. Promoting engagement in cognitively stimulating activities using an intelligent socially assistive robot, In: 2010 IEEE/ASME International Conference on Advanced Intelligent Mechatronics; Montreal, QC, Canada. 2010. p. 533–538. doi: 10.1109/AIM.2010.5695806

[15] Ahmad MI, Mubin O, Shahid S, et al. Robot's adaptive emotional feedback sustains children's social engagement and promotes their vocabulary learning: a long-term child–robot interaction study. Adapt Behav. 2019;27:243–266. doi: 10.1177/1059712319844182

[16] Doherty K, Doherty G. Engagement in HCI: conception, theory and measurement. ACM Comput Surv. 2018;51:99.

[17] Oertel C, Castellano G, Chetouani M, et al. Engagement in human-agent interaction: an overview. Front Robot AI. 2020;7:92. doi: 10.3389/frobt.2020.00092

[18] O'Brien HL, Toms EG. What is user engagement? a conceptual framework for defining user engagement with technology. J Am Soc Inf Sci Technol. 2008;59(6):938–955. doi: 10.1002/asi.v59:6

[19] Wangelin BC, Bradley MM, Kastner A, et al. Affective engagement for facial expressions and emotional scenes: the influence of social anxiety. Biol Psychol. 2012;91(1):103–110. doi: 10.1016/j.biopsycho.2012.05.002

[20] Mauri M, Magagnin V, Cipresso P, et al. Psychophysiological signals associated with affective states. Annu Int Conf IEEE Eng Med Biol Soc. 2010;2010:3563–3566.

[21] Koslouski JB, Wilson-Mendenhall CD, Parsafar P, et al. Measuring emotional well-being through subjective report: a scoping review of reviews. BMJ Open. 2022;12(12):e062120. doi: 10.1136/bmjopen-2022-062120

[22] Law M, Sutherland C, Ahn HS, et al. Developing assistive robots for people with mild cognitive impairment and mild dementia: a qualitative study with older adults and experts in aged care. BMJ Open. 2019;9:e031937. doi: 10.1136/bmjopen-2019-031937

[23] Carros F, Meurer J, Löffler D, et al. Exploring human-robot interaction with the elderly: results from a ten-week case study in a care home. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20; Honolulu, HI, USA. New York: Association for Computing Machinery; 2020. p. 1–12. doi: 10.1145/3313831.3376402

[24] Moro C, Lin S, Nejat G, et al. Social robots and seniors: a comparative study on the influence of dynamic social features on human–robot interaction. Int J Soc Robot. 2019;11:5–24. doi: 10.1007/s12369-018-0488-1

[25] Fiske ST, Cuddy AJ, Glick P. Universal dimensions of social cognition: warmth and competence. Trends Cogn Sci (Regul Ed). 2007;11(2):77–83. doi: 10.1016/j.tics.2006.11.005

[26] Liu SX, Shen Q, Hancock J. Can a social robot be too warm or too competent? Older chinese adults' perceptions of social robots and vulnerabilities. Comput Human Behav. 2021;125:106942. doi: 10.1016/j.chb.2021.106942

[27] Christoforakos L, Gallucci A, Surmava-Große T, et al. Can robots earn our trust the same way humans do? a systematic exploration of competence, warmth, and anthropomorphism as determinants of trust development in HRI. Front Robot AI. 2021;8:640444. doi: 10.3389/frobt.2021.640444

[28] Marx C, Könczöl C, Altmanninger A, et al. The critical robot: impact of performance feedback on intrinsic motivation, self-esteem and psychophysiology in human–robot interaction. Int J Soc Robot. 2024;16:1751–1764. doi: 10.1007/s12369-024-01147-9

[29] Mollahosseini A, Abdollahi H, Mahoor MH. Studying effects of incorporating automated affect perception with spoken dialog in social robots. In: 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN); Nanjing, China. 2018. p. 783–789. doi: 10.1109/ROMAN.2018.8525777

[30] Tozadore DC, Pinto AH, Romero RA. Variation in a humanoid robot behavior to analyse interaction quality in pedagogical sessions with children. In: 2016 XIII Latin American Robotics Symposium and IV Brazilian Robotics Symposium (LARS/SBR); Recife, Brazil. 2016. p. 133–138. doi: 10.1109/LARS-SBR.2016.29

[31] Spatola N, Belletier C, Chausse P, et al. Improved cognitive control in presence of anthropomorphized robots. Int J Soc Robot. 2019;11:463–476. doi: 10.1007/s12369-018-00511-w

[32] Donnermann M, Lein M, Messingschlager T, et al. Social robots and gamification for technology supported learning: an empirical study on engagement and motivation. Comput Human Behav. 2021;121:106792. doi: 10.1016/j.chb.2021.106792

[33] de Haas M, Vogt P, Krahmer E. The effects of feedback on children's engagement and learning outcomes in robot-assisted second language learning. Front Robot AI. 2020;7:101. doi: 10.3389/frobt.2020.00101

[34] Vivas AB, Ypsilanti A, Ladas AI, et al. Enhancement of visuospatial working memory by the differential outcomes procedure in mild cognitive impairment and Alzheimer's disease. Front Aging Neurosci. 2018;10:364. doi: 10.3389/fnagi.2018.00364

[35] Ghafurian M, Chandra S, Hutchinson R, et al. Systematic review of social robots for health and wellbeing: a personal healthcare journey lens. ACM Trans Human-Robot Interact. 2024;14:19.
[36] Ragno L, Borboni A, Vannetti F, et al. Application of social robots in healthcare: review on characteristics, requirements, technical solutions. Sensors (Basel). 2023;23:6820. doi: 10.3390/s23156820
[37] Johal W. Research trends in social robots for learning. Curr Robot Rep. 2020;1(3):75–83. doi: 10.1007/s43154-020-00008-3
[38] Alam A. Social robots in education for long-term human-robot interaction: socially supportive behaviour of robotic tutor for creating robo-tangible learning environment in a guided discovery learning interaction. ECS Trans. 2022;107(1):12389. doi: 10.1149/10701.12389ecst
[39] Błażejowska G, Gruba Ł, Indurkhya B, et al. A study on the role of affective feedback in robot-assisted learning. Sensors. 2023;23(3):1181. doi: 10.3390/s23031181
[40] Konijn EA, Hoorn JF. Robot tutor and pupils' educational ability: teaching the times tables. Comput Educ. 2020;157:103970. doi: 10.1016/j.compedu.2020.103970
[41] Kennedy J, Baxter P, Belpaeme T. The robot who tried too hard: social behaviour of a robot tutor can negatively affect child learning. In: 2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI); Portland, OR, USA. 2015. p. 67–74.
[42] Hadfield J, Chalvatzaki G, Koutras P, et al. A deep learning approach for multi-view engagement estimation of children in a child-robot joint attention task. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); Macau, China. 2019. p. 1251–1256. doi: 10.1109/IROS40897.2019.8968443
[43] Belkaid M, Kompatsiari K, Tommaso DD, et al. Mutual gaze with a robot affects human neural activity and delays decision-making processes. Sci Robot. 2021;6(58):eabc5044. doi: 10.1126/scirobotics.abc5044
[44] Ben-Youssef A, Varni G, Essid S, et al. On-the-fly detection of user engagement decrease in spontaneous human-robot interaction using recurrent and deep neural networks. Int J Soc Robot. 2019;11:815–828. doi: 10.1007/s12369-019-00591-2
[45] Jain S, Thiagarajan B, Shi Z, et al. Modeling engagement in long-term, in-home socially assistive robot interventions for children with autism spectrum disorders. Sci Robot. 2020;5:eaaz3791. doi: 10.1126/scirobotics.aaz3791
[46] Poltorak N, Drimus A. Human-robot interaction assessment using dynamic engagement profiles. In: 2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids); Birmingham, UK. 2017. p. 649–654. doi: 10.1109/HUMANOIDS.2017.8246941
[47] Ranti C, Jones W, Klin A, et al. Blink rate patterns provide a reliable measure of individual engagement with scene content. Sci Rep. 2020;10:8267. doi: 10.1038/s41598-020-64999-x
[48] Maffei A, Angrilli A. Spontaneous blink rate as an index of attention and emotion during film clips viewing. Physiol Behav. 2019;204:256–263. doi: 10.1016/j.physbeh.2019.02.037
[49] Ravandi BS, Khan I, Gander P, et al. Deep learning approaches for user engagement detection in human-robot interaction: a scoping review. Int J Hum–Comput Interact. 2025;1–19. doi: 10.1080/10447318.2025.2470277
[50] Abuhamdeh S. Investigating the 'flow' experience: key conceptual and operational issues. Front Psychol. 2020;11:158. doi: 10.3389/fpsyg.2020.00158
[51] Salamat Ravandi B. Gamification for personalized human-robot interaction in companion social robots. In: 2024 12th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW); Glasgow, UK. 2024. p. 106–110. doi: 10.1109/ACIIW63320.2024.00021
[52] Robinson NL, Turkay S, Cooper LAN, et al. Social robots with gamification principles to increase long-term user interaction. In: Proceedings of the 31st Australian Conference on Human-Computer-Interaction; Fremantle, WA, Australia. New York: Association for Computing Machinery; 2020. p. 359–363.
[53] So W-C, Cheng C-H, Law W-W, et al. Robot dramas may improve joint attention of chinese-speaking low-functioning children with autism: stepped wedge trials. Disabil Rehabil Assist Technol. 2020;18:195–204. doi: 10.1080/17483107.2020.1841836
[54] Riedmann A, Schaper P, Lugrin B. Integration of a social robot and gamification in adult learning and effects on motivation, engagement and performance. AI Soc. 2024;39:369–388. doi: 10.1007/s00146-022-01514-y
[55] Markelius A, Sjöberg S, Bergström M, et al. Differential outcomes training of visuospatial memory: a gamified approach using a socially assistive robot. Int J Soc Robot. 2024;16:363–384. doi: 10.1007/s12369-023-01083-0
[56] Paas F, van Merriënboer JJG. Cognitive-load theory: methods to manage working memory load in the learning of complex tasks. Curr Dir Psychol Sci. 2020;29(4):394–398. doi: 10.1177/0963721420922183
[57] Vivas AB, Ypsilanti A, Ladas AI, et al. Enhancement of visuospatial working memory by the differential outcomes procedure in mild cognitive impairment and Alzheimer's disease. Front Aging Neurosci. 2018;10:364. doi: 10.3389/fnagi.2018.00364
[58] Bradley MM, Lang PJ. Measuring emotion: the self-assessment manikin and the semantic differential. J Behav Ther Exp Psychiatry. 1994;25(1):49–59. doi: 10.1016/0005-7916(94)90063-9
[59] Spatola N, Kühnlenz B, Cheng G. Perception and evaluation in human–robot interaction: the human–robot interaction evaluation scale (hries) – a multicomponent approach of anthropomorphism. Int J Soc Robot. 2021;13:1517–1539. doi: 10.1007/s12369-020-00667-4
[60] Bishay M, Preston K, Strafuss M, et al. Affdex 2.0: a real-time facial expression analysis toolkit. In: 2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG); Waikoloa Beach, HI, USA. IEEE; 2023. pp. 1–8. doi: 10.1109/FG57933.2023.10042673
[61] Kompatsiari K, Ciardo F, Tikhanoff V, et al. It's in the eyes: the engaging role of mutual gaze in HRI. Int J Soc Robot. 2021;13(3):525–535.
[62] Kompatsiari K, Ciardo F, De Tommaso D, et al. Measuring engagement elicited by eye contact in human-robot interaction. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); Macau, China. 2019. p. 6979–6985. doi: 10.1109/IROS40897.2019.8967747

[63] Ayala N, Martin Calderon CA, Sharma N, et al. Examining the utility of blink rate as a proxy for cognitive load in flight simulation. In: Proceedings of the 2024 Symposium on Eye Tracking Research and Applications; Glasgow, UK. New York: Association for Computing Machinery; 2024.

[64] Zhao F, Gaschler R, Schnotz W, et al. Regulating distance to the screen while engaging in difficult tasks. Frontline Learn Res. 2020;8(6):59–76. doi: 10.14786/flr.v8i6.663

[65] Jirak D, Aoki M, Yanagi T, et al. Is it me or the robot? a critical evaluation of human affective state recognition in a cognitive task. Front Neurorobot. 2022;16:882483. doi: 10.3389/fnbot.2022.882483

[66] Namba S, Sato W, Matsui H. Spatio-temporal properties of amused, embarrassed, and pained smiles. J Nonverbal Behav. 2022;46:467–483. doi: 10.1007/s10919-022-00404-7

[67] Horiuchi H, Tanaka H. Concentration evaluation based on blinks of eyes. International Symposium on Affective Science and Engineering (ISASE). 2023;ISASE2023:1–4.

[68] Camilleri A, Dogramadzi S, Caleb-Solly P. A study on the effects of cognitive overloading and distractions on human movement during robot-assisted dressing. Front Robot AI. 2022;9:815871. doi: 10.3389/frobt.2022.815871