



Differential Outcomes Training of Visuospatial Memory: A Gamified Approach Using a Socially Assistive Robot

Alva Markelius^{1,3} · Sofia Sjöberg¹ · Martin Bergström¹ · Bahram Salamat Ravandi¹ · Ana B. Vivas² · Imran Khan¹ · Robert Lowe^{1,4}

Accepted: 27 November 2023 / Published online: 22 December 2023
© The Author(s) 2023

Abstract

This study represents the first stage of evaluating whether cognitive training interventions may be facilitated by the presence of a socially assistive robot (SAR) and gamification. Our experimental setup involves using a SAR providing feedback to a gamified visuospatial working memory task, administered according to a differential outcomes training (DOT) protocol. The study's main objective was to investigate whether performance and attitude towards the task would be affected by different robotic setups (none, simulated or physical) and in relation to different challenge levels. We measured performance accuracy on the gamified visuospatial memory task and self-reported affective ratings, which are relevant for assessing attitude towards the task and providing indicators to the potential for using a SAR for a longer-term cognitive intervention. Additionally, we conducted exploratory analyses of eye movement strategies for memory encoding during the task. The results demonstrated a significant differential outcomes effect (DOE) on memory performance accuracy, regardless of Robot type and Challenge level, providing evidence that a DOE can still be obtained when a SAR interacts with participants. Moreover, the results from the affective ratings revealed that participants accompanied by the physical robot reported lower levels of stress and increased levels of control. Our results demonstrate, for the first time, a DOE using a SAR in a gamified context. This result, coupled with positive subjective reporting of the human–robot interactive experience of participants, demonstrates the potential for using a SAR to: (i) promote positive attitudes for a DOT-based cognitive intervention, without (ii) negatively affecting task performance.

Keywords Human–robot interaction · Socially assistive robots · Gamification · Cognitive intervention · Differential outcomes training

1 Introduction

Presently, dementia is one of the primary causes of cognitive impairment among older people and the seventh leading cause of mortality among all diseases worldwide

[1]. Alzheimer's disease (AD) is the most common form of dementia and contributes to 60–70% of dementia cases [1], and is typically preceded by Mild Cognitive Impairment (MCI) a clinical stage between healthy ageing and dementia [2]. According to recent estimates, the number of people living with MCI and dementia is expected to rise from 55 million in 2022 to 139 million by 2050 [3]. This projected increase in persons living with dementia (PwD) has societal, economic, and health-related consequences with increases in cost of care further contributing to existing barriers to accessing care for persons with dementia [4]. Additionally, the recent COVID-19 pandemic has affected the ability of seniors (those who are more susceptible to memory loss) to travel: a problem that is expected to continue into the future [5]. A crucial challenge that must be addressed, therefore, is how to make existing treatments accessible, scalable, and cost-effective to PwD and their caregivers.

✉ Alva Markelius
ajkm4@cam.ac.uk

✉ Robert Lowe
robert.lowe@ait.gu.se

¹ Department of Applied IT, University of Gothenburg, Gothenburg, Sweden

² Psychology Department, CITY College, University of York Europe Campus, Thessaloniki, Greece

³ Leverhulme Centre for the Future of Intelligence, University of Cambridge, Cambridge, UK

⁴ RISE Research Institutes of Sweden, Gothenburg, Sweden

Given that there are currently no pharmacological treatments to cure dementia [6], a complementary strategy may be to attempt to mitigate impairments in older adults through cognitive-behavioural training. Patients with MCI may be a good target for non-pharmacological solutions, such as cognitive interventions that may serve to mostly retain every day cognitive functioning. Thus, short- and long-term cognitive training is currently gaining traction as an important intervention for persons with MCI for its potential efficacy of delaying progression to dementia, or possibly preventing its onset altogether [7]. One approach to making these interventions more accessible and scalable is to digitise them (usage of technological media for delivery of treatments and therapies [8]). Existing digitised cognitive interventions, such as gamified memory training, have been identified as potential therapeutic interventions [9], which can be used to scale treatments up to wider populations, improving accessibility to, adherence and efficacy of, treatments. Treatment adherence is of particular relevance since several studies have shown low medical treatment adherence for chronic conditions [10] and for older adults in particular [11]. Lack of treatment adherence has been identified as a significant public health issue [12] and low rate of treatment adherence has been identified with higher associated costs [13]. On the other hand, engagement with gamified cognitive interventions of a long duration (e.g. several weeks) may still be relatively low particularly when carried out in individual unsupervised settings such as in the home. PwD often require extensive nursing or caregiver supervision due to the progressive nature of the condition and the challenges it presents in terms of cognitive and functional abilities. These roles are demanding as they often require specialized skills, patience, and empathy to provide adequate care and support [14]. While not replacing the role of human caregivers, digitised assistance, such as that provided by Socially Assistive Robots (SARs), might complement these roles by providing assistance with routine tasks [15]. Integrating SARs into dementia care settings might thus result in improved outcomes and treatment adherence for PwD by complementing, not replacing, human caregivers.

In this study we propose a new digitised cognitive training setup consisting of a visuospatial task with simple game-like elements, both with or without a (physical or simulated) SAR. As an interactive partner, the SAR provides task-related feedback with the aim of enhancing engagement and motivation for the task. While physical robots, in some contexts, may be preferable for use as compared to simulated versions, e.g. for learning [16], simulated robots (and virtual agents) are expected to be more easily accessible in home settings. This pertains to the fact that they can be accessible through most personal computer interfaces, as compared to physical robots that are associated with high costs and low availability. Potentially, both forms could be combined

within interventions, e.g. simulated versions within affordable mobile technologies periodically backed up by use of the physical robot in a clinical facility. Thus, in this study we investigate the use of both a simulated and physical robot to test both versions' efficiency in the context of using a gamified task to enhance visuospatial memory. Our digitised (gamified) memory training task utilises a differential outcomes training (DOT) [17] methodology for the purpose of enhancing visuospatial memory. DOT is a well-studied paradigm for both clinical and non-clinical research. It is typically applied as an experimental, laboratory-based, learning/memory training protocol. The procedure standardly involves single-session reward-based training (as compared to multi-session intervention-based training) and is characterised by presenting unique (differential) reward feedback to correctly remembered responses to specific stimuli (rather than the same reward feedback regardless of stimuli). The task we have selected is a gamified version of a task developed by Vivas et al. [18] who conducted a study using DOT to enhance visuospatial memory in PwD, MCI, and healthy older adults. Detailed description of DOT is presented in Sect. 1.3.

The current study has two objectives: Firstly, to assess whether or not different robotics (simulated or physical) setups can be integrated within a known memory training protocol (DOT); Secondly, to assess the participants' self-reported affective experience towards the setups and in relation to performance accuracy, as a proxy for long term acceptance and viability. This is the first time, to the authors' knowledge, that the combined effects of DOT, gamification and SARs on memory have been investigated together. This study serves as a non-clinical validation experiment with healthy young adults. Its focus is on investigating parameters relevant to longer-term engagement in cognitive interventions for participants with MCI. The hypotheses of the study are:

1. The memory training protocol will be effective (higher performance accuracy under differential outcomes relative to non-differential outcomes training) over all the setups (different robotics setups and non-robotics control).
2. Participants' affective experience will differ as a function of setup.
3. Participants' affective experience will correlate with overall memory performance.

To test the above, we conducted an experiment to evaluate the three setups (robot, simulated, control) according to memory accuracy performance and affective experience. Memory performance was measured on differential outcomes training (DOT) and non-differential outcomes training (Non-DOT) conditions. Affective experience was measured according

to self-reported data collected through a Self Assessment Manikin Scale (SAM-scale) questionnaire [19]. In addition, eye movements were recorded throughout the experiment to explore potential individual differences in eye movement strategies. We were interested in exploring and understanding individual differences in strategies during encoding of the locations. It has been previously shown that eye movement strategies during encoding can influence memory performance (e.g. [20]) and therefore we wanted to investigate if potential individual differences may interact with the effectiveness of the DOT.

The rest of this paper is set out as follows: Sect. 1.1 provides background on SARs in cognitive training and provides an overview of the specific training procedure employed; Sect. 1.2 describes gamification approaches relevant to cognitive tasks and training; Sect. 1.3 provides an introduction to Differential Outcomes Training; Sect. 2 describes the methodology including a full overview of the gamified visuospatial memory task; Sects. 3 and 4 follow with results and discussion, respectively.

1.1 Socially Assistive Robots in Cognitive Training

The use of SARs in therapeutic and caregiving contexts has seen an increase in recent years [21], with a steady shift away from strictly-academic environments and industrial applications, to home, and consumer-based, markets [22]. Possibly owing to a combination of an ageing population [23], rising incidences of diseases and disorders such as dementia [3] and Autism Spectrum Disorder (ASD) [24], the rising costs of healthcare [25] and high turnover of clinical staff [26, 27], researchers and medical practitioners have been exploring alternative approaches to deliver treatments and care.

Arguably one of the core research applications for these types of robots [28], SARs, has been used in interactions with persons with special needs, such as those with developmental disorders or neurodivergence. For PwD, for example, [29] explored using a PARO robot as a companion, both within a group setting at a day-care centre, and individually in their respective homes. Participants reported reduced levels of anxiety and improved mood at home after 12 weeks of interacting with the PARO robot. Because of its animal-like shape and toy-like features, this robot is typically implemented for applications and target groups in need of companions for company and comfort (as opposed to e.g. cognitive training as in the present study). Similar positive effects of companion-type SARs have also been found in [30, 31]. In the former, the Ryan robot engaged with 6 adults with mild-to-moderate dementia in a senior living facility, through activities such as playing games, showing photos, reminding them about their daily schedule, and having conversations with them. Participants reported reduced levels of depressive symptoms and improvements in their quality of life. In the latter, the

Mario-robot was used to mitigate and reduce stress for people with dementia, with results indicating supported resilience to stress to a majority of the participants as well as facilitation of positive and meaningful social support. The latter two robots exhibit a more humanoid form, with their primary applications extending beyond mere comfort and companionship, as is often the case in SARs with pet-like (animal) embodiments [29, 32, 33]. These humanoid robots are designed to engage in more advanced social interactions, dialogues, and conversations. The focus of their applications are usually different from comfort and companionship, instead aiming to provide enhanced assistance to specific target groups. Their intended targets include addressing issues such as stress or depressive symptoms in seniors or persons with dementia, rather than just offering (short-term) companionship.

The benefits of pet-like (animal) SARs are oftentimes restricted to the short term, owing to their limited communication or expressive modalities, and lack of human-like attributes [29, 32, 33]. These limitations are likely to hinder their long-term efficacy as socially assistive partners [34, 35] (in other words, with sufficient interaction with the agent to expose the limited repertoire of (non-adaptive) behaviours that it is capable of expressing). On the other hand, despite their potential for inducing eeriness (i.e. the “uncanny valley” [36]) in the short-term, SARs with humanoid embodiments can play a crucial role in addressing the limitations of pet-like robots, with current evidence showing how humanoid SARs result in greater levels of trust [37, 38], usability, acceptability, and enjoyment in those interacting with them [39]. Such improvements may come from meeting participants’ unconscious expectations of social interactions, through familiarity of more obvious, expressive social interactions, or through a wider range of (multimodal) social behaviours. Ergo, the initial “eeriness” associated with short-term exposure to human-like robots (i.e. the uncanny valley [36]) may be overcome [40] (or may exist much later, or not at all, in some persons [41]), leading to long-term benefits that outweigh the (potential) temporary discomfort in the short-term. However, given the focus of short-term interactions with SARs in caregiving contexts presently, these proposed long-term benefits are still yet to be understood from the existing literature.

Audiovisual reinforcing feedback has been demonstrated to be of crucial importance to improving engagement and learning [42, 43]. SARs with affective interactive capabilities have previously been associated with positive affective responses in therapeutic contexts [44] and with improved learning and engagement with feedback [45]. This has also been investigated in the context of older adults and PwD. For example, Cruz-Sandoval and Favela [46] reported positive affective responses to interacting with a social robot as well as therapeutic benefits in patients with moderate-stage dementia. Sung et al. [47] showed improved activity participation in older adults living in senior centres with longer-term robot-

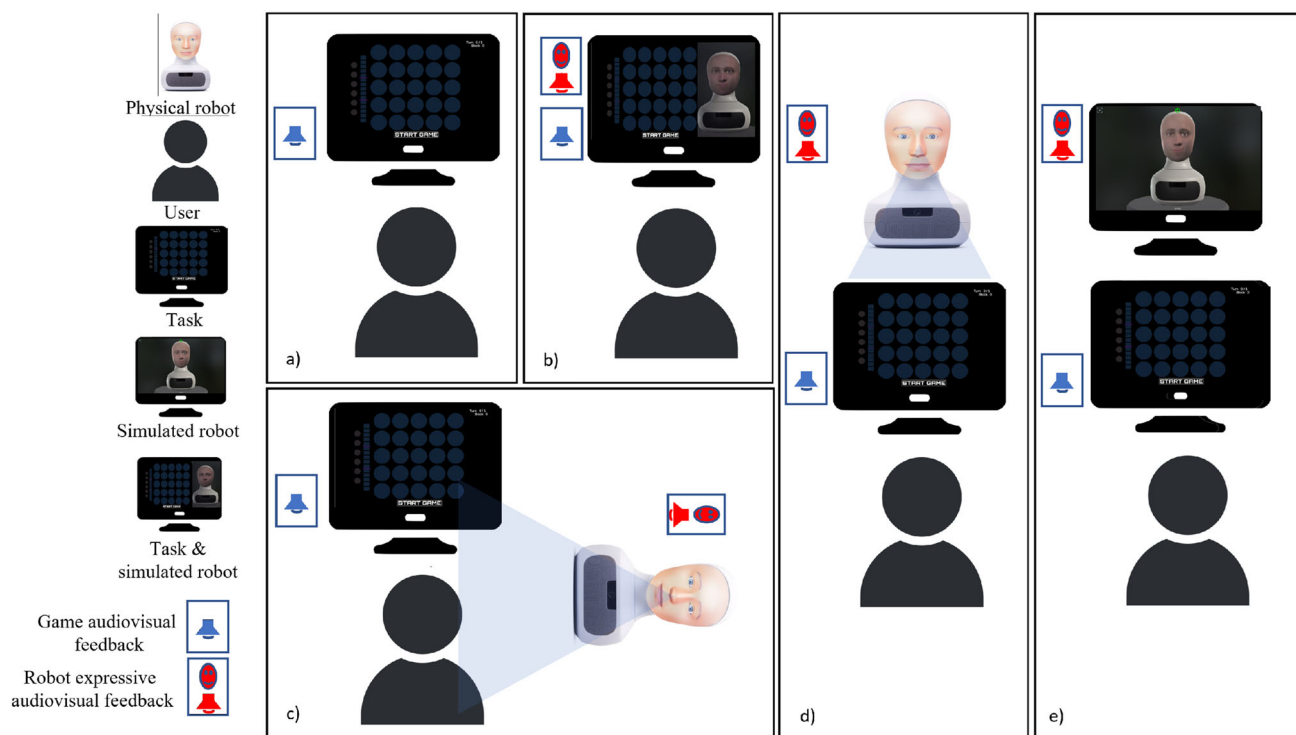


Fig. 1 Overview of potential experimental setups in line with the setup presented and implemented in this study. All setups have in common the inclusion of a digitised task and different ways of incorporating a SAR (except a) showing the setup with task only). **a**, **d** and **e** are the setups

assisted therapy. Andriella et al. [48] described a framework and presented a cognitive robotic system designed to assist patients with mild dementia during brain-training sessions with promising results. Besides showing the potential positive effects of human–robot interaction in memory training, the authors also emphasised the need for further research on improving engagement. Assistive robots have great potential as engaging tools for user interaction in the context of providing both home and on-site treatments with safer, non-contagious interaction and specialised assistance. Similar attempts have been made to use robotic assistance to increase the effectiveness of social and cognitive training interventions and make them more accessible in the context of people with dementia. A study by Chan et al. [49] showed significant benefits of using a SAR for game engagement; specifically, feedback in the form of instructive phrases increased the participants’ attention to the game.

The current investigation implements both a simulated version and a physical version of the Furhat robot,¹ as well as a non-agent (control) condition. We sought to evaluate whether simulated or physical robots could affect participants’ affective reactions towards the gamified memory training and whether their presence affected the effective-

ness of the memory training. The rationale of using Furhat as the SAR within this study is mainly because of its inherent customisability, encompassing attributes ranging from facial appearance and vocal characteristics to tone modulation and facial gesturing. This research serves as an initial step towards a deployment of a SAR to keep patients with MCI or dementia engaged in memory training. The customisable attributes of Furhat, allowing tailoring of its characteristics, holds significant value, enabling the tailoring of assistance during interventions in consonance with the specific needs and preferences of the respective patients. This includes, for example, making the assistance and interventions adaptable to be suitable and more effective across different social or cultural norms or standards or simply individual personal preference. Ensuring individual preferences are met (e.g. through co-design) would thus not only make the interaction with the SAR more enjoyable but also hopefully contribute to more effective interventions.

This work represents the first stage of several ongoing investigations that concern use of the Furhat robot for engaging participants in the task. It is imperative that the robot does not distract the participant from the task but that its presence is nevertheless felt. We are, therefore, required to consider *where* the robot is placed, *when* the robot interacts, and *how* the robot interacts, during the task. Figure 1 shows example

implemented in this study and the others are visualised for illustrating potential variations that can be implemented depending on application and different needs

¹ <https://furhatrobotics.com/>.

setups regarding the placement of Furhat, participant (‘User’) and computerized task. The present study investigates three of our suggested possible setups, to gain insights into how the placement, interaction timing and feedback given by the robot affects the task performance and DOT-effect (DOE). The simulated version of the robotic agent allows for its being embedded within the screen (Fig. 1b) and placed therefore close to the task area, which may thereby allow for quick glances at the (simulated) robot whilst retaining focus on the task. The physical version of the robot allows for a triangulated placement (Fig. 1c) whereby Furhat can look both at task and participant during and between trials thereby allowing for “connection events” [50] such as ‘directed gaze’ (turning to look at the same object/task) and ‘mutual facial gaze’ and so a different mode of embodied interaction. Our choice of setups in this experiment (Fig. 1a, d, e) reflect a desire to initially test the effects of a minimally distracting robot (in terms of placement, time of interaction and mode of interaction): the robot is placed close to the screen and gives verbal and facially expressive feedback only at the end of each trial (see Sect. 1.3), which mirrors the outcome result displayed on the screen. In ongoing investigations we are testing the effects of greater Furhat interactions both during the task trials and at the end of blocks of trials and using different connection events (specifically in relation to Fig. 1c).

1.2 Gamification

Gamification entails the use of game-like features to render monotonous tasks or applications more engaging. Examples of common gamification elements such as points [51], progress bars [52], and challenge levels [52] have become staples in numerous gamified tasks, including learning activities. These elements have demonstrated their effectiveness in enhancing engagement and improving performance in the tasks. Points, for example, have been shown to have an effect on game performance when perceived as a more entertaining game element [51]. Participants’ attitude towards challenge levels as a game element has also been shown to be effective; [52], for example, found that a majority of participants, independent of personality type, reported positive attitudes regardless of the level. The inclusion of game-like features for clinical interventions for memory training can allow for more engaging and entertaining interventions, which may increase compliance and adherence to the intervention. This in turn can lead to greater intervention efficacy, particularly in a home-based setting where there is little or no supervision. The incorporation of gamification into cognitive training for people with MCI provides essential elements that enhance motivation, as it allows for effortless diversification and the introduction of fresh components, preventing excessive repetition or tedium. This includes the possibility of in-game advice and assistance which is crucial for settings with less

supervision [53]. Sailer et al. [9] emphasises that gamification is aimed at having a direct impact on learning practices and attitudes and a beneficial, indirect effect on learning outcomes. Gamification is gaining traction as an approach for making more engaging interventions for both children and adults, clinical and non-clinical, not least in the context of learning and memory [54–56]. In 2020, the Food and Drug Administration (FDA) authorised a video game-based prescription treatment (EndeavorRx) for children with attention deficit hyperactivity disorder (ADHD), constituting the first time such a gamified digital therapy received FDA approval [57]. Whilst gamified cognitive training for MCI and PwD [58] and detection of early stage dementia [59] already exists, EndeavourRX’s results improve the profile of gamification. These results also potentially represent an inflection point in the use of evidence-based gamified cognitive therapies for various neurological disorders (including dementia and related cognitive impairments). In the present study, the process of gamifying an established visuospatial memory training task entailed the integration of standard game-inspired elements, as previously alluded to. The main gamification present in this study thus relates to the memory training task, which is fully explained in detail in Sect. 2.1. The purpose of the SAR is to assist, facilitate and engage participants carrying out the task, and is thus not necessarily seen as a gamification element in itself per se. However, given that avatars and virtual assistants are not uncommon as game elements, certain elements of the SAR could be argued to also belong to the gamification because of its similarities to such avatars.

1.3 Differential Outcomes Training

The specific memory training procedure employed in the present study, the Differential Outcomes Training (DOT), is a non-invasive, relatively easy-to-implement protocol that has been shown to be effective in improving learning and memory in several non-clinical (e.g., children and older adults) and clinical (e.g., people with MCI, and dementia) populations. DOT is typically applied in experimentally controlled laboratory settings and, again most typically, is implemented according to the following phases:

1. Encoding phase: one or more (sample) stimuli ‘S’ are presented (e.g. an image on screen);
2. Memory phase: a delay of several seconds (blank screen or screen with distractor stimuli) occurs;
3. Response phase: one or more response (‘R’) options (requiring selection of another stimulus, e.g. image, on screen) are presented.

The above sequence (a learning unit or ‘trial’) is repeated many times and at the end of each trial, if the ‘correct’

response is selected, for example, if the response stimulus chosen matches with the sample stimulus ('match to sample' setup), rewarding feedback is signalled. This feedback is typically image-based and may relate to some real world reward to be received after the experiment. Participants are required to learn S–R associations over the allocated number of trials such that S1 may be learned to be associated with R1 in order to get the reward (R2 yields no reward when S1 is presented) and S2 is learned to require a response R2 (not R1) to yield a reward (see Fig. 2). The trials using the DOT protocol, as compared to non-DOT control conditions, entail a specific reward being given for a particular S–R pair that differs from (is differential to) rewards given for each other S–R pair. As an example, a controlled setup with stimuli presented as images of a blue pill or a pink pill, might be followed by response options in the form of an image-to-select depicting either a sunny morning or a starry night. So what might be learned, in plain language, could be "I take the blue cholesterol pill at night, and the pink blood-pressure pill in the morning". Learning these associations then is enhanced by rewarding these two rules *differentially*, e.g. with social praise in the 'blue pill→night' pairing, and with a monetary reward in the 'pink pill→morning pairing'.

The faster learning found in DOT has been theorized [17] to owe to the use of differential expectations for differential outcomes. The outcome specific expectations may be triggered following the initial (sample) stimulus (S) presentation and provides additional information (serving as internalized discriminative stimuli) for correct responding (see Fig. 2 for details). The simple manipulation of arranging the outcomes so that they are unique and specific to the new stimulus to be learned has been shown to be effective in improving discriminative learning, and delayed memory recognition in animals and humans (e.g., [18, 60–62]). The effects in learning and memory found with DOT are robust and of a medium-large effect size, as shown in a recent meta-analysis [63], which support the great potential of this protocol for future clinical applications.

2 Methods

Experimentally, the DOT has been shown to enhance visuospatial working memory in people with MCI and dementia [18]. In the present study, we further tested for the first time a gamified version of a visuospatial working memory task with DOT developed by Vivas et al. [18]. This is an important first step to develop other gamified tasks using DOT that can be applied in clinical and non-clinical (e.g., schools) settings to enhance learning and memory. Since the key manipulation of the DOT is the arrangement of feedback (outcomes), we believe that it also has the potential to be further developed in the context of SARs. The present study is a stepping stone

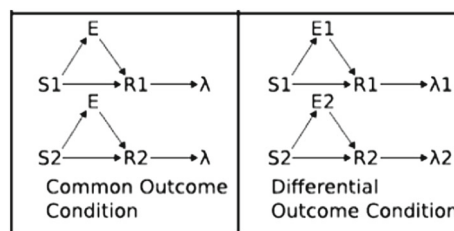


Fig. 2 Abstract depiction of Common and Differential Outcomes learning components. For each trial of a DOT experimental task, a given stimulus (S1 or S2) is followed by two response options (R1 and R2) in a given learning trial but only one response is correct for the stimulus, e.g. S1 requires R1 (not R2) as the response and S2 requires R2 (not R1) as the response in order to receive the rewarding feedback/outcome (λ). Left. Common Outcomes components. A stimulus (S1 or S2) is associated with a response (R1 or R2) but also an expectation (E) in this case carries no informational value and does not facilitate discrimination between response possibilities (given that R1 and R2 are presented in a learning trial). Right. Differential Outcomes components. Specific stimuli, and specific responses, are connected to specific outcomes, e.g. $S1 \rightarrow R1 \rightarrow \lambda1$ and $S2 \rightarrow R2 \rightarrow \lambda2$. If Differential Outcomes components exist, the stimulus (S1 or S2) presentation at the beginning of a learning trial is hypothesised to trigger specific expectations (E1 or E2, respectively) for the concomitant outcomes and strengthen the tendency for the associated response (a differential outcomes effect). An alternative control to the Common Outcome condition, 'Non-Differential Outcomes training' (NDOT), entails random outcomes of those possible, e.g. $\lambda1$ or $\lambda2$, presented following the correct response. As for Common Outcomes, the NDOT learned expectations carry no additional information to help discriminate response selection (as $\lambda1$ and $\lambda2$ are equally likely to be presented for a given correct R). The S–R route or process is referred to as the *retrospective route* (implicated in rehearsal strategies and working memory), whereas the S–E–R route or process is referred to as the *outcome expectancy route* (hypothesised to be necessary for differential outcomes based learning). Figure adapted from Peterson and Trapold [64]

in this direction, by employing for the first time a cognitive training setting involving a gamified DOT task and a simulated and physical robot.

This study implemented a gamified version of an existing visuospatial working memory (Differential Outcomes Training) task [18] that is intended to be adapted for use in a clinical intervention. For a full overview of the difference of the two versions see Appendix A. Additionally, this study investigated the use of a socially assistive robot, providing feedback while participants carried out the task. The gamified memory task had three challenge levels (easy, medium, hard), each consisting of 24 trials (See Sect. 2.1, Fig. 4) with participants completing each level sequentially (for a total of 72 trials). The order of challenge levels was not counter-balanced to follow a more realistic and natural increase in difficulty as is appropriate for an actual intervention. Section 2.1 provides a detailed description of the gamified task. Visualisations of the task interface, stimuli and rewards can be seen in Appendix B. An overview of the setup of the experimental conditions (concerning robot, simulated robot, control) can be seen in Fig. 3.

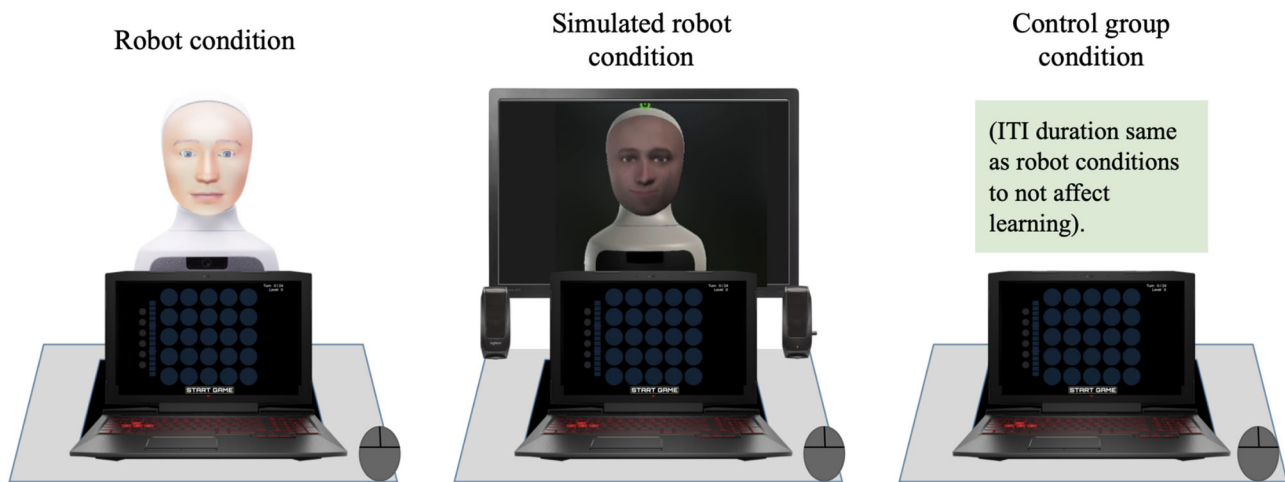


Fig. 3 The three versions of the setup from the participants' point of view with respect to *Robot type* variable: physical robot, simulated robot and the no robot condition (ITI, inter trial interval)

The experiment followed a $2 \times 3 \times 3$ factorial mixed design with the independent variables being *Training group* (Differential or Non-differential outcomes), *Robot type* (None, Simulated robot or Physical robot) and *Challenge level* (Easy, Medium, Hard). *Training group* and *Robot type* were the between-subject factors, whilst *Challenge level* was the within-subject factor. The experiment was carried out with an expected medium to large effect size for *Training group*, with respect to McCormack et al.'s [63] meta-analysis (power = 0.8), (power calculation with MorePower 6.0.4) [65] with at least 72 participants. Dependent variables were quantitatively measured and analysed based on data gathered from the gamified task (Accuracy: percentage of correct responses over each level), together with evaluation based on measures of affective ratings using the Self Assessment Manikin Scale (SAM). Furthermore, eye tracking data (using iMotions software) was analysed for identifying eye movement strategies during the memorisation phase (encoding) of the gamified task (see Sect. 2.1 and Fig. 4).

2.1 The Gamified Task

The visuospatial working memory task used in this study represents a gamified version of the task proposed by Vivas et al. [18]. The differences between our task and that of Vivas et al., including use of simple game-like elements (e.g. points bar, explicit rewards, grid representation) can be seen in Table 2 (Appendix A). The task setup consisted of a 5×5 grid made up of grey cells and a score bar on the left side of the screen filling up incrementally as the participant was rewarded for correct responses. For depiction of the gamified task interface (start screen) and stimuli used see Appendix B. The gamified task was played in a session made up of a series of trials and with difficulty settings fixed for each level. Each trial included a (sample) stimuli phase, an inter

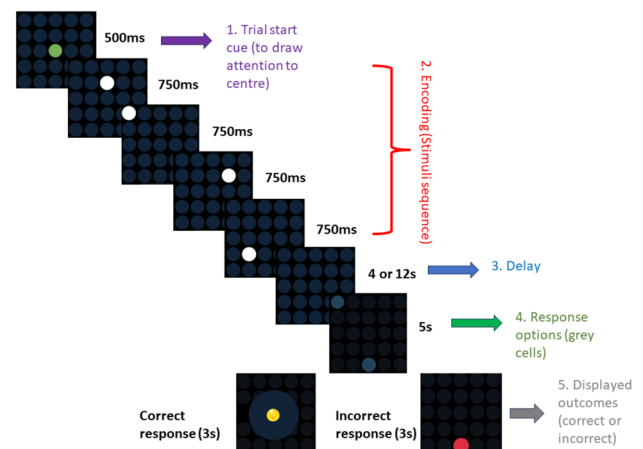


Fig. 4 Process diagram showing the steps of a single trial. Following the trial cue stimulus (1), the trial consists of: (2) an encoding phase (where participants are to remember the locations of the white cells—‘sample stimuli’), (3) a delay phase (designed to tax working memory), and (4) a response phase (participants must select a response location that matches to one of the sample stimuli presented in the encoding phase), following which (5) reward outcome or not (incorrect) is presented. For simplicity of presentation here, only target stimuli—white illuminated cells—are depicted and no distractor stimuli, e.g. red or blue cells. (Color figure online)

stimulus interval (ISI)—or delay—phase, a response phase and an outcome phase. A process diagram of a single trial is shown in Fig. 4 where for simplicity distractor cells are omitted (see Appendix B). A video showing sample trials can be found in Online Resource 1.

The task consisted of 3 levels of 24 trials each, increasing through the challenge levels (Easy, Medium, Hard). Here, difficulty refers to the manipulation of two parameters: Length of stimuli sequence (4, 6 or 8 white illuminating cells of which one was a target cell) and number of response options (two, three or four). Additionally, red and blue distractor stimuli cells appeared as every other stimulus in the stimuli sequence and every stimulus throughout the delay phase

following the stimuli sequence. The role of the distractors was to limit the extent to which participants could focus on rehearsal (working memory) strategies, related to retrospective processing (see Fig. 2) thereby encouraging the role of the outcome expectancy route (see Fig. 2) whose attentional/response selection role requires differential outcome expectations. Whether a distractor cell was red or blue was pseudo-randomly computed, with a minimum of 2 blue cells and a maximum of 4 blue cells. The length of the distractor sequence following the stimuli sequence consisted of either eight or twelve distractor stimuli (one at a time), decided by a pseudo random sequence ensuring the two lengths were presented with an equal amount of occurrences. Target (to be remembered) and distractor (during the delay) cells were defined by their colour, and the participant's task was to react accordingly to instructions received before starting the task. Specifically, the sequence of locations to be remembered appeared always in white, while during the delay distracting locations appeared in other colours.

The distractors were used so as limit the scope for rehearsal-type strategies for remembering stimuli locations and to ensure participants would attend to the screen/task throughout the trials. Distractors were either 'active' (required a response) or 'passive' (were to be ignored). For the active distractor, participants were required to press the space-bar on their keyboard for every blue cell. Timely/'missed' hits to the blue location were followed by a 'positive' or 'negative' audio feedback, respectively. For the passive distractor, if a cell illuminated red the player was instructed to ignore it.

After the delay, the response options were presented in the form of grey illuminated cells (see Fig. 4). A timer bar appeared beneath the grid and the participants had 5 s to click on one of the grey cells before 'timeout'. The correct response was the cell that had been presented as a white stimulus during the current trial encoding (stimuli sequence) phase and this was only true for one of the response cell options in the response phase. If the response was correct, the player was rewarded with an outcome out of four possible unique outcomes (one coin, three coins, one diamond or three diamonds). The difference in amounts of coins and diamonds (one or three) affected the score bar increase (designed to motivate participants) but its main role was to allow for 4 different outcomes (for DOT) that were thereby differential according to type or reward magnitude. Use of pseudo-monetary rewards was considered a gamified component (adapting the Vivas et al. [18] task). Outcomes differential by reward magnitude have been used in previous Differential outcomes experimentation in both animals (e.g. [64]) and humans (e.g. [66]).

In the DOT condition the unique outcomes were bound to specific locations (individual cells) on the grid. The spatial distribution altering the specific output is thus individual cells, not parts of the grid containing several cells. In the

Non-DOT condition, the different rewards were not bound to specific locations, but instead dependent on a pseudo random sequence. For example, in the DOT condition, every time the top right cell was selected as the correct response it would always result in a one diamond reward, and another location on a different trial would always show three diamonds; whereas in the Non-DOT condition a particular cell that was correctly responded to would always result in a reward but it would be any of the four possible rewards chosen at random. For both conditions, the number of occurrences of each outcome was balanced and was used as reward for correct response an equal amount of times over the session. Locations for rewards were randomly selected for each challenge level and thereafter fixed until the next challenge level was reached. They were also randomised for each participant.

2.2 Participants

78 adults aged between 19–46 years old ($M = 25.59$, $SD = 4.86$) participated in the study, where 39 participants were women and 39 were men. We included young and middle aged adults without cognitive impairment as self-reported by the participants. Since the study was non-clinical, the criteria for inclusion of participants was to be cognitively and neurologically healthy, as self-reported by the participants. The sampling procedure was carried out at the university campus of University of Gothenburg (Sweden) at the Knowledge Lab. All the volunteering participants provided informed written consent, and the study was conducted in accordance with the WMA Declaration of Helsinki. To enhance internal validity, all condition groups were randomised. The participants were compensated with a cinema ticket (worth approximately 180 SEK).

2.3 Equipment

The gamified task was carried out on a laptop (HP Omen) with a keyboard and a bluetooth connected mouse. To capture the participants' eye movements, Smart Eye's² AI-X eye tracker was used, connected to iMotions³ Software (version 9.2). For the participants in the None robot condition, only the aforementioned equipment was used. In the simulated robot condition, the simulated Furhat was displayed on a monitor, mounted on a pedestal behind the laptop. In the physical robot condition the Furhat robot was placed in the same position controlling for size difference with respect to the simulated version. Speakers were connected to the monitor to enhance the robot's audibility.

² <https://smarteye.se/>.

³ <https://imotions.com/>.

2.4 Pilot Study

A pilot study with six participants was conducted before the first initialization of the experiment allowing for evaluation and improvement of the different components of the setup. The experiment was revised after the pilot study by adding a warm-up level (3x3 grid), an onscreen display of how many trials each level consisted of (and had been completed at a given stage), and a randomised amount of blue cells displayed during each trial (ranging from 2–4). The six participants in the pilot study were not among the 78 participants included in the experimental analysis.

2.5 Experimental Procedure

Participants conducted the gamified task via a laptop facing either the simulated or the physical Furhat (or none, depending on condition). The gamified task was carried out by the participants individually after receiving instructions both in text form and orally. The participants in the *Robot type* conditions additionally got introduced to the simulated/physical robot as their companion. To get familiarised with the task, the participants played a test version of 10 shorter and easier trials (fewer cells in the task grid) to ensure their understanding of the task. The task then consisted of a total of 72 trials divided equally over three challenge levels (Easy, Medium, Hard), which increased in difficulty and lasted approximately 30 min in total. For a visualisation of sample trials see Online Resource 1. After each trial during the robot conditions, the participants received vocal performance feedback from Furhat (See Appendix C). The participants in the ‘no robot’ condition did not receive any vocalised feedback, but had instead the same inter-trial interval time (ITI) to control for any potential learning effects. The sequence of trials (each challenge level) were presented in blocks of 24 trials, with a break in between to reduce fatigue. During each break between the three levels of the task, the participants filled in the SAM-scale to report their affective ratings (*Valence*, *Arousal* and *Dominance* dimensions) in relation to the task and setup. The three dimensions in the SAM-scales were phrased as follows: *Valence* (happy/unhappy), *Arousal* (stressed/calm) and *Dominance* (in control/not in control), using a Likert scale from 1 to 5. (See Appendix D for a visualisation of the SAM-scale).

2.6 Data Collection and Analysis

Quantitative data was collected for the gamified task and included task accuracy (number of correct responses) for every challenge level of the task. Additionally, data collected from the SAM-scale questionnaire were analysed quantitatively for participants’ self-reported affective ratings after each *Challenge level*. ANOVAs were conducted to analyse

both accuracy and affective scores data with *Training group* (DOT/Non-DOT) and *Robot type* (none, simulated, physical) as the between-subjects factors, and *Challenge level* (easy, medium, hard) as the within-subjects factor. Pearson correlations were conducted to test for the association between overall accuracy and self-reported affective ratings. The eye tracking data (gaze saccade amplitude and frequency) were subjected to cluster analyses to investigate distinct eye movement strategies, which may have affected the encoding part of the task (i.e. when the stimuli sequence was presented – see Fig. 4 white cell sequence, and video in Online Resource 1). This data was obtained from the iMotions software was conducted with SKLearn toolkit version 0.23.2⁴ for KMeans. The statistical significance level was set at .05.

3 Results

This section is divided into three parts: the first part presents the results of the participants’ performance on the visuospatial memory training task; the second part accounts for the results of the affective ratings as self reported with the SAM-scale; the third section concerns an exploratory eye movement analysis aimed at investigating whether particular memory (stimuli encoding) strategies were used by participants.

3.1 Task Performance Accuracy

A summary of mean task accuracy across conditions can be seen in Table 1 and a visualisation of the difference in task accuracy over the different conditions is provided in Fig. 5. Performance data were subjected to a 3 x 2 x 3 mixed ANOVA with *Robot type* (none, simulated and physical) and *Training group* (DOT and Non-DOT) as the between-subject factors and *Challenge level* (1, 2 and 3) as the within-subject factor.

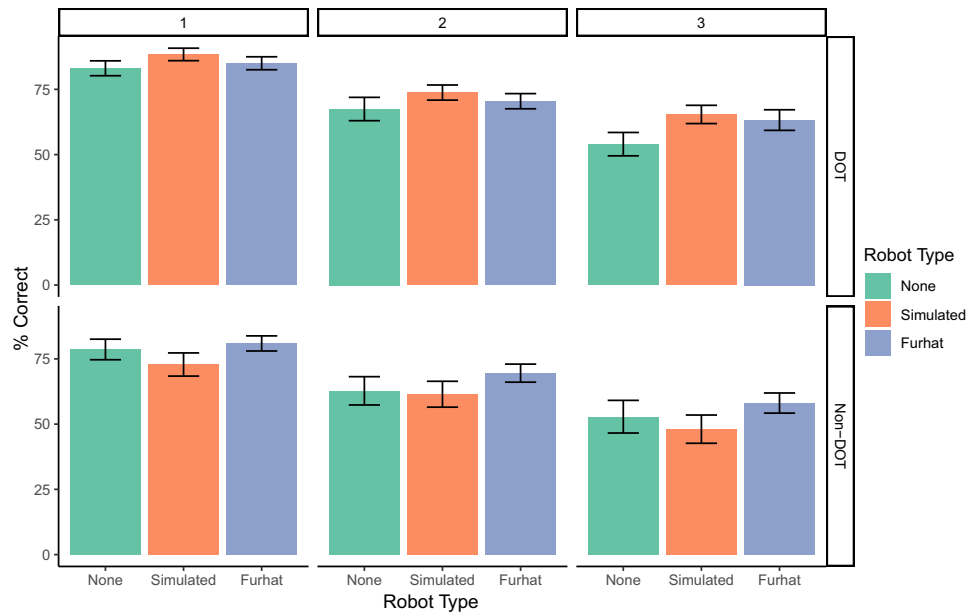
Results showed significant main effects of *Training group* and level, $F(1, 72) = 6.13, p = .016, \eta^2 = .078$, and $F(2, 144) = 158.22, p < .0001, \eta^2 = .687$, respectively. That is, performance accuracy was higher overall in the DOT group (72.23%) than in the Non-DOT group (65.02%). Least Significant Difference (LSD) post-hoc comparisons for level, also showed that overall performance accuracy decreased with increasing levels (81.44%, 67.54% and 56.89% for levels 1, 2 and 3, respectively), all $p < .001$. All the other effects and interactions did not reach statistical significance, $ps > .05$. In summary, participants conducting the visuospatial task under the DOT condition had significantly higher performance accuracy than those in the Non-DOT condition. Challenge level of the task also significantly affected participants’ performance, which declined with increased difficulty.

⁴ <https://scikit-learn.org/>.

Table 1 Mean accuracy, percent correct responses (SD) as a function of *Training group* and *Robot type* over each Challenge level of the memory task

Robot type	Training group	Level 1	Level 2	Level 3	Total
None	DOT	83.08 (10.29)	67.46 (16.11)	54.00 (15.15)	68.18 (18.48)
None	Non-DOT	78.62 (14.21)	62.77 (19.55)	52.85 (22.56)	64.74 (21.47)
Simulated	DOT	88.39 (8.64)	73.77 (10.46)	65.39 (12.61)	75.85 (14.18)
Simulated	Non-DOT	72.85 (16.01)	61.46 (17.88)	48.08 (19.51)	60.80 (20.18)
Physical	DOT	85.00 (8.95)	70.46 (10.49)	63.23 (14.23)	72.90 (14.43)
Physical	Non-DOT	80.92 (10.44)	69.54 (12.47)	58.08 (13.90)	69.51 (15.29)

Fig. 5 Mean accuracy for the gamified task (percent of correct responses) and standard error for each condition over the three 8-trial blocks (challenge levels) where 1 was the lowest level of challenge/difficulty and 3 the highest level



3.2 Self Assessment Manikin-Scale

The following section presents the results of the self-reported affective ratings of the SAM-scale (dimensions of valence, arousal and dominance). Visualisations of the results can be seen in Figs. 6, 7 and 8.

3.2.1 Valence

Affective ratings for *Valence* were subjected to a 3 x 2 x 3 mixed ANOVA with *Robot type* (none, simulated and physical) and *Training group* (DOT and Non-DOT) as the between-subject factors and *Challenge level* (1, 2 and 3) as the within-subject factor (see Fig. 6). Results showed significant main effects of *Training group* and *Challenge level*, and a significant *Robot type* by *Training group* by *Challenge level* interaction: $F(1, 71) = 5.73, p = .019, \eta^2 = .075$, $F(2, 142) = 22.76, p < .001, \eta^2 = .243$, and $F(4, 142) = 3.55, p = .009, \eta^2 = .091$, respectively. That is, participants in the DOT group (3.58) reported more positive affect than in the Non-DOT group (3.19). LSD post-hoc comparisons for the *Challenge level* factor also showed more positive affect for Level 1 (3.82) relative to both Level 2 (3.26) and 3 (3.07),

$ps < .001$, which did not significantly differ. To analyse the interaction we conducted separate 3 (*Robot type*) x 3 (*Challenge level*) mixed ANOVAs for each *Training group*. In the Non-DOT group, there was only a significant main effect of *Challenge level*, $F(2, 72) = 8.97, p < .001, \eta^2 = .199$; whereas in the DOT group the main effect of *Challenge level* and the *Robot type* by level interaction reached statistical significance, $F(2, 70) = 18.09, p < .001, \eta^2 = .341$ and $F(4, 70) = 3.76, p < .001, \eta^2 = .177$, respectively. To analyse the interaction in the DOT group, we further conducted separate repeated measures ANOVAs for each *Robot type* group with *Challenge level* as the within subject factor. The interaction was due to significant main effects of *Challenge level* in both the no robot and the physical robot groups, $F(2, 24) = 19.20, p < .001, \eta^2 = .615$, and $F(2, 24) = 6.28, p = .006, \eta^2 = .344$; whereas the main effect of *Challenge level* did not reach statistical significance in the simulated robot group, $F(2, 22) = 2.41, p = .113, \eta^2 = .179$. That is, while participants in the no robot and the physical robot groups with DOT reported more positive affect for Level 1 relative to both Level 2 and 3, in the simulated robot group with DOT, affective ratings were not influenced by *Challenge level*. In summary, participants in the DOT-condition reported significantly higher

valence than those in Non-DOT condition. Generally, participants reported higher valence during the first challenge level as compared to the other levels.

3.2.2 Arousal

Affective ratings for *Arousal* were subjected to a 3 x 2 x 3 mixed ANOVA (see Fig. 7). Results showed significant main effects of *Robot type* and *Challenge level*, $F(2, 71) = 3.01, p = .056, \eta^2 = .078$, and $F(2, 142) = 8.37, p < .001, \eta^2 = .106$, respectively. The interactions *Robot type* by *Challenge level*, and *Robot type* by *Training group* by *Challenge level* were also statistically significant, $F(4, 142) = 4.24, p = .003, \eta^2 = .107$, and $F(4, 142) = 2.79, p = .029, \eta^2 = .073$. LSD post hoc comparisons for the main effects showed that participants reported lower *Arousal* ratings in the physical robot group (2.72) relative to both the no robot (3.17, $p = .044$) and the simulated robot (3.20, $p = .032$) groups, which did not differ from each other.

Arousal ratings were also lower in the Level 1 condition (2.77) relative to both the Level 2 (3.13, $p = .001$) and Level 3 (3.18, $p = .002$) conditions, which did not differ from each other. To analyse the 2-way interaction, we conducted separate one-way ANOVAs for each Level condition. Results showed a significant main effect of *Robot type* only for the Level 1, $F(2, 74) = 10.12, p < .001, \eta^2 = .215$. LSD post-hoc comparisons showed that *Arousal* ratings were significantly lower in the physical robot group (2.16) relative to both the no robot (3.00, $p = .001$) and the simulated robot (3.20, $p < .001$) groups, which did not differ from each other. To analyse the 3-way interaction, we conducted separate 3 (*Robot type*) x 3 (*Challenge level*) ANOVAs for each *Training group* condition. While in the DOT group only the main effect of *Challenge level* reached significance, in the Non-DOT the interaction *Robot type* by *Challenge level* was also significant, $F(4, 72) = 6.75, p < .001, \eta^2 = .273$. To further analyse this interaction we conducted separate repeated measures ANOVA for each *Robot type* group: the main effect of *Challenge level* was only significant for the physical robot group, $F(2, 24) = 33.78, p < .001, \eta^2 = .738$. LSD post-hoc comparisons showed that *Arousal* ratings were significantly lower in Level 1 (1.77) relative to both Level 2 (2.92, $p < .001$) and Level 3 (3.08, $p < .001$), which did not differ from each other. In sum, participants in the physical robot group reported lower *Arousal* than the other groups during the first level of the task, and a significant interaction effect between *Challenge level* and *Robot type* was found in the Non-DOT condition.

3.2.3 Dominance

Finally, affective ratings for *Dominance* were submitted to a 3 x 2 x 3 mixed ANOVA (see Fig. 8). Results showed a

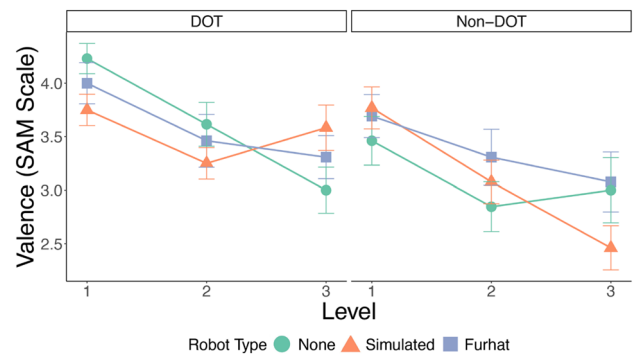


Fig. 6 Mean (with standard error bars) scores of the self assessment manikin scale for *Valence* (happy/unhappy)

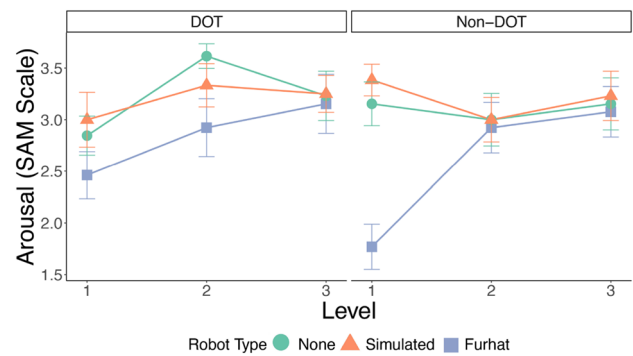


Fig. 7 Mean (with standard error bars) scores of the self assessment manikin scale for *Arousal* (stressed/calm)

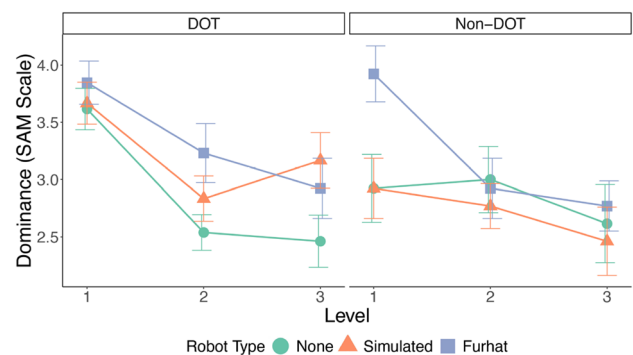


Fig. 8 Mean (with standard error bars) scores of the self assessment manikin scale for *Dominance* (being in control/not in control)

significant main effect of *Challenge level*, $F(2, 142) = 21.20, p < .001, \eta^2 = .230$. LSD post-hoc comparisons showed that *Dominance* ratings were significantly higher in Level 1

Condition (3.84) relative to both the Level 2 (2.88, $p < .001$) and Level 3 (2.72, $p < .001$) conditions, which did not differ from each other. There was a non-significant tendency for a *Robot type* by *Training group* by *Challenge level* interaction, $F(4, 142) = 2.07, p = .088, \eta^2 = .011$. Thus, we further conducted separate 3 x 2 ANOVAs for each *Challenge level* condition; while none of the effects or their interaction reached statistical significance in Levels 2 and 3, the

main effects of *Robot type* and *Training group* were significant for Level 1, $F(2, 71) = 3.26$, $p = .044$, $\eta^2 = .084$ and $F(1, 71) = 4.12$, $p = .046$, $\eta^2 = .055$. That is, participants in the DOT group reported higher *Dominance* ratings (3.71) than those in the Non-DOT group (3.26). LSD post-hoc comparisons also showed that participants in the physical robot group reported higher *Dominance* ratings (3.86) than those in the no robot (3.27, $p = .027$) and the simulated robot (3.29, $p = .035$) groups. In sum, dominance ratings were higher during the first level of the game as well as for participants in the physical robot group compared to no robot and simulated robot. They were also higher for participants in DOT condition compared to Non-DOT.

3.2.4 Correlation with Performance

To test the hypothesis that overall performance would be correlated with the affective ratings, we collapsed the data across *Challenge level* conditions and conducted Pearson bivariate correlations between the overall accuracy and the overall affective ratings for *Valence*, *Arousal* and *Dominance* in the total sample. The results showed significant positive correlations between overall accuracy and overall *Valence*, $r(77) = .265$, $p = .02$, and *Dominance*, $r(77) = .446$, $p < .001$, ratings. That is, more positive affect and higher perceived *Dominance* were associated with better overall performance. In addition, we found a significant negative correlation between overall *Dominance* and *Arousal* ratings, $r(77) = -.360$, $p = .001$. That is, higher perceived *Dominance* was associated with lower levels of *Arousal*. In summary, a significant correlation was found between higher reported *Valence* (happiness) and *Dominance* (being in control) and overall performance on the visuospatial memory training task. Furthermore, reporting lower *Arousal* (stress) was negatively correlated with higher *Dominance* (being in control).

3.3 Eye Movement Strategies

Eye tracking data were obtained from the iMotions software over the entire experimental procedure. The initial motivation to analyse the data was to ensure participants were not distracted during the experiment and to explore whether they were focusing on the score bar, cells and other gamified task elements. Interestingly, visual inspection of the eye-tracking data revealed that participants may be using different eye movement strategies during the encoding of the locations: some participants appeared to fixate in the centre whereas other appeared to follow the location sequence. Since different eye movements strategies (e.g., maintaining fixation) while encoding visuospatial information have been associated with memory performance [20] and could also prove insightful in relation to robot social engagement and interac-

tion, we decided to conduct an ad hoc exploratory evaluation of eye movements strategies.

We extracted gaze saccade amplitude (rapid eye movement between one focal point and another) and number of gaze saccades during the memory encoding phase and defined two eye movement strategies: *Fixators* (FS) - fixating gaze in the middle part of the grid; *Saccaders* (SS) - following the sequence of stimuli with the gaze. They were quantified by: SS = higher value of saccade amplitude and number of saccades, and FS = lower value of saccade amplitude and number of saccades (as determined by thresholds specified by analysing eye movement strategies and their respective typical number of saccades and amplitude lengths). After producing a Pandas data profiling report⁵ of the two dimensions it could be established that there was no linear correlation between the two dimensions (Pearson's r approximately = 0) and they would be likely to help delineate the two strategies. The eye movement strategies were explored in two different ways: a case study of two participants who were first identified through visual inspection as representative of each eye movement strategy defined above, and who also showed a noteworthy difference in performance; and a population based study. Figure 9 shows the results of the case study analysis and the two participants using either FS or SS. Each data point represents one trial of the first level of the task (24 trials in total) and there is a clear division between the FS and the SS.

To test whether the case study could be generalised to the whole sample, we conducted a cluster analysis on the full population sample. Two variables were entered into the analysis: i) Mean saccade amplitude, and ii) number of saccades. To distinguish the participants belonging to either FS or SS groups the clustering was based on mean saccade values above a set threshold to be identified as either FS or SS, thus enabling further analysis of particular eye movement strategies and their impact on memory performance and robot engagement and interaction. The thresholds were determined by extreme values, FS as having a number of saccades below 10 and a mean saccade amplitude of less than 5, and SS as having a number of saccades above 15 and a mean saccade amplitude of more than 6. They were identified by having observably different eye movement strategies and typical number of gaze saccades and amplitude lengths. The analysis was limited to include only the first level of the task and the data was split into 4 blocks of 6 trials each. We conducted k-means cluster analysis ($k=2$) to minimize within-group variance and maximize between-group variance. Figure 10 shows how the clusters are formed throughout the level (corresponding silhouette scores for k-means are 0.56, 0.64, 0.65 and 0.69) resulting in two observably separable clusters at the end of the level: one consisting of 12

⁵ <https://pypi.org/project/pandas-profiling/>.

Fig. 9 Case study of 2 participants' eye movement strategies for each trial over the first level of the task (values of saccade amplitude and number of saccades for 24 trials). One participant is maintaining fixation at the center during stimulus presentation (FS) and the other is shifting fixation to follow the stimuli sequence

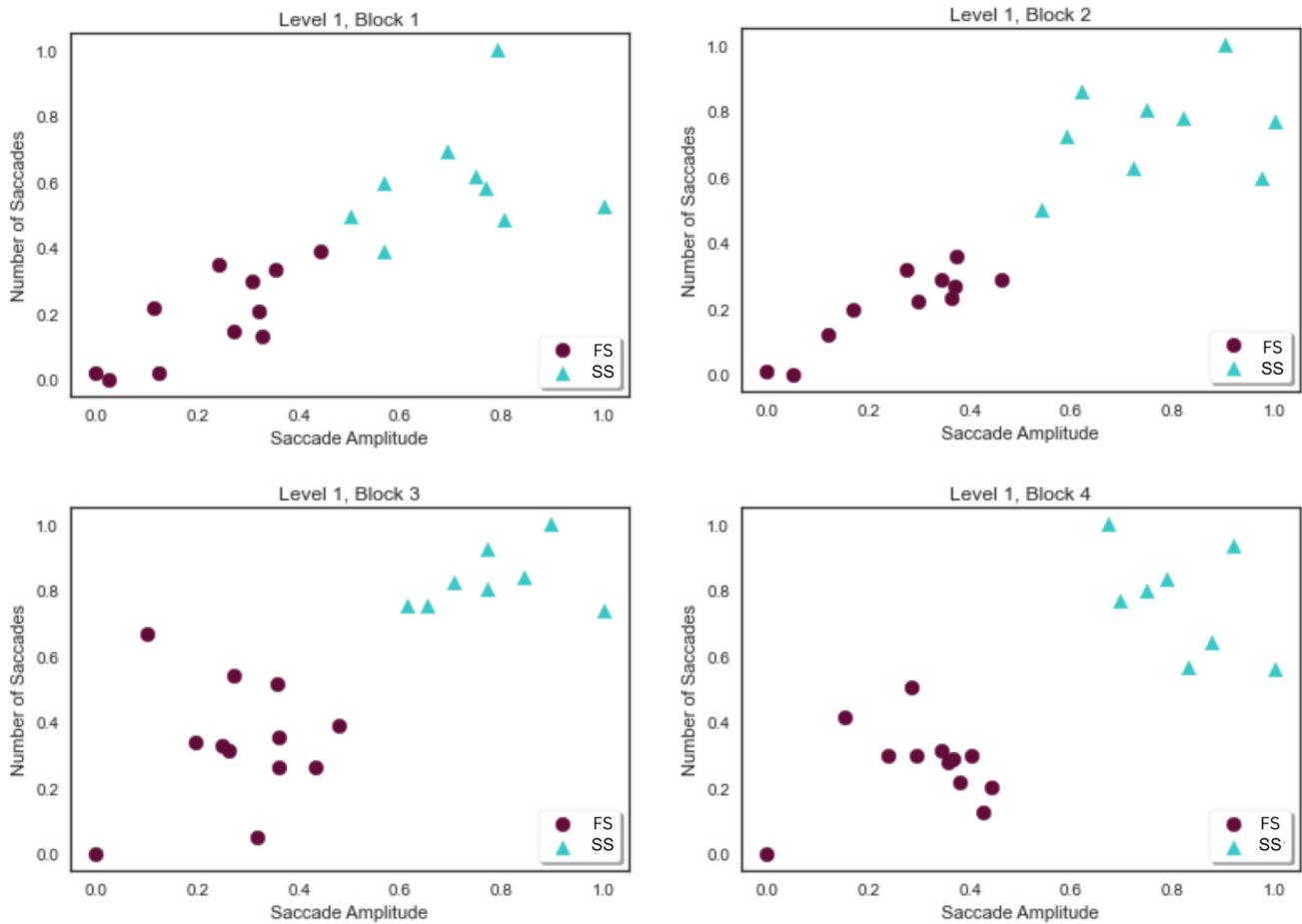
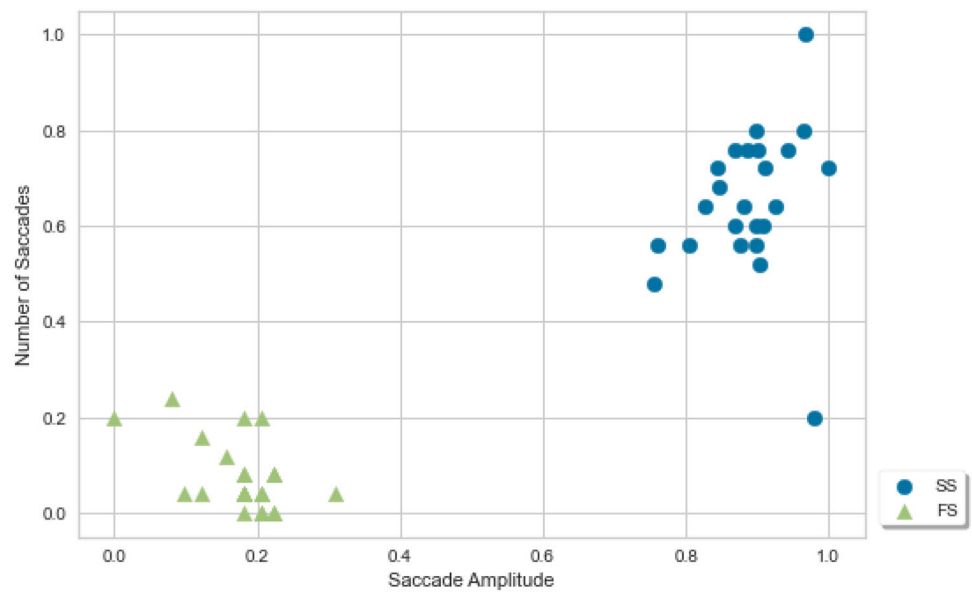


Fig. 10 Eye movement strategies for 4 blocks of 6 trials during Level 1 of the gamified task. Participants are maintaining fixation at the center (FS) and shifting fixation (SS) to follow the stimuli

participants and a second of 8 participants. We then conducted a t-test to compare the two strategy groups (FS and SS) with respect to overall memory performance (task accu-

racy). The results showed a near significant tendency for a better memory performance in the FS (73% accuracy) than in the SS group (62% correct responses), $t(18) = 1.94, p = .069$.

4 Discussion

The aim of this study was to adapt a gamified version of the visuospatial working memory task of Vivas et al. [18] and test its effectiveness within an experiment using a robotic setup, and to explore the effects of SAR feedback as an engaging element. We hypothesised that the memory training performance accuracy would be higher under DOT relative to Non-DOT conditions over all the different robotics conditions. We also hypothesised that participants' affective experience would differ as a function of setup and correlate with overall memory performance. The results showed that the Differential Outcomes Training (DOT) was effective in improving overall working memory performance in all the setups including the physical social assistive robot (SAR) and the simulated one. That is, overall memory performance was significantly better in the differential outcome group relative to the non-differential outcomes group and not contingent upon *Robot type* or *Challenge Level*. This result is consistent with that of Vivas et al. [18] (response accuracy in DOT participants being significantly greater than that found in non-DOT participants) where in our study healthy, younger adults were the participants. This finding constitutes a first step for incorporating a SAR in a non-disruptive way into a cognitive training procedure rooted in DOT for interventions with persons with cognitive impairments.

4.1 Affective Ratings

With regard to affective responses, there were significant interactions found between the *Robot type*, *Training group* and *Challenge level* conditions. Specifically, with *Valence* scores, there was a significant three-way interaction: in the differential outcomes condition, participants in the no robot (control) and physical robot condition reported more positive affect in the easiest level relative to the other two challenge levels, while for the simulated robot group, *Valence* was not affected by challenge level. For *Arousal* scores, participants, overall, reported lower *Arousal* in the physical robot group as compared to both no robot and simulated robot groups, as well as for level 1 compared to level 2 and 3. These effects were further modulated by a significant three-way interaction. That is, participants reported significantly lower *Arousal* for the easiest level only in the physical robot condition and under non-differential outcomes condition. Finally, with regard to *Dominance* scores (in control/not in control), there were significant main effects of *Robot type* and *Training group* but only for the easiest level of difficulty. Specifically, participants reported feeling more in control (higher *Dominance*) in the physical robot group, relative to both the control and simulated robot groups, and in the differential outcome group relative to the non-differential outcome group.

That significant findings applied only to the first challenge level is noteworthy since this challenge level was always presented first to the participants. On this basis, we cannot know whether an increase in 'positive' affective states in robot conditions was contingent upon the challenge not being strong or whether a novelty effect obtained. In the case of the latter effect, it would be important for future studies to allow for a range of audiovisual interactions to occur and with respect to different stages in the task, e.g. not just at the end of trials, but during trials, and at the end of blocks of trials. Catering for the *when* and the *how* of the SAR interaction is the subject of an ongoing investigation. The rationale behind using the SAM-scale was mainly related to its intuitive and visual nature making it suitable to measure affective states in complex experimental settings like the one employed in this study. The SAM-scale is a widely used measure of subjective affective states across dimensions of valence, arousal and dominance [67], which are highly relevant to measure participants' reactions towards the task, setup and robot. It has previously been used in similar gamified contexts [68] and its pictorial nature makes it accessible and intuitive to use across different social and cultural contexts [67]. An additional factor is that, considering the ultimate goal is to examine these approaches in older individuals with dementia or MCI, having methods that are straightforward to compare and consistent over time is advantageous, particularly when other measures might be impractical or challenging to duplicate.

4.2 Eye Movement Strategies

The findings from the exploratory cluster analyses with eye movements also support the existence of two distinct eye movement strategies during encoding: Maintaining fixation during the stimuli presentation at encoding (Fixation Strategy, FS) versus shifting fixation (Saccading Strategy, SS) to follow the sequence of stimuli. Furthermore, the FS was associated with better memory performance, although this tendency was not statistically significant. Our findings suggest that participants may adopt two distinct eye movement strategies: SS where participants shift fixation to follow the sequence of locations during encoding and FS where participants maintain fixation at the centre during the stimuli presentation. There is also some preliminary (close to statistical significance) evidence to support that an FS may be related with better memory performance. This finding is in agreement with previous research investigating eye movements in a similar visuospatial neuropsychological task, namely the Corsi block test (e.g., see [20] for a review). It has been suggested that maintaining fixation may be a more adaptive strategy that facilitates chunking of the location sequence, while shifting fixation may on the other hand disrupt the retinotopic representation of the locations in working memory. Thus, future studies should investigate in

a systematic manner the relationship between eye movement strategies and effectiveness of the visuospatial memory training. In our study, we could not explore the eye movement strategies as a function of robot setup due to the small sample size in each cluster. Future studies should investigate how the interaction with SARs at different phases of the training task may further affect eye movement strategies (e.g., by capturing attention), and potentially the effectiveness of the intervention. Given that the eye-tracking analyses were secondary to the main two objectives of this study, we focused only on the first challenge level in our experiment, which also yielded most significant results with affective analyses. This decision was also driven by the challenges posed by the manual annotation process in processing the eye-tracking data. Furthermore, the results should be interpreted with caution due to the relatively small sample size of each cluster. An in-depth investigation of how eye movement strategies may have interacted and changed as a function of difficulty level was beyond the scope of this study. Future studies can be aimed at investigating how eye movement strategies during encoding of stimuli may influence the effectiveness of the DOT over different challenge levels.

4.3 Cognitive Interventions with Socially Assistive Robots

The key finding of this study is that a gamified version of an existing memory cognitive training task was effective in improving memory when used in combination with both a physical and simulated SAR. The focus at this stage of our investigation was on *when* the robot interacted, which was at the end of trials providing audiovisual feedback consistent with the differential or non-differential outcomes displayed on screen. This mode of interaction proved non-disruptive; notwithstanding, the two robot setups appeared to modulate the affective responses of the participants as a function of difficulty, so that more *Dominance* and less *Arousal* were reported in the easiest level with the presence of the physical robot. This combined result – non-disruptive performance effect of SAR and positive affective states in the presence of SAR – is promising in relation to utilising a SAR in longer-term training interventions. Affective scores were also correlated with overall performance, so that better memory performance was associated with higher positive affect (*Valence*) and feeling more in control (*Dominance*). However, our study cannot establish causality in this association, and so more research is needed to understand how the use of physical and simulated SAR may improve the effectiveness of gamified cognitive training via modulation of affective experience, particularly in relation to difficulty level.

It is not a given that use of humanoid robots as assistants in longer-term interventions is beneficial for adherence or for the intervention, per se. In a recent study by [69], it was

reported that elderly citizens mostly did not view the Furhat robot with which they interacted positively. The participants reported that a body-less robot head could be scary and even harmful for people with dementia. However, it was also noted that the participants: i) “were engaged in the interaction”, ii) were mostly not positive towards any robots (not just robot heads) prior to the study, iii) only engaged with the robot in a one-off 10 min conversation. The view of a robot head as scary, or uncanny, might be mitigated by more exposure. For example, perceived Anthropomorphism and feelings of uncanniness were found to improve with repeated exposure to, and familiarity with, a robot [70, 71]. Moreover, in the context of cognitive intervention, Furhat’s capacity to be customisable across numerous modalities (including its face, voice, (micro-)behaviours or gestures) may mitigate feelings of uncanniness. Embodied interaction contexts (e.g. physical spatial proximity, or complex algorithms for (personalised) interaction) can be leveraged as part of participatory action research strategies (i.e. co-design of robot or intervention contexts) with end-users such as individuals interacting with the robot as part of cognitive intervention therapy. Taking this approach, we argue, will provide repeated exposure to, and opportunities to become familiar with, the robot as well as providing users with a sense of empowerment [72]. In turn, this may facilitate positive long-term affective attitudes towards the presence of Furhat as part of cognitive interventions. Finally, in order to avert participant confusion in interacting with a robot head, we consider that SARs would be most applicable for participants that have Mild Cognitive Impairment (MCI) as opposed to full dementia (particularly beyond early stage).

Another important aspect to be further addressed in future studies is how the complex nature of multimodal and multi-sensory social communication in humans — including facial, vocal, bodily, haptic, and spatial [73] modalities — may form unconscious expectations of social interactions in human–robot interaction scenarios. This multimodal approach to social interaction must also be considered in the context of interactions with, and by, SARs. Though this study has only leveraged a single dimension of (non-adaptive) social interaction (voice), the humanoid SAR employed in this study (Furhat) is endowed with further interaction modalities that can be exploited, including human-like facial expressions or gestures, bodily (head) movements, and voice types. If we are also to consider proximity between social agents as a modality of social interaction (i.e. one that can convey social or affective context [74]), the physical location of the SAR (with respect to both distance from the participant, and its position relative to the task or user) can also be manipulated as part of this set up (see Fig. 1). In sum, the potential for exploring multiple interaction modalities, in combination with the limitations from pet-like robots, justifies our approach of using a humanoid robot in this context.

In summary, the present study should be viewed as the first, necessary step for evaluating the use of SARs in the context of a cognitive intervention (focused on visuospatial memory training). This work constitutes a critical foundation for future development of our SAR cognitive intervention research; specifically, given that adapting or altering behaviour of a SAR is a crucial feature for SARs in long-term applications [28], the task-related feedback given by the SAR (through voice, and/or facial expressions) can instead be adaptive, for example, by personalising feedback with respect to a user's emotional (affective) status, which may improve users' affective responses to the SAR, as [75] has recently shown. Future work intends to investigate the effects of this type of affect-based adaptive interaction on task-related performance and affective responses. Our findings show that a DOE was obtained in the presence of the SAR and the affective responses of being more in control and (i.e. higher *Dominance* on the SAM scale) as well as being more calm (i.e. lower *Arousal* values on the SAM scale). Due to the way the SAM scale was presented in our study (Appendix D), the latter finding may also be interpreted as lower levels of stress. This may have implications for how SARs can be used to help improve and sustain engagement to keep patients from discontinuing longer-term cognitive interventions spanning over several weeks or months. Intervention adherence is a noteworthy problem for clinical interventions (as discussed in Sect. 1), and digitising the delivery of cognitive interventions could thus be a step towards mitigating and approaching this issue. Loss of engagement for long-term interventions could potentially be prevented by including elements such as gamification and SARs providing feedback. However, this requires that the effects of the cognitive interventions themselves are not compromised by the presence of a robot.

4.4 Gamifying the Cognitive Task

In this study, we utilised a number of basic gamification elements (score bar, attention-grabbing flashing scores, audiovisual feedback, grid-based task layout) accompanied by an artificial avatar (physical, or simulated, SAR), with certain similarities to game-like avatars. Considering the significant DOT-effect observed on our task (which gamified an existing visuospatial memory training task), one can reasonably infer that the process of gamifying DOT does not appear to undermine the task's inherent effectiveness. In other words, the gamification elements included in the task did not compromise the DOT-effect or learning for participants in this study. It would be of high interest for future studies of the task to investigate the effectiveness also for the target population, that is how effective gamification is for seniors with, or without MCI or dementia. The gamified setup presented in this study is application-agnostic, and the benefits of using gamification as well as robot feedback for cognitive

digital interventions could prove useful for a range of cognitive deficits and learning purposes, ranging from children to elderly. It would also be of interest to investigate each of the gamification elements separately, to establish their roles in enhancing learning and engagement in the task, respectively. However, the main focus of the present study was on the use of a SAR in a gamified context, and whether, through minimal but task-relevant interaction, it can be used in a non-disruptive way with a gamified task as well as to increase engagement.

The SAR (Furhat), rather than having a task-neutral interactive role, had the task-relevant role of providing outcome-consistent feedback in every trial of the experiment. However, it could be argued that its role was redundant (audiovisual outcome-specific feedback was provided in each condition). The Furhat feedback was intended to provide additional clarification and assurance to the participants and the SAM scale results indicated that those conditions with the robot provided that (e.g. greater calm/lowered arousal). An alternative role of Furhat could be to provide audiovisual feedback *in place of* of the screen-based audiovisual feedback. However, Furhat is intended to serve as an assistant or companion to the game rather than be an integral part of it and so this approach was deemed inconsistent with our, at least initial, aims. More gamification could also have been deployed in this task. However, the selected gamified elements were intended to be simple so as to maintain a reasonable mapping of the gamified task to the original task of Vivas et al. [18] (see Table 2, Appendix A). In ongoing research, in collaboration with research partners, the task is being gamified to a much larger degree and in relation to a themed cognitive intervention to be trialed on persons with Mild Cognitive Impairment. Thereby the research presented in this article provides an interface between the work of Vivas et al. [18] and the full gamified cognitive intervention intended for clinical application, which permits us to test different scenarios relatively quickly mapping the scientific method onto the intervention possibilities.

5 Conclusion

In conclusion, we have demonstrated for the first time how DOT may be applied in gamified contexts including that with minimal but task-meaningful interactions with a SAR. The results are promising in relation to the potential for use of SARs in longer-term cognitive interventions for persons with a range of different cognitive impairments. Maintaining engagement and treatment adherence for cognitive interventions, in particular for PwD and MCI, has previously constituted a significant problem. The results in this study indicate that SARs could lead to more positive affective experiences of the cognitive intervention task, while not compromising the effects of the task. Insofar as this increased positivity may

lead to increased engagement and adherence to longer-term cognitive training such social interaction of SARs may be valuable. Future studies will initially focus on MCI as they are functionally (cognitively) closer to the group studied here and might respond best to interventions with SARs on this basis.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12369-023-01083-0>.

Acknowledgements The authors would like to thank Brain+ (Denmark), Demensförbundet (Sweden) and Swedish Foundation for International Cooperation in Research and Higher Education (REF: IB2021-8964) for helping to finance work reported in this paper.

Author Contributions The first two authors contributed equally to this work. AM: Methodology, Software, Formal analysis, Validation, Visualization, Data curation, Writing - Original Draft, Review and Editing, Project administration. SS: Methodology, Investigation, Resources, Writing - Original Draft, Visualization. MB: Software, Formal analysis, Data Curation, Investigation. BSR: Conceptualisation, Methodology, Software, Validation, Visualization. IK: Writing - Original Draft, Review and Editing, Visualization, Validation. AV: Conceptualisation, Formal analysis, Writing - Original Draft, Review and Editing. RL: Conceptualisation, Methodology, Writing - Original Draft, Review and Editing, Supervision, Funding acquisition.

Funding This research was partly financed by Brain+ (ACTTDCS – "Attentional Control Training and Transcranial Direct Current Stimulation to Treat Dementia", REF: E115073), Demensförbundet and the Swedish Foundation for International Cooperation in Research and Higher Education (STINT, REF: IB2021-8964). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Therefore, the authors of this study declare no conflicts of interest.

Data Availability The datasets and code generated during and analysed during the current study are available in the Open Science Framework (OSF) repository, identifier: DOI: 10.17605/OSF.IO/86VWJ ([url: https://osf.io/86vwj/](https://osf.io/86vwj/)).

Declarations

Conflict of interest All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Consent to Participate All volunteering participants whose data were used in the work presented in this paper provided informed written consent to participate in the study. The study was conducted in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki declaration and its later amendments.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix A: Task Comparison

See Table 2 .

Table 2 Comparison of task-based features found in Vivas et al. [18] visuospatial memory task and the task used in this study

Experimental Feature	Vivas et al. [18]	The current study
Subjects	AD/MCI/HC (age controlled)	HC (19–46 y)
Grid Configuration	3 × 3 (8 cells active)	5 × 5 (24 cells active)
Number of Trials	48	72
Block Size	16 Trials	8 Trials
Challenge Levels	1	3
Number of Sample Stimuli	4	4, 6, 8
Number of Response Options	2	2, 3, 4
Use of Distractor Task Stimuli	No	Yes
Inter Stimulus Interval (ISI)	2 s or 15 s (random alternating per trial)	4 s or 12 s (random alternating per trial)
Inter Trial Interval	Approx. 0ms	Approx. 0ms
Response Option Timeout	Unlimited	5 s
Stimulus Presentation Duration	750 ms	750 ms
Stimulus Duration Gap	0 ms	0 ms (but with distractor cells)
Differential Outcomes Types	1 per stimulus, image of landscape	1 per stimulus, 1 or 3 coins, 1 or 3 diamonds
Reward Types	Real world reward, via Raffle	Score bar linked to cinema ticket winnings
Outcome Presentation Duration	3 s	3 s
Outcome Presentation Location	Independent of sample stimulus presentation	Independent of sample stimulus presentation

Appendix B: Visualisations of the Gamified Task

See Figs. 11, 12, 13, 14 and 15.

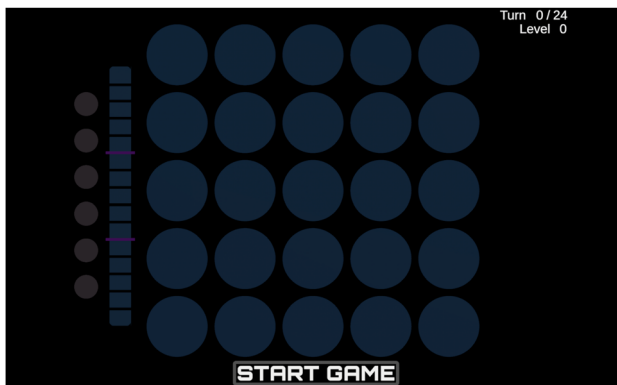


Fig. 11 Starting view of the task grid. The grid consisted of 5×5 cells where stimuli were presented. The score bar to the left would be filled up during the task if participants correctly chose the response (target) stimulus. The information in the upper right corner indicated number of trials taken and challenge level

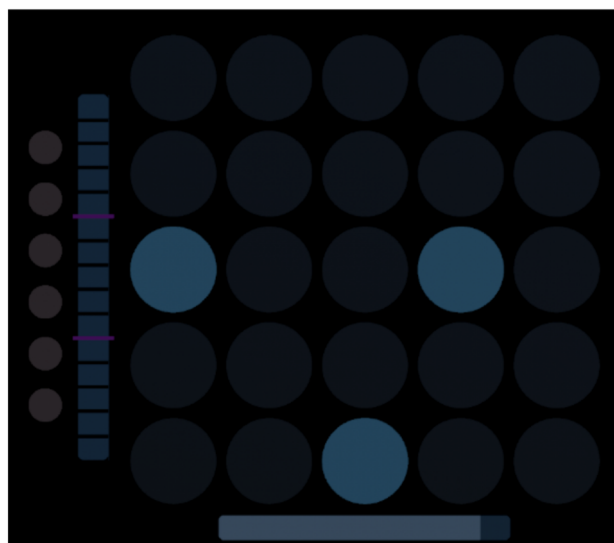


Fig. 13 Visualisation of response options, including time bar below the grid indicating the response time limit for participants

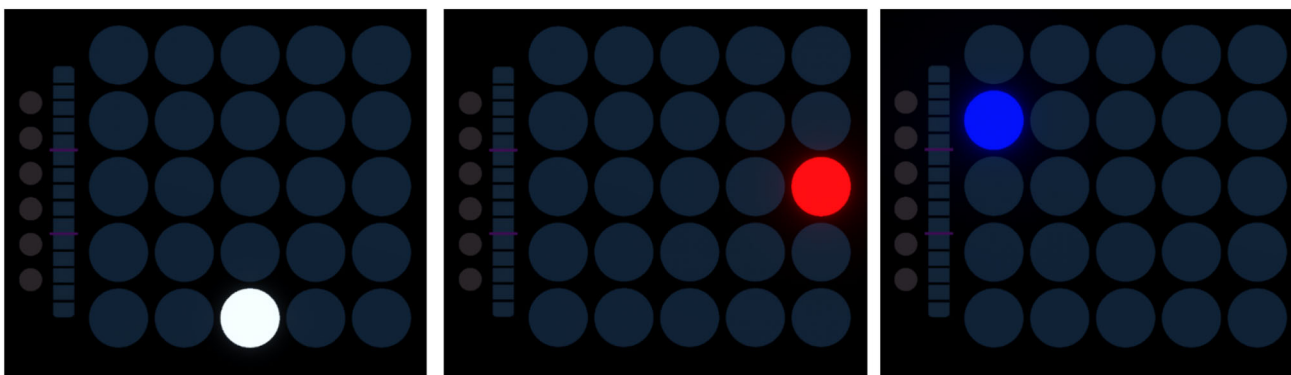


Fig. 12 Visualisation of stimuli, including distractors. White stimuli are part of the sequence to be memorised. Red distractors are ignored while participants have to press the keyboard spacebar immediately following the presentation of blue cells (active distractors). (Color figure online)

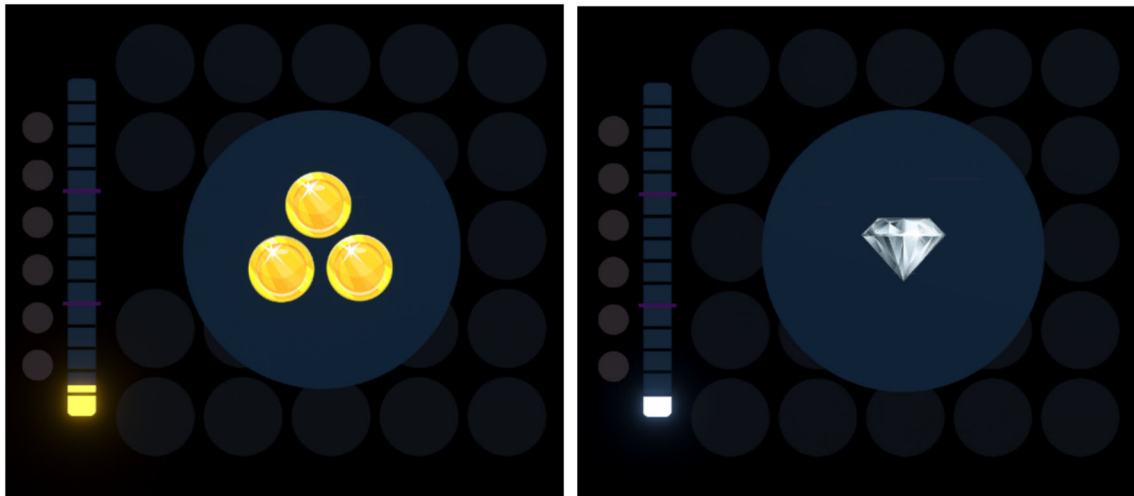


Fig. 14 Examples of reward outcomes displayed on the computer screen when getting a correct answer

Fig. 15 The different rewards that were possible to receive. The rewards were differential both regarding type (diamond or coin) as well as magnitude (3 or 1)



Appendix C: Feedback Script for Robot

The following outcomes were vocalised by the physical or the simulated Furhat depending on the outcome the participant received, if they had an incorrect answer, or if the participant did not give an answer in time. The different outcomes were randomised in regard to the respective outcomes that the participant received.

Outcome 1 - One coin:

- “Yes, one coin.”
- “Huh, you got a coin.”
- “Hmm, good work.”

Outcome 2 - Three coins:

- “Wow, excellent! Three coins.”
- “Whoa, you got three coins!”
- “Oh yes, jack-pot!”

Outcome 3 - One diamond:

- “Aha, one diamond!”
- “Okay! You got a diamond.”
- “Yes, well done.”

Outcome 4 - Three diamonds:

- “Wow, you got three diamonds!”
- “Outstanding! So many diamonds.”
- “Yess! Brilliant!”

Outcome 5 - Incorrect answer:

- “Incorrect, keep fighting!”
- “Try to stay focused!”
- “Noo, try again!”
- “Come on, you can do it!”

Outcome 6 - Time out:

- “Uh, time out!”
- “Time out, be faster!”
- “Oh no, pick up speed.”
- “Huh, catch up.”

Appendix D: SAM-Scale Visualisation

See Fig. 16.

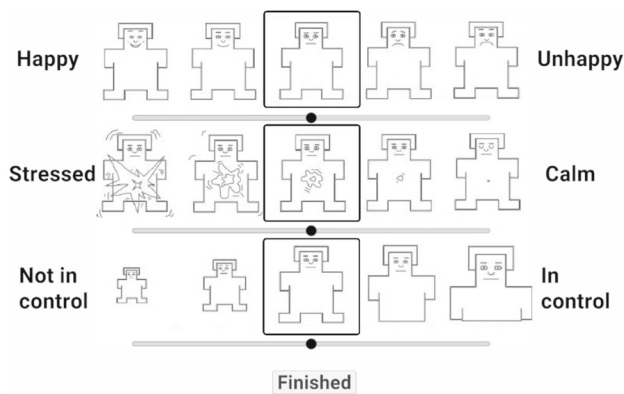


Fig. 16 SAM-scale as seen by the participants on the screen after each challenge level of the task

References

- World Health Organization (2022) Ageing and health. <https://www.who.int/news-room/fact-sheets/detail/ageing-and-health>. Accessed on 02 May 2022
- Hampel H, Lista S (2016) The rising global tide of cognitive impairment. *Nat Rev Neurol* 12(3):131–132. <https://doi.org/10.1038/nrneurol.2015.250>
- Alzheimer's Disease International (2020) Dementia statistics. <https://www.alzint.org/about/dementia-facts-figures/dementia-statistics>. Accessed on 05 Feb 2022
- Pedroza P et al (2022) Global and regional spending on dementia care from 2000–2019 and expected future health spending scenarios from 2020–2050: an economic modelling exercise. *eClinicalMedicine*. <https://doi.org/10.1016/j.eclinm.2022.101337>
- Cuffaro L et al (2020) Dementia care and COVID-19 pandemic: a necessary digital revolution. *Neurol Sci* 41(8):1977–1979. <https://doi.org/10.1007/s10072-020-04512-4>
- Alzheimer's Society (2022) Drug treatments and medication for Alzheimer's disease. <https://www.alzheimers.org.uk/about-dementia/treatments/dementia-drugs/drug-treatments-and-medication-alzheimers-disease>. Accessed on 15 Apr 2023
- Jeong Hong Y, Hye Jang E, Hwang J, Hoon Roh J, Lee J-H (2015) The efficacy of cognitive intervention programs for mild cognitive impairment: a systematic review. *Curr Alzheimer Res* 12(6):527–542. <https://doi.org/10.2174/1567205012666150530201636>
- Torous J et al (2021) The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry* 20(3):318–335. <https://doi.org/10.1002/wps.20883>
- Sailer M, Homner L (2020) The gamification of learning: a meta-analysis. *Educ Psychol Rev* 32(1):77–112. <https://doi.org/10.1007/s10648-019-09498-w>
- Fernandez-Lazaro CI et al (2019) Adherence to treatment and related factors among patients with chronic conditions in primary care: a cross-sectional study. *BMC Family Pract*. <https://doi.org/10.1186/s12875-019-1019-3>
- Walsh CA et al (2019) The association between medication non-adherence and adverse health outcomes in ageing populations: a systematic review and meta-analysis. *Br J Clin Pharmacol* 85(11):2464–2478. <https://doi.org/10.1111/bcp.14075>
- Bosworth HB et al (2011) Medication adherence: a call for action. *Am Heart J* 162(3):412–424. <https://doi.org/10.1016/j.ahj.2011.06.007>
- Sokol MC, McGuigan KA, Verbrugge RR, Epstein RS (2005) Impact of medication adherence on hospitalization risk and health-care cost. *Med Care* 43(6):521–530. <https://doi.org/10.1097/01.mlr.0000163641.86870.af>
- Etters L, Goodall D, Harrison BE (2008) Caregiver burden among dementia patient caregivers: a review of the literature. *J Am Acad Nurse Pract* 20(8):423–428. <https://doi.org/10.1111/j.1745-7599.2008.00342.x>
- Wangmo T, Lipps M, Kressig RW, Ienca M (2019) Ethical concerns with the use of intelligent assistive technology: findings from a qualitative study with professional stakeholders. *BMC Med Ethics*. <https://doi.org/10.1186/s12910-019-0437-z>
- Belpaeme T, Kennedy J, Ramachandran A, Scassellati B, Tanaka F (2018) Social robots for education: a review. *Sci Robot*. <https://doi.org/10.1126/scirobotics.aat5954>
- Trapold MA (1970) Are expectancies based upon different positive reinforcing events discriminably different? *Learn Motiv* 1(2):129–140. [https://doi.org/10.1016/0023-9690\(70\)90079-2](https://doi.org/10.1016/0023-9690(70)90079-2)
- Vivas AB et al (2018) Enhancement of visuospatial working memory by the differential outcomes procedure in mild cognitive impairment and Alzheimer's disease. *Front Aging Neurosci* 10:364. <https://doi.org/10.3389/fnagi.2018.00364>
- Bradley MM, Lang PJ (1994) Measuring emotion: the self-assessment manikin and the semantic differential. *J Behav Therapy Exp Psychiatry* 25(1):49–59. [https://doi.org/10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9)
- Hodgson TL, Ezard G, Hermens F (2019) Eye movements in neuropsychological tasks. *Process Visuospatial Atten Work Mem*. https://doi.org/10.1007/7854_2019_98
- Kabacińska K, Prescott TJ, Robillard JM (2021) Socially assistive robots as mental health interventions for children: a scoping review. *Int J Soc Robot* 13:919–935. <https://doi.org/10.1007/s12369-020-00679-0>
- Tung F-W (2016) Child perception of humanoid robot appearance and behavior. *Int J Hum Comput Interact* 32(6):493–502. <https://doi.org/10.1080/10447318.2016.1172808>
- Rudnicka E et al (2020) The World Health Organization (WHO) approach to healthy ageing. *Maturitas* 139:6–11. <https://doi.org/10.1016/j.maturitas.2020.05.018>
- Russell G et al (2022) Time trends in autism diagnosis over 20 years: a UK population-based cohort study. *J Child Psychol Psychiatry* 63(6):674–682. <https://doi.org/10.1111/jcpp.13505>
- World Health Organization (2022) Global spending on health: rising to the pandemic's challenges. <https://www.who.int/publications/i/item/9789240064911>. Accessed on 06 Apr 2023
- Aiken LH, Sloane DM, Bruyneel L, Van den Heede K, Sermeus W (2013) Nurses' reports of working conditions and hospital quality of care in 12 countries in Europe. *Int J Nurs Stud* 50(2):143–153. <https://doi.org/10.1016/j.ijnurstu.2012.11.009>
- Nashwan AJ et al (2021) Comparing the impact of COVID-19 on nurses' turnover intentions before and during the pandemic in Qatar. *J Pers Med* 11(6):456. <https://doi.org/10.3390/jpm11060456>
- Matarić MJ, Scassellati B (2016) In: Siciliano B, Khatib O (eds) Socially assistive robotics. Springer Handbook of Robotics. Springer, Cham, pp 1973–1994. https://doi.org/10.1007/978-3-319-32552-1_73
- Liang A et al (2017) A pilot randomized trial of a companion robot for people with dementia living in the community. *J Am Med Dir Assoc* 18(10):871–878. <https://doi.org/10.1016/j.jamda.2017.05.019>
- Abdollahi H, Mollahosseini A, Lane JT, Mahoor MH (2017) A pilot study on using an intelligent life-like robot as a companion for elderly individuals with dementia and depression. *IEEE, Birmingham*, pp 541–546
- D'Onofrio G et al (2019) MARIO project: validation and evidence of service robots for older people with dementia. *J Alzheimer's Dis* 68(4):1587–1601. <https://doi.org/10.3233/JAD-181165>

32. Bemelmans R, Gelderblom GJ, Jonker P, de Witte L (2015) Effectiveness of robot *Paro* in intramural psychogeriatric care: a multicenter quasi-experimental study. *J Am Med Dir Assoc* 16(11):946–950. <https://doi.org/10.1016/j.jamda.2015.05.007>
33. Sefidgar YS et al (2016) Design and evaluation of a touch-centered calming interaction with a social robot. *IEEE Trans Affect Comput* 7(2):108–121. <https://doi.org/10.1109/TAFFC.2015.2457893>
34. Broadbent E, Stafford R, MacDonald B (2009) Acceptance of healthcare robots for the older population: review and future directions. *Int J Soc Robot* 1(4):319–330. <https://doi.org/10.1007/s12369-009-0030-6>
35. Libin A, Cohen-Mansfield J (2004) Therapeutic robot for nursing home residents with dementia: preliminary inquiry. *Am J Alzheimer's Dis Other Dement* 19(2):111–116
36. Mori M, MacDorman K, Kageki N (2012) The Uncanny valley [From the Field]. *IEEE Robot Autom Mag* 19(2):98–100. <https://doi.org/10.1109/MRA.2012.2192811>
37. Powers A, Kiesler S, Fussell S, Torrey C (2007) Comparing a computer agent with a humanoid robot. *ACM, Arlington*, pp 145–152. <https://doi.org/10.1145/1228716.1228736>
38. Natarajan M, Gombolay M (2020) Effects of anthropomorphism and accountability on trust in human robot interaction. *ACM, Cambridge*, pp 33–42. <https://doi.org/10.1145/3319502.3374839>
39. Papadopoulos I, Koulouglioti C, Lazzarino R, Ali S (2020) Enablers and barriers to the implementation of socially assistive humanoid robots in health and social care: a systematic review. *BMJ Open* 10(1):e033096. <https://doi.org/10.1136/bmjopen-2019-033096>
40. Kim B, de Visser E, Phillips E (2022) Two uncanny valleys: re-evaluating the uncanny valley across the full spectrum of real-world human-like robots. *Comput Human Behav* 135:107340. <https://doi.org/10.1016/j.chb.2022.107340>
41. MacDorman KF, Entezari SO (2015) Individual differences predict sensitivity to the uncanny valley. *Interact Stud Soc Behav Commun Biol Artif Syst* 16(2):141–172. <https://doi.org/10.1075/is.16.2.01mac>
42. de Haas M, Vogt P, Krahmer E (2020) The effects of feedback on children's engagement and learning outcomes in robot-assisted second language learning. *Front Robot AI*. <https://doi.org/10.3389/frobot.2020.00101>
43. Brown L, Kerwin R, Howard AM (2013) Applying behavioral strategies for student engagement using a robotic educational agent. In: 2013 IEEE international conference on systems, man, and cybernetics, pp 4360–4365. <https://doi.org/10.1109/SMC.2013.744>
44. Logan DE et al (2019) Social robots for hospitalized children. *Pediatrics*. <https://doi.org/10.1542/peds.2018-1511>
45. Ahmad MI, Mubin O, Shahid S, Orlando J (2019) Robot's adaptive emotional feedback sustains children's social engagement and promotes their vocabulary learning: a long-term child–robot interaction study. *Adapt Behav* 27(4):243–266. <https://doi.org/10.1177/1059712319844182>
46. Cruz-Sandoval D, Favela J (2019) Incorporating conversational strategies in a social robot to interact with people with dementia. *Dement Geriatr Cognit Disord* 47(3):140–148. <https://doi.org/10.1159/000497801>
47. Sung H-C, Chang S-M, Chin M-Y, Lee W-L (2015) Robot-assisted therapy for improving social interactions and activity participation among institutionalized older adults: a pilot study. *Asia-Pac Psychiatry* 7(1):1–6. <https://doi.org/10.1111/appy.12131>
48. Andriella A, Torras C, Alenya G (2020) Cognitive system framework for brain-training exercise based on human–robot interaction. *Cognit Comput* 12(4):793–810. <https://doi.org/10.1007/s12559-019-09696-2>
49. Chan J, Nejat G (2010) Promoting engagement in cognitively stimulating activities using an intelligent socially assistive robot. *IEEE*, pp 533–538
50. Rich C, Ponsleur B, Holroyd A, Sidner CL (2010) Recognizing engagement in human–robot interaction. In: *Proceeding of the 5th ACM/IEEE international conference on human–robot interaction—HRI'10*. <https://doi.org/10.1145/1734454.1734580>
51. Lopez CE, Tucker CS (2019) The effects of player type on performance: a gamification case study. *Comput Hum Behav* 91:333–345. <https://doi.org/10.1016/j.chb.2018.10.005>
52. Denden M, Tlili A, Essalmi F, Jemni M (2017) Educational gamification based on personality. In: 2017 IEEE/ACS 14th international conference on computer systems and applications (AICCSA), pp 1399–1405. <https://doi.org/10.1109/AICCSA.2017.87>
53. Groznik V, Sadikov A (2019) Gamification in cognitive assessment and cognitive training for mild cognitive impairment. In: *Augmented reality games II: the gamification of education, medicine and art*, pp 179–204 (2019)
54. Briffa M et al (2020) Improved students' performance within gamified learning environment: a meta-analysis study. *Int J Educ Res* 8(1):223–244
55. Gooch D, Thompson P, Nash HM, Snowling MJ, Hulme C (2016) The development of executive function and language skills in the early school years. *J Child Psychol Psychiatry* 57(2):180–187. <https://doi.org/10.1111/jcpp.12458>
56. Su C, Fan K-K, Xiao P-W (2015) The effects of learning styles and meaningful learning on the learning achievement of gamification health education curriculum. *EURASIA J Math Sci Technol Educ*. <https://doi.org/10.12973/eurasia.2015.1413a>
57. Canady VA (2020) Company moves forward with nondrug treatment for ADHD. *Mental Health Wkly* 30(26):7–8. <https://doi.org/10.1002/mhw.32426>
58. Groznik V, Sadikov A (2019) Gamification in cognitive assessment and cognitive training for mild cognitive impairment. In: *Augmented reality games II*, pp 179–204. https://doi.org/10.1007/978-3-030-15620-6_8
59. Zeng Z, et al (2018) Towards long-term tracking and detection of early dementia. In: *Proceedings of the 3rd international conference on crowd science and engineering*. <https://doi.org/10.1145/3265689.3265719>
60. Esteban L, Plaza V, López-Crespo G, Vivas AB, Estévez AF (2014) Differential outcomes training improves face recognition memory in children and in adults with down syndrome. *Res Dev Disabil* 35(6):1384–1392. <https://doi.org/10.1016/j.ridd.2014.03.031>
61. Esteban L, Vivas AB, Fuentes LJ, Estévez AF (2015) Spatial working memory is enhanced in children by differential outcomes. *Sci Rep* 5(1):17112. <https://doi.org/10.1038/srep17112>
62. Plaza V, López-Crespo G, Antúnez C, Fuentes LJ, Estévez AF (2012) Improving delayed face recognition in Alzheimer's disease by differential outcomes. *Neuropsychology* 26(4):483. <https://doi.org/10.1037/a0028485>
63. McCormack JC, Elliffe D, Virués-Ortega J (2019) Quantifying the effects of the differential outcomes procedure in humans: a systematic review and a meta-analysis. *J Appl Behav Anal* 52(3):870–892. <https://doi.org/10.1002/jaba.578>
64. Peterson GB, Trapold MA (1980) Effects of altering outcome expectancies on pigeons' delayed conditional discrimination performance. *Learn Motiv* 11(3):267–288. [https://doi.org/10.1016/0023-9690\(80\)90001-6](https://doi.org/10.1016/0023-9690(80)90001-6)
65. Campbell JI, Thompson VA (2012) Morepower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behav Res Methods* 44(4):1255–1265. <https://doi.org/10.3758/s13428-012-0186-0>
66. Rittmo J, Karlsson R, Gander P, Lowe R (2020) Vicarious value learning: knowledge transfer through affective processing on a social differential outcomes task. *Acta Psychol* 209:103134

67. Morris JD (1995) Observations: Sam: the self-assessment manikin; an efficient cross-cultural measurement of emotional response. *J Advert Res* 35(6):63–68
68. Xie T, Cao M, Pan Z (2020) Applying self-assessment manikin (SAM) to evaluate the affective arousal effects of VR games, pp 134–138
69. Thunberg S, Arnelid M, Ziemke T (2022) Older adults' perception of the Furhat robot, pp 4–12
70. Paetzel M, Castellano G (2019) Let me get to know you better: can interactions help to overcome uncanny feelings? pp 59–67
71. Paetzel M, Perugia G, Castellano G (2020) The persistence of first impressions: the effect of repeated interactions on the perception of a social robot, pp 73–82
72. Ostrowski AK, Breazeal C, Park HW (2021) Long-term co-design guidelines: empowering older adults as co-designers of social robots, pp 1165–1172
73. Stivers T, Sidnell J (2005) Introduction: multimodal interaction. *Semiotica* 2005(156):1–20. <https://doi.org/10.1515/semi.2005.2005.156.1>
74. Varga S, Krueger J (2013) Background emotions, proximity and distributed emotion regulation. *Rev Philos Psychol* 4(2):271–292. <https://doi.org/10.1007/s13164-013-0134-7>
75. Churamani N, Axelsson M, Caldir A, Gunes H (2022) Continual learning for affective robotics: a proof of concept for wellbeing. arXiv preprint [arXiv:2206.11354](https://arxiv.org/abs/2206.11354)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.