Giorgi, Ioanna, Rajapakse, Sachini, Palomino, Marco and Masala, Giovanni Luca (2025) *Older adults' perceptions of robots with differing intelligence in social multi-robot interactions*. ACM Transactions on Human-Robot Interaction, 15 (1).

Published: 18 November 2025
Online AM: 29 September 2025
Accepted: 12 September 2025
Revised: 24 June 2025
Received: 13 September 2024

RESEARCH-ARTICLE

Older Adults' Perceptions of Robots with Differing Intelligence in Social Multi-Robot Interactions

IOANNA GIORGI, University of Kent, Canterbury, Kent, U.K.

SACHINI RAJAPAKSE, University of Kent, Canterbury, Kent, U.K.

MARCO PALOMINO, University of Plymouth, Plymouth, Devon, U.K.

GIOVANNI LUCA MASALA, University of Kent, Canterbury, Kent, U.K.

Open Access Support provided by:

University of Kent

University of Plymouth

https://doi.org/10.1145/3769854
EISSN: 2573-9522

Older Adults' Perceptions of Robots with Differing Intelligence in Social Multi-Robot Interactions

IOANNA GIORGI and SACHINI RAJAPAKSE, School of Computing, University of Kent, Canterbury, United Kingdom of Great Britain and Northern Ireland
MARCO PALOMINO, School of Engineering, Computing and Mathematics, University of Plymouth, Plymouth, UK, and School of Natural and Computing Sciences, University of Aberdeen, Aberdeen, UK
GIOVANNI L. MASALA, School of Computing, University of Kent, Canterbury, United Kingdom of Great Britain and Northern Ireland

Social robots present significant potential to support ageing societies, yet their integration into older adults' lives still requires substantial investigation. Existing research has explored multiple factors that influence older adults' attitudes towards robots, but few have directly manipulated perceptions of robot intelligence. We address this gap on the premise that older adults may experience reduced confidence in managing technology, making them more sensitive to cues of robot competence, as the basis for their trust and acceptance. This study adopts a *human multi-robot* approach, situating social interactions within a group where human-robot and robot-robot interactions co-exist. Unlike research limited to dyadic interactions or that deploys multiple robots in isolation to provide complementary assistance or different points of engagement, we model group (triadic) interactions in a structured experiment where participants played a cognitive game simultaneously with two social robots of different cognitive competence. This design was intended to leverage inter-robot comparisons when forming older adults' perceptions, rendering the manipulation of perceived intelligence more salient. Using a mixed-method approach, we explored how the intelligence construct affected trust, usability and the willingness to adopt social robots, including preference for continued interaction. Our findings showed participants placed greater trust and intent to use the robot with higher constructed intelligence, though many favoured the robot that appealed to them aesthetically or displayed human-like fallibility, which some perceived as intelligence. The interaction increased participants' willingness to use robots in other proxy (sensitive, health-related) contexts, suggesting that perceptions formed in one context may transfer to others, potentially reinforcing acceptance of robots broadly with continued interactions.

CCS Concepts: • **Human-centered computing → HCI design and evaluation methods**; • **Social and professional topics → Seniors**; • **Computing methodologies → Artificial intelligence**;

The author M.P. was academically employed at the University of Plymouth when the study was conducted. G.M. was a visiting academic at the University of Plymouth, under the Interreg 2 Seas Mers Zeeën AGE'In project (2S05-014).
This research was funded by the Interreg 2 Seas Mers Zeeën AGE'In project (2S05-014) at the University of Plymouth.
Authors' Contact Information: Ioanna Giorgi (corresponding author), School of Computing, University of Kent, Canterbury, United Kingdom of Great Britain and Northern Ireland; e-mail: i.giorgi@kent.ac.uk; Sachini Rajapakse, School of Computing, University of Kent, Canterbury, United Kingdom of Great Britain and Northern Ireland; e-mail: sachinihimara@gmail.com; Marco Palomino, School of Engineering, Computing and Mathematics, University of Plymouth, Plymouth, UK, and School of Natural and Computing Sciences, University of Aberdeen, Aberdeen, UK; e-mail: marco.palomino@abdn.ac.uk; Giovanni L. Masala, School of Computing, University of Kent, Canterbury, United Kingdom of Great Britain and Northern Ireland; e-mail: g.masala@kent.ac.uk.

ACM 2573-9522/2025/11-ART25
https://doi.org/10.1145/3769854

Additional Key Words and Phrases: Robot intelligence, Human multi-robot interaction, Human-Robot Interaction (HRI), Social robots, Silver care

ACM Reference format:
Ioanna Giorgi, Sachini Rajapakse, Marco Palomino, and Giovanni L. Masala. 2025. Older Adults' Perceptions of Robots with Differing Intelligence in Social Multi-Robot Interactions. *ACM Trans. Hum.-Robot Interact.* 15, 1, Article 25 (November 2025), 28 pages.
https://doi.org/10.1145/3769854

1 Introduction

Social robots are increasingly recognised as valuable support resources in the healthcare sector, particularly for older adults. The rising elderly population and the accompanying shortage of caregivers have spurred research into the potential of robots to assist with tasks such as mobility, cognitive stimulation, household tasks and companionship, addressing issues like isolation [1], cognitive decline [2] and physical limitations [3]. Older adults were found receptive to using robots for daily assistance [3, 4] and effectively supporting their management of chronic conditions by providing reminders and monitoring health parameters [5].

The successful adoption and use of robots as intended depend heavily on cultivating positive attitudes towards them. However, studies show that the acceptance of and trust in robots is more subtle among the older population and tends to decrease with age [6]. While human-centric aspects cannot be neglected, much remains to be uncovered about the specific robot-related factors that attract older adults' acceptance and use of robots [7]. Such factors may be utilitarian (usefulness, practicality, perceived intelligence) or hedonic (user experience, with no direct relation to task-specific goals) [8]. Both are equally important. For instance, older adults tend to prefer robots that are perceived as friendly, reliable and capable [9]. Robots that are seen as intelligent and helpful are more likely to be trusted and adopted by older users [10]. However, the robot's physical design is also a strong aspect that dictates how older adults act around robots. Notably, robots designed with human-like features and behaviours are better received by older adults, who find them more relatable and engaging [11]. Additionally, the degree of interactivity the robot has with the user was also seen to significantly affect older adults' interaction with it [12]. These factors greatly shape their user experience, potentially influencing their perceptions just as much as the robot's functional value in completing tasks.

1.1 Robot Intelligence and Older Adults

Using robots in contexts beyond their designed capabilities can lead to negative perceptions and reduced acceptability [12], as people expect robots to behave and perform appropriately for the given task. One such expectation is that the robot appears intelligent. The perception of intelligence contributes to the robot being considered genuine, more likeable and more realistic [8]. The intelligence of a robot, defined by its ability to understand, learn and respond to human behaviour appropriately [13], plays a crucial role in user satisfaction, trust and willingness to use or rely on it. In fact, for older adults, robot intelligence and proactive behaviours are granted for the robot to be considered use-worthy [14].

Robots that are perceived as competent and reliable are more likely to build trust among users [15]. This trust is reinforced when robots consistently perform tasks correctly and provide accurate information, which is particularly important for older adults who may rely on these technologies in sensitive contexts [16]. Likewise, if the robot's behaviour does not align with the older adults' expectations of an intelligent agent (e.g., making mistakes in tasks or failing to understand instructions), it could lead to decreased trust and a negative perception of the robot [17]. However, older adults often have specific and contradictory expectations of what an 'intelligent' robot should be capable of, encompassing several aspects of conversational ability, task execution, emotional intelligence, and appearance, which make designing acceptable robot assistants or companions particularly challenging [6, 18].

Moreover, perceived robot intelligence can influence how enjoyable older adults find the robot, which is a strong predictor of their intention-to-use social robots [19]. For example, when robots display adaptive social behaviours, they are often seen as more intelligent and trustworthy, leading to more engaging and effective interactions [20]. This adaptability, including the ability to express emotions and understand social cues, enhances user acceptance and promotes continued use and trust [21]. Moreover, through higher levels of intelligence, among other factors (e.g., deeper anthropomorphism), robots can evoke social presence in humans [22]. The more competent participants believe the robot to be, the more they attribute human-like or living qualities to it, thus feeling as if they are interacting with an aware and responsive being rather than just a machine [23].

These studies collectively demonstrate that the perceived intelligence of robots, including both cognitive and emotional dimensions, is pivotal in shaping how older adults perceive and interact with them. For elderly users in particular, the *perception* of robot intelligence appears central to acceptance and sustained use. Older adults often experience reduced confidence in managing technology and daily tasks, which makes them more sensitive to cues that the robot understands context, adapts to their pace, and responds appropriately to situations [10, 19]. A robot that is perceived as 'merely social' but not intelligent risks being dismissed as superficial or infantilising, whereas one that demonstrates competence and context-awareness is more likely to be trusted as a genuine partner in daily life [24]. For example, [25] showed that robots responding with expected conversational flow were perceived more positively by older adults, leading to greater engagement, than those providing ambiguous responses. This work illustrates how manipulating the apparent competence of a robot can directly influence older adults' acceptance. However, there remain relatively few studies that explicitly manipulate the level of perceived intelligence in elderly–robot interaction, pointing to an important gap for future research.

1.2 Human Multi-Robot Interaction

The concept of simultaneous group interactions in multi-human-multi-robot systems is an emerging area of research, given advancements in the physical, cognitive and computational abilities of robots [26]. This research looks at how robots can coordinate among themselves to interact with one or multiple humans in a more fluid and natural manner. Different from dyadic interactions, this imposes requirements for more complex control strategies between robots that may be dissimilar in many forms, and to mediate multiple interactions occurring simultaneously and avoid or resolve potential conflicts [26]. Moreover, the interaction in such systems does not always conform to the standard definition of interaction as 'a process involving reciprocal stimulation or response between the agents' [27], in that reciprocity is not fundamental.

One way to categorise such **human-robot interaction (HRI)** systems is in terms of team structure and size: (i) single human, single robot; (ii) single human, multiple robots; (iii) multiple humans, single robot; (iv) multiple humans, multiple robots [26]. For the current study, we have focused on the single-human-multiple-robots approach. This team includes human-robot and robot-robot interactions. Studies have shown that, in such interactions, the human (operator) experiences a large cognitive burden [28] and the psychophysiological state is directly correlated with the number of robots involved in the interaction [29]. The studies have, thus, suggested that the robots must display a certain degree of autonomy. In fact, the typical application of such systems is that of (semi)autonomous robots executing task-specific instructions with only some supervision or input from a human operator, particularly for research-and-rescue operations [30], assembly tasks [31] or human-swarm interactions [32]. In other approaches, humans do not assume the role of the operator at all, and robots operate with the most autonomy. For example, studies have used multiple robots to provide navigation instructions to a human [33], including indoor [34, 35] and hazardous environments [36, 37] or in creative applications such as turn-taking storytelling or drama-playing [38, 39].

The current applications of human multi-robot systems primarily emphasise coordination and collaborative control to perform tasks jointly with some human intervention. However, simultaneous, cohesive group interactions that feel truly organic and natural remain limited. Even those investigations of multi-robot systems in social and health domains [40] targeting older adults and the less abled [41–43], or children [44, 45], focus mainly on effective task execution—such as one robot handling cognitive reminders while another assists with mobility—or on providing multiple points of engagement, but they do not mimic natural group dynamics where the robots engage in a coordinated, simultaneous interaction with the user. Additionally, some of the proposed solutions have not been tested empirically with their targeted population.

1.3 The Present Study

In this study, we focus specifically on older adults, as they present a distinct set of needs and expectations. This is part of a broader vision to incorporate social robots into the lives of seniors, to support age-related challenges and address the growing need for social attention [44]. Thus, our study seeks to explore whether, and in what ways, perceived robot intelligence impacts older adults' perceptions and attitudes towards robots. Older adults may perceive, interpret and respond differently to robot behaviour, due to explicit (i.e., conscious) and implicit (i.e., unconscious) constructs [46], some of which are age-related, like alterations in hearing, vision, cognitive ability and social-emotional motivations. Studying how robot intelligence may impact older adults' trust and acceptability of robots can offer valuable insights into their technological design, to match the values and needs of this demographic. In turn, it may offer an understanding of the lens through which they view and appreciate robot intelligence. Previous empirical evidence suggests that where younger users equate a robot's intelligence with accuracy and rapidity, older adults attribute warmth, trustworthiness, empathic prompting and fallback strategies to indications of intelligence [47]. However, in contexts of health outcomes, reliability often supersedes displays of friendliness, indicating that evaluations of robot intelligence are fluid [48]. Therefore, robot design necessitates more studies and careful investigation.

While investigating the impact of perceived robot intelligence among older adults, we intentionally approach this from a human multi-robot interaction design. This is a novel but promising paradigm in HRI for designing proximal, intuitive and natural group interactions, where situational awareness can help better understand how variations in robot attributes impact user perceptions and behaviours. For example, a similar approach showed that children displayed different interaction patterns when engaging with multiple robots of varying intelligence [49]. Other studies compared dyadic interactions (1 robot–1 human) with triadic interactions (1 human–2 robots) and concluded that people attribute different perceptions to the robots based on the condition [50]. Such findings imply that in group interactions involving multiple robots, humans often use inter-robot comparative cues when forming perceptions and attitudes towards them. Thus, we also adopted a human multi-robot approach to leverage group mechanics to make the intelligence manipulation more salient. We preliminary assumed that participants would show clearer differences in their evaluations if they interacted simultaneously with two robots, which were purposefully designed to exhibit different levels of intelligence. On the one hand, this could generate inter-robot comparative cues during the experiment, thereby potentially enhancing the perceived difference in their levels of intelligence. On the other hand, it would allow us to collect within-subject evaluations by having each participant experience both robots at the same time. The latter is aimed at increasing our sample size for the experimental conditions of robot intelligence, yielding more data points and potentially reducing the margin of error. This is an important consideration, given that our target end users were older adults, a population who is generally more difficult to recruit for studies. As a result, sample sizes in such research are often smaller due to these enrolment challenges. In our within-subject design, the participants serve as their own controls, thereby reducing the noise associated with between-subject variability, especially in small sample sizes and potentially increasing the sensitivity of the statistical tests to detect significant differences between the conditions if they exist.

It is important to clarify what is defined as *intelligence* in the context of this study. Since intelligence is generally a complex construct of multiple dimensions, here, we operationalised intelligence as the robot's capacity to provide accurate responses during the group interaction, alongside its ability to detect and correct the other robot that consistently made errors. Moreover, to isolate intelligence as the key attribute under investigation, we sought to minimise any confounding factors that could bias participants' evaluations, such as robot attractiveness and task suitability, by choosing social robots with similar appearances. Social robots are particularly and equally well suited for tasks involving interactive and collaborative engagements. Furthermore, to increase the ecological validity of our group interaction design, we used multi-party interactive scenarios, like game playing, to better reflect naturally occurring social dynamics. This was also aimed at increasing participants' comfort, engagement and satisfaction, resulting in richer, more reliable data. To the best of our knowledge, no studies have experimented with group interactions of older adults simultaneously with multiple *autonomous* robots, which feel natural. This interactive dimension may have a significant impact on how older adults perceive the robots' autonomy, competence and attributes in social contexts.

Lastly, building on the role of perceived intelligence, we aimed to investigate if these perceptions would directly influence behavioural choices in older adults. Specifically, we gave participants the option to continue working with only one of the robots in a subsequent task, which was described in advance. Studies have shown that when participants receive the task first, their choice of robot will be conditioned by the nature of the task (e.g., whether the task is more analytical or social) [51]. Hence, to investigate whether intelligence played a role in participants' choice, we chose a task of increased sensitivity, such as one involving mental well-being. We hypothesised that, in this context, participants would favour the robot with higher cognitive performance and that this would correlate with it being perceived as more intelligent. This can help examine how intelligence affects trust and reliance on robot partners.

1.4 Research Hypotheses

Given the theoretical and methodological insights above, we formulated the following hypotheses:

H1: A robot with higher constructed intelligence will receive better perceptions, trust and stronger intention to use it, with these positive effects growing over time more significantly compared to the robot with lower constructed intelligence.

H2: Older adults' trust in the robot, perceived usefulness and intention to use it within the tested interaction contexts will predict their intention to use the robot in similar everyday situations.

H3: Older adults who interact simultaneously with two differently intelligent robots are more likely to continue interacting with the robot they perceive as more competent, particularly in contexts regarding their health.

2 Methodology of the Present Study

The study employs a quantitative, deductive design, complemented by qualitative findings. Ethical approval was granted by the Plymouth Ethics Online System at the University of Plymouth.

The experimental procedure, titled 'AGE IN Robot Home' (project ID 3162), received approval in November 2021 for all pilot studies related to the project, after amendments were made following recommendations from the Research Ethics and Integrity Committee. Accordingly, written informed consent, including publication rights for case details, was obtained from all participants.

2.1 Setting and Sampling

Healthy older adult participants aged between 60 and 80 (mean = 69.5, SD = 7.16) were recruited through purposive and snowball sampling with the help of the Plymouth Community Homes. The population involved 17 females and 6 male participants. Participants unable to consent or with any known history of cognitive decline, or neurological or psychiatric disorders, were excluded from the study. Sixteen participants had previous experience with a range of our social robots, having taken part in our earlier focus groups and pilot studies. All participants reported dexterity in using mobile phones and computers; 47.8% held a University degree, 21.7% had completed A-levels and 30.5% had completed only primary school.

The interaction took place in the *Robot Home* laboratory resembling a living room at the University of Plymouth. Participants signed ethics consent. They were informed that audio and videotaping would take place during the study and that data protection was in place, including confidentiality and complete data anonymisation. Before the interaction, the researchers introduced the robots to the participants and instructed participants how to interact with the robots used in the study, for example, to speak loudly and articulate their answers clearly and to act naturally. Ongoing technical support was offered before each interaction began, for example, on using the tablet application (experimental design, Section 2.2.1) or smartwatch (Section 2.2.2). All interactions took place without the intervention of the experimenter and any remote control of the robots.

Participants were asked to complete three questionnaires, one before the interaction and one after each part of the experiment. Mood was assessed at the beginning and end of the intervention to examine any changes. The data from the questionnaires were used to produce our quantitative and qualitative findings presented in this work.

2.2 Study Design

To explore our research hypotheses, we designed a two-part experimental protocol. Two social robots, NAO and Buddy, were selected to engage in multi-party cognitive game playing (Task 1), followed by a one-to-one mindful breathing exercise (Task 2), with each participant. NAO is a humanoid developed by Aldebaran (formerly Softbank Robotics), used in healthcare and education for social engagement applications [52] (Figure 1, robot on the left-hand side). The robot stands 22.6 inches tall, weighs 12.1 pounds and features 25 df. Buddy is an open and scalable mobile social robot, developed by Blue Frog Robotics to assist, entertain and educate [53] (Figure 1, robot on the right-hand side). One of Buddy's main features is its ability to simulate a range of emotions through a touchscreen located at its head. It is 24 inches tall and has 4 df (head movements and wheels). Both robots can support natural interactions with humans, and all the designed interactions in the study (game playing, meditating) occur through natural language. These behaviours and features may enhance the robots' naturalness, perception as autonomous agents and their easiness of use [54]. The rationale for not using identical robots is to avoid potential confusion and help participants clearly distinguish between them. In particular, within the questionnaires, each set of questions was accompanied by images of the respective robots, so participants could easily identify and differentiate each robot. Having identical robots could potentially have blurred these distinctions, thus, weakening the manipulation or the reliability of the participants' evaluations. Instead, the different embodiments were intended to reduce cognitive overload and potential misattributions.

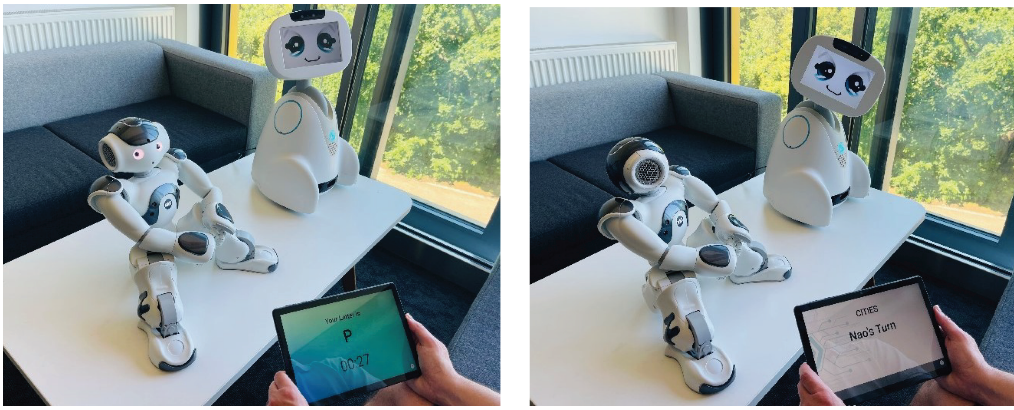

Fig. 1. The setup of the multi-party game with the two robots (NAO: left, Buddy: right) and the human participant. The game is self-directed by the participant through the tablet-based application, the interaction occurs autonomously and in natural language between the involved parties. The robots simulate a 'natural' behaviour, for example, by turning their heads when talking to each other before facing the participant again, or performing animated gestures.

2.2.1 Task 1: Cognitive Game. The game selected for Task 1 of the interaction was a variation of the classic word game 'City, Country, River'. The objective of the game is to score the most points by quickly coming up with words that start with a randomly generated letter for each given category. This game was chosen for it being known to promote memory and recall, language skills and mental agility, while also entertaining and encouraging social interaction and competitive play.

The game was played in a group between one participant and two robots simultaneously and autonomously, i.e., no remote control (Figure 1). The human participant was given a tablet to control the game and a printed sheet to record their answers. The tablet ran the game application, which communicated data in the background with each robot via TCP/IP sockets. The game lasted for four rounds (four distinct letters), with the selected categories being 'City', 'Animal' and 'Object'. To begin the game, the participant would press a button on the tablet to generate the first letter at random, starting a 60-second countdown (Figure 2). The players continued filling in the categories until the time ran out. At the end of the countdown, each robot and the participant compared their answers, with the participant self-directing this order through the tablet. According to the rules, blank spaces or incorrect answers were worth 0 points, correct answers 5 points and unique answers among all parties were worth 10 points. The participant recorded their total score on the printed sheet each round, while the robots' scores were recorded in an electronic scoring system running in the background. The player with the highest score would win the game.

In this experimental design, we use the term *constructed intelligence* to refer to the robot's designed capacity to provide accurate responses and to detect and correct errors made by the other robot during group game-playing. Thus, we manipulated the constructed intelligence of the robots between two conditions: (i) *high constructed intelligence or 'smart' condition* and (ii) *low constructed intelligence or 'non-smart' condition.* We highlight that the terms *smart robot* and *non-smart condition* are applied by the researchers *post hoc* to distinguish between the experimental manipulations. Specific to this study, one of the robots was purposefully designed to provide perfect answers for each category in the game, often using rare or difficult words (*smart condition*), while the other robot (*non-smart condition*) intentionally made mistakes, like selecting words that did not belong to the category (e.g., *as animal, I chose an alien*) or misspelling words to start with the generated letter (*Pucharest* instead of *Bucharest*). The *smart* robot corrected the other robot when it made a

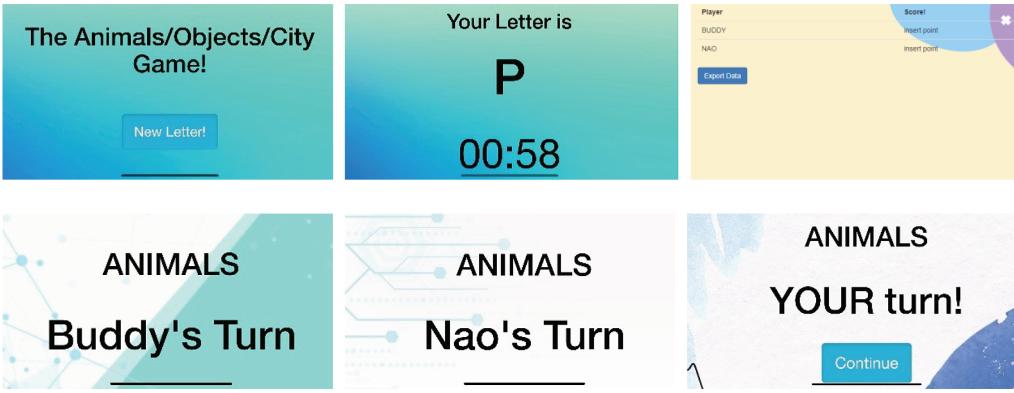

Fig. 2. Our modified 'City, Country, River' game running on a tablet app, with three categories 'Animals, Objects, Cities'.

mistake, and explained the correction accordingly. The fallible robot acknowledged its mistakes. The participants were blind to the conditions and the terms (*smart, intelligence*), and their ratings were based on self-perception. The interaction was designed to appear as if the robots were naturally talking among themselves. For example, the robots would turn their heads to face each other when talking or performing animated gestures (Figure 1: right). Then, they would turn to face the participant and maintain 'eye contact' given a people-tracking ability. Intentional eye contact is argued to be a desirable feature for older adults, which may enhance a robot's social relevance [55].

To ensure consistency and eliminate variability of the experimental conditions among the participants, the entire game was pre-structured and identical both in content and length of interaction. We used the same letters and rounds of game categories for each participant. This aimed to minimise differences in exposure, while keeping the robots' behaviour autonomous and as intended each time. Moreover, since the participant controlled the turn-taking mechanism through the tablet, by pressing the screen to advance the game, the robots would generate their scripted answers without processing or reacting to participants' speech. This was done to prevent misrecognition errors and to control for variability in the interaction design, both of which could influence participants' perceptions of the robots. The only item loosely controlled by the experimenter remotely was calculating the final score of the game. This was displayed to the participant on the tablet at the end of the game, but no winner was announced explicitly, as this behaviour was not autonomous.

As mentioned previously, we matched the robots in size, expressive modality and overall design to minimise morphology bias, while still choosing to use different robots to maintain clarity in participant evaluations. Moreover, we counterbalanced the robot-intelligence assignment across participants, such that each robot appeared in both conditions for half of the sample. That is, each robot served in the *smart* condition for half the sample and in the *non-smart* condition for the other half. This aimed to evenly distribute potential biases or order effects across the study, if they existed. Collectively, these strategies help partially mitigate fixed associations.

A real example of a triadic (two-robot one-human) interaction (Task 1)

Participant: The new letter is P
NAO: Okay
Buddy: Got it!
Timer runs out
NAO: For the object, I chose a pen.

Buddy: My choice of object is the pedestal.
Participant: I chose a phone.
NAO: As an animal, I chose the pigeon.
Buddy: My choice is the peacock.
Participant: I also chose pigeon.
NAO: As a city, I chose Pucharest. **NAO makes a mistake**
Buddy: **Turns to face NAO, followed by NAO turning to face Buddy** That's wrong! Bucharest is spelt with a "B"! My choice is Penzance.
NAO: **Facing Buddy** Oops! I feel very silly!
Both turn back to face the participant
Participant: My choice is Plymouth.

2.2.2 Task 2: Mindfulness Meditation. Upon completing Task 1 of the interaction and the respective questionnaire, the participants were informed that Task 2 of the study would be a mental well-being session to help relax their minds and reduce stress. They were then asked to choose only *one* of the robots to continue this intervention. We expected that knowing the task beforehand could potentially influence their choice and that this choice may correlate with the chosen robot's behaviour in part 1, i.e., the manipulation of constructed intelligence. The participants were asked to motivate their choice, and their answers were recorded for the analysis.

Task 2 consisted of a guided meditation and breathing session. Mindfulness meditation is thought to lower heart rate, which decreases anxiety, and stress and improves calmness and focus [56]. Such therapies are endorsed by the National Institute for Health and Care Excellence and may prove beneficial for older adults. The researcher asked the participants to wear a smartwatch, which would measure their heart rate. The robot would take the participant's unique readings from the app connected to the smartwatch and communicate the value to the participant (e.g., *your current heart rate is 80 bpm*). Then, the robot would guide participants through the meditation session, starting by playing a piece of calming background music. The robot's instructions would mirror the elements of a traditional mindfulness session, including initial relaxation, focusing on the breath, observing sensations, acknowledging thoughts and gradually transitioning back to the present moment. The robot would modulate its voice, slow down its speech and pause at appropriate moments to allow for deeper reflection and relaxation. The script, speed and pauses were replicated from human-directed meditations and similarly executed by the robot. At the end of the mediation, the robot would take a second reading from the smartwatch and compare the two readings (e.g., *your heart rate improved by 5 bpm*).

Finally, the robot would ask the participants if they felt more relaxed after the meditation and provide positive reinforcement regardless of their response. The answers were recorded for the analysis.

Mindfulness meditation (Task 2)

music starts
As you are seated, make yourself comfortable **long pause**. Take a few moments to be still in your body **long pause**. Allow yourself to be aware of where your mind is at just now **long pause**. Take a moment to notice your thoughts **long pause**. Gently **short pause** close your eyes **pause** or softly gaze ahead **pause** finding a point of focus that is not distracting **long pause**. Bring your awareness to your breath **pause**. Start by taking two full breaths **very long pause**. [...]
[...] When your mind pulls you away or distracts you during this practice **pause**, you can come back to this part of your breath **pause**, as an anchor, or a point of focus **long pause**. Breathe in **long pause** and breathe out **very long pause**. As you reach the end of this practice **short pause**, congratulate yourself for taking this time to be present with your breath **long pause**. You can now open your eyes **pause** or simply bring your awareness back to me.
music ends

3 Data Analysis

Self-reported measures were performed with data collected across three questionnaires: (i) pre-interaction questionnaires; (ii) after Task 1—cognitive game/entertainment with both robots simultaneously; and (iii) after Task 2—meditation with only the chosen robot. These questionnaires were designed ex-novo to align with our aims, guided by constructs commonly measured in HRI literature, e.g., perceived trust, fit-for-use and intention to use. The items are inspired and consistent with scales like the Godspeed Questionnaire Series [57], Trust in Automation [58] and Technology Acceptance constructs [59], but they are modified for the context of the study and the target demographic. All items and their phrasing are reported in full under each construct (measure) in the quantitative and qualitative analyses sections, along with the specific time point at which they were administered (pre-interaction, post-Task 1, post-Task 2).

The quantitative analysis is performed with the entire sample (N = 23), regardless of the manipulation check of the constructed robot intelligence. This is because explicit awareness of the manipulation is not a prerequisite for effective manipulation in this study. Despite the robot conditions existing, this manipulation targets participants' experiential and behavioural impressions. Social cognition research (e.g., [60]) argues about *implicit processing,* where exposure to systematically different behaviours leads to attitudinal or perceptual shifts. This phenomenon has also been used or replicated in HRI studies (e.g., [61]). Thus, our analysis aims to quantify how our manipulation design may affect the intended measures (perceptions, trust, intention to use, etc.) implicitly. In particular, since intelligence is a complex construct, it may encompass more dimensions beyond the accuracy of responses, and its perception may be dependent on the participants' intrinsic characteristics. The analyses were conducted in R (version 4.2.1, R Core Team, 2020) and Python (3.8). The packages used were *lme4* package for fitting the Linear Mixed-Effects Models in R, *pingouin* for mixed-model ANOVA, *sklearn (scikit-learn)* for *k*-means clustering and *statsmodels* for Ordinary Least Squares regression analysis. Associations were explored using Spearman's rank correlations.

3.1 Perceived Robot Attributes

We measured the participants' perceptions of robots in general in the pre-interaction questionnaire using the following 'Robots are' items as a baseline in a 7-point Likert agreement scale, where 1 indicated strong disagreement and 7 indicated strong agreement:

- Robots are: (i) reliable; (ii) can count on; (iii) consistent; (iv) capable; (v) competent; (vi) meticulous.
- Robots are: (i) sincere; (ii) genuine; (iii); authentic; (iv) ethical; (v) respectable; (vi) principled.
- Robots are: (i) knowledgeable; (ii) responsible; (iii) intelligent; (iv) sensible.

These items were re-evaluated after Task 1 for both robots to measure the participants' perceptions immediately following their interaction with them. The results were aggregated on the robots' constructed intelligence condition (*smart/non-smart*), regardless of the type of robot. The purpose of the evaluation was to assess the effect of perceived intelligence on the perceived robot attributes.

The items were then re-evaluated post-Task 2 only for the robot selected to continue the interaction. The purpose of the evaluation was to assess the effect of the task and time on the participants' perceived robot attributes. This was compared against the baseline (pre-interaction) and the perceptions only of the participants who chose that robot after Part 1.

3.1.1 Task 1: Cognitive Game. The collected data contained measurements for multiple attributes across two conditions (*smart* vs. *non-smart*) and 2 time points (baseline or pre-interaction and

Task 1). Mixed-design ANOVA was used to assess the effects of condition, time and interactions for each individual attribute.

Our analysis showed that certain attributes displayed notable time effects. The perception of 'sincerity' improved significantly from baseline to Task 1 ($F(1,88) = 1{,}225.0$, $p = 0.018$) and moderate improvements were observed for the attributes like 'genuine' ($F(1,88) = 225.0$, $p = 0.042$) and 'authentic' ($F(1,88) = 196.0$, $p = 0.045$). Since the data were not normally distributed, we corroborated the results with the Wilcoxon signed-rank test, which confirmed a statistically significant time effect on the said attributes (sincere: $p = 0.0033$; genuine: $p = 0.0221$; authentic $p = 0.0246$). However, none of the attributes showed significant interaction effects between condition and time, suggesting that these improvements occurred for both robots, regardless of their designed intelligence. No statistically significant effects were observed for the other attributes, but their ratings remained stable within the medium range (mean = 3.83–5).

3.1.2 Task 2: Mindfulness Meditation. Similarly, we assessed the effects of individual attributes across two conditions and three time points after Task 2. Individual attributes showed no significant differences in condition and interaction effects, with only marginal time effects. Nevertheless, we explored the combined effects of all attributes, given that the lack of significant individual effects may arise due to several factors related to statistical power, effect size, consistency of effects and averaging of the variability in attributes, thus masking their true effect.

We used a mixed-model ANOVA to assess the combined impact of condition, time and their interaction. Time was treated as repeated measures and condition as between-subject. Calculations using Cohen's f derived from the ANOVA F-values indicated that the model was well-powered at 92%. The confidence in detecting time effects was the highest, with a statistical power of 99.9%, while there was a reasonable chance of detecting condition effects, with a statistical power of 62.1%. The results revealed a significant condition effect ($p = 0.0027$), with the robots in the *smart* condition being perceived more positively overall compared to those in the *non-smart* condition. Moreover, the time effect was significant ($p < 0.001$), with perceptions improving over time, for both robots. The interaction effect *condition* * *time* was also moderately but statistically significant ($p = 0.0487$), indicating that the effect of robot condition on mean points depended on the time point, with the difference in perceptions between *smart* and *non-smart* robot conditions becoming more pronounced as the interaction progressed. Given some deviations in the normality of the data, we further confirmed the ANOVA results with a Generalised Linear Mixed Model using a Gamma family with a log link function, which is suitable for continuous, positively skewed data.

When interpreting these results, it is important to highlight that in Task 2 (meditation), there is no experimented manipulation of the robot's condition. The selected robot by the participant performs the task identically, regardless of whether it was assigned in the *smart* or *non-smart* condition in Task 1 (cognitive game). Therefore, analyses of condition and interaction effects are based on the robot's condition during Task 1. Hence, the results indicate that participants who selected the robot in the *smart* condition to continue the interaction after Task 1 rated the robot more positively than those participants who selected the robot in the *non-smart* condition. While mean ratings increased for both conditions relative to pre-interaction baseline measures, the improvement after Task 2 was more pronounced for the robot with higher constructed intelligence (*smart*).

We used further descriptive statistics in Figure 3 to illustrate the differences observed between the conditions after Task 2. The rationale to report only these findings is based on the results of the ANOVA, followed by a Tukey HSD *post hoc* test, which confirmed a statistically significant difference between the *smart* and *non-smart* conditions at Task 2 ($p = 0.0038$), but not at Task 1 ($p = 0.3123$). In Task 1, participants evaluated both robots, resulting in a well-powered within-subject design with paired data for each participant. Since no significant differences were found

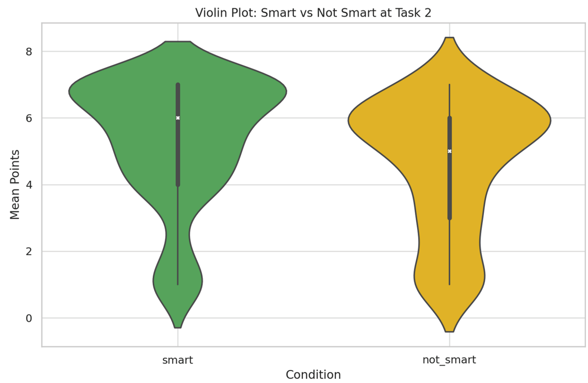

Fig. 3. The results illustrate that a divergence in participant perceptions emerged by Task 2. The distribution shape shows that robots in the *smart* condition received consistently higher ratings at this point, indicating that the difference in evaluations became apparent only after Task 2. The distribution of mean points in the *smart* condition is more concentrated at higher values. The median is higher, and the density is more pronounced in the upper range of points. The distribution is wider and more variable for the *non-smart* condition, with a lower median. The density shows a broader spread of scores, indicating greater variability in ratings. The overall shape suggests that robots with lower constructed intelligence receive a wider range of scores, with a tendency towards lower ratings compared to those with higher constructed intelligence.

between the robot conditions at the end of Task 1, any subsequent differences observed in Task 2 are likely due to the continuation of interaction and evolving perception, instead of pre-existing biases. This interpretation is consistent with the mixed-model ANOVA results, which showed that differences in perceptions between the robots became more pronounced and significant over time.

The results confirmed that differences in perceived robot attributes became more pronounced over time. The robot in the *smart* condition consistently received higher ratings, indicating that its perceived intelligence had a lasting and increasing impact on users' views of its reliability and suitability for tasks. This supports our hypothesis H1 in terms of overall perceptions.

3.2 Perceived Trust

We measured the perceived trust of the participants in the pre-interaction questionnaire using three items as a baseline on a Likert agreement scale from 1 (Strongly disagree) to 7 (Strongly agree):

- I would trust robots in entertainment contexts.
- I would trust robots if they helped me with cognitive ability exercises.
- I would trust robots if they supported me with my mental well-being.

Items 1 and 2 were re-evaluated following Task 1 to measure participants' perceived trust in both robots immediately after the interaction (replacing 'robots' with 'Buddy' and 'NAO' in the queried items) compared to the baseline. In the experimental context, this evaluation sought to identify any correlation or mediation effects between the robot's designed intelligence and the participants' perceived trust. The results were analysed under combined intelligence conditions (*smart/non-smart*), regardless of the specific robot.

Item 3 was re-evaluated exclusively after Task 2 to assess participants' perceived trust in the selected robot in the experimental context of mental well-being compared to the baseline trust. The intention was to assess any effects of the task or time on the participants' perceived trust and any variations attributable to perceived intelligence.

3.2.1 Task 1: Cognitive Game. We performed Linear Mixed-Effects Regression Modelling. Preliminary results with a mixed model that included random effects and fixed effects, where the participant ID was treated as a random effect, i.e., intercepts vary across participants, whereas time and condition were treated as fixed effects, revealed that neither time, condition, nor their combination and interaction had a significant impact on entertainment or cognitive trust scores. This model assumes the effects of time and condition to be the same for every participant (homogeneous effects). However, given the observed high variability of the intercepts, we decided to fit a model with random intercepts and slopes for 'time' and 'condition', which allow the effects of such predictors to vary across participants (heterogeneous effect). This assumes that each participant has unique responses to time and condition allowing for individual differences in how they impact trust. The rationale of adding random slopes is to try to capture individual differences in how people react to the same condition (*smart* vs. *non-smart*) or time points (pre- vs. post-interaction) and to improve model fit. Evaluations using Akaike Information Criterion and Bayesian Information Criterion revealed that the model with random slopes had the lowest scores, indicating that this model fits the data best, thus capturing individual variability better.

For the entertainment context, the model with random slopes revealed a significant effect of the *smart* condition ($p = 0.009$), with higher trust scores, but no significant effect of time ($p = 0.729$). For the cognitive context, the model with random slopes also showed a significant effect of the *smart* condition ($p = 0.007$) and no significant effect of time ($p = 0.914$). The significant condition effect reveals that, when participants specifically evaluated each of the robots they interacted with, they consistently rated the robot in the *smart* condition as more trustworthy than the robot in the *non-smart* condition, confirming our hypothesis H1 regarding the trust construct. Nevertheless, the lack of time effect indicated that their trust did not significantly change over time (H1 not supported). In other words, the interaction did not substantially alter their overall perception of robots *in general*, but they did perceive the robots with higher constructed intelligence more favourably than their lower constructed intelligence counterparts after the interaction. However, it is important to note that a single and short interaction may be insufficient to observe such an effect in time.

3.2.2 Task 2: Mindfulness Meditation. Preliminary analysis showed positive but no statistically significant differences in trust ratings between the *smart* and *non-smart* robot conditions after the interaction, despite a statistically significant difference having been observed after Task 1. It is plausible that these differences were not evident after Task 2 because the robots performed the task identically and correctly. Hence, conditions of differing intelligences were not present during Task 2, and the robots were rated based on the interaction alone, potentially 'closing' the trust gap between the robots. For this reason, we investigated the evolution of trust after Task 2. This allowed us to explore whether and how the longer interaction, together with a consistent robot behaviour, may have influenced trust, potentially explaining why the distinction between the robots' perceived trustworthiness became less pronounced. We examined participants' trust ratings before and after Task 2, focusing on the direction of their trust changes (no change, increase, or decrease). We aimed to assess whether post-trust was driven primarily by initial trust levels (anchoring effect) or whether it was potentially recovered or strengthened through consistent, reliable robot behaviour in Task 2.

We applied *k*-means clustering with the number of clusters ($K = 3$) determined using the Elbow method heuristic. Descriptive statistics were used to illustrate each cluster's pre- and post-trust scores (Figure 4). These were statistically significant:

— Cluster 0 participants (8) began with *high trust* and experienced a further increase in trust after Task 2 (Figure 4). All participants in this cluster chose to continue interacting with

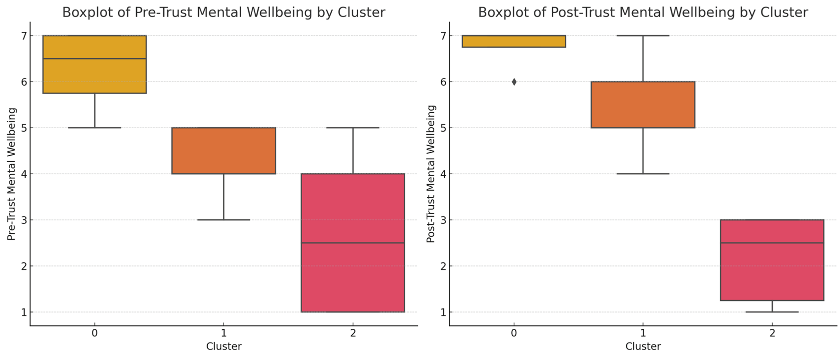

Fig. 4. (*Cluster 0)* Median pre-trust score is around 6.5, with a relatively tight distribution (small IQR and few outliers). Median post-trust score is around 7, with a tight distribution (small IQR and few outliers). (*Cluster 1)* Median pre-trust score is around 4, with a moderate spread. Median post-trust score is around 5, with a moderate spread. (*Cluster 2*) Median pre-trust score is around 2.5, with a broader distribution. Median post-trust score is around 2.5, with a slightly more concentrated distribution after the interaction, suggesting reduced variability, but no systematic decline.

the robot in the *smart* condition after Task 1. Five (62.5%) of these participants had perceived the robot as more intelligent. All of them trusted the robot after Task 1 (100%). Seven (87.5%) participants felt relaxed after the meditation in Task 2. The trust increase after Task 2 might indicate that, for this group, pre-existing positive expectations, including high initial trust and perceived robot intelligence, likely created a positive feedback loop where trust was not only sustained but strengthened given longer interactions and consistent robot performance.

— Participants in Cluster 1 began with *moderate* trust. All of them chose to continue interacting with the robot in the *non-smart* condition, even though only 3 (33.3%) perceived it as intelligent. Despite this, over half (55.6%) reported trusting the robot already after Task 1. Their trust levels increased further after Task 2 (Figure 4). This suggests that for this group, trust was not solely dependent on perceived intelligence. Instead, it may reflect a baseline willingness to trust or a broader positive disposition towards the robot they selected. The subsequent increase in trust after Task 2 indicates that the robot's reliable performance in this task, potentially evidenced by unanimous reports of feeling relaxed after the interaction, can reinforce or validate initial trust, even if the robot was not perceived as especially intelligent. In other words, participants appear to have extended trust based on factors other than intelligence, possibly personal preference, comfort or perceived suitability, and then had that trust confirmed by the robot's consistent behaviour.

— Cluster 2 involved (6) participants with *low and diverse trust*, whose trust remained low and inconsistent, but less widely distributed after the interaction. This variability is reflected in the mixed composition of the group, with participants having chosen the robot either in the *smart* or *non-smart* condition following Task 1. Only half (3) perceived their robot as more intelligent, and just two participants (33.3%) reported trust in it after Task 1. The diversity in trust levels and lack of consistent increase suggest that the interaction did not uniformly affect trust for this group. Instead, it was likely influenced by individual predispositions rather than the robot's characteristics or performance. This highlights the role of pre-existing trust as a potential stronger determinant of post-task trust.

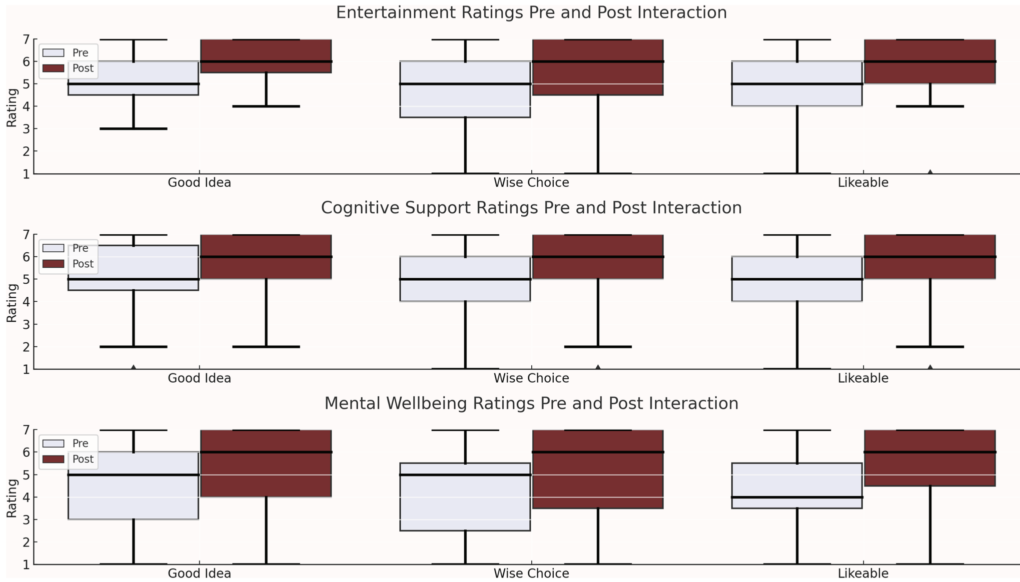

Fig. 5. The box plots show the spread and central tendency of the ratings before and after the interaction for each category. Significant changes can be observed in the mental well-being category, while entertainment and cognitive ratings show less variation. The changes were statistically significant for all categories.

3.3 Perceived Fit-for-Use

Fit-for-use refers to the extent to which a robot is perceived to be appropriate, effective and suitable for fulfilling a specific task or purpose in a given context. In this study, fit-for-use is measured as participants' attitudes towards the use of robots in different roles (such as entertainment, cognitive support or mental well-being) assessed along three dimensions: whether using the robot is considered a good idea, a wise choice and likeable. The following statements were evaluated using a Likert scale from 1 (Strongly disagree) to 7 (Strongly agree):

— Using robots for entertainment is (i) a good idea; (ii) a wise choice; (iii) likeable
— Using robots for cognitive support is (i) a good idea; (ii) a wise choice; (iii) likeable
— Using robots for mental well-being is (i) a good idea; (ii) a wise choice; (iii) likeable

Items 1 and 2 (entertainment and cognitive support) were re-assessed after Task 1 to measure participants' perceptions of fit-for-use in these contexts immediately following their interaction with the robots. The results are analysed and presented in terms of the robots' perceived intelligence condition. Item 3 (mental well-being) was re-measured after Task 2.

Descriptive statistics revealed improvements across all categories and dimensions after the interaction (Figure 5). Post-interaction ratings showed narrower interquartile ranges, indicating more consistent and higher ratings compared to pre-interaction scores. The most notable improvement occurred in the mental well-being category, suggesting that the interaction positively influenced participants' perceptions of fit-for-use in this area. These results were statistically significant across all categories.

A mixed-effects linear regression model was used to investigate whether participants' trust in robots and the robot condition (*smart* or *non-smart*) influenced post-interaction scores across fit-for-use dimensions (*good idea*, *wise choice*, *likeable*) in the categories of entertainment, cognitive and mental well-being. The mixed-effects (multilevel) model was used to handle different numbers of observations per robot condition, since the entertainment and cognitive categories involved within-subject evaluations for both robots simultaneously, whereas mental well-being ratings were collected under between-subject designs (i.e., participants continued interacting with only one robot, which previously was either in the *smart* or *non-smart* condition). Such a model is also not biased by the intercept variability. The model was fitted with the Restricted Maximum Likelihood. The dependent variable was the post-interaction score aggregated across all categories and fit-for-use dimensions. Instead, the post-trust scores for each participant averaged across the categories (entertainment, cognitive, well-being), the robot condition (*smart*, *non-smart*) and their interaction (trust x condition) served as independent variables. In this analysis, the pre-trust scores served as a baseline covariate that did not vary with robot condition, since participants reported trust in robots in general before the interaction. Our analysis indicated that the post-trust score was a significant positive predictor of post-interaction scores ($\beta = 0.88$, $p < 0.001$), meaning higher trust was associated with higher perceived fit-for-use ratings. Moreover, baseline trust (pre-trust) also positively influenced post-ratings, meaning that participants with higher existing trust tended to give higher fit-for-use ratings after the interaction, regardless of the robot. When examining the effect of robot condition, we found a significant negative main effect ($\beta = -0.87$, $p = 0.030$). Specifically, when post-interaction trust was low (trust = 1 in the Likert scale), participants rated the robot in the *smart* condition lower than the *non-smart* condition on fit-for-use dimensions. However, at higher post-trust levels, this was reversed, with participants rating the *smart* robot condition more favourably than the *non-smart* condition. This was further supported by the significant and positive interaction effect between trust and *smart* robot condition ($\beta = 0.17$, $p = 0.020$), which indicated that fit-for-use ratings for the *smart* condition increased more sharply with greater trust compared to the *non-smart* condition. Our model was highly significant overall, and the inclusion of predictors significantly improved model fit compared to an intercept-only (null) model (Likelihood Ratio Test: $\chi^2(6) = 62.90$, $p < 0.001$). It converged successfully and has reasonable group variance, indicating mixed-effects were appropriate. These results support our hypothesis H1 in terms of fit-for-use.

3.4 Intention-to-Use Robots

The intention-to-use robots was measured as the willingness of older adults to use the robot over the upcoming weeks in several contexts of daily living. We designed the following item ex-novo and asked participants to indicate their level of agreement on a 7-point Likert agreement scale for each of the contexts:

— I would be willing to use robots in the following contexts in the next weeks: (i) daily house routines; (ii) medicines prescriptions; (iii) physical health support; (iv) mental well-being support; (v) entertainment; (vi) cognitive training support.

This measure was used to assess the associations between the level of trust in the robot with the proxy variable of the use of robots in different contexts. We assumed that trust differences in the experimental contexts (i.e., entertainment, cognitive support, mental well-being) could potentially generalise to differences in the perceived trust of older adults for other contexts of daily living, such as house routines, physical health support and medicine prescriptions, due to an anchoring effect [62] (i.e., where a decision is based on a situation that is more or less related).

We performed repeated measures ANOVA for the intention-to-use scores across different domains pre- and post-interaction. In the following results, variability refers to the spread or consistency of participants' responses within each category, with +/− values representing either an increase or decrease in variability. No statistically significant differences between the pre- and post-intervention scores were found for the categories of daily house use and entertainment, but these differences were highly significant for physical health ($p = 0.032$, reduced variability −0.81), mental well-being ($p = 0.001$, reduced variability −0.49) and cognitive training ($p = 0.013$, reduced variability −0.47). Due to a missing value in the rating of intention-to-use robots in medicine for one of the participants in the pre-interaction, we performed both a mixed-effects ANOVA to handle missing/unbalanced data and a repeated measures ANOVA by removing the entry from the data as an outlier. In the mixed-effect ANOVA, we considered the intercept (post-interaction average), time effect (pre vs. post-difference) and group variance (participants' variability). Both tests confirmed a significant difference between the pre- and post-interaction scores also for this domain ($p = 0.009$, $p = 0.015$, respectively). Although the variance increased slightly (+0.32), the upper quartiles became more consistent. The results suggest a potential positive impact of the interaction on the participants' willingness to use robots, particularly in areas related to health and well-being, leading to more uniform responses across participants.

Moreover, we analysed the associations between the different categories using Spearman's Rank correlation matrix. Different from the ANOVA results, here, we examine how consistently responses in different categories move together (Spearman's ρ). All categories in Figure 6 refer to evaluations made post-interaction, with the coefficients representing the degree of change of the associations post-interaction compared to their associations before the interaction (i.e., +/− values refer to Δ Spearman ρ from pre- to post-interaction). We focused mainly on the correlations between the experimental categories (entertainment, cognitive support, mental well-being) with the close context categories (daily use, physical health, medicine). The results indicated that the intention-to-use robots in entertainment correlated more strongly with the intention-to-use robots daily in the house ($\Delta\rho = +0.43$) after the interaction ($\rho = 0.70$). Both the intention-to-use robots in cognitive training and mental well-being revealed an increase in their association with the intention-to-use robots for medicine, respectively +0.22 and +0.12 ($\rho = 0.74$ and $\rho = 0.69$). Cognitive training and mental well-being were more positively associated with one another after the interaction than before ($\Delta\rho = +0.38$, $\rho = 0.95$). Mental well-being also positively impacted physical health, with their post-interaction association strengthening by +0.14 ($\rho = 0.88$). The results suggest that the interaction may have reinforced the perceived usefulness of robots across multiple related domains.

Further strong positive associations were observed in how *trust* and intention-to-use robots were related after interaction, compared to before. In other words, we quantified the changes in the associations between trust and intention to use from pre- to post-interaction, represented as the change in correlation coefficients (i.e., Δ Spearman ρ). Separate calculations for trust in robots depending on *smart* and *non-smart* conditions showed notably higher correlations for the *smart* condition, especially when investigating trust in mental well-being and cognitive support with the intention-to-use robots in mental well-being and medicine contexts. For example, trust in the robot in the *smart* condition (mental well-being) and the intention to use it revealed strong and statistically significant monotonic relationships for mental well-being ($\rho = 0.877$, $p < 0.001$,) and medicine ($\rho = 0.887$, $p < 0.001$). In relative terms, the correlation between trust and intention-to-use robots was strengthened by 0.043 for mental well-being and 0.320 for medicine post-interaction, but for robots in the *non-smart* condition, it decreased by 0.072 for mental well-being and 0.044 for medicine. Similarly, the correlations were strengthened when considering reported trust for cognitive support and intention-to-use robots with higher constructed intelligence in medicine ($\Delta\rho = +0.147$, $\rho = 0.582$), but decreased for those with lower constructed intelligence by 0.174 ($\rho = 0.286$). These results provide promising evidence that trust in the robot leads to a greater intention to use it (H1) and produces an anchor effect in related contexts of daily living (H2). This may suggest that robots with higher constructed intelligence could enhance their acceptance across various domains, including those more sensitive. However, these results should be read with caution due to the sample size.

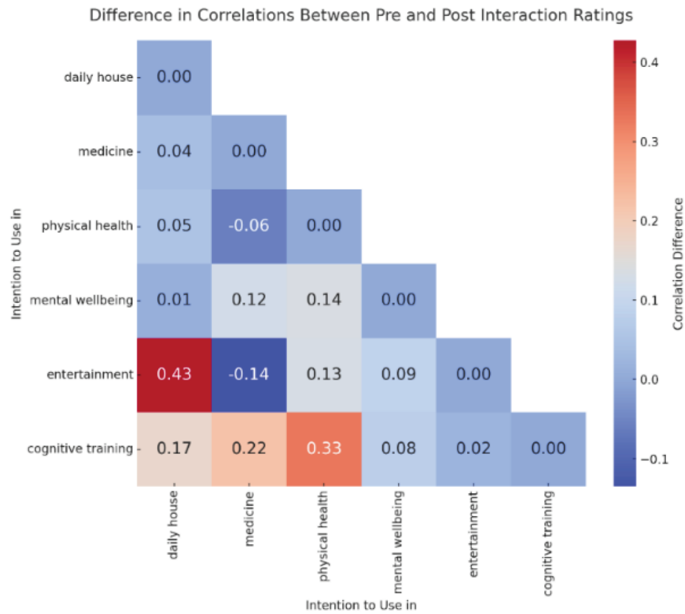

Fig. 6. Spearman's correlation matrix shows the differences in correlation coefficients between pre- and post-interaction ratings across six categories: daily house, medicine, physical health, mental well-being, entertainment and cognitive training. Only the main diagonal and the lower triangle are shown to highlight unique pairwise relationships. Notable observations include a significant increase in the correlation between entertainment and daily house ($\Delta\rho = +0.43$) and substantial increases between cognitive training and both medicine ($\Delta\rho = +0.22$) and physical health ($\Delta\rho = +0.33$). Some correlations decreased slightly, such as between entertainment and medicine ($\Delta\rho = -0.14$).

3.5 *Post Hoc* Analysis of the Robot's Morphology Bias

Despite swapping the robot conditions for half the sample of participants to mitigate any fixed associations between robot appearance and perceived intelligence, there is no guarantee that morphology bias did not emerge. Therefore, we conducted a *post hoc* analysis based on robot agent (NAO or Buddy), instead of robot condition (*smart vs. non-smart*).

3.5.1 Robot Attributes. We ran a combination of univariate ANOVA, Welch's test and Mann-Whitney U test, where appropriate, after confirming assumptions of normality (Shapiro-Wilk test) and variance homogeneity (Levene's test) for each attribute. Our results showed no statistical differences between the robot agents on any attributes across all three tests (all p-values > 0.5). We also run PERMANOVA to test whether the multivariate profiles combining all robot attributes (i.e., combined effect) differ between Buddy and NAO. This analysis also showed no statistically significant differences between the robots (pseudo-F = 0.1916, p-value = 0.881, sample size = 46, groups = 2, number of permutations = 999). Thus, participants' impressions of robot attributes were not affected by the robots' morphology.

3.5.2 Perceived Trust. We ran a one-way ANOVA treating robot as a single factor with two levels (Buddy and NAO) and testing if trust differed by robot. The results showed no statistically significant differences in trust scores in either entertainment ($F = 0.661$, $p = 0.425$), cognitive support ($F = 0.000$, $p = 1.0$) or mental well-being ($F = 0.042$, $p = 0.840$). These confirm that, on average, participants rated both robots similarly across the measured trust construct.

3.5.3 Perceived Fit-for-Use. We ran a Mann-Whitney U test to compare the differences between Buddy (*smart* condition) and NAO (*smart* condition) for individual scores of fit-for-use dimensions (likeable, good idea, wise choice) across entertainment, cognitive support and mental well-being domains. Since the observations are paired, comparing Buddy and NAO in their *smart* condition inherently involves an indirect comparison of the robots in their *non-smart* configurations. The results showed no significant differences based on the robot agent (all p-values > 0.34). The MANOVA test comparing the mean vectors of the combined variables (i.e., scores across multiple domains and response types: Entertainment, Cognitive, Mental; Good, Wise, Like) between groups also showed no significant differences with all tests consistently producing $p > 0.05$ (e.g., Wilks' Lambda = 0.4605, $F(9,13) = 1.6923$, $p = 0.1882$; Pillai's Trace = 0.5395, $F(9,13) = 1.6923$, $p = 0.1882$). Thus, the overall fit-for-use profiles do not differ between Buddy and NAO robots.

3.5.4 Intention-to-Use Robots. We used the Wilcoxon signed-rank test to verify whether the intention-to-use scores (after Task 1) differed between the robots (Buddy vs. NAO). The test showed that across all six domains (daily house, entertainment, cognitive training, mental well-being, medicine), there were no statistically significant differences in participants' scores between Buddy and NAO when both were in the *smart* condition (all p-values > 0.35). Due to paired comparisons, the results apply also to the *non-smart* conditions.

We further evaluated whether intention-to-use scores at the end of the interaction (after Task 2) differed by the chosen robot (Buddy and NAO). The Mann-Whitney U test showed no statistically significant differences (all p-values > 0.08) in intention-to-use scores across all domains between participants who chose to continue interacting with Buddy, compared to those who chose NAO.

Overall, our *post hoc* analysis confirms the validity of our manipulated intelligence condition. Significant differences were found between the *smart* and *non-smart* robot conditions across all measured outcomes, which were not confounded by the robot agent (Buddy or NAO), since no differences were found based on robot agent. This shows little to no bias in the study design and findings, indicating that the manipulation condition is the primary driver of user attitudes and perceptions in this experiment. Future work could also include baseline perception measures.

4 Qualitative Findings

We asked the participants to motivate their choice of robot after Task 1 and whether the meditation helped them relax after Task 2. This was not intended as a manipulation check. Rather, the responses are used to corroborate the quantitative results and understand if participants perceived our designed conditions as intended. We added the following binary evaluations in the questionnaires administered immediately after the respective interactions:

- I felt that I could *trust* the chosen robot more. (True/False)
- I felt that the chosen robot was *more intelligent* than the other. (True/False)
- I felt *relaxed* after meditating with the robot. (True/False)

Items 1 and 2 were queried after Task 1 and after participants selected their preferred robot. The questions were accompanied by images of the robots to help participants remember them by name. Item 3 was queried after Task 2.

4.1 Factors Affecting Robot Choice

We ran a crosstabulation to analyse the relationships between the categorical variables of *designed robot intelligence* (i.e., condition), *reported robot intelligence* and *reported trust*, for the chosen robot after Task 1. Each of the variables can take two values: Smart or Not Smart (constructed robot intelligence), Intelligent or Not Intelligent (reported robot intelligence) and Trusted or Not trusted

Table 1. Crosstabulation of Robot Design, Perceived Intelligence and Trust, Measured after Task 1 and after the Participants Had Chosen One Robot to Continue the Interaction

Cluster	Variable combination	Count
1	Smart, Intelligent, Trusted	7
2	Smart, Not Intelligent, Trusted	3
3	Smart, Not Intelligent, Not Trusted	3
4	Not Smart, Intelligent, Trusted	3
5	Not Smart, Not Intelligent, Trusted	2
6	Not Smart, Not Intelligent, Not Trusted	5

The counts indicate the number of instances each combination occurred in the dataset.

(reported trust). The frequency distribution of the variable combinations is summarised in Table 1 and reported as clusters of participants for simplicity.

These findings showed mixed support for our H3. Only half of the participants' perceptions of the robot aligned with our designed expectations (Clusters 1, 2, 6). We analysed the participants' qualitative responses to the questionnaire to understand their choice and interpret the resulting variations in perceived trust and intelligence. These revealed ambivalent yet unsurprising factors and expectations between the clusters (Table 1).

Participants chose to continue interacting with the robot given primarily the following rationales:

(1) *Utilitarian Factors*: The robot's perceived reliability, knowledge and stable behaviour, which was particularly noted by six participants in Cluster 1 and one participant in Cluster 2. All chose the robot in the *smart* condition. Their trust was cognitive, i.e., based on rational thinking [63].

Reliability, knowledge and stable behaviour
Nao's voice is squeaky and Nao is very fallible. Buddy's voice is much more reliable as was his behaviour. (chose Buddy)
More knowledgeable! (chose Buddy)
Seems wiser! (chose Buddy)
Buddy made more mistakes, NAO was more interesting in that it "reacted" more. (chose NAO)
NAO answered all questions. (chose NAO)
NAO. This robot replied with more correct answers, in fact, I think all were correct. (chose NAO)

(2) *Hedonic Factors*: Sensory experience and aesthetic appeal, such as expectations of human resemblance or personal (often contradicting) preferences of morphological features. This was more prominent for participants in Clusters 3 and 6 and the remaining participants in Clusters 2. Nevertheless, most participants perceived the robot as neither intelligent nor trustworthy, despite preferring it.

Expectations of appearance
I prefer how Nao looks (chose NAO)
He has a cute face and I could understand him easier. He seemed happier than NAO. (chose Buddy)
Looks more human. (chose NAO)
NAO is very cute and clever. (chose NAO)
Morphology is more similar to a human. (chose NAO)
Probably because it's more a "human" shape (chose NAO)
I chose Nao as he is cute, yet looks more like I imagine a robot should look, although it was a difficult choice to make as Buddy looks good too. (chose NAO)
Buddy is rather "creeps" – I think it's the eyes. (chose NAO)
He's a friendly little fellow! Sadly I disliked NAO the first time I met him, so I'm guilty of bias. (chose Buddy)

Table 2. Contingency Table Showing the Frequency with Which Each Robot was Chosen to Continue the Interaction to Test for Morphology Bias

	Manipulated intelligence		**Perceived intelligence**	
Robot chosen	*Smart*	*Non-smart*	*More intelligent = TRUE*	*More intelligent = FALSE*
Buddy	7	5	7	5
NAO	6	5	5	6
Total	13	10	12	12

(3) *Affective Factors*: A particularly notable finding was that participants in Clusters 4 and 5 preferred the robot with lower constructed intelligence, viewing its tendency to make and admit mistakes as more natural and human-like, which sometimes led to a personal affinity with the robot. They also chose the robot in the *non-smar*t condition. Some (3 participants) perceived this attribute as a mark of intelligence (indicated by the reported intelligence variable), and to all 5 participants, it invoked trust in the robot (reported trust). This trust was emotional, i.e., based on affect [63].

To err is more human (!)
Nao seems to be patronising. (chose Buddy)
Because he gets it wrong sometimes. (chose Buddy)
Buddy was more human-like admitting mistakes. (chose Buddy)
I think Buddy was more fun and took defeat well. (chose Buddy)
I chose NAO because (individually) I felt protective towards him. (chose NAO)

The crosstabulation results do not show a clear systematic bias favouring either robot. They also suggest that trust can be either cognitive (i.e., based on perceived competence) or affective (i.e., based on emotional connection).

Lastly, to test for morphology bias, we ran a contingency table (Table 2) based on the robot agent instead of the condition to quantify how many times each robot was chosen to continue the interaction, and whether it was designed with higher intelligence.

The Chi-square test ($\chi^2 = 0.00$, p-value = 1.0) confirmed that the observed frequencies matched exactly the expected frequencies (Buddy [6.7826, 5.2174], NAO [6.2147, 4.7826]), with the p-value of 1.0 verifying the null hypothesis of no association between the robot choice and it being in the *smart* condition. Similarly, there was no association between the robot choice and it being reported as the more intelligent robot ($\chi^2 = 0.0399$, p-value = 0.8416). Therefore, the robot preference was not confounded by the morphology, with both robots equally likely to be perceived and reported as intelligent.

4.2 Impact of Interaction on Mood

Throughout the study, participants' emotional states were monitored across three phases to explore any trends in emotional changes related to the tasks. Before the interaction, participants reported low levels of negative emotions like anger, sadness, loneliness, panic and fear, with mean scores around 1.0–1.2, indicating these feelings were not prominent. In contrast, positive emotions such as happiness (mean score of 4.17), relaxation (3.96), calmness (3.65) and enjoyment (4.30) were higher, suggesting a generally positive initial emotional state. Low levels of anxiety (2.57) and nervousness (2.13) were also noted.

After completing both tasks, there was a significant increase in positive emotions, such as happiness, enjoyment, relaxation and calmness (all means over 5), indicating improved tranquillity and well-being. Negative emotions like anger and sadness remained at their lowest rating. Notably, anxiety and nervousness decreased to mean scores of 1.43 and 1.30, respectively, showing reduced distress. Participants also showed enhancements in confidence after the interaction. Negative experiences about the interaction, like disappointment, discomfort and confusion, decreased, indicating a meaningful improvement in the interaction experience across both the cognitive game and mental well-being tasks. These changes were all statistically significant. Furthermore, only six participants did not report feeling relaxed following the mindfulness meditation, suggesting that the majority experienced a positive impact on their mental and emotional well-being from the intervention.

While a direct causal analysis between intelligence and mood was not conducted, the observed improvements in participants' mood appear to be task-related and occur for both robots. This complementary context suggests that repeated or meaningful engagements with robots positively impact older adults. It may indicate that older adults look beyond robotic fallibility and respond to more human-like qualities, such as making and admitting mistakes. In turn, successful error recovery, e.g., performing a subsequent task well, may serve as a mitigating factor. However, these interpretations remain exploratory and warrant further investigation through dedicated analyses.

5 Discussion

This study aimed to investigate how older adults perceive robots that differ in their levels of intelligence, focusing specifically on how perceived robot intelligence affects trust, attitudes, and the intention to use these robots in various contexts. Unlike most HRI studies focusing on dyadic interactions, this research uses a human multi-robot approach, where participants simultaneously engage with two social robots of differing intelligence. To our knowledge, it is the first study to mimic natural *group interactions* between multiple robots and older adults. This approach may offer a more realistic understanding of how older adults perceive and act around robots in scenarios that mirror real-world social dynamics, such as social companionship or healthcare. The within-subject design of the study also allows participants to compare directly the differently intelligent robots within the same interaction, serving as their own control. This design reduces variability between participants and potentially increases the sensitivity of the results, allowing for more precise measurements of how different robot attributes (such as intelligence) affect user perceptions.

Our analysis showed that, indeed, the perceived intelligence of the robots had a significant effect on participants' evaluations of robot attributes, their trust and their intention-to-use robots in their lives, meeting most of our expectations as designers.

While trust and intelligence appeared to be positively related, their relationship remains complex. On the one hand, trust seemed to be mediated by higher perceived intelligence. Our mixed-method analysis combining quantitative ratings and qualitative evaluation provides some converging evidence that participants' perceptions of robot intelligence primarily reflected cognitive capability rather than ease of use or likeability. For example, participants' perceptions of robots' attributes did not significantly differ across both robots, indicating no bias based on appearance or voice. Over time and tasks, the ratings for the robot with higher constructed intelligence improved significantly compared to that with lower constructed intelligence, suggesting that observed cognitive behaviours influenced participants' evolving perceptions. Similarly, trust was a significant predictor of the fit-for-use ratings, with a stronger positive relationship for the robot with higher constructed intelligence, indicating that perceived cognitive intelligence amplified trust and perceived suitability. The participants' qualitative evaluations revealed that the majority explicitly cited cognitive competence, problem-solving ability or task-related effectiveness as key factors driving their preference for the robot with higher intelligence. *Post hoc* analyses confirmed that differences in robot morphology or sensory features did not dominate or confound perceptions of intelligence. Nevertheless, trust remained strongly dependent on older adults' pre-existing attitudes and beliefs about robots. In other words, their trust will depend heavily on their initial perspectives, though interactions with robots can still impact it. This suggests that a single interaction may be insufficient to significantly alter their views and that ongoing exposure to robots may be necessary for older adults to shift their perceptions over time [64].

Our analysis of older adults' intention-to-use robots seems to support this idea, as we noticed that interacting with the robots impacted how participants viewed the robot in other related areas. Specifically, their intention-to-use robots after interacting with them amplified across multiple related areas beyond the tested contexts of entertainment, cognitive support and mental well-being, stretching to sensitive tasks. The correlations between certain contexts in which older adults would generally be willing to use robots even before interacting with them became significantly stronger after the interaction. For example, cognitive training and mental well-being were increasingly associated with medicine use (+0.22 and +0.12, respectively), with their post-interaction correlations exceeding 0.80. This pattern may be devoted to anchoring bias [62]. The findings suggest that meaningful and continuous interactions with robots have the potential to shift older adults' willingness to incorporate robotic assistance into various and increasingly sensitive aspects of their daily lives. In terms of trust, the data suggested that higher trust is necessary for more sensitive contexts, but this trust can be built and reinforced through related, positive interactions with the robots over time.

Corroborating quantitative findings with qualitative insights indicated further complex and multifaced factors affecting trust. For example, quantitative Clusters 0 and 2 showed that older adults with a general predisposition to trust robots tend to maintain high trust levels (Cluster 0), while those with low initial trust remain sceptical (Cluster 2), confirming how pre-perceptions are difficult to alter. This is to be expected given that human-related factors are found to influence how people interact with others [65]. When analysing their self-reported responses, it is notable that participants in Cluster 0, who began with high trust and further increased it, also predominantly belonged to the groups that favoured the robot's utilitarian attributes, such as reliability, stability and knowledge. However, two participants were the exception: instead, they started with general low trust, which did not improve much, despite perceiving the robot with higher constructed intelligence as the one more intelligent and also reported trusting it. This suggests that while some older adults may prefer or expect a competent robot, their ingrained human-related factors, like pre-existing beliefs, play a significant role in their attitudes towards robots: for some, accepting robots may come naturally, whereas others remain more resistant no matter their design [64, 65].

On another (thwarted) hand, participants had their own agendas for favouring or trusting the robot they chose. Despite the quantitative analysis showing a consistent effect of the robot intelligence and no measurable bias due to the robots' physical appearance, the participants' choice of robot was still majorly driven by sensory experience and aesthetic appeal, such as expectations of human-like qualities or personal, often conflicting, inclination towards the robots' physical features. Half of this category of participants preferred the competent robot, not because they perceived it as more intelligent (many did not), but due to its appearance. Interestingly, participants drawn to the robot's hedonic features aligned with those in quantitative Cluster 2, characterised by low pre- and post-interaction trust—all Cluster 2 participants reported choosing the robot for its hedonic attributes. We speculate that when trust was not established and predispositions to distrust robots were held, older adults prioritised the robot's appearance to make their choice.

The study also identified a category of participants who fell somewhere in the middle—those who initially held moderate levels of trust but were able to increase their trust after interacting with the robot (quantitative Cluster 1). This is by far the most striking insight drawn from the study. All of these participants chose to interact with the robot with lower constructed intelligence, and the majority also acknowledged that its designed intelligence was inferior to the other robot. However, their trust strengthened after the interaction, and most explicitly reported trusting the robot. These participants predominantly preferred the less intelligent robot for its affective attributes. In particular, they appeared to value human-like traits like making and admitting mistakes. The fallibility of the robot enhanced its trustworthiness for these participants, who developed a personal connection with it. The results align with previous research on HRI that underscores the importance of emotional intelligence and social presence in building and maintaining trust [22].

Overall, the intervention had a clear positive impact on participants' emotions and mood, increasing feelings of calmness and enjoyment. Most participants reported feeling relaxed after the mindfulness meditation, regardless of whether they viewed the robot as intelligent. Since both robots performed this task equally well, this may suggest that longer exposure to robots could help older adults move past initial perceptions and become more accepting of them over time, if robots can recover from fallibility. However, it is not surprising that when given a choice, participants gravitated towards the robot they preferred, as they tend to have diverse expectations and value different aspects of the robot, which was evident in their qualitative responses. Thus, this could well explain why they received the intervention positively.

In summary, the intelligence of the more competent robot, or lack thereof, was acknowledged by more than half of the participants (14). Quantitative insights revealed that intelligence impacted trust, fit-for-use and intention-to-use robots. However, only 7 out of 12 participants who interacted with the robot with higher intelligence chose to continue interacting with it in subsequent tasks because of this factor (*utilitarian attributes*: knowledgeable, reliable, stable). The remaining 16 participants fell into either of the two categories (i) those whom the robot appealed to due to its looks (*hedonic attributes:* likeability, appearance—11 participants), none of whom considered their chosen robot intelligent, despite five interacting with the more competent robot, and (ii) those who favoured the fallible robot for this perceived human-like quality (*affective attributes*: e.g., fallibility—five participants), with three considering this as a sign of intelligence. For most participants, the intervention had a clear positive impact. These results collectively highlight the complexity of older adults' interactions with robots and the challenge of navigating the balance between the different factors affecting their acceptance and ongoing engagement with such robots.

Although our sample size was moderate, the within-subject design strengthens the validity of the study's insights. Each of our 23 older adult participants interacted with two robots and rated both after the same interaction, generating two sets of data per participant and resulting in a total of 46 data points. This approach is particularly valuable given the challenges of enrolling older adults or other vulnerable groups for such studies. Our findings could serve as a foundation for future research with larger populations, to gather more robust data and draw conclusive evidence that generalises to the wider older adult demographic. Moreover, although our results revealed evident differences, a single interaction may not be sufficient to fully capture shifts in participants' attitudes and trust over time, as indicated, for instance, by the lack of time effects. Thus, longer or repeated exposure of older adults to robots might be necessary for these differences to emerge over time, particularly in more sensitive contexts. For example, longitudinal studies that track changes in attitudes and trust over extended periods would provide valuable data on the sustainability of the positive impacts observed in short-term interventions [66]. This may also help mitigate any risks of rejection at first sight or the novelty of the interaction wearing off [65]. Lastly, it is important to replicate natural and spontaneous interactions between robots and older adults away from controlled lab environments, for example, in participants' own homes or healthcare settings, which could lead to results that better reflect real-world HRIs.

Acknowledgement

The authors extend their gratitude to Plymouth Community Homes for their assistance in recruiting participants for the pilot study and to the participants for generously sharing their time and insights.

References

[1] O. E Lee, H. Lee, A. Park, and N. G. Choi. 2022. My precious friend: Human-robot interactions in home care for socially isolated older adults. *Clinical Gerontologist* 47, 1 (2022), 161–170. DOI: https://doi.org/10.1080/07317115.2022.2156829

[2] Norina Gasteiger, Ho Seok Ahn, Chiara Gasteiger, Christopher Lee, Jongyoon Lim, Christine Fok, Bruce A. Macdonald, Geon Ha Kim, and Elizabeth Broadbent. 2021. Robot-delivered cognitive stimulation games for older adults: Usability and acceptability evaluation. *ACM Transactions on Human-Robot Interaction*, 10, 4 (2021), 1–18. DOI: https://doi.org/10.1145/3451882

[3] K. Zsiga, G. Edelmayer, P. Rumeau, O. Péter, A. Tóth, and G. Fazekas. 2013. Home care robot for socially supporting the elderly: Focus group studies in three European countries to screen user attitudes and requirements. *International Journal of Rehabilitation Research. Internationale Zeitschrift Fur Rehabilitationsforschung. Revue Internationale de Recherches de Readaptation* 36, 4 (2013), 375–378. DOI: https://doi.org/10.1097/MRR.0b013e3283643d26

[4] C.-A. Smarr, A. Prakash, J. M. Beer, T. L. Mitzner, C. C. Kemp, and W. A. Rogers. 2012. Older adults' preferences for and acceptance of robot assistance for everyday living tasks. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 56, 1 (2012), 153–157. DOI: https://doi.org/10.1177/1071181312561009

[5] L. Pu, W. Moyle, C. Jones, and M. Todorovic. 2019. The effectiveness of social robots for older adults: A systematic review and meta-analysis of randomized controlled studies. *The Gerontologist* 59, 1 (2019), e37–e51. DOI: https://doi.org/10.1093/geront/gny046

[6] A. S. Gessl, S. Schlögl, and N. Mevenkamp. 2019. On the perceptions and acceptance of artificially intelligent robotics and the psychology of the future elderly. *Behaviour & Information Technology* 38, 11 (2019), 1068–1087. DOI: https://doi.org/10.1080/0144929X.2019.1566499

[7] H. L. Bradwell, K. J. Edwards, R. Winnington, S. Thill, and R. B. Jones. 2019. Companion robots for older people: Importance of user-centred design demonstrated through observations and focus groups comparing preferences of older people and roboticists in South West England. *BMJ Open* 9, 9 (2019), e032468. DOI: https://doi.org/10.1136/bmjopen-2019-032468

[8] M. M. De Graaf and S. B. Allouch. 2013. Exploring influencing variables for the acceptance of social robots. *Robotics and Autonomous Systems* 61, 12 (2013), 1476–1486. DOI: https://doi.org/10.1016/j.robot.2013.07.007

[9] M. M. de Graaf, S. Ben Allouch, and J. A. Van Dijk. 2015. What makes robots social?: A user's perspective on characteristics for social human-robot interaction. In *Social Robotics: 7th International Conference (ICSR '15)*. Springer International Publishing, 184–193. DOI: https://doi.org/10.1007/978-3-319-25554-5_19.

[10] E. Broadbent, R. Stafford, and B. MacDonald. 2009. Acceptance of healthcare robots for the older population: Review and future directions. *International Journal of Social Robotics* 1 (2009), 319–330. DOI: https://doi.org/10.1007/s12369-009-0030-6

[11] D. S. Syrdal, K. Dautenhahn, S. N. Woods, M. L. Walters, and K. L. Koay. 2007. Looking good? Appearance preferences and robot personality inferences at zero acquaintance. In *AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics*, Vol. 86, 230–234.

[12] K. Dautenhahn. 2007. Socially intelligent robots: Dimensions of human–robot interaction. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 362, 1480 (2007), 679–704. DOI: https://doi.org/10.1098/rstb.2006.2004

[13] T. B. N. Da Silveira and H. S. Lopes. 2023. Intelligence across humans and machines: A joint perspective. *Frontiers in Psychology* 14 (2023), 1209761. DOI: https://doi.org/10.3389/fpsyg.2023.1209761

[14] G. Cortellessa, R. De Benedictis, F. Fracasso, A. Orlandini, A. Umbrico, and A. Cesta. 2021. AI and robotics to help older adults: Revisiting projects in search of lessons learned. *Paladyn, Journal of Behavioral Robotics* 12, 1 (2021), 356–378. DOI: https://doi.org/10.1515/pjbr-2021-0025

[15] L. Christoforakos, A. Gallucci, T. Surmava-Große, D. Ullrich, and S. Diefenbach. 2021. Can robots earn our trust the same way humans do? A systematic exploration of competence, warmth, and anthropomorphism as determinants of trust development in HRI. *Frontiers in Robotics and AI* 8 (2021), 640444. DOI: https://doi.org/10.3389/frobt.2021.640444

[16] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. Chen, E. J. De Visser, and R. Parasuraman. 2011. A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors* 53, 5 (2011), 517–527. DOI: https://doi.org/10.1177/0018720811417254

[17] C. Esterwood and L. P. RobertJr., 2023. Three strikes and you are out!: The impacts of multiple human–robot trust violations and repairs on robot trustworthiness. *Computers in Human Behavior* 142 (2023), 107658. DOI: https://doi.org/10.1016/j.chb.2023.107658

[18] R. A. Søraa, G. Tøndel, M. W. Kharas, and J. A. Serrano. 2023. What do older adults want from social robots? A qualitative research approach to human-robot interaction (HRI) studies. *International Journal of Social Robotics* 15, 3 (2023), 411–424. DOI: https://doi.org/10.1007/s12369-022-00914-w

[19] M. Heerink, B. Kröse, V. Evers, and B. Wielinga. 2010. Assessing acceptance of assistive social agent technology by older adults: The Almere model. *International Journal of Social Robotics* 2 (2010), 361–375. DOI: https://doi.org/10.1007/s12369-010-0068-5

[20] C. D. Kidd and C. Breazeal. 2008. Robots at home: Understanding long-term human-robot interaction. In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 3230–3235. DOI: https://doi.org/10.1109/IROS.2008.4651113.

[21] I. Leite, A. Pereira, G. Castellano, S. Mascarenhas, C. Martinho, and A. Paiva. 2012. Modelling empathy in social robotic companions. In *Advances in User Modeling: UMAP 2011 Workshops*. Lecture Notes in Computer Science, Vol. 7138, Springer, Berlin, Heidelberg, 135–147. DOI: https://doi.org/10.1007/978-3-642-28509-7_14

[22] N. Chen, X. Liu, Y. Zhai, and X. Hu. 2023. Development and validation of a robot social presence measurement dimension scale. *Scientific Reports* 13, 1 (2023), 2911. DOI: https://doi.org/10.1038/s41598-023-28817-4

[23] C. Bartneck, T. Kanda, O. Mubin, and A. Al Mahmud. 2009. Does the design of a robot influence its animacy and perceived intelligence? *International Journal of Social Robotics* 1 (2009), 195–204. DOI: https://doi.org/10.1007/s12369-009-0013-7

[24] M. M. De Graaf, S. B. Allouch, and T. Klamer. 2015. Sharing a life with Harvey: Exploring the acceptance of and relationship-building with a social robot. *Computers in Human Behavior* 43 (2015), 1–14. DOI: https://doi.org/10.1016/j.chb.2014.10.030

[25] Toshiaki Nishio, Yuichiro Yoshikawa, Kazuki Sakai, Takamasa Iio, Mariko Chiba, Taichi Asami, Yoshinori Isoda, and Hiroshi Ishiguro. 2021. The effects of physically embodied multiple conversation robots on the elderly. *Frontiers in Robotics and AI* 8 (2021), 633045. DOI: https://doi.org/10.3389/frobt.2021.633045

[26] A. Dahiya, A. M. Aroyo, K. Dautenhahn, and S. L. Smith. 2023. A survey of multi-agent human–robot interaction systems. *Robotics and Autonomous Systems* 161 (2023), 104335. DOI: https://doi.org/10.1016/j.robot.2022.104335

[27] A. Apa. 2022. Dictionary of Psychology. Retrieved September 2024 from https://dictionary.apa.org/social-interactions

[28] C. L. Gittens. 2024. Remote HRI: A methodology for maintaining COVID-19 physical distancing and human interaction requirements in HRI studies. *Information Systems Frontiers: A Journal of Research and Innovation* 26, 1 (2024), 91–106. DOI: https://doi.org/10.1007/s10796-021-10162-4

[29] G. Podevijn, R. O'grady, N. Mathews, A. Gilles, C. Fantini-Hauwel, and M. Dorigo. 2016. Investigating the effect of increasing robot group sizes on the human psychophysiological state in the context of human–swarm interaction. *Swarm Intelligence* 10 (2016), 193–210. DOI: https://doi.org/10.1007/s11721-016-0124-3

[30] J. Wang and M. Lewis. 2008. Assessing cooperation in human control of heterogeneous robots. In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction*, 9–16. DOI: https://doi.org/10.1145/1349822.1349825

[31] B. Sellner, F. W. Heger, L. M. Hiatt, R. Simmons, and S. Singh. 2006. Coordinated multiagent teams and sliding autonomy for large-scale assembly. *Proceedings of the IEEE* 94, 7 (2006), 1425–1444. DOI: https://doi.org/10.1109/JPROC.2006.876966

[32] A. Kolling, P. Walker, N. Chakraborty, K. Sycara, and M. Lewis. 2015. Human interaction with robot swarms: A survey. *IEEE Transactions on Human-Machine Systems* 46, 1 (2015), 9–26. DOI: https://doi.org/10.1109/THMS.2015.2480801

[33] H. Yedidsion, J. Deans, C. Sheehan, M. Chillara, J. Hart, P. Stone, and R. J. Mooney. 2019. Optimal use of verbal instructions for multi-robot human navigation guidance. In *Social Robotics: 11th International Conference (ICSR '19)*, Springer International Publishing, 133–143. DOI: https://doi.org/10.1007/978-3-030-35888-4_13

[34] X. Z. Tan, S. Reig, E. J. Carter, and A. Steinfeld. 2019. From one to another: How robot-robot interaction affects users' perceptions following a transition between robots. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 114–122. DOI: https://doi.org/10.1109/HRI.2019.8673304.

[35] P. Khandelwal and P. Stone. 2014. Leading the way: An efficient multi-robot guidance system. In *2014 AAAI Fall Symposium Series*.

[36] J. Penders, L. Alboul, U. Witkowski, A. Naghsh, J. Saez-Pons, S. Herbrechtsmeier, and M. El-Habbal. 2011. A robot swarm assisting a human fire-fighter. *Advanced Robotics* 25, 1–2 (2011), 93–117. DOI: https://doi.org/10.1163/016918610X538507

[37] J. Saez-Pons, L. Alboul, and J. Penders. 2011. Experiments in cooperative human multi-robot navigation. In *2011 IEEE International Conference on Robotics and Automation*. IEEE, 1–4. DOI: https://doi.org/10.1109/ICRA.2011.5980592.

[38] I. Leite, M. McCoy, M. Lohani, D. Ullman, N. Salomons, C. Stokes, S. Rivers, and B. Scassellati. 2015. Emotional storytelling in the classroom: Individual versus group interaction between children and robots. In *Proceedings of the*

10th Annual ACM/IEEE International Conference on Human-Robot Interaction, 75–82. DOI: https://doi.org/10.1145/2696454.2696481
[39] J. Swaminathan, J. Akintoye, M. R. Fraune, and H. Knight. 2021. Robots that run their own human experiments: Exploring relational humor with multi-robot comedy. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. IEEE, 1262–1268. DOI: https://doi.org/10.1109/RO-MAN50785.2021.9515324.
[40] R. Li, M. A. Oskoei, and H. Hu. 2013. Towards ROS-based multi-robot architecture for ambient assisted living. In *2013 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 3458–3463. DOI: https://doi.org/10.1109/SMC.2013.590.
[41] P. Benavidez, M. Kumar, S. Agaian, and M. Jamshidi. 2015. Design of a home multi-robot system for the elderly and disabled. In *2015 10th System of Systems Engineering Conference (SoSE)*. IEEE, 392–397. DOI: https://doi.org/10.1109/SYSOSE.2015.7151907
[42] V. Villani, B. Capelli, C. Secchi, C. Fantuzzi, and L. Sabattini. 2020. Humans interacting with multi-robot systems: A natural affect-based approach. *Autonomous Robots* 44 (2020), 601–616. DOI: https://doi.org/10.1007/s10514-019-09889-6
[43] Ramón Barber, Francisco J. Ortiz, Santiago Garrido, Francisco M. Calatrava-Nicolás, Alicia Mora, Adrián Prados, José Alfonso Vera-Repullo, Joaquín Roca-González, Inmaculada Méndez, and Óscar Martínez Mozos. 2022. A multirobot system in an assisted home environment to support the elderly in their daily lives. *Sensors* 22, 20 (2022), 7983. DOI: https://doi.org/10.3390/s22207983
[44] C. Lytridis, C. I. Papadopoulou, G. A. Papakostas, V. G. Kaburlasos, V. A. Nikopoulou, M. D. Kerasidou, and N. Dalivigkas. 2020. Robot-assisted autism spectrum disorder (ASD) interventions: A multi-robot approach. In *2020 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*. IEEE, 1–4. DOI: https://doi.org/10.23919/SoftCOM50211.2020.9238273.
[45] N. Efthymiou, P. P. Filntisis, P. Koutras, A. Tsiami, J. Hadfield, G. Potamianos, and P. Maragos. 2022. Childbot: Multi-robot perception and interaction with children. *Robotics and Autonomous Systems* 150 (2022), 103975. DOI: https://doi.org/10.1016/j.robot.2021.103975
[46] Sung-En Chien, Li Chu, Hsing-Hao Lee, Chien-Chun Yang, Fo-Hui Lin, Pei-Ling Yang, Te-Mei Wang, and Su-Ling Yeh. 2019. Age difference in perceived ease of use, curiosity, and implicit negative attitude toward robots. *ACM Transactions on Human-Robot Interaction*, 8, 2 (2019), 1–19. DOI: https://doi.org/10.1145/3311788
[47] Y. L. Wang and C. W. Lo. 2025. The effects of response time on older and young adults' interaction experience with Chatbot. *BMC Psychology* 13, 1 (2025), 150. DOI: https://doi.org/10.1186/s40359-025-02459-9
[48] I. Giorgi, F. A. Tirotto, O. Hagen, F. Aider, M. Gianni, M. Palomino, and G. L. Masala. 2022. Friendly but faulty: A pilot study on the perceived trust of older adults in a social robot. *IEEE Access* 10 (2022), 92084–92096. DOI: https://doi.org/10.1109/ACCESS.2022.3202942
[49] N. Caruana, R. Moffat, A. Miguel-Blanco, and E. S. Cross. 2023. Perceptions of intelligence & sentience shape children's interactions with robot reading companions. *Scientific Reports* 13, 1 (2023), 7341. DOI: https://doi.org/10.1038/s41598-023-32104-7
[50] A. Müller and A. Richert. 2024. Egocentric robots in a human-centric world? Exploring group-robot-interaction in public spaces. arXiv:240718009. DOI: https://doi.org/10.48550/arXiv.2407.18009
[51] M. J. Kim, A. Getu, H. Sharp, and E. Wiese. 2021. Advice from robots: Would you choose a robot that looked more or less human? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 65, 1 (2021), 117–121. DOI: https://doi.org/10.1177/1071181321651121
[52] NAO the humanoid and programmable robot, Aldebaran Robotics. Retrieved June 2025 from https://www.aldebaran.com/en/nao
[53] Buddy, The First Smart, Mobile and Emotional Robot, Blue Frog Robotics. Retrieved June 2025 from https://www.bluefrogrobotics.com/buddy-en
[54] T. Vandemeulebroucke, K. Dzi, and C. Gastmans. 2021. Older adults' experiences with and perceptions of the use of socially assistive robots in aged care: A systematic review of quantitative evidence. *Archives of Gerontology and Geriatrics* 95 (2021), 104399. DOI: https://doi.org/10.1016/j.archger.2021.104399
[55] A. Abubshait and E. Wiese. 2017. You look human, but act like a machine: Agent appearance and behavior modulate different aspects of human–robot interaction. *Frontiers in Psychology* 8 (2017), 1393. DOI: https://doi.org/10.3389/fpsyg.2017.01393
[56] NHS. Mindfulness. Retrieved October 2025, from https://www.nhs.uk/mental-health/self-help/tips-and-support/mindfulness/
[57] C. Bartneck. 2023. Godspeed questionnaire series: Translations and usage. In *International Handbook of Behavioral Health Assessment*. Springer International Publishing, Cham, 1–35.
[58] M. Körber. 2019. Theoretical considerations and development of a questionnaire to measure trust in automation. In *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018) Volume VI: Transport Ergonomics and Human Factors (TEHF), Aerospace Human Factors and Ergonomics 20*, Springer International Publishing, 13–30. DOI: https://doi.org/10.1007/978-3-319-96074-6_2

[59] N. Marangunić and A. Granić. 2015. Technology acceptance model: A literature review from 1986 to 2013. *Universal Access in the Information Society* 14 (2015), 81–95. DOI: https://doi.org/10.1007/s10209-014-0348-1

[60] R. H. Fazio and M. A. Olson. 2003. Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology* 54, 1 (2003), 297–327. DOI: https://doi.org/10.1146/annurev.psych.54.101601.145225

[61] N. Surdel, Y. E. Bigman, X. Shen, W. Y. Lee, M. F. Jung, and M. J. Ferguson. 2024. Judging robot ability: How people form implicit and explicit impressions of robot competence. *Journal of Experimental Psychology: General* 153, 5 (2024), 1309.

[62] A. Furnham and H. C. Boo. 2011. A literature review of the anchoring effect. *The Journal of Socio-Economics* 40, 1 (2011), 35–42. DOI: https://doi.org/10.1016/j.socec.2010.10.008

[63] E. Glikson and A. W. Woolley. 2020. Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals* 14, 2 (2020), 627–660.

[64] C. Esterwood, K. Essenmacher, H. Yang, F. Zeng, and L. P. Robert. 2021. A meta-analysis of human personality and robot acceptance in human-robot interaction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–18. DOI: https://doi.org/10.1145/3411764.3445542

[65] M. Paetzel, G. Perugia, and G. Castellano. 2020. The persistence of first impressions: the effect of repeated interactions on the perception of a social robot. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 73–82. DOI: https://doi.org/10.1145/3319502.3374786

[66] Y. C. Chen, S. L. Yeh, W. Lin, H. P. Yueh, and L. C. Fu. 2023. The effects of social presence and familiarity on children–robot interactions. *Sensors* 23, 9 (2023), 4231. DOI: https://doi.org/10.3390/s23094231

Received 13 September 2024; revised 24 June 2025; accepted 12 September 2025