Exploring Human-Robot Trust Through the Investment Game: An Immersive Space Mission Scenario

Emy Arts*
Universität Hamburg
Hamburg, Germany
8arts@informatik.uni-hamburg.de

Sebastian Zörner*
Universität Hamburg
Hamburg, Germany
8zoerner@informatik.uni-hamburg.de

Kavish Bhatia
Universität Hamburg
Hamburg, Germany
8bhatia@informatik.uni-hamburg.de

Glareh Mir
Universität Hamburg
Hamburg, Germany
8gmir@informatik.uni-hamburg.de

Florian Schmalzl
Universität Hamburg
Hamburg, Germany
6schmalz@informatik.uni-hamburg.de

Ankit Srivastava
Universität Hamburg
Hamburg, Germany
8srivast@informatik.uni-hamburg.de

Brenda Vasiljevic
Universität Hamburg
Hamburg, Germany
8vasiljevic@informatik.uni-hamburg.de

Tayfun Alpay
Universität Hamburg
Hamburg, Germany
alpay@informatik.uni-hamburg.de

Annika Peters
Universität Hamburg
Hamburg, Germany
apeters@informatik.uni-hamburg.de

Erik Strahl
Universität Hamburg
Hamburg, Germany
strahl@informatik.uni-hamburg.de

Stefan Wermter
Universität Hamburg
Hamburg, Germany
wermter@informatik.uni-hamburg.de

ABSTRACT

As robots become more advanced and capable, developing trust is an important factor of human-robot interaction and cooperation. However, as multiple environmental and social factors can influence trust, it is important to develop more elaborate scenarios and methods to measure human-robot trust. A widely used measurement of trust in social science is the *investment game*. In this study, we propose a scaled-up, immersive, science fiction Human-Robot Interaction (HRI) scenario for intrinsic motivation on human-robot collaboration, built upon the investment game and aimed at adapting the investment game for human-robot trust. For this purpose, we utilise two Neuro-Inspired Companion (NICO) - robots and a projected scenery. We investigate the applicability of our space mission experiment design to measure trust and the impact of non-verbal communication. We observe a correlation of 0.43 ($p = 0.02$) between self-assessed trust and trust measured from the game and a positive impact of non-verbal communication on trust ($p = 0.0008$) and robot perception for anthropomorphism ($p = 0.007$) and animacy ($p = 0.00002$). We conclude that our scenario is an appropriate method to measure trust in human-robot interaction and also to study how non-verbal communication influences a human's trust in robots.

*Authors contributed equally to this research. Remaining authors in alphabetical order.

HAI '20, November 10–13, 2020, Sydney, Australia

ACM ISBN 978-1-4503-8054-6/20/11.
https://doi.org/10.1145/3406499.3415078

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; *HCI design and evaluation methods*; **Empirical studies in HCI**;

KEYWORDS

human-robot interaction, human-robot trust, investment game, non-verbal communication

ACM Reference Format:
Emy Arts, Sebastian Zörner, Kavish Bhatia, Glareh Mir, Florian Schmalzl, Ankit Srivastava, Brenda Vasiljevic, Tayfun Alpay, Annika Peters, Erik Strahl, and Stefan Wermter. 2020. Exploring Human-Robot Trust Through the Investment Game: An Immersive Space Mission Scenario. In *Proceedings of the 8th International Conference on Human-Agent Interaction (HAI '20), November 10–13, 2020, Virtual Event, NSW, Australia.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3406499.3415078

1 INTRODUCTION

As robot capabilities become more and more sophisticated, we not only want them to solve increasingly complex tasks independently but ultimately aid humans in their day-to-day life. Moreover, such social robots should act in a way that is reliable, transparent, and builds trust in their capabilities as well as their intentions [17]. As soon as humans and robots autonomously work in a team on collaborative tasks, trust becomes essential for effective human-robot interaction [11]. This shows the need for a deeper understanding of what makes us willing to cooperate with robots and which factors enhance or destroy trust during interactions. We approach this topic by adopting the investment game [5], a widely used experiment to measure trust in *human-human* collaboration. In the investment game, trust is measured as the amount of money a person is willing to give to an anonymous counterpart, in the prospect of a future profit. While others have used it in an HRI setting, some report limitations and differences when applying it to human-robot collaboration (which we elaborate on in Section 2). We therefore adapt the original investment game towards a persuasive HRI cooperative scenario by scaling up both the robotic agent as well as the environment. With scaling up we allude to the progression towards a human-like interaction: a realistic cooperative scenario as opposed to an abstract exchange of money. We do this by removing the ability of the participant to make choices based on domain knowledge and introducing a plausible currency for both humans as well as robotic agents, along with a weighted choice between two trustees. The result is an HRI scenario, concealed as a futuristic, immersive space ship adventure containing multiple rounds of the investment game for participants to develop intrinsic motivation to collaborate with the robots. In this scenario, we utilise two fully autonomous Neuro-Inspired Companion (NICO) [26] humanoid robots that advise the participant (who acts as the ship's commander), a voice-controlled ship AI that guides the experiment, and a large curved projector screen that simulates the inside of the ship's cockpit with an interactive video feed (see Figure 1). During the experiment, participants encounter four different challenges (e.g., engine failures, impending asteroids) where they can question the two robots, that always provide two diverging solutions. Subsequently, the participants are asked to make a choice by distributing the ship's resources between the two robots and themselves, which we evaluate as a quantitative measurement for trust. The immersive setup allows us to control the emergence, destruction, and reconstruction of trust in the robotic companions throughout the game. To improve how the robot is perceived by the participant and to ensure an experience which results in a more human-like interaction, we add non-verbal cues to our robots such as eye gaze towards the participants, facial expressions and gestures (see Section 3.2.3 for details). Such features of non-verbal communication (NVC), generally defined as 'unspoken dialogue' [10], have previously been shown to account for over 60% of the meaning in communication for human interactions [46], as they allow us to communicate mental states such as thoughts and feelings [1]. They are also thought to play an important role in human-*robot* interaction, as the implicit, robotic, non-verbal communication improves the efficiency and transparency of the interaction, leading to better cooperation between human subjects and robots [8]. As non-verbal communication is essential to both human-human and human-robot trust [10, 14], we strive to measure the effect of NVCs in our HRI scenario to assess how well it simulates a natural interaction. Therefore, we utilise our novel investment game scenario to investigate two research questions related to both evaluating trust as well as the impact of NVCs on trust:

(1) Does our variant of the investment game provide a reliable measurement for human-robot trust?
(2) Does non-verbal Communication (NVC) affect human-robot trust positively?

After surveying the latest research on measuring trust in human-robot interaction and its shortcomings (Chapter 2) we describe our approach (Chapter 3) and introduce an empirical study to evaluate our hypotheses (Chapter 4). We discuss the results as well as the limitations of this study (Chapter 5) and conclude our findings (Chapter 6) with an outlook on further research.

Figure 1: The experimental setup. On the table there are three compartments, with the one in the front (closest to the participant) containing the total amount of 7 energy cells to distribute.

2 TRUST AND THE INVESTMENT GAME

One of the biggest challenges in human-robot interaction is to develop a more natural relationship with robots. Previous research shows that people refrain from accepting, tolerating, and using robotic agents in everyday tasks, mainly because robots still appear like intruders [53]. A survey by the institute DemoSCOPE (N = 1007) has found that while 50% would accept information from a robot, only 16% would be willing to work in a team with one [2]. A considerable portion of the general population still fears robots and artificial intelligence, caused by a range of concerns about the negative impact on interpersonal relationships and potential job displacement [19, 31]. This begs the question of what could aid in easing humans into collaboration with a robot. As robots become more advanced and take greater responsibility in social jobs such as in the education sector [22, 28, 38] or in the healthcare industry [32, 37], this requires humans to be able to trust them. Whereas human-human trust has been extensively studied, human-robot trust poses new and complex research challenges. Prominent factors that influence trust in a robot are robot performance and characteristics [6, 23], and the timing and magnitude of robot errors [42, 43]. When it comes to trust, the prediction and predictability of behaviour are fundamental [52]. Constructs such as emotional empathy, shared attention, and mental perspective-taking are essential to understand, recognise, and predict human behaviour, as well as adhere to people's expectations of appropriate behaviour given circumstances [9]. This behavioural prediction is also transferred to human-robot trust [52], as humans require building a mental model, thus anthropomorphising the machine. During the first encounter, humans tend to apply social norms to robots just as they do to humans [41]. In contrast to this however stands the "uncanny valley" phenomenon: when a robot exhibits aesthetic characteristics too similar to a human, this can subtly alter trusting behaviour negatively [33]. To quantitatively measure human-human trust, previous work relies heavily on the *investment game* (also referred to as the *trust game*) [5], an economic experiment derived from game theory. Berg et al. introduced the investment game in 1995, where a subject (trustor) invests money in a counterpart. They can decide which fraction p of their monetary resource will be sent, that is then multiplied by a predetermined factor of three. The receiving person (trustee) is free to keep the whole of the tripled amount or can opt to send a fraction q of the received sum back to the trustor. The trust is quantitatively measured as the amount of money invested by the trustor in the trustee. As trust games have been established to measure trust between humans, some researchers have also used these games to empirically measure trust between humans and robots, to varying degrees of success. The amount invested by the trustor turns out to reflect a mixture of the generalised trust (a stable individual characteristic) and the specific trust towards the trustee [21]. So while most studies kept the original setup, some extended the environment towards a virtual reality setup [21], settings with multiple robots [18, 54] or switched the roles so that the human becomes the trustee dependant on the robot's willingness to invest [47]. Other variants such as the *Give-Some Game* slightly change the rules towards an economic analogue of the prisoner's dilemma [13, 14]. A different approach [20] specifically fosters participants to get to know each other before the experiment, instead of the double-blind procedure originally proposed, thereby opening up possibilities to study the influence of social interaction on trust [24]. As previously mentioned, in every social interaction involving trust, predictability is essential. This is where non-verbal communication (NVC) plays a major role [14]. Various studies show supportive evidence that implicit robotic non-verbal communication improves the efficiency and transparency of interaction [8] and report increased measures of trustworthiness when displaying non-verbal cues [13]. Robotic arm gestures have been shown to reinforce anthropomorphism, liveliness and sympathy [44, 45] — regardless of gesture congruency [46]. In fact, a lack of social cues of a robot may cause the participant to employ unwanted *testing* behaviour where they try to outwit the machine [36]. A lot of research has gone into the study of non-verbal communication via the investment game in human-agent interaction [16, 21, 24, 36, 53, 54]. However, only a few of them have used robots that can be considered anthropomorphic and humanoid, which leaves doubt to whether the trust measured is comparable to human-human trust. As a matter of fact, to the best of our knowledge, there has not yet been any research definitively confirming whether the investment game is indeed suitable for measuring human-robot trust. While it is a valid, established trust measuring experiment, the original version lacks certain features to make it suitable for a human-robot interaction scenario: a plausible currency for both humans as well as robotic agents and a human-like interaction without the possibility to make choices based on domain knowledge. The current work addresses this gap and aims to create a scenario that provides these features under which trust in robots can be built and destroyed, in order to clearly measure the correlation between the trust experienced by a human, and the trust that is displayed in the trust game.

Figure 2: The Neuro-Inspired COmpanion (NICO)

3 HRI SCENARIO DESIGN

3.1 An Immersive Extension of the Investment Game

We base our study design around a variant of the investment game, in which two robotic counsellors compete for investments from the human participant. However, in contrast to previous competitive variants [21], our design allows the human subject to allocate their investment proportionally between the two robots and themselves. Motivated by the goal to avoid prior experience in the game as an influence for player investments, we deliberately exaggerate the design of our game scenario: in our space mission, the participants impersonate the commander of a space ship with the task to deliver important cargo to a distant planet. For this mission, they are accompanied by two robotic officers. Throughout their journey through outer space, they encounter challenges such as asteroid fields and ship malfunctions that require immediate intervention and collaborative solutions. The robotic officers counsel the participant by proposing actions to take, and the participant decides by allocating energy resources towards the actions. The robotic officers however provide contrary solutions to solve the challenge. Further more, their advice is designed to be incomprehensible technical jargon, leaving the participant with no other choice than basing their decision on the impression of the officer's persona. The scenario involves four different instances of such events which represent different scenes of the experiment. By design, their first investment (regardless of how it is distributed) is unsuccessful and on each of the following investments they receive positive feedback instead. This enables us to observe the effects of building and destroying trust. Our scenario setup entails two important requirements: i) making the participant reliant on the robots' expertise to foster cooperation, and ii) ensuring that the invested currency has an inherent value to both the participant and the robots. We achieve the former by designing a challenging scenario setting of a space journey: all participants have negligible expertise regarding space travel – the robotic officers, however, are introduced as specifically designed to advise in interstellar travel, thus should be perceived as more knowledgeable in the subject matter. This allows us to circumvent participants making decisions based on their previous experiences, leaving the participant primarily reliant on the robots' advice. Our second requirement is to employ a currency that is considered valuable for both the trustor and the trustee. As we anticipate that participants do not perceive money as valuable currency for robotic agents, we adapt the currency with fictional *energy cells*, represented by cubes. These energy cells have a value to the player as they function as a resource that can provide the ship's engine with the extra power to reach the destination planet faster. On the other hand, the robotic officers require such energy to execute their solutions, ensuring safety on the journey. We thereby create a currency that is perceived as valuable to both trustor and trustee. Lastly, the participants can proportionally choose how much they invest, i.e., they can arbitrarily distribute their energy cells between both robots and themselves. However, as 7 cells are provided in total, participants are unable to distribute all energy cells evenly among the 3 options (officer A, officer B, ship engine), effectively forcing them to voice a preference. These three aspects - the inability of the participants to make choices based on domain knowledge, a shared currency between human trustor and robot trustee, and the weighted choice between two agents - allow us to meet the above-mentioned requirements for a suitable human-robot interaction scenario.

3.2 Experimental Setup

An overview of the experimental setup can be seen in Figure 1. The participant is seated in the cockpit of the ship (depicted by the interactive video feed), containing the two robots and a table where they can distribute the energy cells. One of the main goals of our design is to achieve an immersive and enjoyable experience for the participants. Besides concealing our research question, our scenario needs to establish enough involvement to allow trust-building towards the robots. For this purpose, we developed a fully autonomous system that only requires the experimenter's intervention in case of larger failures such as speech recognition errors. Through a state machine implemented in ROS [50], the following components are interconnected and synchronised:

3.2.1 The environment. The environment mainly consists of four projectors aimed at a curved canvas in front of the participant [4]. This provides the scenery for the science fiction scenario by displaying images and video, simulating the inside view of a space ship cockpit. The canvas shows the journey through the galaxy with transition videos between scenes and provides visual feedback as a warning for the problems which the participant faces during the mission. Loudspeakers behind the canvas are used for the ship's voice and special sound effects such as engine noise and alarm sounds. Turquoise ambient lighting and dry ice fog create an atmospheric environment throughout the game, while red lights are used occasionally to indicate potentially dangerous encounters.

3.2.2 The robots. The two robot officers, non-descriptively named *732-A* and *732-B*, are located at a maximum angular distance to each other and the participant. We chose their names to be as neutral and unrelated to any prior experience of participants as possible. We utilise NICO (Neuro-Inspired COmpanion) [26, 27], an open-source social robotics platform for humanoid robots (see Figure 2) designed by the *Knowledge Technology* group at the *University of Hamburg*. NICO is a child-sized humanoid robot that has a range of programmable capabilities, accessible and customise-able through the Robot Operating System (ROS) [40]. It has 10 degrees-of-freedom in the torso (head and arms) and 22 degrees-of-freedom in the hands (under-actuated, 8 motors) with additional joints for fingers, which allows for fine-grained gestures and body language. It is also capable of displaying a range of facial expressions through LED matrices in its eyebrows and mouth. Both of these robots were further integrated with loudspeakers in their torsos to produce enhanced speech.

3.2.3 Non-verbal communication. We equipped one of the robots with a set of non-verbal cues that show evidence to improve the transparency of the interaction and reinforce the spoken word [9]. These cues are: gaze direction via head movements towards the participant and the other robot, 4 different facial expressions (happiness, sadness, surprise, anger), as well as gestures towards the participant such as pointing, saluting or beat gestures. The other robot adheres to a *minimal* set of neutral, alternating head and arm movements to keep the illusion of life [36], such as looking down at the allocated energy cells and turning their head towards the speaker. We alternate the condition between participants in order to control for potential biases.

3.2.4 The vision system. An RGB-camera is placed behind the participant to track the movements of the robots and the participant. Another RGB-camera is placed on top of the canvas to track participant expressions and movements. These cameras are used by the experimenter to observe the participant and to monitor the experiment. On the table, in front of the players, there are three heptagonal-shaped compartments holding the energy cells. All compartments have seven quadratic markers on which the energy cells must be placed for successful allocation. At the beginning of the game, all seven energy cells are placed in the commander's compartment. An additional RGB-camera is mounted on top of the commanding table near the ceiling to count and track energy cubes allocation and de-allocation from the robot compartments. A picture of the commanding table taken by this camera can be seen in Figure 3. Object detection is used to handle the energy cell counting during allocation as well as to confirm when the robot's compartments are empty before proceeding to the scene. After a request from the state machine, the object detection algorithm processes an image (taken from the RGB-camera mounted on top of the commanding table) using the *OpenCV* library [7], to detect the number of energy cells allocated to each heptagon-shaped compartment.

3.2.5 The speech systems. Interactive dialogue via spoken words is a cornerstone to enable natural human-like human-robot interaction [29, 49]. We therefore built the space ship AI named *Wendigo* as a closed dialogue manager utilising the SMACH[1] state management library, the Automatic Speech Recognition system DOCKS2 developed by Twiefel and Möller [51], and the Amazon Polly[2] Speech Synthesis system. The participants can directly interact with *Wendigo* and the robotic officers via a microphone located in the middle

[1]http://wiki.ros.org/smach (accessed 19/06/2020)
[2]https://aws.amazon.com/polly/ (accessed 19/06/2020)

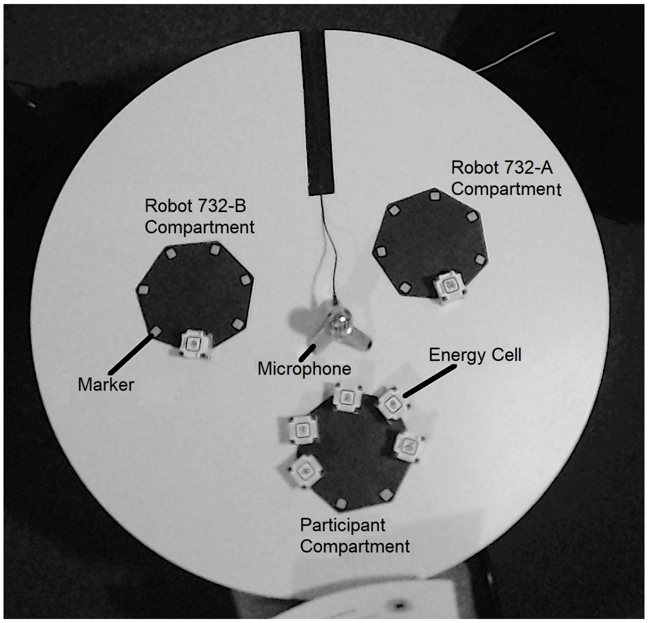

Figure 3: Top view of the commanding table. One energy cell is assigned to each robot and the participant kept five.

of the commanding table. The dialogue is restricted in allowing the participants to only pick questions from a predefined list and confirming that they are ready to go on with the experiment. Both NICO robot officers exhibit the same voice persona represented by embodied loudspeakers, allowing for a natural sound-source localisation.

3.3 Protocol & Game Scenes

As formulated in Chapter 3.2, we strive to automate the experiment procedure as much as possible. In the remaining human interventions, we follow a written protocol to diminish experimenter bias. The participants are welcomed and brought to the anteroom, where they are asked to fill out a consent form along with a pre-experiment questionnaire regarding their age, sex, personality (Big Five Inventory-2 Short [48]), former experience with robots and computers, risk propensity [35], and trust propensity [34]. They are then introduced to the space mission task, their role as the commander, and the two robotic officers who accompany them on their journey. The experimenter then guides the participant towards the experiment room and lets them familiarise themselves with the cockpit environment, the energy cells, the allocation compartments and a list of possible questions that can be asked to the robots throughout the game. The experimenter steps back out of the cockpit, and both the space ship AI *Wendigo* and the robotic officers introduce themselves. *Wendigo* conducts an introductory round of the cube allocation concealed as a system check, to acquaint the participant with the experiment procedure and reveal possible misunderstandings. The experimenter then enters again to answer any remaining questions before the start of the main experiment. In the actual experiment, the ship AI *Wendigo* guides the participant through four consecutive scenes, each of which repeats the following structure:

(1) *Wendigo* draws attention to a challenge (Scene 1: malfunctioning navigation system, Scene 2: interfering asteroids, Scene 3: inactive autopilot, Scene 4: leaking cooling system).
(2) Both robotic officers advertise their solution for which they require energy cells.
(3) The participant asks a question from the list of predefined options, to which the robotic officers reply one after another and in a randomised order.
(4) The participant is asked to distribute the energy cells as they see fit, and say '*Wendigo*, I am done! 'when they are done. *Wendigo* then provides feedback on the decision outcome (Scene 1: negative, Scene 2 - 4: positive).
(5) The participant places all energy cells back in their own compartment. After which the state machine autonomously transitions to the next scene.

During the experiment, the experimenter takes free-form observation notes about the body-language cues and speech of the participant, as well as any noteworthy occurrence during the experiment that could influence the participant's data. Following the experiment, the participant is provided with the post-study questionnaire that asks the participants to evaluate each robot regarding their persona. For this purpose, we use the Godspeed questionnaire [3] except for the *Perceived Safety* category since the participant did not have to physically interact with the robots and kept their distance throughout the experiment. The post-study questionnaire also asks the participant to rate the trustworthiness [6] and performance of each robot and choose which robot they preferred as an assistant, as well as provide feedback about shortcomings, immersion, and their overall experience during the experiment.

4 RESULTS

The study was conducted with 53 participants, of whom 45 finished the experiment successfully without any technical issue or language barrier. In the following sections, we will discuss the general population statistics, the results of the trust game and the effect of non-verbal communication (NVC) on trust and evaluate the general perception of the robots.

4.1 Population Statistics

The following population statistics apply to the 45 participants who completed the experiment without complications. The experiment was mainly advertised in a computer science department to people with at least some experience and familiarity with computers and robots, who are comfortable with participating in a science-fiction game and could understand and speak English fairly well. While 29% of the participants had worked with robots previously as developer, 42% had never interacted with a robot prior to the experiment. Our participants' mean age (M = 26.8, SD = 7.0) lies in the range of young adults and 38% of them identify as female. All of the participants were familiar with computers and 51% of them have programming experience. The population of participants were compared to the population of the general population with results obtained from other studies with a Welch's t-test for independent samples on descriptive statistics with significance level

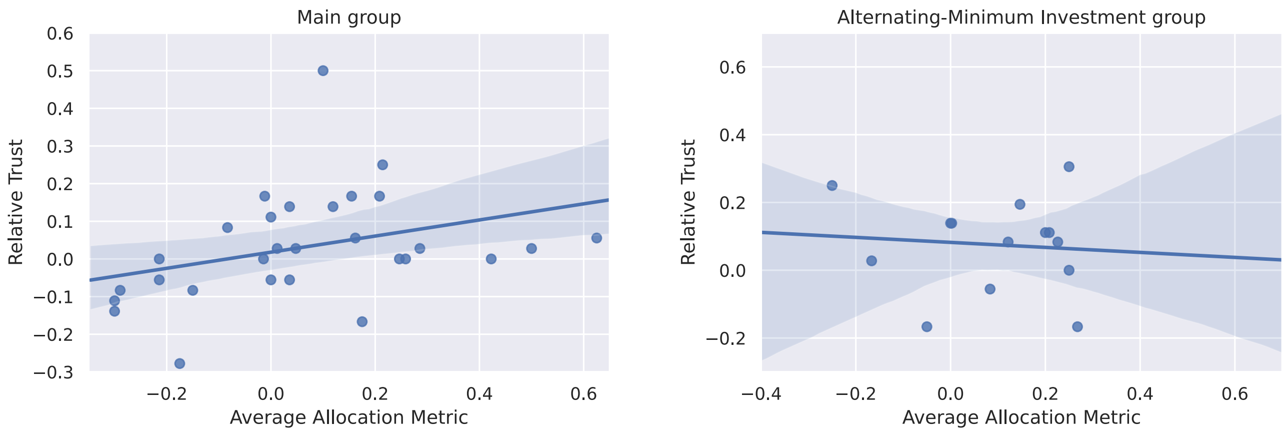

Figure 4: Correlation of relative trust and allocation metric for the two participant groups.

0.01. Based on the personality questionnaire (Section 3) results, the participants had average scores for extroversion ($M = 4.68$, $SD = 1.30$), agreeableness ($M = 5.22$, $SD = 1.03$) and neuroticism ($M = 3.62$, $SD = 1.78$). However they scored below-average in conscientiousness ($M = 4.79$, $SD = 0.99$) and above-average in openness ($M = 5.50$, $SD = 1.04$) compared to the general German population of a similar age group [30]. The trust and risk propensity questionnaires showed that our participants were less prone to take risks ($M = 4.05$, $SD = 1.32$) than the general population [35] yet more prone to trust [34] ($M = 2.93$, $SD = 0.61$).

4.2 Metrics and Grouping Criteria

We introduce two metrics specific to our scenario that allow us to quantify the differences in the trust placed between the robots.

4.2.1 Allocation Metric: Measures the investment displayed via energy cells allocated to each single robot. The allocation metric is calculated as $A(R) = \frac{cubes(R_2) - cubes(R_1)}{cubes(R_2) + cubes(R_1)}$ where $cubes(R)$ stands for the energy cells allocated to one of the robots $R \in \{R_1, R_2\}$. $A(R) < 0$ indicates a preference for R_1, $A(R) > 0$ a preference for R_2, while the magnitude in the differences is indicated by $|A(R)|$.

4.2.2 Relative Trust Metric: Measures the trust expressed in each robot according to the post-questionnaire. Relative trust is calculated as $T(R) = trust(R_2) - trust(R_1)$ where $trust(R)$ is the value obtained from the different trustworthiness Likert items in the post-interaction questionnaire, normalised to lie within $[0, 1]$. As before, $T(R) > 0$ indicates a preference for R_2 or a preference for R_1 otherwise, and the magnitude in the differences is indicated by $|T(R)|$.

Inspecting both the Allocation Metric and the Relative Trust metric over consecutive scenes, we can segment the participants into two groups:

4.2.3 The Alternating-Minimum Investment Group ($N = 16$): During the exploratory data analysis, two outstanding gameplay patterns were observed. These two patterns are defined by specific behaviour throughout the game, participants that showed either one or both of these behaviours were grouped together:

- *Minimum Investment Behaviour*: This behaviour resembles a lack of engagement in the game. Three of the participants investing less than one-third of the available cubes were considered disengaged. A threshold of fewer than 10 energy cells allocated in total throughout the four scenes was considered as a criterion for this group.
- *Alternating Investment Behaviour*: The energy cell allocation results indicated that some participants changed their minds about the robot they trusted more throughout the game. A group of 14 participants changed their mind at every scene as they would alternate between either allocating more energy cells to one robot or the other, or allocating an equal amount to both robots. These alternating participants did not particularly trust or prefer one robot over another to invest in throughout the game.

Further analysis of the alternating-minimum investment group showed that there is no link between these patterns and one specific robot, nor the NVC variable. As such, this behaviour did not depend on the content of speech or appearance of either of the robots.

4.2.4 The Main Group ($N = 29$): This is the group of participants that did not show either of the two aforementioned behaviours: the majority of the participants. With a Mann-Whitney U test for independent samples we found that these participants had no notable differences to the alternating-minimum investment group with regards to risk and trust propensity. They, however, obtained a lower score in Neuroticism ($p = 0.024$) in the personality questionnaire than the alternating-minimum investment group.

4.3 Transferability of the Investment Game

The aim of our study is to verify that our scaled-up version of the investment game can be used to measure trust in HRI. The results were evaluated separately on the main group ($N = 29$) and the alternating-minimum investment group ($N = 16$). For this the coherence between measured trust and self-assessed trust was evaluated by means of the Spearman test for correlation on the previously introduced metrics: the allocation metric represents the measured trust and the relative trust metric represents the self-assessed trust. A statistically significant correlation can be observed for the main group ($correlation = 0.43$, $p = 0.02$), however not for the alternating-minimum investment group ($correlation = -0.24$, $p = 0.37$). A comparison between both groups can be seen in Figure 4. In the standard human-human investment game, the amount of money invested by the trustor represents the trust in the trustee. As such, the observed correlation supports the hypothesis that our variation of the investment game between human and robot works much like the investment game between two humans. The fact that alternating-minimum investment behaviour was found also in a simple setting [36] and that there was no relation between the alternating behaviour of the participants and the robot characteristics show that the setting had no impact on the effectiveness of the trust game. This supports our hypothesis, that our scaled-up version of the investment game can indeed be used as a measure of trust.

4.4 Impact of Non-verbal communication (NVC) on the Perception of the Robot

After ensuring that it is indeed possible to measure trust in human-robot interaction with our investment game, as shown by the correlation, we further look into the impact of NVC on trust in the robot but also on other characteristics of the robot. As has been mentioned previously, NVC plays a significant role in human interaction but also in the efficiency and transparency of the interaction between humans and robots [11]. In our case, we find that these non-verbal cues have indeed made an impact on the trust in the robot as well as on its perceived anthropomorphism and animacy. We analyse the main group which didn't show alternating-minimum investment behaviour ($N = 29$) where it has been established that the game does measure trust. For this main group, the non-verbal communication of the robot had an impact on the number of energy cells received. This impact was observed in the first scene, the only scene where the participant had no previous disappointment related to any of the robots, but had already gotten to know the robot. In this scene, the robot that showed non-verbal communication obtained a significantly higher amount of energy cells compared to the other. The one-sided Wilcoxon test for independent samples between the distribution of the energy cells for the robot with NVCs and the robot with minimal NVC (MNVC) confirmed this ($p = 0.0008$). Independent of the gameplay choices, for all participants ($N = 45$) the robot showing NVC seemed more human-like and animated. As can be seen in Figure 5, the Godspeed values for anthropomorphism ($p = 0.008$) and animacy ($p = 0.00001$) are significantly distinct when comparing the NVC / MNVC conditions with a Mann-Whitney U test, whereas this is not the case for likeability ($p = 0.23$) and intelligence ($p = 0.24$). The observed values for anthropomorphism support our hypothesis that the NVC robot

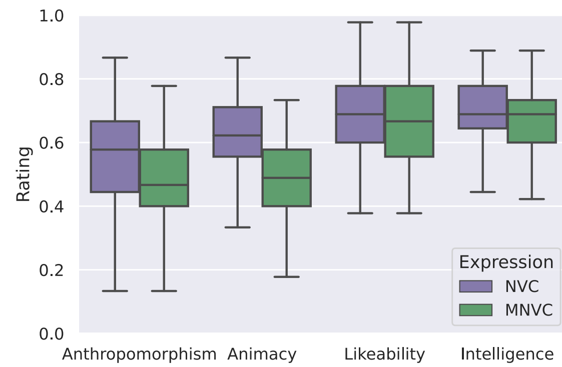

Figure 5: Effect of non-verbal communication (NVC) and minimal non-verbal communication (MNVC) on Godspeed items

invokes more trust, which is consistent with findings of similar studies. Mortham et al. [52] state that the perceived anthropomorphism of the robot increases the trust in the robot, especially for non-specialist humans, as the human needs to create a mental model for the robot to trust it. Furthermore, an increase in NVC leads to an increase in motion which subsequently leads to more perceived animacy [39]. However, likeability seems to not be affected by the use of NVC, potentially because the quantity and type of gestures used for non-verbal communication vary with culture [15]. Thus the degree to which a robot moves does not necessarily influence the likeability of the robot, as this is a personal preference that can vary across participants. Consistent with previous research [12], there are no perceived differences in intelligence either.

Our results show a correlation between trust measured by the investment game and the self-reported trust from the questionnaire. This gives us evidence that the scaled-up investment game can be used as a tool for measuring human-robot trust and therefore it can have practical applications in future experiments to study the impact of different variables (such as NVCs) between robots on how trustworthy the human perceives them. We anticipate that this serves as a positive example of extending socioeconomic experiments to a human-robot social interaction setting. Our experiment revolves around three main characteristics: the weighted choice between two agents, the inability of the participants to make choices based on domain knowledge, and the additional incentive for interaction between the trustor and the trustee. Maintaining these characteristics our scaled-up investment game can be adapted to a variety of situations and environments where trust but also NVCs play a big role in people's well being and productivity. Such environments comprise, but are not limited to, a work environment or a public service environment. In our study a futuristic environment was chosen as the majority of people know robots from media and science fiction stories [25]. We hypothesise that this is not a limiting factor for the replication of our study, although this should be subject to further research.

5 DISCUSSION AND FUTURE WORK

Overall, our results provide evidence that our variant of the investment game provides a reliable measure for human-robot trust and that non-verbal communication affects human-robot trust positively. However, there are some drawbacks which we discuss further in the next section.

5.1 Gameplay Behaviour

In our study, we found different game play behaviours that identify the two groups on which results were compared. The alternating-minimum investment group, as the name suggests, either alternated their investment of invested little in the robots which shows no engagement in the game. We were not able to measure a significant trust correlation for this group of participants, whereas the main group showed a significant result for this correlation. We hypothesise that the participants in the alternating-minimum investment group could have been alternating their strategies to infer the experiment research question or to simply test the system, this could be due to the fact of the experiment being advertised in a computer science department. Some participants might also not like the experimental setup or not feel immersed enough to participate. This type of behaviour indicates an insufficient foundation to establish trust for some of the participants, though further research is necessary to study their particular motivations. Mota et al. [36] observed that when a human needs to judge the trustworthiness of a robot, they draw on past social experiences with humans or try to build social experience with the robot. Due to insufficient shared social cues between humans and robots, humans are mostly incapable of determining the trustworthiness of a robot based on past experiences. In our experiment, almost half of the participants had never interacted with a robot previously. Also, building social experience with the robot was enforced by making the participant ask them one question before each round of cube allocation. Since only one question could be asked per round, and there were only 4 rounds, alternating behaviour could be a consequence of a failed attempt to build social experience. From this perspective, adding more rounds to the game could potentially lead to the behaviour regularising over time. For the small number of participants who showed non-engaging behaviour, this could be a result of misunderstanding the rules of the game, the relative worth of the energy cubes, or a general aversion to decision-making or to the presented scenario. This could also be considered as an attempt to delay decision-making until enough social experience has been built between the participant and the robots. It is worth mentioning that the Alternating-Minimum Investment group in the personality test showed higher scores for neuroticism compared to the main group.

5.2 Limitations

While we were able to observe and measure trust through the player's investments, a number of limitations could be improved on in future studies. Since our robots functioned fully autonomously, the natural language interface sometimes malfunctioned due to the user or machine errors, potentially failing the objective to promote user engagement. The participants who had to repeat themselves, some even multiple times, must have at least experienced a break in the immersion or, more severe, a sense of frustration that might have biased their results. Future experiments could investigate the effect of simplified design choices on our measurements, for example by employing a wizard-of-oz setup rather than an autonomous one. The processing time of the many parts of the experimental setup sometimes leads to slight delays between user action and robot reaction, which similarly could have lead to a break of the immersion and frustration. Our study is limited to the NICO robots and we have encountered some technical limitations including the lack of a larger range of different facial expressions and a wider range of human-like movements. Moreover, NICO has a childlike appearance and it is unclear as to how the perceived robot age can affect human perception of honesty and reliability, even though we introduced the NICOs as specialists in the complex field of space exploration. It is important to note that we compared the use of non-verbal communication (NVC) against the use of minimal non-verbal communication (MNVC). There is currently no widely established baseline or notion of *minimal* NVC and the impact of our interpretation and subsequent design choices on the participants is an open question. Our study showed that the mere presence of NVC has a positive impact on both the trust in the robot and the perceived characteristics of it. Future studies could investigate how different gestures affect trust, as there is no clear consensus of which social cues translate to "reliable" or "unreliable", and no obvious way to categorise these cues.

6 CONCLUSION

We provided an elaborate HRI scenario to model the building of trust more closely to human relationships than in the original investment game. Our experimental setup includes social interaction, non-verbal communication, a shared goal, and intrinsic motivation, thereby allowing participants to collaborate with robots more realistically than in the original investment game, measuring trust reliably. The environmental variables that our scenario (and its life-like agents) add to the data are a natural reflection of the many factors, internal and external, that influence human trust and how different levels of trust affect human behaviour in different contexts, modelling aspects of human-robot trust that the original trust game does not cover. We found a correlation between the self-assessed trust and the trust measured from the game for the majority of participants (main group). These same participants allocated more energy cells to the robot with non-verbal communication (NVC) in the first scene of the game. We were therefore able to replicate the positive effect of non-verbal communication on trust and robot perception. The Godspeed values for anthropomorphism and animacy were increased by NVC for all participants. Future research should comprise an investigation of the gameplay behaviours observed and could explore the effects of the use of different robots in this setup. Moreover, a similar setup can be used in future studies as a platform for studying trust and other potential factors that influence trust.

ACKNOWLEDGMENTS

This study would not have been possible without Ahmed Abdelghany, Connor Gäde, Vadym Gryshchuk, Matthew Ng, Shahd Safarani, Nilesh Vijayrania, Christoper Glenn Wulur, and Sophia Zell. The authors also gratefully acknowledge partial support from the German Research Foundation (DFG) under Project CML (TRR-169).

REFERENCES

[1] Nalini Ambady and Max Weisbuch. 2010. Nonverbal behavior. *Handbook of social psychology* 1 (2010), 464–487. https://psycnet.apa.org/record/2010-03505-013

[2] DemoSCOPE Angestellte Schweiz. 15/2/2019 (accessed 20/6/2019). Roboter wie Pepper übernehmen immer mehr Tätigkeiten in unserem Alltag. https://de-1statista-1com-100b48btl40ef.emedien3.sub.uni-hamburg.de/statistik/daten/studie/1005815/umfrage/akzeptanz-von-roboter-dienstleistungen-in-der-schweiz/.

[3] Christoph Bartneck, Dana Kulic, Elizabeth Croft, and Susana Zoghbi. 2008. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics* 1 (2008), 71–81. https://doi.org/10.1007/s12369-008-0001-3

[4] Johannes Bauer, Jorge Dávila-Chacón, Erik Strahl, and Stefan Wermter. 2012. Smoke and mirrors—virtual realities for sensor fusion experiments in biomimetic robotics. In *2012 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. IEEE, IEEE, 114–119. https://doi.org/10.1109/MFI.2012.6343022

[5] Joyce Berg, John Dickhaut, and Kevin McCabe. 1995. Trust, Reciprocity, and Social History. *Games and Economic Behavior* 10, 1 (1995), 122–142. https://doi.org/10.1006/game.1995.1027

[6] Jasmin Bernotat, Friederike Eyssel, and Janik Sachse. 2019. The (Fe)male Robot: How Robot Body Shape Impacts First Impressions and Trust Towards Robots. *International Journal of Social Robotics* 11 (05 2019), 1–13. https://doi.org/10.1007/s12369-019-00562-7

[7] G. Bradski. 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (2000).

[8] Cynthia Breazeal, Cory D Kidd, Andrea Lockerd Thomaz, Guy Hoffman, and Matt Berlin. 2005. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *2005 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 708–713. https://doi.org/10.1109/IROS.2005.1545011

[9] Cynthia Breazeal, Atsuo Takanishi, and Tetsunori Kobayashi. 2008. Social Robots that Interact with People. https://link.springer.com/content/pdf/10.1007/978-3-540-30301-5_59.pdf

[10] Judee K Burgoon, Laura K Guerrero, and Kory Floyd. 2016. *Nonverbal communication.* Routledge.

[11] J. Casper and R. R. Murphy. 2003. Human-robot interactions during the robot-assisted urban search and rescue response at the World Trade Center. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 33, 3 (2003), 367–385. https://ieeexplore.ieee.org/document/1200160

[12] Amol Deshmukh, Bart Craenen, Alessandro Vinciarelli, and Mary Ellen Foster. 2018. Shaping Robot Gestures to Shape Users' Perception: The Effect of Amplitude and Speed on Godspeed Ratings. In *Proceedings of the 6th International Conference on Human-Agent Interaction (HAI '18)*. Association for Computing Machinery, New York, NY, USA, 293–300. https://doi.org/10.1145/3284432.3284445

[13] David DeSteno, Monica Y Bartlett, Jolie Baumann, Lisa A Williams, and Leah Dickens. 2010. Gratitude as moral sentiment: emotion-guided cooperation in economic exchange. *Emotion* 10, 2 (2010), 289.

[14] David DeSteno, Cynthia Breazeal, Robert Frank, David Pizarro, Jolie Wormwood, Leah Dickens, and Jin Lee. 2012. Detecting the Trustworthiness of Novel Partners in Economic Exchange. *Psychological science* 23 (11 2012). https://doi.org/10.1177/0956797612448793

[15] Joseph A DeVito, Susan O'Rourke, and Linda O'Neill. 2000. *Human communication.* Longman.

[16] John Duffy. 2008. Trust in Second Life. *University of Pittsburgh, Department of Economics, Working Papers* 78, 388 (01 2008). https://doi.org/10.4284/0038-4038-78.1.53

[17] H. Felzmann, E. Fosch-Villaronga, C. Lutz, and A. Tamo-Larrieux. 2019. Robots and Transparency: The Multiple Dimensions of Transparency in the Context of Robot Technologies. *IEEE Robotics Automation Magazine* 26, 2 (2019), 71–78. https://ieeexplore.ieee.org/document/8684252

[18] Ceenu George, Malin Eiband, Michael Hufnagel, and Heinrich Hussmann. 2018. Trusting Strangers in Immersive Virtual Reality. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion.* 1–2.

[19] Vasile Gherheş. 2018. Why Are We Afraid of Artificial Intelligence (Ai)? *European Review Of Applied Sociology* 11, 17 (2018), 6–15. https://doi.org/10.1515/eras-2018-0006

[20] Edward L Glaeser, David I Laibson, Jose A Scheinkman, and Christine L Soutter. 2000. Measuring trust. *The quarterly journal of economics* 115, 3 (2000), 811–846. https://doi.org/10.1162/003355300554926

[21] Joanna Hale, Madeleine EM Payne, Kathryn M Taylor, Davide Paoletti, and Antonia F De C Hamilton. 2018. The virtual maze: A behavioural tool for measuring trust. *Quarterly Journal of Experimental Psychology* 71, 4 (2018), 989–1008. https://doi.org/10.1080/17470218.2017.1307865

[22] Ibrahim Hameed, Girts Strazdins, Håvard Hatlemark, Ivar Jakobsen, and John Damdam. 2018. *Robots That Can Mix Serious with Fun.* 595–604. https://doi.org/10.1007/978-3-319-74690-6_58

[23] Peter Hancock, Deborah Billings, Kristin Schaefer, Jessie Chen, Ewart de Visser, and Raja Parasuraman. 2011. A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. *Human factors* 53 (10 2011), 517–27. https://doi.org/10.1177/0018720811417254

[24] Kerstin Haring, Yoshio Matsumoto, and Katsumi Watanabe. 2013. How Do People Perceive and Trust a Lifelike Robot? *Lecture Notes in Engineering and Computer Science* 1. http://www.iaeng.org/publication/WCECS2013/WCECS2013_pp425-430.pdf

[25] Aike C. Horstmann and Nicole C. Krämer. 2019. Great Expectations? Relation of Previous Experiences With Social Robots in Real Life or in the Media and Expectancies Based on Qualitative and Quantitative Assessment. *Frontiers in Psychology* 10 (2019), 939. https://doi.org/10.3389/fpsyg.2019.00939

[26] Matthias Kerzel, Theresa Pekarek-Rosin, Erik Strahl, Stefan Heinrich, and Stefan Wermter. 2020. Teaching NICO How to Grasp: An Empirical Study on Crossmodal Social Interaction as a Key Factor for Robots Learning from Humans. *Frontiers in Neurorobotics* 14 (2020), 28. https://doi.org/10.3389/fnbot.2020.00028

[27] Matthias Kerzel, Erik Strahl, Sven Magg, Nicolás Navarro-Guerrero, Stefan Heinrich, and Stefan Wermter. 2017. NICO—Neuro-inspired companion: A developmental humanoid robot platform for multimodal interaction. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 113–120. https://www2.informatik.uni-hamburg.de/wtm/publications/2017/KSMNHW17/NICO_RO-MAN_2017-PREPRINT-2017.pdf

[28] Svetlana Kubilinskiene, Inga Zilinskiene, Valentina Dagiene, and Vytenis Sinkevičius. 2017. Applying Robotics in School Education: a Systematic Review. *Baltic Journal of Modern Computing* 5 (03 2017). https://doi.org/10.22364/bjmc.2017.5.1.04

[29] Vladimir A Kulyukin. 2006. On natural language dialogue with assistive robots. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. 164–171. https://doi.org/10.1145/1121241.1121270

[30] Frieder R Lang, Dennis John, Oliver Lüdtke, Jürgen Schupp, and Gert G Wagner. 2011. Short assessment of the Big Five: Robust across survey methods except telephone interviewing. *Behavior research methods* 43, 2 (2011), 548–567.

[31] Yuhua Liang and Seungcheol Austin Lee. 2017. Fear of autonomous robots and artificial intelligence: Evidence from national representative data with probability sampling. *International Journal of Social Robotics* 9, 3 (2017), 379–384. https://doi.org/10.1007/s12369-017-0401-3

[32] Deirdre E. Logan, Cynthia Breazeal, Matthew S. Goodwin, Sooyeon Jeong, Brianna O'Connell, Duncan Smith-Freedman, James Heathers, and Peter Weinstock. 2019. Social Robots for Hospitalized Children. *Pediatrics* 144, 1 (2019). https://doi.org/10.1542/peds.2018-1511

[33] Maya B Mathur and David B Reichling. 2016. Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley. *Cognition* 146 (2016), 22–32. https://doi.org/10.1016/j.cognition.2015.09.008

[34] Roger C Mayer and James H Davis. 1999. The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of applied psychology* 84, 1 (1999), 123.

[35] Ree M. Meertens and René Lion. 2008. Measuring an Individuals Tendency to Take Risks: The Risk Propensity Scale. *Journal of Applied Social Psychology* 38, 6 (2008), 1506–1520. https://doi.org/10.1111/j.1559-1816.2008.00357.x

[36] Roberta C Ramos Mota, Daniel J Rea, Anna Le Tran, James E Young, Ehud Sharlin, and Mario C Sousa. 2016. Playing the 'trust game'with robots: Social strategies and experiences. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Columbia University, NY, 519–524. https://ieeexplore.ieee.org/document/7745167

[37] Toshiharu Mukai, S. Hirano, Hiromichi Nakashima, Y. Kato, Y. Sakaida, S. Guo, and Hosoe Shigeyuki. 2010. Development of a nursing-care assistant robot RIBA that can lift a human in its arms. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 5996–6001. https://doi.org/10.1109/IROS.2010.5651735

[38] Michelle Neumann. 2019. Social Robots and Young Children's Early Language and Literacy Learning. *Early Childhood Education Journal* (10 2019). https://doi.org/10.1007/s10643-019-00997-7

[39] Domenico Parisi and Matthew Schlesinger. 2002. Artificial life and Piaget. *Cognitive Development* 17, 3-4 (2002), 1301–1321.

[40] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y Ng. 2009. ROS: an open-source Robot Operating System. In *ICRA workshop on open source software*, Vol. 3. Kobe, Japan, 5. https://www.ros.org/

[41] Tage S Rai and Daniel Diermeier. 2015. Corporations are cyborgs: Organizations elicit anger but not sympathy when they can think but cannot feel. *Organizational Behavior and Human Decision Processes* 126 (2015), 18–26. https://doi.org/10.1016/j.obhdp.2014.10.001

[42] Alessandra Rossi, Kerstin Dautenhahn, Kheng Koay, and Joe Saunders. 2017. Investigating Human Perceptions of Trust in Robots for Safe HRI in Home Environments. 375–376. https://doi.org/10.1145/3029798.3034822

[43] Alessandra Rossi, Kerstin Dautenhahn, Kheng Lee Koay, and Michael L. Walters. 2017. How the Timing and Magnitude of Robot Errors Influence Peoples' Trust of Robots in an Emergency Scenario. In *Social Robotics*. Springer International

Publishing, Cham, 42–52.

[44] Maha Salem, Friederike Eyssel, Katharina Rohlfing, Stefan Kopp, and Frank Joublin. 2013. To err is human (-like): Effects of robot gesture on perceived anthropomorphism and likability. *International Journal of Social Robotics* 5, 3 (2013), 313–323. https://doi.org/10.1007/s12369-013-0196-9

[45] Maha Salem, Katharina Rohlfing, Stefan Kopp, and Frank Joublin. 2011. A friendly gesture: Investigating the effect of multimodal robot behavior in human-robot interaction. In *2011 Ro-Man*. IEEE, 247–252. https://doi.org/10.1109/ROMAN.2011.6005285

[46] Shane Saunderson and Goldie Nejat. 2019. How robots influence humans: A survey of nonverbal communication in social human–robot interaction. *International Journal of Social Robotics* 11, 4 (2019), 575–608. https://doi.org/10.1007/s12369-019-00523-0

[47] E. Schniter, T.W. Shields, and D. Sznycer. 2020. Trust in humans and robots: Economically similar but emotionally different. *Journal of Economic Psychology* 78 (2020), 102253. https://doi.org/10.1016/j.joep.2020.102253

[48] Christopher J. Soto and Oliver P. John. 2017. Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality* 68 (2017), 69–81. https://doi.org/10.1016/j.jrp.2017.02.004

[49] Dimitris Spiliotopoulos, Ion Androutsopoulos, and Constantine D Spyropoulos. 2001. Human-robot interaction based on spoken natural language dialogue. In *Proceedings of the European workshop on service and humanoid robots*. 25–27. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.28.6747&rep=rep1&type=pdf

[50] Stanford Artificial Intelligence Laboratory et al. 23/5/2018 (accessed 20/6/2019). Robotic Operating System. Retrieved March 2, 2005 from https://www.ros.org

[51] Johannes Twiefel, Timo Baumann, Stefan Heinrich, and Stefan Wermter. 2014. Improving domain-independent cloud-based speech recognition with domain-dependent phonetic post-processing. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*. https://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/view/8554/8604

[52] Robert H Wortham and Andreas Theodorou. 2017. Robot transparency, trust and utility. *Connection Science* 29, 3 (2017), 242–248. https://doi.org/10.1080/09540091.2017.1313816

[53] Debora Zanatto. 2019. When do we cooperate with robots? *University of Plymouth, Dissertations* (01 2019). https://pearl.plymouth.ac.uk/bitstream/handle/10026.1/14677/2019zanatto10513659phd_full.pdf

[54] Debora Zanatto, Massimiliano Patacchiola, Jeremy Goslin, Serge Thill, and Angelo Cangelosi. 2020. Do Humans Imitate Robots? An Investigation of Strategic Social Learning in Human-Robot Interaction. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20)*. Association for Computing Machinery, New York, NY, USA, 449–457. https://doi.org/10.1145/3319502.3374776