AFFECTIVE AND SOCIAL SIGNALS FOR HRI

EDITED BY: Hatice Gunes, Ginevra Castellano and Bilge Mutlu

PUBLISHED IN: Frontiers in Robotics and AI, Frontiers in Psychology and Frontiers in Computer Science

frontiers Research Topics

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-88963-454-5
DOI 10.3389/978-2-88963-454-5

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

AFFECTIVE AND SOCIAL SIGNALS FOR HRI

Topic Editors:
Hatice Gunes, University of Cambridge, United Kingdom
Ginevra Castellano, Uppsala University, Sweden
Bilge Mutlu, University of Wisconsin-Madison, United States

Designing robots with socio-emotional skills is a challenging research topic still in its infancy. These skills are important for robots to be able to provide not only physical, but also social support to human users, and to engage in and sustain long-term interactions with them in a variety of application domains that require human-robot interaction, including healthcare, education, entertainment, manufacturing, and many others. The availability of commercial robotic platforms and developments in collaborative academic research provide us a positive outlook, however, the capabilities of current social robots are quite limited. The main challenge is understanding the underlying mechanisms of the humans in responding to and interacting with real life situations, and how to model these mechanisms for the embodiment of naturalistic, human-inspired behaviors via robots. To address this challenge successfully requires an understanding of the essential components of social interaction including nonverbal behavioral cues such as interpersonal distance, body position, body posture, arm and hand gestures, head and facial gestures, gaze, silences, vocal outbursts and their dynamics. To create truly intelligent social robots, these nonverbal cues need to be interpreted to form an understanding of the higher level phenomena including first-impression formation, social roles, interpersonal relationships, focus of attention, synchrony, affective states, emotions, and personality, and in turn defining optimal protocols and behaviors to express these phenomena through robotic platforms in an appropriate and timely manner. Achieving this goal requires the fields of psychology, nonverbal behavior, vision, social signal processing, affective computing, and HRI to constantly interact with one another. This Research Topic aims to foster such interactions and collaborations by bringing together the latest works and developments from across a range of research groups and disciplines working in these fields.

The Research Topic is a collection of 14 articles that span across five research themes. Three articles co-authored by Terada and Takeuchi, Jung et al., and Kennedy et al. explore the design of "social and affective cues" for robots and investigate their effects on human-robot interaction. Mirnig et al., Bremner et al., and Strait et al. investigate people's "perceptions of robots" in different settings and scenarios, such as when robots make errors. Articles by Lee et al., Leite et al., and Heath et al. investigate the factors that shape "dialogic interaction with robots," such as interaction context. The articles under the theme "social and affective therapy" by Rouaix et al., Rudovic et al., and Matsuda et al. report on how individuals from clinical populations, such as those with dementia, autism, and other pervasive developmental disorders (PDDs), interact with robots in therapeutic scenarios. Finally, Miklósi et al. and Durantin et al. offer "new perspectives in human-robot interaction" with a focus on reframing social interaction and human-robot relationships. We are excited about sharing this rich collection with the scientific community and about its contributions to the human-robot interaction literature.

Citation: Gunes, H., Castellano, G., Mutlu, B., eds. (2020). Affective and Social Signals for HRI. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88963-454-5

Table of Contents

SOCIAL AND AFFECTIVE CUES

PERCEPTIONS OF ROBOTS

DIALOGIC INTERACTION WITH ROBOTS

SOCIAL AND AFFECTIVE THERAPY

NEW PERSPECTIVES IN HUMAN-ROBOT INTERACTION

frontiers
in Psychology

ORIGINAL RESEARCH
published: 22 May 2017
doi: 10.3389/fpsyg.2017.00724

Check for updates

Emotional Expression in Simple Line Drawings of a Robot's Face Leads to Higher Offers in the Ultimatum Game

Kazunori Terada and Chikara Takeuchi*

Informatics Course, Department of Electrical, Electronics and Computer Engineering, Faculty of Engineering, Gifu University, Gifu, Japan

In the present study, we investigated whether expressing emotional states using a simple line drawing to represent a robot's face can serve to elicit altruistic behavior from humans. An experimental investigation was conducted in which human participants interacted with a humanoid robot whose facial expression was shown on an LCD monitor that was mounted as its head (Study 1). Participants were asked to play the ultimatum game, which is usually used to measure human altruistic behavior. All participants were assigned to be the proposer and were instructed to decide their offer within 1 min by controlling a slider bar. The corners of the robot's mouth, as indicated by the line drawing, simply moved upward, or downward depending on the position of the slider bar. The results suggest that the change in the facial expression depicted by a simple line drawing of a face significantly affected the participant's final offer in the ultimatum game. The offers were increased by 13% when subjects were shown contingent changes of facial expression. The results were compared with an experiment in a teleoperation setting in which participants interacted with another person through a computer display showing the same line drawings used in Study 1 (Study 2). The results showed that offers were 15% higher if participants were shown a contingent facial expression change. Together, Studies 1 and 2 indicate that emotional expression in simple line drawings of a robot's face elicits the same higher offer from humans as a human telepresence does.

Keywords: robot, facial expression, emotion, altruistic behavior, human-robot interaction

OPEN ACCESS

Edited by:
Hatice Gunes,
University of Cambridge, UK

Reviewed by:
Andrej Košir,
University of Ljubljana, Slovenia
Magalie Ochs,
UMR7296 Laboratoire des Sciences de l'Information et des Systèmes (LSIS), France

****Correspondence:***
Kazunori Terada
terada@gifu-u.ac.jp

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Psychology

Received: *10 January 2017*
Accepted: *21 April 2017*
Published: *22 May 2017*

Citation:
Terada K and Takeuchi C (2017) Emotional Expression in Simple Line Drawings of a Robot's Face Leads to Higher Offers in the Ultimatum Game. Front. Psychol. 8:724. doi: 10.3389/fpsyg.2017.00724

1. INTRODUCTION

Recently, there has been increasing interest and progress in robotic emotional expressions. A wide variety of methods for achieving emotional expression have been proposed (Bethel and Murphy, 2008), including facial expressions (Bartneck, 2003; Breazeal, 2004; Kanoh et al., 2004; Itoh et al., 2006; Matsui et al., 2010), speech (Kim et al., 2009a,b), body movement (Shimokawa and Sawaragi, 2001; Bethel and Murphy, 2007), and colors (Sugano and Ogata, 1996; Kim et al., 2009a,b; Terada et al., 2012). Leaving aside discussion regarding a robot's ability to possess genuine emotions, implementing a display of emotion in robots could be useful not only by increasing their friendliness but also by helping them to influence people without explicit language (Breazeal, 2003, 2004).

There have been studies on the effect of robotic emotions on human behavior (Cassell and Thorisson, 1999; Bickmore and Picard, 2005; Leyzberg et al., 2011); these focused on the task-oriented effects of emotions. Leyzberg et al. (2011) showed that robots that express emotions elicited better human teaching. A long-term experiment conducted by Bickmore and Picard (2005) showed that an agent with relational behavior, including social-emotional responses, contributed to increasing participants' positive attitude about exercise. While these studies revealed that robots with emotions positively affect human behavior, the nature, and essential function of these emotions have not been discussed. In the present study, we focused on the social functional aspect of emotions and experimentally investigated the effect of emotional expression as depicted through a simple line drawing of a face on human economic behavior.

Emotions control the behavior of an agent. For example, fear increases heart rate and muscle tension and drives an agent to escape from a situation; consequently, fear helps in avoiding dangerous situations. Emotions affect not only one's own behavior but also that of others. An angry individual, for example, usually obtains concessions from a competitor in a conflict situation (van Kleef et al., 2004; Sinaceur and Tiedens, 2006; van Kleef and Côté, 2007; van Dijk et al., 2008; van Kleef et al., 2008; Sell et al., 2009; Fabiansson and Denson, 2012; Reed et al., 2014). Positive emotions are considered to have evolved to maintain cooperative relationships (Trivers, 1971; Alexander, 1987; Frank, 1988; Scharlemann et al., 2001; Brown and Moore, 2002; Brown et al., 2003; Mehu et al., 2007; Reed et al., 2012; Mussel et al., 2013).

Altruism is a behavior that reduces the actor's wealth while increasing the wealth of the recipient, whereas cooperation is a process in which agents work together to gain common or mutual wealth. However, altruism can be considered to be asynchronous cooperative behavior by considering direct or indirect reciprocity (Nowak and Sigmund, 2005). In order to produce altruistic behavior, one must ignore the loss of one's own wealth. Positive emotions such as happiness and kindness that are elicited from another's facial expressions presumably compensate for the loss. Therefore, emotion is more important for long-term or indirect reciprocal relationships than short-term (one-shot) cooperative tasks. We used altruistic behavior as a measure of the function of the robot's facial expression because our focus is on the long-term human-robot relationship.

Researchers have been investigating whether people have a tendency to cooperate with robots (Nishio et al., 2012; Torta et al., 2013; Sandoval et al., 2016). Decision making in economic games such as the prisoners' dilemma and the ultimatum game is used to measure the cooperative attitude of participants. Nishio et al. (2012) have studied how the appearance of agents (computer, humanoid, android, or human) affects participants' cooperativeness. They conclude that although the appearance of agents does not affect cooperativeness, conversation with a human-like agent (android) leads people to be more cooperative. Torta et al. (2013) reported that rejection scores in the ultimatum game are higher in the case of a computer opponent than in the case of a human or robotic opponent, indicating that people might treat a robot as a reciprocal partner. Sandoval et al. (2016) showed that participants who interacted with a robot showed significantly less cooperation than when they interacted with a human in the prisoner's dilemma. Further, participants offered significantly less money in the ultimatum game to the robot than to the human agent, indicating that people tend to cooperate more with a human agent than with a robot.

From the above discussion, the following prediction could be derived: if robots offer emotional expression, people behave more cooperatively toward them. There are a few studies that examine the effect of the emotional expression of robots on human cooperative behavior in terms of economic behavior (de Melo et al., 2010, 2011). de Melo et al. (2010) conducted an experiment in which participants play the iterated prisoner's dilemma against two different virtual agents that play the tit-for-tat strategy but communicate different goal orientations (cooperative vs. individualistic) through their patterns of facial displays. They showed that participants were sensitive to differences in the facial displays and cooperated significantly more with the cooperative agent. de Melo et al. (2011), in another study, reported that participants conceded more to a virtual agent that expresses anger than to one that expresses happiness in a negotiation task.

The studies of de Melo et al. (2010, 2011) used human-like virtual character agents. In our study, we used a real robot with a simple line drawing of a face to remove realistic and biological human features from the agent's face (Terada et al., 2013). Most of the robots that are used in human-robot interaction studies have sophisticated facial expression mechanisms (Breazeal, 2004; Itoh et al., 2006; Matsui et al., 2010; Becker-Asano and Ishiguro, 2011; Mazzei et al., 2011). The underlying assumption is that mimicking real human facial expressions induces humans to emotionally respond as they would when interacting with a real human. However, studies have revealed that line drawing facial expressions are recognized to the same extent as a realistic face (Katsikitis, 1997; Britton et al., 2008), affect human altruistic behavior even they are slightly different (Brown and Moore, 2002), and are processed in the human brain in the same way as a human face (Britton et al., 2008) .

In the present study, we investigated whether a simple line drawing of a face is useful in human-robot interaction in terms of human-robot cooperative relationships. Terada et al. (2013) have showed that emotional expression by robots led people to behave more altruistically toward the robots even though the emotion was represented by simple line drawings. However, it is unclear whether this effect is the same extent as that of human-human interaction. In the present paper, we first show the results of human-robot condition reported in Terada et al. (2013) as Study 1. We then show the results of human-human condition (Study 2) and compare the results of these two studies.

The ultimatum game has been used to measure human altruistic behavior (Güth and Tietz, 1990; Sanfey et al., 2003; Oosterbeek et al., 2004; Xiao and Houser, 2005; van Dijk et al., 2008; Yamagishi et al., 2009). It is played by two players, a proposer and a responder, who are given the opportunity to split an allotment of money. The proposer has the right to divide the money and offer an amount to the responder. If the responder accepts the proposal, both players keep the money. If the responder rejects the proposal, neither player receives the money. The findings of a meta-analysis of 37 papers with 75 results from ultimatum game experiments showed that on average, the proposer offers 40% of the money to the responder, and 16% of the offers are rejected (Oosterbeek et al., 2004).

In our study, all participants were assigned to be the proposer and were instructed to decide their offer within 1 min by controlling a slider bar. In the decision period, a change in the responder's facial expression was shown to the proposer (only in the change of facial expression condition), which is not a normal procedure in the ultimatum game. The communication before the decision is treated as cheap talk, which is costless and unverifiable preplay statements about private information and non-credible threats about future actions (Croson et al., 2003). Croson et al. (2003) showed that threats of future actions influenced bargaining outcomes.

The goal of the present study was to explore whether communication using the facial expression of robots is effective in establishing human-robot cooperative relationships. We used the offer in the ultimatum game as the measurement of cooperative attitude of human toward a robot. As a result, the effectiveness of facial expression of robots in human-robot cooperative relationship could be evaluated in terms of economic value.

Studies 1 and 2 were both conducted in accordance with the recommendations of the Ethical Guidelines for Medical and Health Research Involving Human Subjects provided by the Ministry of Education, Culture, Sports, Science, and Technology and the Ministry of Health, Labor, and Welfare in Japan with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Medical Review Board of Gifu University Graduate School of Medicine.

2. STUDY 1

2.1. Method

2.1.1. Participants

Twenty-six healthy graduate and undergraduate students (15 male, 11 female, $M_{\text{age}} = 19.62$ years, $SD_{\text{age}} = 3.85$ years, age range: 18–24 years) participated in the experiment. Participants were recruited through advertising on posters and via e-mail at the university. They were informed that they would be paid with a JPY 500 (approximately USD 5) book coupon for their time. All were ignorant of the purpose of the experiment.

2.1.2. Experimental Design

A single-factor two-level between-participants experimental design was used. Participants were randomly assigned to either a "change of facial expression" or a "static face" condition. All participants assumed the role of the proposer and were asked to determine their offer within 1 min by controlling the slider bar. The only difference between the two conditions was whether the corners of the mouth of the line drawing shown on an LCD monitor mounted on the robot moved upward or downward according to the position of the slider bar. In the initial state, a straight line segment represented the line drawing mouth.

2.1.3. Apparatus

The ultimatum game was played once. The proposer was given 100 points, which corresponds to JPY 1000 (approximately USD 10), as the amount to divide. The proposer was given 1 min to determine the offer (*decision phase*). During the decision phase, the proposer adjusted the offer by controlling the slider bar. Participants were informed that the game would be played against a humanoid robot that might react to the participant's offer through an LCD monitor mounted on the robot.

A GUI was used to determine the offer and to communicate the emotional state of the responder (see **Figure 1**). The proposer was asked to decide the offer within 1 min by moving a slider bar on the GUI, which was controlled by a gamepad connected to the computer.

Static Face Condition

The line drawing face did not change during the proposal period.

Change of Facial Expression Condition

The corners of the mouth of the line drawing moved upward and downward according to the position of and one second after the movement of the slider bar. This delay was inserted to prevent the participants from assuming that the responder was merely a simple computer program; an immediate mouth movement completely contingent on the proposer's action might strongly indicate artificiality. The software's calculation rate was 60 fps, the same as the monitor used to display the GUI.

Figure 2 shows the control points of a Bézier curve, which represented the line drawing of the mouth. The points P3 and P4 are the static points. The Y-coordinates of the points P0, P1, and P2 changed according to the position of the slider bar. **Figure 3** illustrates examples of the facial expressions shown to the proposer as a function of the proposer's offer $x \in [0, 100]$. If the slider bar moved to the right, the offer decreased and negative facial expressions, such as those shown in **Figures 3A,B**, were displayed.

Figure 4 shows the experimental system. We mounted an LCD monitor on a Robovie-X, a commercially available robot. Line drawings of facial expressions were shown on the mounted LCD monitor, which was connected to a laptop computer via a

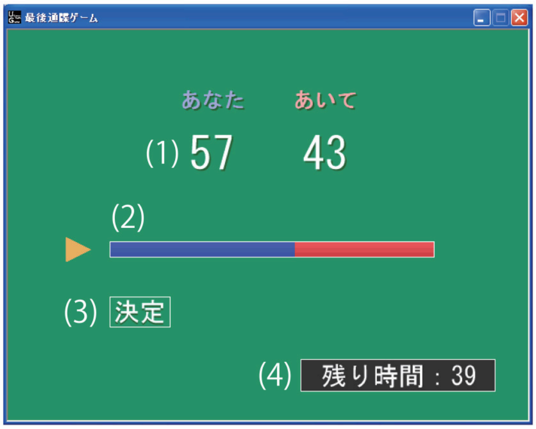


FIGURE 1 | Graphical user interface used by the proposer to determine the offer: (1) numerical representation of the offer, (2) slider bar to change the offer, (3) button for final decision, and (4) time remaining.

USB cable. The laptop computer was also used to display the GUI, and a gamepad for controlling the slider bar on the GUI was connected to the laptop.

2.1.4. Procedure

In the experiment room, participants were asked to read an instruction sheet that stated the rules of the ultimatum game, how to use the interface, and that "the response of the responder will be shown on the head display." In addition, they were informed that they were assigned to be the proposer and that they would win additional money according to their score in the game.

After the proposal, the participants were not immediately informed of the responder's acceptance/rejection: they were first asked to complete a questionnaire to avoid the questionnaire responses being affected by the responder's decision. After completing the questionnaire, the participants were informed that they had all played as proposers against a computer program, and they were paid with an additional JPY 500 (approximately USD 5) book coupon, the amount of money that would be given if a 50:50 offer was accepted.

2.1.5. Measurement and Analysis

The offer was recorded every 0.5 s. After the game, participants were asked to answer four 7-point Likert scale questions (0 = "definitely no" to 7 = "definitely yes"):

- *Q1. Did you perceive emotions in the picture shown on the head of the robot?*
- *Q2. Did you consider the responder's emotions when deciding your offer?*

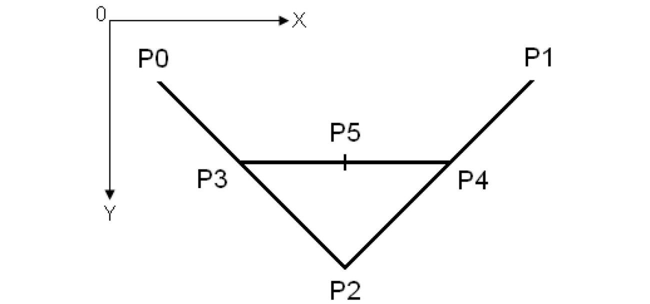


FIGURE 2 | Control points of the Bézier curve used to represent the mouth.

After answering the post-questionnaires, participants were asked whether they realized that they had been playing against a computer program.

The one-way analysis of variance (ANOVA) was used if the data were normally and homogeneously distributed. The Welch's ANOVA was used if the data were normally distributed, but the assumption of homogeneity of variance was violated. The Mann–Whitney U-test was used if the data were homogeneously distributed, but the assumption of normality was rejected. The Brunner–Munzel test was used if both the assumption of normality and the homogeneity of variance were violated.

2.2. Results

The mean durations for deciding an offer were 28.31 s (SD = 17.27) and 18.69 s (SD = 14.26) in the change of facial expression and static face conditions, respectively. The one-way ANOVA, $F_{(1,\,24)} = 2.40, p = 0.13$, indicated that the difference was not statistically significant.

Figure 5 presents the mean final offers over participants in both conditions. Welch's ANOVA, $F_{(1,\,15.71)} = 6.22$, $p < 0.05$, showed that offers were higher in the change of facial expression condition ($M = 51.62$, $SD = 6.91$) than in the static face condition ($M = 38.69$, $SD = 17.35$).

Figure 6 displays the results of the post-experiment questionnaire. The Mann–Whitney U-tests, $U = 18.5$, $z = 3.46$, $p < 0.001$, revealed that ratings for perceiving emotions from the line drawing were significantly higher in the change of facial expression condition than in the static face condition. The one-way ANOVA, $F_{(1,\,24)} = 30.03$, $p < 0.001$, revealed that ratings for the consideration of emotions were significantly higher in the change of facial expression condition than in the static face condition.

Ten out of 13 participants in the change of facial expression condition realized that they had played against a computer program that generates a simple mouth movement completely contingent on the participants' action.

2.3. Discussion

The results show that offers were higher in the change of facial expression condition than in the static face condition, confirming that emotional expression by robots led participants to behave more altruistically toward the robots even though the emotion

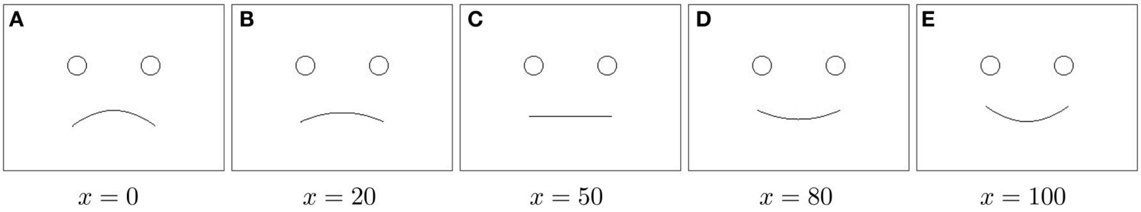


FIGURE 3 | Examples of the facial expressions displayed to the proposer as a function of the proposer's offer $x \in [0, 100]$. (A) $x = 0$. **(B)** $x = 20$. **(C)** $x = 50$. **(D)** $x = 80$. **(E)** $x = 100$.

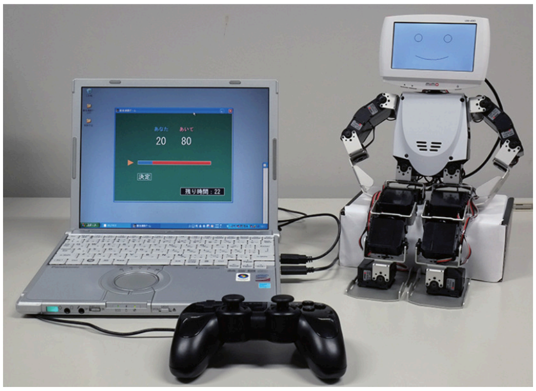


FIGURE 4 | System used in our experiment.

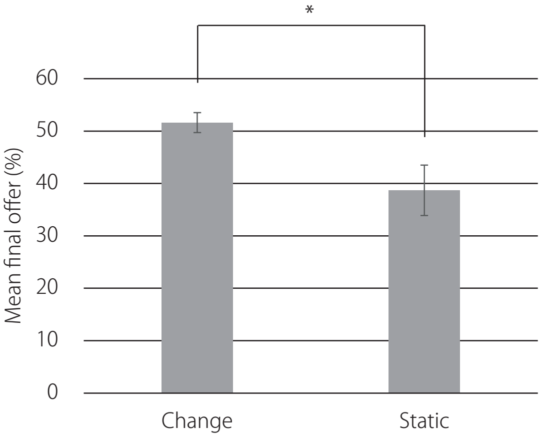


FIGURE 5 | Mean final offers over participants in both conditions. Error bars indicate standard errors. *$p < 0.05$.

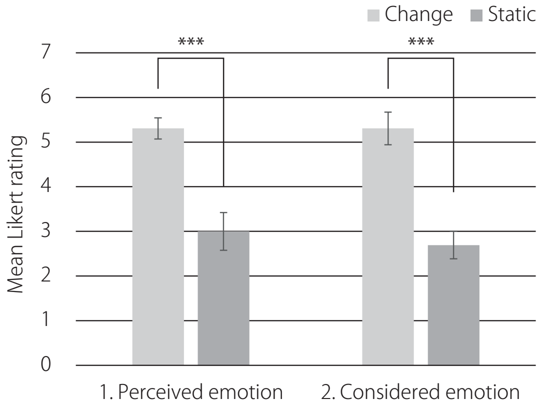


FIGURE 6 | Post-experiment questionnaire. Error bars indicate standard errors. ***$p < 0.001$.

was represented by simple line drawings. The results of the post-experiment questionnaires support the behavioral result that the 12.92% gap between the two conditions was caused by the emotions that participants recognized from the change of facial expressions exhibited by the line drawing. Participants in the change of facial expression condition gave higher ratings, an average of 5.30, to the question "Did you perceive emotions in the picture located on the upper right of the GUI?" than did participants in the static face condition. There was a large gap, an average of 2.30, in the Q1 rating between the two conditions, which indicates that perceiving emotions caused the participants' altruistic behavior.

The conditions differed only in whether the corners of the mouth of the line drawing in the GUI changed. However, we did not explicitly inform participants that the line drawing symbolized a face or that the position of the bar represented the position of the corners of the mouth. The participants arbitrarily attributed a facial property to the geometric line drawings and attributed emotions to variable Bézier curves. According to Ekman (2003), a convex mouth shape, in which the corners of the lips curl downward, indicates sadness, and a concave mouth shape, in which corners of the lips move upward, indicates happiness. Although, we did not identify the emotions that participants perceived from the line drawings, the universality of facial expressions supports the assumption that participants recognized sadness when they were shown a convex mouth and happiness when they were shown a concave mouth.

Our results show that although a substantial number of participants (78%) in the change of facial expression condition realized that the mouth movement was controlled by a computer program, the effect of facial expression was still observed. de Melo et al. (2011) reported similar findings from a study in which participants were involved in a negotiation with computer agents. Taken together, these findings imply that facial expressions are effective in inducing people to cooperate with robots even though they know that the expressions are controlled by a program.

A meta-analysis of 75 results from ultimatum game experiments revealed that the proposer usually offers 40% of the money to the responder (Oosterbeek et al., 2004). However, participants in the change of facial expression condition offered an average of 51.62% of the money. This indicates that the offer increased by approximately 10% when people were shown changes of facial expression corresponding to their offer. By contrast, participants in the static face condition offered an average of 38.69%. This value roughly corresponds to that offered in the earlier studies that included no emotional interaction in their experimental setting.

There are two potential reasons why participants in the change of facial expression condition offered approximately 50:50, which is a fair offer. The first is the impression that the responder has the capability to respond emotionally, which is formed by the dynamic change of facial expression in response to the participant's operation. In this case, the facial expression itself does not have an absolute meaning: simply perceiving adaptivity or the ability to respond to the user's input might be lead to a fair offer. The second reason is a neutral face. In our experimental setting, a neutral face, in which the mouth was represented by a straight line, was displayed to participants when the offer was 50%. Participants could adjust the slider bar to make the facial expression neutral. Further investigation, in which a neutral face does not correspond to a 50% offer, is needed to test these two hypotheses.

Our results do not identify whether positive or negative emotion contributed to an increase in the offer. It is known that expressing anger can elicit concessions from others (van Kleef et al., 2004; Sinaceur and Tiedens, 2006; van Kleef and Côté, 2007; van Dijk et al., 2008; van Kleef et al., 2008; Sell et al., 2009; Fabiansson and Denson, 2012; Reed et al., 2014), while happiness can elicit altruism (Scharlemann et al., 2001; Brown and Moore, 2002; Brown et al., 2003; Mehu et al., 2007; Mussel et al., 2013). These findings suggest that both the negative and positive expressions shown in our experiment might have contributed to the proposer raising the offer.

Croson et al. (2003) showed that threats of future actions influenced bargaining outcomes. The negative emotional expression that was contingently presented when a low offer was proposed might have played the role of cheap talk.

3. STUDY 2

Study 2 was conducted to compare the result of Study 1 with those from a study in which participants played the game against a human responder in a teleoperation setting through a computer display. The aim of this study was to determine whether the altruistic behavior induced by the robot's facial expression is also induced by a facial expression controlled by a human.

3.1. Method

3.1.1. Participants and Experimental Design

Forty healthy graduate and undergraduate students (35 male, 5 female, $M_{age} = 21.38$ years, $SD_{age} = 1.51$ years, age range: 18–23 years) participated in the experiment. All participants were ignorant of the purpose of the experiment.

As in Study 1, a single-factor two-level between-participants experimental design was used. Participants were randomly assigned to either a "static face" or a "change of facial expression" condition.

3.1.2. Apparatus

The apparatus used was identical to that used in Study 1 except that the facial expression was shown on the upper right area of the GUI as shown in **Figure 7**.

3.1.3. Procedure

The procedure was identical to that used in Study 1, except for the following changes. The experiment was conducted on two participants who knew each other. The two participants came to the experiment together and were taken to different rooms. In their different rooms, they were asked to read the instruction paper, and *both participants*were informed that they were assigned to be the *proposer*. Thus, all participants played the role of the proposer without knowing it. They were informed that "the response of your partner will be shown on the upper right area of the interface." The facial expression was automatically changed based on the position of the slider bar controlled by the participant, as in Study 1.

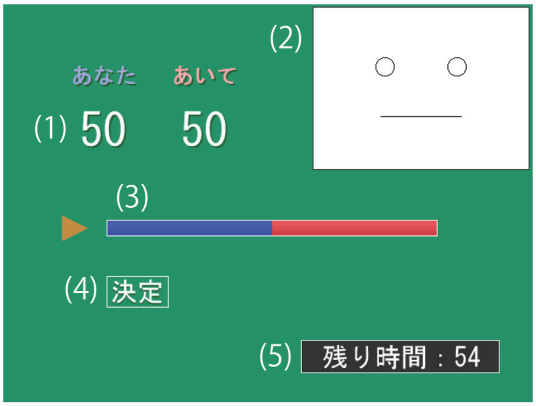


FIGURE 7 | Graphical user interface used in Study 2.

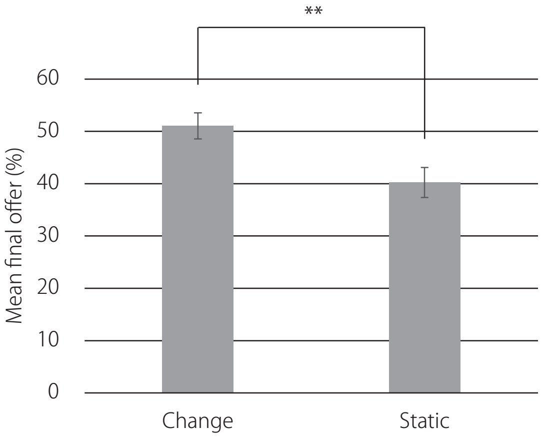


FIGURE 8 | Mean final offers averaged over participants in each of the two conditions. Error bars indicate standard errors. $^{**}p < 0.01$.

3.2. Results

The data of one participant in each of the two conditions were excluded because they reported that they realized that they were playing against a computer program.

The mean durations spent deciding the amount of the offer were 50.31 s ($SD = 12.55$) and 46.47 s ($SD = 14.68$) in the facial expression change condition and static face condition, respectively. The Mann–Whitney U-tests, $U = 162$, $z = 0.54$ $p = 0.59$, show that no statistically significant difference was observed.

Figure 8 presents the mean final offers averaged over participants in each of the two conditions. Error bars indicate standard errors of the mean value. The Mann–Whitney U-tests, $U = 92, z = 2.61, p < 0.01$, show that offers were higher in the facial expression change condition ($M = 51.05, SD = 10.88$) than in the static face condition ($M = 40.21, SD = 12.48$).

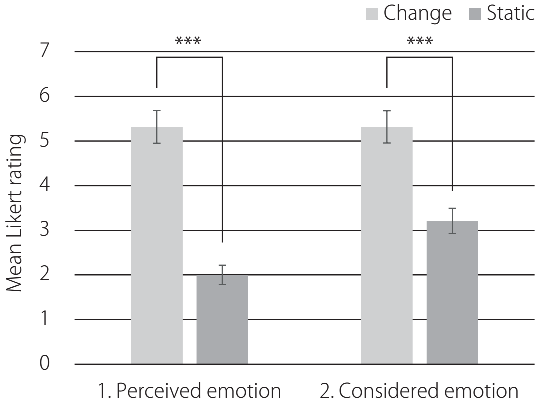


FIGURE 9 | Post-experiment questionnaire. Error bars indicate standard errors. ***$p < 0.001$.

Figure 9 shows the results of the post-experiment questionnaire. The Brunner-Munzel test, $W = 9.7, p < 0.01$, revealed that the ratings for perceiving emotions from the line-drawing of a face were significantly higher in the facial expression condition than in the static face condition. The Mann–Whitney U-tests, $U = 47.5, z = 3.95, p < 0.001$, revealed that ratings for ratings for considering this emotion were significantly higher in the facial expression condition than in the static face condition.

3.3. Discussion

The results show that offers were higher in the change of facial expression than in the static face condition, confirming that emotional expression given by an online responder through an avatar face composed of simple lines led participants to behave more altruistically to the responder. The results of the post-experiment questionnaires support the behavioral result.

4. GENERAL DISCUSSION

The behavioral and questionnaire results for Study 2 were similar to those of Study 1, confirming that emotional expression conveyed through simple line drawings representing a robot's face has the function of eliciting altruistic behavior from humans to the same extent as human telepresence.

However, it appears that the duration of time spent deciding the offer amount for a human responder was longer than that for a robot. This indicates that those participants who played the game against a human took more time to find the point of compromise. Despite this fact, interestingly, the mean final offers were almost the same between Study 1 and Study 2. It is possible that humans have a cognitive tendency to treat robots as non-negotiable partners and that this leads to a shorter duration of time spent exploring the point of compromise. However, the facial expression of the robot might have suppressed this cognitive tendency and led the participants to be more altruistic.

The results of our studies are consistent with those of previous studies (de Melo et al., 2010, 2011). The studies of de Melo et al. (2010, 2011) and ours all showed that the emotional expressions of artificial agents are effective in inducing humans to cooperate. de Melo et al. (2010, 2011) used a human-like virtual agent, whereas we used a real robot with a simple line drawing depicting its face. This implies that sophisticated human-likeness is not necessarily needed for a cooperative relationship to develop between robots and humans. This might be because facial expressions, even the face is a line drawing, are processed subcortically (Johnson, 2005; Britton et al., 2008). Nishio et al. (2012) have conducted experiments with an android robot that has a highly human-like appearance and concluded that the appearance of the agent does not affect cooperativeness. From these results, we would suggest that the ability to interact is more important than a human-like appearance for an artificial agent to develop a cooperative relationship with a human.

A substantial number of studies have shown that in economic games played by humans, facial expressions affect the decision to cooperate or not regardless of the type of game [e.g., ultimatum games (Mussel et al., 2014), prisoner's dilemma (Reed et al., 2012), dictator games (Brown and Moore, 2002), and trust games (Tortosa et al., 2013)]. Furthermore, whereas de Melo et al. (2010) used the prisoner's dilemma, we used an ultimatum game, and both studies show that the emotional expressions of artificial agents are effective in inducing humans to cooperate. Overall, it is possible that the emotional expressions of artificial agents are useful for building cooperative relationships with humans regardless of the type of game. However, long-term field study should be conducted to investigate whether the emotional expression contributes to the initiation and maintenance of real human-robot cooperative relationships.

Some limitations occur in the present study. First, while our participants were selected from a small, culturally homogeneous population and the gender ratio was not controlled, studies have suggested that culture (Russell, 1994; Hess et al., 2000; Mandal and Ambady, 2004) and gender (Hess et al., 2000; Mussel et al., 2014) influence the expression and interpretation of emotions. Larger and more diverse samples should be used to examine gender and cultural effects on human-robot cooperative relationships mediated by emotional expressions. Second, we used a small humanoid robot, and its facial expression was shown on an LCD monitor that was mounted as its head. Further, investigation using various types of robots such as life-sized humanoid robots and robots with sophisticated facial expression mechanisms should be performed to generalize the findings.

AUTHOR CONTRIBUTIONS

KT designed the experiments; CT performed the experiments; KT and CT analyzed the data; and KT wrote the manuscript.

ACKNOWLEDGMENTS

This study was partially supported by JSPS KAKENHI No. 15H02735 and "Cognitive Interaction Design" (No.26118005). We would like to thank IEEE for permission to partially reprint from Terada et al. (2013).

REFERENCES

Alexander, R. D. (1987). *The Biology of Moral Systems*. New York, NY: Aldine De Gruyter.

Bartneck, C. (2003). "Interacting with an embodied emotional character," in *Proceedings of the 2003 International Conference on Designing Pleasurable Products and Interfaces, DPPI '03* (New York, NY: ACM), 55–60.

Becker-Asano, C., and Ishiguro, H. (2011). "Evaluating facial displays of emotion for the android robot geminoid f," in *IEEE Workshop on Affective Computational Intelligence (WACI)* (Paris), 1–8.

Bethel, C. L., and Murphy, R. R. (2007). "Non-facial/non-verbal methods of affective expression as applied to robot-assisted victim assessment," in *Proceedings of the 2nd ACM/IEEE International Conference on Human-Robot Interaction, HRI '07* (New York, NY: ACM), 287–294.

Bethel, C. L., and Murphy, R. R. (2008). Survey of non-facial/non-verbal affective expressions for appearance-constrained robots. *IEEE Trans. Syst. Man Cybernet. C Appl. Rev.* 38, 83–92. doi: 10.1109/TSMCC.2007.905845

Bickmore, T. W., and Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Trans. Comp. Hum. Interact.* 12, 293–327. doi: 10.1145/1067860.1067867

Breazeal, C. (2003). Emotion and sociable humanoid robots. *Int. J. Hum. Comp. Stud.* 59, 119–155. doi: 10.1016/S1071-5819(03)00018-1

Breazeal, C. (2004). Function meets style: insights from emotion theory applied to HRI. *IEEE Trans. Syst. Man. Cybernet. C Appl. Rev.* 34, 187–194. doi: 10.1109/TSMCC.2004.826270

Britton, J. C., Shin, L. M., Barrett, L. F., Rauch, S. L., and Wright, C. I. (2008). Amygdala and fusiform gyrus temporal dynamics: responses to negative facial expressions. *BMC Neurosci.* 9:44. doi: 10.1186/1471-2202-9-44

Brown, W. M., and Moore, C. (2002). Chapter 3: smile asymmetries and reputation as reliable indicators of likelihood to cooperate: an evolutionary analysis," in *Advances in Psychology Research,* Vol. 11 (New York, NY: Nova Science Publishers), 59–78.

Brown, W. M., Palameta, B., and Moore, C. (2003). Are there nonverbal cues to commitment? an exploratory study using the zero-acquaintance video presentation paradigm. *Evol. Psychol.* 1, 42–69. doi: 10.1177/147470490300100104

Cassell, J., and Thorisson, K. R. (1999). The power of a nod and a glance: envelope vs. emotional feedback in animated conversational agents. *Appl. Artif. Intell.* 13, 519–538.

Croson, R., Boles, T., and Murnighan, J. K. (2003). Cheap talk in bargaining experiments: lying and threats in ultimatum games. *J. Econ. Behav. Organ.* 51, 143–159. doi: 10.1016/S0167-2681(02)00092-6

de Melo, C. M., Carnevale, P., and Gratch, J. (2010). "The influence of emotions in embodied agents on human decision-making," in *Intelligent Virtual Agents, 10th International Conference, IVA 2010, Philadelphia, PA, USA. Proceedings*, eds J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, and A. Safonova (Berlin; Heidelberg: Springer), 357–370. Available online at: http://www.springer.com/us/book/9783642158919#otherversion=9783642158926

de Melo, C. M., Carnevale, P., and Gratch, J. (2011). "The effect of expression of anger and happiness in computer agents on negotiations with humans," in *The 10th International Conference on Autonomous Agents and Multiagent Systems, Vol. 3, AAMAS '11*, ed S. C. Richland (Taipei: International Foundation for Autonomous Agents and Multiagent Systems), 937–944.

Ekman, P. (2003). *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life*. New York, NY: Times Books.

Fabiansson, E. C., and Denson, T. F. (2012). The effects of intrapersonal anger and its regulation in economic bargaining. *PLoS ONE* 7:e51595. doi: 10.1371/journal.pone.0051595

Frank, R. H. (1988). *Passions within Reason: The Strategic Role of the Emotions*. New York, NY: W.W. Norton & Co.

Güth, W., and Tietz, R. (1990). Ultimatum bargaining behavior : a survey and comparison of experimental results. *J. Econ. Psychol.* 11, 417–449.

Hess, U., Blairy, S., and Kleck, R. E. (2000). The influence of facial emotion displays, gender, and ethnicity on judgments of dominance and affiliation. *J. Nonverb. Behav.* 24, 265–283. doi: 10.1023/A:1006623213355

Itoh, K., Miwa, H., Nukariya, Y., Zecca, M., Takanobu, H., Roccella, S., et al. (2006). "Mechanisms and functions for a humanoid robot to express human-like emotions," in *Proceedings of the 2006 IEEE International Conference on Robotics and Automation* (Orlando, FL), 4390–4392.

Johnson, M. H. (2005). Subcortical face processing. *Nat. Rev. Neurosci.* 6, 766–774. doi: 10.1038/nrn1766

Kanoh, M., Kato, S., and Itoh, H. (2004). "Facial expressions using emotional space in sensitivity communication robot "ifbot"," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2004 (IROS 2004)*, *Vol. 2* (Sendai), 1586–1591.

Katsikitis, M. (1997). The classification of facial expressions of emotion: a multidimensional-scaling approach. *Perception* 26, 613–626.

Kim, E. H., Kwak, S., and Kwak, Y. K. (2009a). "Can robotic emotional expressions induce a human to empathize with a robot?" in *The 18th IEEE International Symposium on Robot and Human Interactive Communication, 2009, RO-MAN 2009* (IEEE), 358–362.

Kim, E. H., Kwak, S. S., Han, J., and Kwak, Y. K. (2009b). "Evaluation of the expressions of robotic emotions of the emotional robot, "mung"," in *Proceedings of the 3rd International Conference on Ubiquitous Information Management and Communication, ICUIMC '09* (New York, NY: ACM), 362–365.

Leyzberg, D., Avrunin, E., Liu, J., and Scassellati, B. (2011). "Robots that express emotion elicit better human teaching," in *HRI '11: Proceeding of the 6th ACM/IEEE International Conference on Human Robot Interaction, HRI '11* (New York, NY: ACM), 347–354.

Mandal, M. K., and Ambady, N. (2004). Laterality of facial expressions of emotion: Universal and culture-specific influences. *Behav. Neurol.* 15, 23–34. doi: 10.1155/2004/786529

Matsui, Y., Kanoh, M., Kato, S., Nakamura, T., and Itoh, H. (2010). A model for generating facial expressions using virtual emotion based on simple recurrent network. *J. Adv. Comput. Intell. Intellig. Inform.* 14, 453–463. doi: 10.20965/jaciii.2010.p0453

Mazzei, D., Lazzeri, N., Billeci, L., Igliozzi, R., Mancini, A., Ahluwalia, A., Muratori, F., and Rossi, D. D. (2011). "Development and evaluation of a social robot platform for therapy in autism," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, 2011* (Boston, MA), 4515–4518.

Mehu, M., Grammer, K., and Dunbar, R. I. (2007). Smiles when sharing. *Evol. Hum. Behav.* 28, 415–422. doi: 10.1016/j.evolhumbehav.2007.05.010

Mussel, P., Göritz, A. S., and Hewig, J. (2013). The value of a smile: facial expression affects ultimatum-game responses. *Judgm. Decis. Making* 8, 381–385. Available online at: http://journal.sjdm.org/12/12817/jdm12817.html

Mussel, P., Hewig, J., Allen, J. J. B., Coles, M. G. H., and Miltner, W. (2014). Smiling faces, sometimes they don't tell the truth: facial expression in the ultimatum game impacts decision making and event-related potentials. *Psychophysiology* 51, 358–363. doi: 10.1111/psyp.12184

Nishio, S., Ogawa, K., Kanakogi, Y., Itakura, S., and Ishiguro, H. (2012). "Do robot appearance and speech affect people's attitude? Evaluation through the ultimatum game," in *The 21th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN2012)* (Paris), 809–814.

Nowak, M. A., and Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature* 437, 1291–1298. doi: 10.1038/nature04131

Oosterbeek, H., Sloof, R., and van de Kuilen, G. (2004). Cultural differences in ultimatum game experiments: evidence from a meta-analysis. *Exp. Econ.* 7, 171–188. doi: 10.1023/B:EXEC.0000026978.14316.74

Reed, L. I., DeScioli, P., and Pinker, S. A. (2014). The commitment function of angry facial expressions. *Psychol. Sci.* 25, 1511–1517. doi: 10.1177/0956797614531027

Reed, L. I., Zeglen, K. N., and Schmidt, K. L. (2012). Facial expressions as honest signals of cooperative intent in a one-shot anonymous prisoner's dilemma game. *Evol. Hum. Behav.* 33, 200–209. doi: 10.1016/j.evolhumbehav.2011.09.003

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? a review of the cross-cultural studies. *Psychol. Bull.* 115, 102–141.

Sandoval, E. B., Brandstetter, J., Obaid, M., and Bartneck, C. (2016). Reciprocity in human-robot interaction: a quantitative approach through the prisoner's dilemma and the ultimatum game. *Int. J. Soc. Robot.* 8, 303–317. doi: 10.1007/s12369-015-0323-x

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science* 300, 1755–1758. doi: 10.1126/science.1082976

Scharlemann, J. P. W., Eckel, C. C., Kacelnik, A., and Wilson, R. K. (2001). The value of a smile: game theory with a human face. *J. Econ. Psychol.* 22, 617–640. doi: 10.1016/S0167-4870(01)00059-9

Sell, A., Tooby, J., and Cosmides, L. (2009). Formidability and the logic of human anger. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15073–15078. doi: 10.1073/pnas.0904312106

Shimokawa, T., and Sawaragi, T. (2001). "Acquiring communicative motor acts of social robot using interactive evolutionary computation," in *IEEE International Conference on Systems, Man, and Cybernetics*, *Vol. 3* (Anchorage, AK), 1396–1401.

Sinaceur, M., and Tiedens, L. Z. (2006). Get mad and get more than even: when and why anger expression is effective in negotiations. *J. Exp. Soc. Psychol.* 42, 314–322. doi: 10.1016/j.jesp.2005.05.002

Sugano, S., and Ogata, T. (1996). Emergence of mind in robots for human interface - research methodology and robot model. *IEEE Int. Conf. Robot. Automat.* 2, 1191–1198.

Terada, K., Takeuchi, C., and Ito, A. (2013). "Effect of emotional expression in simple line drawings of a face on human economic behavior," in *The 22th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2013)* (Gyeongju), 51–56.

Terada, K., Yamauchi, A., and Ito, A. (2012). "Artificial emotion expression for a robot by dynamic color change," in *The 21th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2012)* (Paris), 314–321.

Torta, E., van Dijk, E., Ruijten, P. A. M., and Cuijpers, R. H. (2013). *The Ultimatum Game as Measurement Tool for Anthropomorphism in Human–Robot Interaction.* Cham: Springer International Publishing.

Tortosa, M. I., Strizhko, T., Capizzi, M., and Ruz, M. (2013). Interpersonal effects of emotion in a multi-round trust game. *Psicológica* 34, 179–198. Available online at: https://www.uv.es/psicologica/articulos2.13/3Tortosa.pdf

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Q. Rev. Biol.* 46, 35–57.

van Dijk, E., van Kleef, G. A., Steinel, W., and van Beest, I. (2008). A social functional approach to emotions in bargaining: when communicating anger pays and when it backfires. *J. Pers. Soc. Psychol.* 94, 600–614. doi: 10.1037/0022-3514.94.4.600

van Kleef, G. A., and Côté, S. (2007). Expressing anger in conflict: when it helps and when it hurts. *J. Appl. Psychol.* 92, 1557–1569. doi: 10.1037/0021-9010.92.6.1557

van Kleef, G. A., Dreu, C. K. W. D., and Manstead, A. S. R. (2004). The interpersonal effects of anger and happiness in negotiations. *J. Pers. Soc. Psychol.* 86, 57–76. doi: 10.1037/0022-3514.86.1.57

van Kleef, G. A., van Dijk, E., Steinel, W., Harinck, F., and van Beest, I. (2008). Anger in social conflict: cross-situational comparisons and suggestions for the future. *Group Decis. Negot.* 17, 13–30. doi: 10.1007/s10726-007-9092-8

Xiao, E., and Houser, D. (2005). Emotion expression in human punishment behavior. *Proc. Natl. Acad. Sci. U.S.A.* 102, 7398–7401. doi: 10.1073/pnas.0502399102

Yamagishi, T., Horita, Y., Takagishi, H., Shinada, M., Tanida, S., and Cook, K. S. (2009). The private rejection of unfair offers and emotional commitment. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11520–11523. doi: 10.1073/pnas.0900636106

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Terada and Takeuchi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in ICT

ORIGINAL RESEARCH
published: 17 March 2017
doi: 10.3389/fict.2017.00003

A First Step toward the Automatic Understanding of Social Touch for Naturalistic Human–Robot Interaction

Merel M. Jung, Mannes Poel, Dennis Reidsma and Dirk K. J. Heylen*

Human Media Interaction, University of Twente, Enschede, Netherlands

Social robots should be able to automatically understand and respond to human touch. The meaning of touch does not only depend on the form of touch but also on the context in which the touch takes place. To gain more insight into the factors that are relevant to interpret the meaning of touch within a social context we elicited touch behaviors by letting participants interact with a robot pet companion in the context of different affective scenarios. In a contextualized lab setting, participants (n = 31) acted as if they were coming home in different emotional states (i.e., stressed, depressed, relaxed, and excited) without being given specific instructions on the kinds of behaviors that they should display. Based on video footage of the interactions and interviews we explored the use of touch behaviors, the expressed social messages, and the expected robot pet responses. Results show that emotional state influenced the social messages that were communicated to the robot pet as well as the expected responses. Furthermore, it was found that multimodal cues were used to communicate with the robot pet, that is, participants often talked to the robot pet while touching it and making eye contact. Additionally, the findings of this study indicate that the categorization of touch behaviors into discrete touch gesture categories based on dictionary definitions is not a suitable approach to capture the complex nature of touch behaviors in less controlled settings. These findings can inform the design of a behavioral model for robot pet companions and future directions to interpret touch behaviors in less controlled settings are discussed.

Keywords: social touch, human–robot interaction, robot pet companion, multimodal interaction, touch recognition, behavior analysis, affective context

OPEN ACCESS

Edited by:
Hatice Gunes,
University of Cambridge, UK

Reviewed by:
Christian Becker-Asano,
Robert Bosch, Germany
Maria Koutsombogera,
Trinity College Dublin, Ireland

Correspondence:
Merel M. Jung
m.m.jung@utwente.nl

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in ICT

Received: *31 December 2016*
Accepted: *27 February 2017*
Published: *17 March 2017*

Citation:
Jung MM, Poel M, Reidsma D and Heylen DKJ (2017) A First Step toward the Automatic Understanding of Social Touch for Naturalistic Human–Robot Interaction. Front. ICT 4:3. doi: 10.3389/fict.2017.00003

1. INTRODUCTION

Touch plays an important role in establishing and maintaining social interaction (Gallace and Spence, 2010). In interpersonal interaction, this modality can be used to communicate emotions and other social messages (Jones and Yarbrough, 1985; Hertenstein et al., 2006, 2009). More recently, the study of social touch was also extended to interaction with humanoid and robotic animals (e.g., Knight et al., 2009; Kim et al., 2010; Yohanan and MacLean, 2012; Cooney et al., 2015). In order to make these interactions more natural, robots should be able to understand and respond to human touch.

A social robot needs to sense and recognize different touch gestures (e.g., Kim et al., 2010; Silvera-Tawil et al., 2012; Altun and MacLean, 2015; Jung et al., 2015, 2016) and should be able to interpret touch in order to respond in an appropriate manner (see **Figure 1**). Perhaps robot seal Paro is the most famous example of a social robot that responds to touch (Wada and Shibata, 2007). Paro is equipped with touch sensors with which it distinguishes between soft touches (which are always interpreted to be positive) and rough touches (which are always interpreted to be negative) (Wada and Shibata, 2007). This interpretation of touch is oversimplified as the complexity of the human tactile system allows for touch behaviors to vary not only depending on the intensity but also based on movement, velocity, abruptness, temperature, location, and duration (Hertenstein et al., 2009). Moreover, the meaning of touch can often not be inferred from the type of touch alone but is also dependent on other factors such as concurrent verbal and non-verbal behavior, the type of interpersonal relationship (Heslin et al., 1983; Suvilehto et al., 2015), and the situation in which the touch takes place (Jones and Yarbrough, 1985). Although previous research (Heslin et al., 1983; Hertenstein et al., 2006, 2009) indicated that there is no one-to-one mapping of touch gestures to a specific meaning of touch, touch can have a clear meaning in a specific context (Jones and Yarbrough, 1985).

The current study focuses on (touch) interaction with a robot pet companion. According to Veevers, a pet companion can fulfill different roles in the life of humans, a pet can facilitate interpersonal interaction or can even serve as a surrogate for interpersonal interaction, and expensive and/or exotic pets can be owned as a status symbol (Veevers, 1985). Furthermore, interaction with pet companions is associated with health benefits, and more recent studies indicate that these effects also extend to interaction with robot pets (Eachus, 2001; Banks et al., 2008). Although touch is a natural way to interact with real pets, currently commercially available robot pets such as Paro (Wada and Shibata, 2007), Hasbro's companion pets,[1] and JustoCat[2] are equipped with only a few touch sensors and do not interpret different types of touch within context.

We argue that the recognition and interpretation of touch consists of three levels: (1) low-level touch parameters such as intensity, duration, and contact area; (2) mid-level touch gestures such as pat, stroke, and tickle; and (3) high-level social messages such as affection, greeting, and play. To automatically understand social touch, research focuses on investigating the connection between these levels. Current studies in the domain of social touch for human–robot interaction focused mainly on highly controlled settings in which users were requested to perform different touch behaviors, one at the time, according to predefined labels (e.g., Cooney et al., 2012; Silvera-Tawil et al., 2012, 2014; Yohanan and MacLean, 2012; Jung et al., 2015, 2016). In this study we focus on the latter two levels as we are interested in the meaning of touch behaviors. To gain more insight into the factors that are relevant to interpret touch behaviors within social context, we opted to elicit touch behaviors by letting participants act out four scenarios in which they interacted with a robot pet companion in different emotional states. Moreover, in contrast to most previous studies, participants could freely act out the given scenarios with the robot pet within the confined space of a living room setting.

In this paper, we present contributions in two areas. First, we explore the use of touch behaviors as well as the expressed social messages and expected robot pet responses in different affective scenarios. Second, we reflect upon the challenges of the segmentation and labeling of touch behaviors in a less controlled setting in which no specific instructions are given on the kinds of (touch) behaviors that should be displayed. We address the first contribution with the following three research questions. (RQ1) What kinds of touch gestures are used to communicate with a robot pet in the different affective scenarios? (RQ2) Which social messages are communicated, and what is the expected response in the different affective scenarios? (RQ3) What other social signals can aid the interpretation of touch behaviors? Furthermore, we reflect upon our effort to segment and label touch behaviors in a less controlled setting with the fourth research question. (RQ4) How well do annotation schemes work in a contextualized lab situation?

The remainder of the article is structured as follows. Related work on the meaning of social touch in both interpersonal and human–robot interaction will be discussed in the next section followed by the description of the materials and methods for the presented study. Then, the results will be provided and discussed in the subsequent sections. Conclusions will be drawn in the last section.

2. RELATED WORK

Previous studies have looked into the meaning of touch in both interpersonal interaction (Jones and Yarbrough, 1985) and human–robot interaction with either a humanoid robot (Kim et al., 2010; Silvera-Tawil et al., 2014; Cooney et al., 2015) or a robot animal (Knight et al., 2009; Yohanan and MacLean, 2012). In a diary study on the use of interpersonal touch, different meanings of touch were categorized based on the participants' verbal translations of the touch interactions (Jones and Yarbrough, 1985). Seven main categories were distinguished: positive affect touches (e.g., support), playful touches (e.g., playful affection), control touches (e.g., attention-getting), ritualistic touches (e.g., greeting), hybrid touches (e.g., greeting/affection), task-related touches (instrumental intrinsic), and accidental touches. Interestingly, there was a lack of reports on negative interpersonal touch interaction. Within these categories, common contextual factors were identified such as the type of touch, any

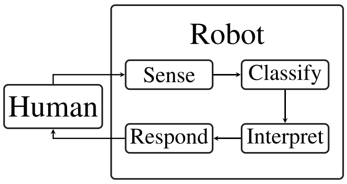


FIGURE 1 | Interaction cycle for a socially intelligent robot that can respond to human touch.

[1] http://joyforall.hasbro.com.

[2] http://justocat.com.

accompanying verbal statement, and the situation in which the touch took place. It was found that depending on the context, a specific form of touch can have multiple meanings and that different forms of touch can have a similar meaning. Furthermore, touch was found to be often preceded, accompanied, or followed by a verbal statement.

In a study on human–robot interaction, participants were asked to indicate which touch gestures they were likely to use to communicate emotional states to a cat-sized robot animal (Yohanan and MacLean, 2012). Gestures that were judged to be likely used were performed sequentially on the robot. Participants expected that the robot's emotional response was either similar or sympathetic to the emotional state that was communicated. The nature of the touch behavior was found to be friendly as no aggressive gestures (e.g., slap or hit) were used even when negative emotions were communicated. Five categories of intent were distinguished based on touch gesture characteristics that could be mapped to emotional states: affectionate, comforting, playful, protective, and restful. Also, video segments of the touch gestures were annotated to characterize the gestures based on the point of contact, intensity, and duration revealing differences between touch gestures and their use in different emotional states. In follow-up research, the touch sensor data recorded in this study (i.e., Yohanan and MacLean, 2012) were used to classify 26 touch gestures and 9 emotional states using random forests (Altun and MacLean, 2015). Between-subjects emotion recognition of 9 emotional states yielded an accuracy of 36%, while within-subjects the accuracy was 48%. Between-subjects touch gesture recognition of 26 gestures yielded an accuracy of 33%. Furthermore, the authors' results indicated that accurate touch gesture recognition could improve affect recognition.

In other work, Kim et al. (2010) instructed participants to use four different touch gestures to give positive or negative feedback to a humanoid robot while playing a game. A model was trained to infer whether a touch gesture was meant as a positive or a negative reward for the robot. It was found that participants used "pat" and "rub" to give positive feedback and "hit" to give negative feedback, while "push" could be used for both although the touch gesture was mostly used for negative feedback. Knight et al. (2009) argued for the importance of body location as contextual factor to infer the meaning of touch. The authors made the distinction between what they called *symbolic gestures*, which have social significance based on the involved body location(s) (e.g., footrub and hug) and body location-independent *touch subgestures* (e.g., pat and poke).

Although previous studies indicate that there is a link between touch gestures and the higher level social meaning of touch, Silvera-Tawil et al. (2014) argued that the meaning of touch could also be recognized directly based on characteristics from touch sensor data and other factors such as the context and the touch location. In their effort to automatically recognize emotions and social messages directly from sensor data without first recognizing the used touch gestures, participants were asked to perform six basic emotions: anger, disgust, fear, happiness, sadness, and surprise on both a mannequin arm with an artificial skin and a human arm. In addition, six social messages were communicated: acceptance, affection, animosity, attention-getting, greeting, and rejection. Recognition rates for the emotions were 46.9% for the algorithm and 51.8% for human classification. The recognition rates for the social messages were found to be slightly higher, yielding accuracies of 49.7 and 62.1% for the algorithm and human classification, respectively.

Some attempts have been made to study touch interaction in a less controlled setting, for example, Noda et al. (2007) elicited touch during the interaction with a humanoid robot by designing a scenario in which participants used different touch gestures to communicate a particular social message such as greeting the robot by shaking hands, playing together by tickling the robot, and hugging the robot to say goodbye at the end of the interaction. Results showed an accuracy of over 60% for the recognition of the different touch behaviors that were performed within the scenario. In another study on the use of touch in multimodal human–robot interaction, participants were given various reasons to interact with a small humanoid robot such as giving reassurance, getting attention, and giving approval (Cooney et al., 2015). The robot was capable of recognizing touch, speech, and visual cues, and participants were free to use different modalities. Also, participants rated videos in which a confederate interacted with the robot using different modalities. Results showed that touch was often used to communicate with the robot and that touch was especially important for expressing affection. Furthermore, playing with the robot and expressing loneliness were deemed more suitable than displaying negative emotions.

To summarize, previous studies illustrate that touch can be used to express and communicate different kinds of affective and social messages (Jones and Yarbrough, 1985; Yohanan and MacLean, 2012; Silvera-Tawil et al., 2014; Cooney et al., 2015). Moreover, touch gestures that were used to communicate were often positive in nature, and their meaning is dependent on the context such as one's emotional state (Jones and Yarbrough, 1985; Yohanan and MacLean, 2012; Cooney et al., 2015). These findings confirm that currently available robot pet companions, such as Paro, which only distinguishes between positive and negative touch, are not sufficiently capable of understanding and responding to people in a socially appropriate way. Furthermore, there are indications that other modalities might be helpful in interpreting the social messages as touch behavior generally does not occur in isolation (Jones and Yarbrough, 1985; Cooney et al., 2015). For the reasons outlined above, we opted to study interactions between a human and a robot pet companion in the context of different emotional states in a contextualized lab situation.

3. MATERIALS AND METHODS

In this study we elicited interactions between a human and a robot pet companion in a lab-build living room setting. Participants were instructed to act as if they would come home in different emotional states (i.e., stressed, depressed, relaxed, and excited). These four emotional states were chosen as they span opposite ends of the valence and arousal scale (see **Figure 2**): stressed (low valence, high arousal), depressed (low valence, low arousal), relaxed (high valence, low arousal), and excited (high valence, high arousal) (Russell et al., 1989). Furthermore, similar emotional states have been used in a more controlled research setting before, and the

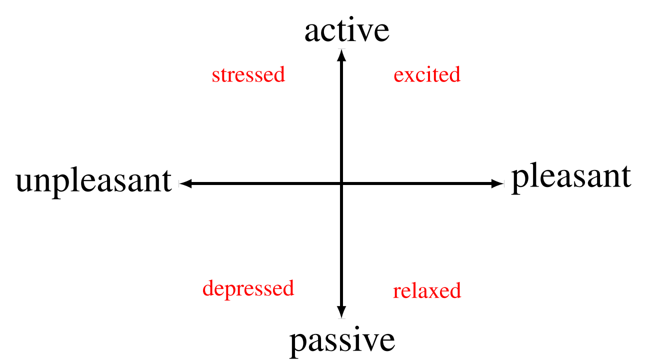


FIGURE 2 | **Mapping of emotional state based on associated valence and arousal levels, model adapted from Russell et al. (1989).**

results from this study indicate that emotional state influences touch behavior as well as the expected robot response (Yohanan and MacLean, 2012). To gain more insight into the factors that are relevant to interpret touch within a social context, we annotated touch behaviors from video footage of the interactions. Also, a questionnaire was administered and interviews were conducted to interpret the high-level meaning behind the interactions and get insight into the responses that would be expected from the robot pet.

3.1. Participants

In total 31 participants (20 males, 11 females) volunteered to take part in the study. The age of the participants ranged from 22 to 64 years (M = 34.3; SD = 12.8), and 28 were right-handed, 2 left-handed, and 1 ambidextrous. All studied or worked at the University of Twente in the Netherlands. Most (21) had the Dutch nationality; others were Belgian, Ecuadorean, English, German (2×), Greek, Indian, Iranian, Italian, and South Korean. This study was approved by the ethics committee of the Faculty of Electrical Engineering, Mathematics and Computer Science of the University of Twente. All research participants provided written informed consent in accordance with the Declaration of Helsinki.

3.2. Apparatus/Materials

3.2.1. Living Room Setting

The living room setting consisted of a space of approximately 23 m^2 containing a small couch, a coffee table, and two plants (see **Figure 3**, left). Two camcorders were positioned facing the couch at an approximately 45° angle to record the interactions (50 fps, 1,080 p). To allow participants to interact freely with the robot pet (i.e., no wires) and have a controlled interaction (i.e., no unpredictable robot behavior), a stuffed animal dog was used as a proxy for a robot pet. The robot pet (35 cm; in a laying position) was positioned on the couch at the far end from the door facing the table (see **Figure 3**).

3.2.2. Questionnaire

The questionnaire was divided into two parts. Part one was completed before the interview was conducted and part two after the interview. Part one consisted of demographics: gender, age, nationality, occupation, and handedness followed by six questions about the reenactment of the scenarios rated on a 4-point Likert scale ranging from 1 (strongly disagree) to 4 (strongly agree). Four questions were about the participants' ability to imagine themselves in the scenarios: "I was able to imagine myself coming home feeling stressed/depressed/relaxed/excited." The other two questions were about the robot pet: "I was able to imagine that the pet was a functional robot" and "I based my interaction with the robot pet on how I interact with a real animal."

The second part consisted of a questionnaire about the expectations of living with a robot pet, which was based on the 11-item Comfort from Companion Animals Scale (CCAS) (Zasloff, 1996). Participants were asked to imagine that they would get a robot pet like the one in the study as a gift. This robot pet can react to touch and verbal commands. Participants were asked to answer the questions about the role they expect the robot pet would play in their life. The questions from the CCAS were adjusted to fit the purpose of the study, for example, the item "my pet provides me with companionship" was changed to "I expect my robot pet to provide me with companionship." Items were rated on a 4-point Likert scale ranging from 1 (strongly disagree) to 4 (strongly agree), as all items were phased positively a higher score indicates greater expected comfort from the robot pet.

3.2.3. Interview

A semi-structured interview was conducted between the first and the second part of the questionnaire. The video footage of their reenactment of the scenarios was shown to the participants, and they were asked to answer the following questions after watching each of the four scenario fragments: (1) "What message did you want to communicate to the robot?" (2) "What response would you expect from the robot?" (3) "How could the robot express this?" The participant, the interviewer, and the computer screen were recorded during the interview using a camcorder.

3.3. Procedure

Upon entering the room in which the study took place, the participant was welcomed and was asked to read the instructions and sign an informed consent form. Then, participants were taken into the hallway where they received the instructions for the example scenario in which they were asked to act out coming home in a neutral mood. If the instructions were clear, participants were asked to interact with a robot pet by acting out four different scenarios, one by one, in which they would come home in a particular emotional state, feeling stressed, depressed, relaxed, or excited. The study had a within-subject design; instructions for each of the scenarios were given to each of the participants in random order. In each scenario, the participant was instructed to enter the "living room," sit down on the couch, and act out the scenario as he/she sees fit. Participants were instructed to focus on the initial interaction as the robot pet would not respond (≈30 s were given as a guideline); however, the duration of the interaction was up to the participant who was instructed to return to the hallway when he or she finished an interaction. When the participant had returned to the hallway at the end of an interaction the next scenario was provided.

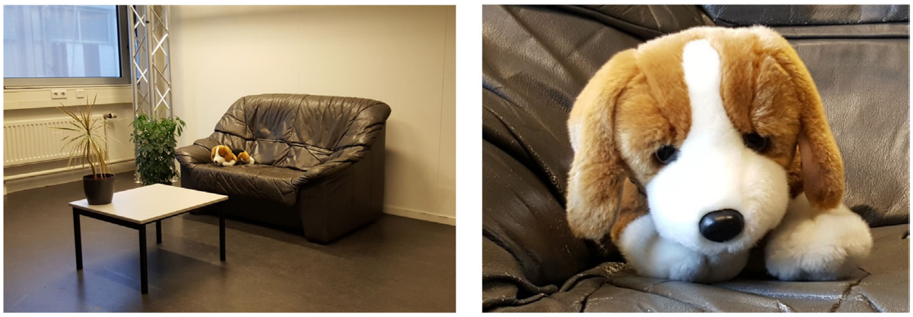


FIGURE 3 | **The living room setting with the robot pet on the couch (left) and the pet up-close (right)**.

After the last scenario, the participant was asked to fill out a questionnaire asking about demographic information and about acting out the scenarios. Then, the video footage of their reenactment of the scenarios was shown to the participant, and an interview was conducted on these interactions. After the interview, the participants completed the second part of the questionnaire about their expectations if they would own a functional robot pet. The entire procedure took approximately 20 min for each participant. At the end of the study, participants were offered a candy bar to thank them for participating.

3.4. Data Analysis

3.4.1. Questionnaire

The questionnaire data were analyzed using IBM SPSS Statistics version 22. The median values and the 25th and 75th percentiles (i.e., Q1 and Q3, respectively) were calculated for the questions about the reenactment of the scenarios. The ratings on the items of the expected comfort from the robot pet scale were summed before calculating these descriptive statistics. Additionally, a Friedman test was conducted to check whether there was a statistical difference between the perceived ability of the participants to imagine themselves in the different scenarios. The significance threshold was set at 0.05, and the exact *p*-value is reported for a two-tailed test.

3.4.2. Annotation of Scenario Videos

The video footage from the two cameras were synced and put together in a split screen video before annotation. Videos were coded by two annotators, which included one of the authors (Merel M. Jung), henceforth "the first coder," using the ELAN[3] annotation software.

For the segmentation of touch behaviors we followed a method that is commonly used to segment signs and co-speech gestures into movement units, which in the simplest form consist of three phases: a preparation phase, an expressive phase, and a retraction phase (Kendon, 1980; Kita et al., 1997). The onset of a movement unit is defined at the first indication of the initiation of a movement that is usually preceded by the departure of the hand's resting position. The end of a movement unit is defined as the moment when the hand makes first contact with a resting surface such as the lap or an arm rest. Similarly, the touch actions were segmented by the first coder from the moment that the participant reached out to the robot pet to make physical contact until the contact with the pet was ended and the hands of the participants returned to the resting position. Per touch action segment, the following information was coded by the two annotators in a single annotation tier: the performed sequence of touch gestures and the robot pet's body part(s) on which each touch gesture was performed (see **Figure 4**). The touch gesture categories consisted of the 30 touch gestures plus their definitions from the touch dictionary of Yohanan and MacLean (2012), which is a list of plausible touch gestures for interaction with a robot pet. Furthermore, based on observations we added an additional category for *puppeteering*, which was defined as "participant puppeteers the robot pet to pretend that it is moving on its own" and to reduce forced-choice we added *other*, which was defined as "the touch gesture performed cannot be described by any of the previous categories." The robot pet's body parts were divided into six categories: head (i.e., back, top, and sides of the head and ears), face (i.e., forehead, eyes, nose, mouth, cheeks, and chin), body (i.e., neck, back, and sides), belly, legs, and tail.

Coding the touch behaviors that were performed during each touch segment proved to be difficult. Both annotators were often unsure when to define the start of a new touch gesture as gestures were often followed up in quick succession. Furthermore, hybrid forms of several touch gestures were often observed resulting in difficulties to categorize the touch behavior into one of the categories. In **Table 1** some of the touch gestures are listed that were frequently observed but that were also difficult to distinguish based on their dictionary definitions. These touch gestures are all of relatively long duration compared to quick gestures such as pat and slap, and all include movement across the contact area. The distinguishing features are based on the gesture's intensity, human contact point (e.g., whole hand vs. fingernails), and the movement pattern (e.g., back and forth or seemingly random). An example of commonly encountered confusion was in cases where the hand was moved repeatedly back and forth on the fur of the robot pet, which indicated the use of a rub gesture, while the use of gentle pressure seemed to indicate a stroke-like gesture.

[3] Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands; http://tla.mpi.nl/tools/tla-tools/elan.

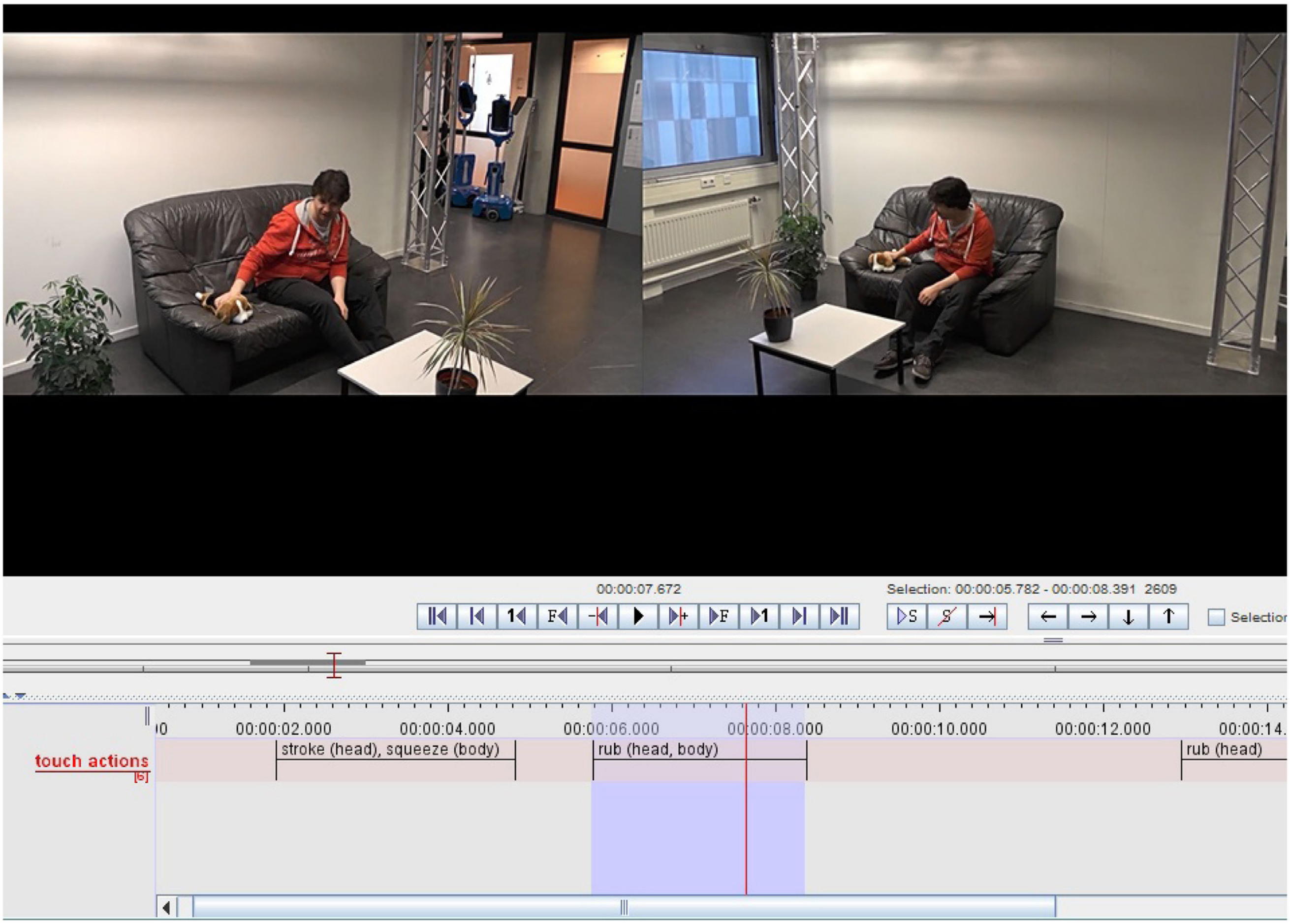


FIGURE 4 | Screenshot of the annotation process showing the annotation tier in which the touch gestures and the body location on the robot pet are annotated.

TABLE 1 | Example touch gesture categories with definitions, adapted from Yohanan and MacLean (2012).

Gesture label	Gesture definition
Massage	Rub or knead the robot pet with your hands
Rub	Move your hand repeatedly back and forth on the fur of the robot pet with firm pressure
Scratch	Rub the robot pet with your fingernails
Stroke	Move your hand with gentle pressure over the robot pet's fur, often repeatedly
Tickle	Touch the robot pet with light finger movements

Furthermore, the use of video footage to code touch gestures made it difficult to determine the exact point of contact, which is the only differentiating feature to distinguish between a rub and a scratch gesture based on these definitions.

Even after several iterations of revisiting the codebook in order to clarify what the distinguishing features of several touch gestures are, it was still not possible to reach an acceptable level of agreement. Difficulties were caused by a mixture of touch events that were hard to observe on video and differences in interpretation by the annotators, which included both the segmentation of individual touch gestures (i.e., within the larger predefined segments) and the assignment of labels, despite the commonly developed annotation scheme. Furthermore, as one touch segment could consist of a sequence of touch gestures, it was difficult to calculate the inter-rater reliability (i.e., Cohen's kappa) as the number of touch gestures could differ per coder. The location of the touch gestures on the robot pet's body was related to the coding of the touch gestures themselves, and therefore it was also not possible to reach an acceptable agreement on this part.

Due to the difficulties described above we decided instead to coarsely describe the interactions in the results section based on the modalities that the participants used to communicate to the robot pet. Also, a Friedman test was conducted to check whether there was a statistical difference between the duration of the interactions in the different scenarios. The significance threshold was set at 0.05, and the exact *p*-value is reported for a two-tailed test. The implications of the findings from the annotation process will be explicated in the discussion section.

3.4.3. Interview

The interview answers were grouped per scenario based on common themes. The data were split into two parts. (1) Information on the social messages (and possible behaviors to express those) that were communicated by the participant to the robot pet. (2) Information on the expected messages and behaviors that were expected to be communicated by the robot pet. Themes were labeled, and the number of participants that mentioned the specific topic was counted for each scenario. Furthermore, the communicated social messages for each scenario were mapped to the expected responses from the robot pet to look for frequently occurring patterns.

4. RESULTS

4.1. Questionnaire

Participants' rating of their ability to imagine themselves in the four different scenarios on a scale ranging from 1 (strongly disagree) to 4 (strongly agree) were the following: stressed (*Mdn* (*Q*1, *Q*3) = 3 (3, 3)), depressed (*Mdn* (*Q*1, *Q*3) = 3 (2, 3)), relaxed (*Mdn* (*Q*1, *Q*3) = 3 (3, 3)), and excited (*Mdn* (*Q*1, *Q*3) = 3 (2, 4)). There was no statistically significant difference between the ratings of the scenarios ($\chi^2(3) = 3.297$, $p = 0.352$). Median (*Q*1, *Q*3) perceived ability to imagine the pet as a functional robot was 2 (2, 2) and the statement "I based my interaction with the robot pet on how I interact with a real animal" was rated at 3 (2, 4). The total scores for the expected comfort from the robot pet ranged from 18 to 37 (*Mdn* (*Q*1, *Q*3) = 30(26, 35)), possible total scores ranged from 11 to 44 where a higher score indicated greater expected comfort.

4.2. Observations from the Scenario Videos

Between the different scenarios there were some differences in the level of interaction with the robot pet (see **Table 2**). Participants often used both touch and speech to communicate with the robot pet. Almost all participants used at least the touch modality to communicate, few exceptions occurred in the low valence scenarios (i.e., stressed and depressed). Examples of touch behaviors that were observed were participants sitting next to the robot pet on the couch while touching it using stroking-like gestures, hugging the pet, and having the robot pet sit on their laps while resting a hand on top of it. Speech was most prevalent in the excited scenario, while it was least prevalent in the depressed scenario. Observed behaviors included participants using speech to greet the robot pet when entering, talk about their day, express their emotional state, and show interest in the pet. Some instances of pet-directed speech were observed as well. Another notable observation was that participants oriented the robot pet to face them indicating that they wanted to make eye contact. Furthermore, some participants incorporated the use of their mobile phone in the scenarios, for example, to indicate that they would be preoccupied with their own activities (e.g., sending text messages to friends), to take a picture of the robot pet or to watch online videos together. Others engaged in fake activities with imaginary objects such as playing catch or watching TV together.

TABLE 2 | Number of participants that engaged in different levels of interaction with the robot pet per scenario.

	Emotional state			
Interaction type	**Stressed**	**Depressed**	**Relaxed**	**Excited**
No interaction	3	3	0	0
Speech only	2	0	0	0
Touch only	8	12	13	7
Touch + speech	18	16	18	24
Sum	31	31	31	31

The duration of an interaction was measured as the time in seconds between the start of the interaction (i.e., opening the door to enter the living room) and the end (i.e., closing the door after leaving the room). Overall, the duration of the interactions ranged between 17 and 112 s. There was a statistically significant difference in the duration of interaction between the four scenarios ($\chi^2(3) = 16.347$, $p = 0.001$). A *post hoc* analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < 0.008$. The median (*Q*1, *Q*3) duration in seconds for each of the scenarios was stressed 41 (29, 55), depressed 42 (32, 55), relaxed 42 (32, 53), and excited 35 (28, 45). The duration of interaction in the excited was significantly shorter compared to the other scenarios: stressed ($Z = -2.968$, $p = 0.002$), depressed ($Z = -3.875$, $p < 0.001$), and relaxed ($Z = 3.316$, $p = 0.001$). The other scenarios did not differ significantly (all *p*'s >0.008).

4.3. Interview

In general, participants mostly watched the whole scenario before answering the questions, while others commented on their behavior right away. Also, some participants mentioned at the beginning that they felt a bit awkward to watch themselves on video. The social messages that were communicated by the participant to the robot pet and messages that were expected to be communicated by the robot pet are listed for each scenario in **Tables 3** and **4**, respectively. **Table 5** shows the mapping between the two most frequently communicated social messages for each scenario and the most common expected responses from the robot pet to these messages. We will further discuss the interview results based on these mappings.

4.3.1. Stressed

In the stressed scenario, most participants wanted to communicate that they were stressed by indicating to the robot pet that they had lots of things to do or that they were preoccupied with something ($n = 17$). Notably, some of these participants involved the robot pet as a way to regulate their emotions by touching the pet as a means of distraction. In response, some of these participants wanted company from the robot pet by staying close and through physical interaction ($n = 6$). Importantly, the pet's behavior should be calm, and the robot should not be too demanding. Other participants wanted support from the robot pet by calming them down and providing comfort ($n = 6$).

TABLE 3 | Social messages that were communicated to robot pet for each scenario.

Emotional state			
Stressed	**Depressed**	**Relaxed**	**Excited**
Express emotional state (17)	Seek emotional support (11)	Enjoy company (14)	Express emotional state (15)
Do not want to interact (6)	Express emotional state (8)	Express emotional state (8)	Actively seek interaction (11)
Acknowledge (3)	Do not want to interact (6)	Acknowledge (7)	Enjoy company (4)
Seek emotional support (3)	Want to interact (3)	No expectations (2)	Do not want to interact (1)
Actively seek interaction (2)	Acknowledge (1) Enjoy company (1) No expectations (1)		

The number of participants is in parentheses.

TABLE 4 | Social messages that the robot pet is expected to communicate for each scenario.

Emotional state			
Stressed	**Depressed**	**Relaxed**	**Excited**
Keep company (8)	Keep company (12)	Keep company (12)	Pick up the mood (24)
Provide emotional support (7)	Provide emotional support (11)	Pick up the mood (7)	Engage in interaction (6)
Focus on own needs (6)	Engage in interaction (4)	Engage in interaction (5)	Show appreciation (1)
Understand the situation (5)	Focus on own needs (2)	Focus on own needs (5)	
Engage in interaction (3)	Ask for attention (1)	No interaction (1)	
Do not understand (2)	Show appreciation (1)	Do not understand (1)	

The number of participants is in parentheses.

TABLE 5 | Breakdown of the most frequently communicated messages to the robot pet and the expected responses for each scenario.

Emotional state	**Communicated message**		**Expected robot pet response**
Stressed	Express emotional state (17)	→	Keep company (6) Provide emotional support (6)
	Do not want to interact (6)	→	Understand the situation (3) Focus on own needs (2)
Depressed	Seek emotional support (11)	→	Provide emotional support (6) Keep company (3)
	Express emotional state (8)	→	Provide emotional support (4) Keep company (3)
Relaxed	Enjoy company (14)	→	Keep company (10)
	Express emotional state (8)	→	Pick up the mood (5)
Excited	Express emotional state (15)	→	Pick up the mood (15)
	Actively seek interaction (11)	→	Pick up the mood (6) Engage in interaction (5)

The number of participants is in parentheses.

In contrast, some participants did not want to interact with the robot pet at all as they preferred to be alone in this situation or did not want to be distracted by the pet ($n = 6$). In response, most participants wanted that the robot pet showed its understanding of the situation by keeping its distance ($n = 3$). Others mentioned that the robot pet should have its own personality and should behave accordingly, which might result in the robot pet asking for attention even if this behavior is undesirable in this situation or that the pet would mind its own business ($n = 2$).

4.3.2. Depressed

In the depressed scenario, participants often communicated to the robot pet that they were looking for comfort in order to feel less depressed ($n = 11$). In response, these participants often wanted comfort from the robot pet ($n = 6$). They wanted the pet to do this by sitting on their lap or right next to them and making sounds. Also, participants specified that the robot pet should not approach them too enthusiastically. Others indicated that the robot pet should keep them company ($n = 3$) by staying close and showing its understanding of the situation.

Other participants just wanted to express how they felt ($n = 8$), for example, by telling the pet why they were feeling depressed. In response most of these participants also expected that the robot pet would either provide emotional support ($n = 4$) or would keep them company ($n = 3$).

4.3.3. Relaxed

In the relaxed scenario, participants often wanted to communicate that they enjoyed the pet's company ($n = 14$), for example, by having the pet sit on their lap or right next to them, touching the robot and talking to it. In response, these participants often wanted company from the robot pet ($n = 10$), for example, by staying close, listen, and engage in physical interaction. Furthermore, the pet's behavior should be calm and should reflect that it enjoys being together with the human (e.g., wagging tail or purring).

Other participants mentioned that they wanted to express that they were feeling relaxed ($n = 8$) such as by telling the pet about their day and that everything was alright. In response most of these participants wanted that the robot pet picked up on their mood by displaying relaxed behavior as well such as by lying down ($n = 5$).

4.3.4. Excited

In the excited scenario, participants often wanted to communicate their excitement to the robot pet ($n = 15$), for example, by touching and talking to the robot. In response, all these participants wanted that the robot pet picked up on their mood by becoming excited as well ($n = 15$). The robot pet could show its excitement by actively moving around, wagging its tail, and making positive sounds.

Other participants wanted to actively interact with the robot pet ($n = 11$) by playing with it or going out for a walk together. In response, most of these participants wanted that the robot pet picked up on their mood as well ($n = 24$) or preferred that the robot pet would actively engage them in play behavior ($n = 6$).

5. DISCUSSION

5.1. Categorization of Touch Behaviors

In this study, we observed participants that freely interacted with a robot pet companion. As a consequence, we observed an interesting but complex mix of touch behaviors such as the use of multiple touch gestures that were alternated, hybrid forms of prototypical touch gestures and combinations of simultaneously performed touch gestures (e.g., stroking while hugging). A previous attempt to annotate touch behaviors was limited to the coding of characteristics of touch gestures that were performed sequentially, which completely eliminates difficulties regarding segmentation and labeling that were encountered in this study (Yohanan and MacLean, 2012). Segmentation and labeling of individual touch gestures based a method borrowed from previous work on air gestures proved not to be straightforward. Although air gestures and touch gestures both rely on the same modality (i.e., movements of the hand(s)) their communicative functions are different. Air gestures, especially sign language, are a more explicit form of communication compared to communication through touch in which there is no one-to-one mapping between touch gestures and their meaning. Furthermore, in less controlled interactions it proved to be difficult to categorize touch behaviors into discrete touch gesture categories based on dictionary definitions, such as the gestures defined in **Table 1**. These results indicate that this approach might not be suitable to capture the nature of touch behavior in less controlled settings.

In accordance with previous findings from Yohanan and MacLean (2012) we frequently observed the use of massage, rub, scratch, stroke, and tickle-like gestures to communicate to the robot pet. As a result valuable information would be lost if these gestures would be collapsed into a single category to bypass the difficulties to clearly distinguish between these gestures. Some of the difficulties were due to the use of video footage to observe touch behavior. For example, the intensity level can only be roughly estimated from video [see also Yohanan and MacLean (2012)], and some details such as the precise point of contact were lost because of occlusion. However, confusions in identifying touch gestures with similar characteristics were also observed in studies where touch behaviors were captured by pressure sensors and algorithms were trained to automatically recognize different gestures (e.g., Silvera-Tawil et al., 2012; Jung et al., 2015, 2016). Moreover, segmentation and categorization of touch behavior based on touch sensor data would still remain challenging.

As the segmentation and categorization of touch behaviors into touch gestures might not be that straight forward in a less controlled setting, it might be more sensible to recognize and interpret social messages directly from touch sensor data as was previously suggested by Silvera-Tawil et al. (2014). Moreover, processing techniques from other modalities such as image processing, speech, and action recognition proved to be transferable to touch gesture recognition (Jung et al., 2015). Therefore, the existing body of literature on the transition toward automatic behavior analysis of these modalities in naturalistic settings might provide valuable insights for the understanding of touch behavior as well (e.g., Nicolaou et al., 2011; Gunes and Schuller, 2013; Kächele et al., 2016).

5.2. Observed Multimodal Behaviors

The following coarse descriptions of interactions with the robot pet from two different participants illustrate the use of multimodal cues in the depressed and excited scenario, respectively.

> Participant walks into the living room and sits down on the couch next to the robot pet. Immediately she picks up the pet and holds it against her body using a hug-like gesture. While holding the pet she tells to the pet that she had a bad day while she makes eye contact from time to time. Then she sits quietly while still holding the pet and making eye contact. Finally, she puts the pet back on the couch and gets up to leave the room.
>
> Participant runs into the living room and slides in front of the couch. He picks up the robot pet from the couch and then sits down on the couch with the pet resting on his leg. Then he talks to the pet using pet-directed speech: 'How are you? How are you? Yes! You're a good dog! Good doggy!'. Meanwhile he touches the pet using stroke-like gestures and looks at it. He then puts the robot pet back on the couch again while he still touches the pet using stroke-like gestures. Finally, he gets up from the couch and leaves the room.

As illustrated in the descriptions above, participants often talked to the robot pet while touching it (see also **Table 2**) indicating that the combination of speech (emotion) recognition and touch recognition might aid the understanding of touch behavior. Although we observed forms of speech that had characteristics of pet-directed speech (e.g., short sentences, repetition, and higher pitched voice), it should be noted that no analysis of the prosodic features of the speech was performed. However, the use of pet-directed speech has been observed previously, for example, Batliner et al. (2006) found that children used pet-directed speech when interacting with Sony's pet robot dog AIBO. A limitation of the current setup is that it did not allow for a detailed analysis on the added value of other social cues such as facial expression, posture, and gaze behavior for the interpretation of touch behavior.

By allowing the participants to freely interact with the robot pet within the confined space of a living room setting we were able to observe behavior that might otherwise not be observed. Social interaction involving objects such as taking pictures of the robot with a mobile phone were also observed by Cooney et al. (2015) who argued that these factors should be investigated to enable rich social interaction with robots. However, it is important to keep in mind that although participants in this study were able to freely interact within the given context, the results are confined to the given interaction scenarios. Furthermore, as the study relied on acted behaviors, participants might have displayed prototypical behaviors to clearly differentiate between the scenarios. However, although participants indicated in the questionnaire and during the interview that they had some difficulties acting out the scenarios with a stuffed animal, social behaviors such as making eye contact while talking (e.g., see descriptions above) were observed indicating that at least most participants treated the pet as a social agent. Additionally, it should be noted that touch was not only used to communicate to the robot pet but also often used to move/puppeteer the robot pet as it was unable to move on its own.

Surprisingly, interactions in the excited scenario were shorter despite the fact that all participants engaged in some form of interaction with the robot pet (see **Table 2**). A possible explanation is that participants often only quickly wanted to convey their excitement compared to other scenarios where they were seeking comfort or quietly sat down together with the robot pet to enjoy each others company (see **Table 3**). Furthermore, previous studies indicate that some emotions are more straightforward than others, for example, anger was found to be easier to express through touch than sadness (Hertenstein et al., 2009). Similarly, excitement might have been easier to convey than the other emotional states in this study.

5.3. Communicated Social Messages and Expected Robot Pet Responses

The interview results showed that the communicated messages and expected robot pet responses differed depending on the affective scenario and individual preference (see **Tables 3** and **4**). Moreover, **Table 5** shows that there is no one-to-one relation between communicated messages and expected responses. For example, variation in expectations from the robot pet in the stressed scenario ranged from actively providing support to staying out of the way meaning that in order to respond in a socially appropriate manner, a robot pet should be able to judge whether the user wants to be left alone and when to engage in interaction. From the interviews it became clear that this is not always clear-cut, in the depressed and stressed scenarios some participants indicated that they did not want to initiate interaction but that they might be open to the robot pet approaching them (sometimes after a while). Participants often wanted to communicate their emotional state to the robot pet, especially in the high arousal scenarios (see **Table 3**); however, it should be noted that the focus on emotional states in the scenarios provided in the study might have biased participants toward expressing this emotional state.

Whether a robot pet should completely adapt its behavior to the user is dependent on the role of the pet. In this study the nature of the bond between the participant and his/her robot pet was not specified. Some participants argued that a robot pet should mimic a real pet with its own personality and needs, which might conflict with the current needs of the user. In contrast, other participants proposed that the robot pet could take the role of therapist/coach, which would focus on the user's needs. Mentioned abilities that such a robot pet should have included cheering you up, providing comfort, talking about feelings, and communicating motivational messages. In the role of a friend the robot pet should also take the user's needs into account, albeit to a lesser extent.

In this study we observed how various people, in this case males and females from the working-age population, interacted with a robot pet companion. However, it should be noted that individual factors such as previous experience with animals, personality, gender, age, and nationality might play an important role in these interactions. Interestingly, even though the robot pet's embodiment clearly resembled a dog, some participants treated the robot pet as a cat. Whether participants treated the robot pet as a dog or a cat seemed to depend on their preference and history with real pets. Additionally, it should be noted that the participants studied or worked in the computer science department and that all were at least to some extent familiar with social robots. As a result some participants took the current state of technology into account when suggesting possible robot behaviors, for example, one participant mentioned that it is non-trivial to build a robot dog that would be able to jump on the couch. The use of a stuffed animal dog as a proxy for a functioning robot pet allowed for a more controlled setup. However, the lack of response from the robot pet resulted in less realistic interactions as the participant had to puppeteer the pet or imagine its response. Furthermore, it is important to note that participants were asked to act as if they were coming home in a particular emotional state. Although this is a common approach in studies on touch behavior (e.g., Hertenstein et al., 2006, 2009; Yohanan and MacLean, 2012; Silvera-Tawil et al., 2014), it is unclear whether the same results would have been found if the emotional states were induced in the participants. Despite the above mentioned considerations we observed an interesting range of interactions and were able to find patterns in the social messages that were communicated and the responses that were expected from the robot pet.

6. CONCLUSION

To gain more insight into the factors that are relevant to interpret touch within a social context we studied interactions between humans and a robot pet companion in different affective scenarios. The study took place in a contextualized lab setting in which participants acted as if they were coming home in different emotional states (i.e., stressed, depressed, relaxed, and excited) without being given specific instructions on the kinds of behaviors that they should display.

Results showed that depending on the emotional state of the user, different social messages were communicated to the robot pet such as expressing one's emotional state, seeking emotional support, or enjoying the pet's company. The expected response from the robot pet to these social messages also varied based on the emotional state. Examples of expected responses were keeping the user company, providing emotional support, or picking up on the user's mood. Additionally, the expected response from the robot pet was dependent on the different roles that were envisioned such as a robot that mimics a real pet with its own personality or a robot companion that serves as a therapist/coach offering emotional support.

Findings from the video observations showed the use of multimodal cues to communicate with the robot pet. Participants often talked to the robot pet while touching it and making eye contact confirming previous findings on the importance of studying touch in multimodal interaction. Segmentation and labeling of touch gestures proved to be difficult due to the complexity of the observed interactions. The findings of this study indicate that the categorization of touch behaviors into discrete touch gesture categories based on dictionary definitions is not a suitable approach to capture the nature of touch behavior in less controlled settings.

Additional research will be necessary to determine whether direct recognition and interpretation of higher level social messages from touch sensor data would be a viable option in less controlled situations. Moreover, as the current results are based on acted scenarios, it is important to verify in future research whether similar behaviors occur in a naturalistic setting in which people would interact with a fully functioning robot pet in their own home. A first step could be to induce emotions in participants and observe their interactions with a responding robot pet in a lab setting. The use of verbal behavior that coincides with touch interaction seems another interesting direction for future studies into the automatic understanding of social touch.

AUTHOR CONTRIBUTIONS

MJ designed the study with assistance from MP and DH; collected the data and analyzed the results with assistance from MP, DR, and DH; and wrote the manuscript with contributions from MP, DR, and DH. All the authors reviewed and approved the manuscript.

ACKNOWLEDGMENTS

The authors would like to acknowledge Josca Snippe for participating in the annotation process. This publication was supported by the Dutch national program COMMIT.

REFERENCES

Altun, K., and MacLean, K. E. (2015). Recognizing affect in human touch of a robot. *Pattern Recognit. Lett.* 66, 31–40. doi:10.1016/j.patrec.2014.10.016

Banks, M. R., Willoughby, L. M., and Banks, W. A. (2008). Animal-assisted therapy and loneliness in nursing homes: use of robotic versus living dogs. *J. Am. Med. Dir. Assoc.* 9, 173–177. doi:10.1016/j.jamda.2007.11.007

Batliner, A., Biersack, S., and Steidl, S. (2006). "The prosody of pet robot directed speech: evidence from children," in *Proceedings of Speech Prosody* (Dresden, Germany), 1–4.

Cooney, M. D., Nishio, S., and Ishiguro, H. (2012). "Recognizing affection for a touch-based interaction with a humanoid robot," in *Proceedings of the International Conference on Intelligent Robots and Systems (IROS)* (Vilamoura-Algarve, Portugal), 1420–1427.

Cooney, M. D., Nishio, S., and Ishiguro, H. (2015). Importance of touch for conveying affection in a multimodal interaction with a small humanoid robot. *Int. J. Humanoid Robot.* 12, 1550002. doi:10.1142/S0219843615500024

Eachus, P. (2001). Pets, people and robots: the role of companion animals and robopets in the promotion of health and well-being. *Int. J. Health Promot. Educ.* 39, 7–13. doi:10.1080/14635240.2001.10806140

Gallace, A., and Spence, C. (2010). The science of interpersonal touch: an overview. *Neurosci. Biobehav. Rev.* 34, 246–259. doi:10.1016/j.neubiorev.2008.10.004

Gunes, H., and Schuller, B. (2013). Categorical and dimensional affect analysis in continuous input: current trends and future directions. *Image Vis. Comput.* 31, 120–136. doi:10.1016/j.imavis.2012.06.016

Hertenstein, M. J., Holmes, R., McCullough, M., and Keltner, D. (2009). The communication of emotion via touch. *Emotion* 9, 566–573. doi:10.1037/a0016108

Hertenstein, M. J., Keltner, D., App, B., Bulleit, B. A., and Jaskolka, A. R. (2006). Touch communicates distinct emotions. *Emotion* 6, 528–533. doi:10.1037/1528-3542.6.3.528

Heslin, R., Nguyen, T. D., and Nguyen, M. L. (1983). Meaning of touch: the case of touch from a stranger or same sex person. *J. Nonverbal Behav.* 7, 147–157. doi:10.1007/BF00986945

Jones, S. E., and Yarbrough, A. E. (1985). A naturalistic study of the meanings of touch. *Commun. Monogr.* 52, 19–56. doi:10.1080/03637758509376094

Jung, M. M., Cang, X. L., Poel, M., and MacLean, K. E. (2015). "Touch challenge '15: recognizing social touch gestures," in *Proceedings of the International Conference on Multimodal Interaction (ICMI)* (Seattle, WA), 387–390.

Jung, M. M., Poel, M., Poppe, R., and Heylen, D. K. J. (2016). Automatic recognition of touch gestures in the corpus of social touch. *J. Multimodal User Interfaces*. 11, 81–96. doi:10.1007/s12193-016-0232-9

Kächele, M., Schels, M., Meudt, S., Palm, G., and Schwenker, F. (2016). Revisiting the EmotiW challenge: how wild is it really? *J. Multimodal User Interfaces* 10, 151–162. doi:10.1007/s12193-015-0202-7

Kendon, A. (1980). Gesticulation and speech: two aspects of the process of utterance. *Relat. Verbal Nonverbal Commun.* 25, 207–227.

Kim, Y.-M., Koo, S.-Y., Lim, J. G., and Kwon, D.-S. (2010). A robust online touch pattern recognition for dynamic human-robot interaction. *Trans. Consum. Electron.* 56, 1979–1987. doi:10.1109/TCE.2010.5606355

Kita, S., van Gijn, I., and van der Hulst, H. (1997). "Movement phases in signs and co-speech gestures, and their transcription by human coders," in *International Gesture Workshop* (Bielefeld, Germany), 23–35.

Knight, H., Toscano, R., Stiehl, W. D., Chang, A., Wang, Y., and Breazeal, C. (2009). "Real-time social touch gesture recognition for sensate robots," in *Proceedings of the International Conference on Intelligent Robots and Systems (IROS)* (St. Louis, MO), 3715–3720.

Nicolaou, M. A., Gunes, H., and Pantic, M. (2011). Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space. *Trans. Affect. Comput.* 2, 92–105. doi:10.1109/T-AFFC.2011.9

Noda, T., Ishiguro, H., Miyashita, T., and Hagita, N. (2007). "Map acquisition and classification of haptic interaction using cross correlation between distributed tactile sensors on the whole body surface," in *International Conference on Intelligent Robots and Systems (IROS)* (San Diego, CA), 1099–1105.

Russell, J. A., Weiss, A., and Mendelsohn, G. A. (1989). Affect grid: a single-item scale of pleasure and arousal. *J. Pers. Soc. Psychol.* 57, 493–502. doi:10.1037/0022-3514.57.3.493

Silvera-Tawil, D., Rye, D., and Velonaki, M. (2012). Interpretation of the modality of touch on an artificial arm covered with an EIT-based sensitive skin. *Robot. Res.* 31, 1627–1641. doi:10.1177/0278364912455441

Silvera-Tawil, D., Rye, D., and Velonaki, M. (2014). Interpretation of social touch on an artificial arm covered with an EIT-based sensitive skin. *Int. J. Soc. Robot.* 6, 489–505. doi:10.1007/s12369-013-0223-x

Suvilehto, J. T., Glerean, E., Dunbar, R. I., Hari, R., and Nummenmaa, L. (2015). Topography of social touching depends on emotional bonds between humans. *Proc. Natl. Acad. Sci. U.S.A.* 112, 13811–13816. doi:10.1073/pnas.1519231112

Veevers, J. E. (1985). The social meaning of pets: alternative roles for companion animals. *Marriage Fam. Rev.* 8, 11–30. doi:10.1300/J002v08n03_03

Wada, K., and Shibata, T. (2007). Living with seal robots – its sociopsychological and physiological influences on the elderly at a care house. *Trans. Robot.* 23, 972–980. doi:10.1109/TRO.2007.906261

Yohanan, S., and MacLean, K. E. (2012). The role of affective touch in human-robot interaction: human intent and expectations in touching the haptic creature. *Int. J. Soc. Robot.* 4, 163–180. doi:10.1007/s12369-011-0126-7

Zasloff, R. L. (1996). Measuring attachment to companion animals: a dog is not a cat is not a bird. *Appl. Anim. Behav. Sci.* 47, 43–48. doi:10.1016/0168-1591(95)01009-2

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Jung, Poel, Reidsma and Heylen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

ORIGINAL RESEARCH
published: 24 April 2017
doi: 10.3389/fict.2017.00006

The Impact of Robot Tutor Nonverbal Social Behavior on Child Learning

James Kennedy[1]*, *Paul Baxter*[2] *and Tony Belpaeme*[1,3]

[1] Centre for Robotics and Neural Systems, Faculty of Science and Engineering, Plymouth University, Plymouth, UK, [2] Lincoln Centre for Autonomous Systems, School of Computer Science, University of Lincoln, Lincoln, UK, [3] ID Lab, Department of Electronics and Information Systems, Ghent University, Ghent, Belgium

Several studies have indicated that interacting with social robots in educational contexts may lead to a greater learning than interactions with computers or virtual agents. As such, an increasing amount of social human–robot interaction research is being conducted in the learning domain, particularly with children. However, it is unclear precisely what social behavior a robot should employ in such interactions. Inspiration can be taken from human–human studies; this often leads to an assumption that the more social behavior an agent utilizes, the better the learning outcome will be. We apply a nonverbal behavior metric to a series of studies in which children are taught how to identify prime numbers by a robot with various behavioral manipulations. We find a trend, which generally agrees with the pedagogy literature, but also that overt nonverbal behavior does not account for all learning differences. We discuss the impact of novelty, child expectations, and responses to social cues to further the understanding of the relationship between robot social behavior and learning. We suggest that the combination of nonverbal behavior and social cue congruency is necessary to facilitate learning.

Keywords: human–robot interaction, robot tutors, social behavior, child learning, nonverbal immediacy

OPEN ACCESS

Edited by:
Hatice Gunes,
University of Cambridge, UK

Reviewed by:
Shin'Ichi Konomi,
University of Tokyo, Japan
Khiet Phuong Truong,
University of Twente, Netherlands

****Correspondence:***
James Kennedy
james.kennedy@plymouth.ac.uk

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in ICT

Received: *31 December 2016*
Accepted: *29 March 2017*
Published: *24 April 2017*

Citation:
Kennedy J, Baxter P and Belpaeme T (2017) The Impact of Robot Tutor Nonverbal Social Behavior on Child Learning. Front. ICT 4:6. doi: 10.3389/fict.2017.00006

1. INTRODUCTION

The efficacy of robots in educational contexts has been demonstrated by several researchers when compared to not having a robot at all and when compared to other types of media, such as virtual characters (Han et al., 2005; Leyzberg et al., 2012; Tanaka and Matsuzoe, 2012; Alemi et al., 2014). One suggestion for why such differences are observed stems from the idea that humans see computers as social agents (Reeves and Nass, 1996) and that robots have increased social presence over other media as they are physically present in the world (Jung and Lee, 2004; Wainer et al., 2007). If the social behavior of an agent can be improved, then the social presence will increase and interaction outcomes should improve further (for example, through social facilitation effects (Zajonc, 1965)), but it is unclear how robot social behavior should be implemented to achieve such aims.

This has resulted in researchers exploring various aspects of robot social behavior and attempting to measure the outcomes of interactions in educational contexts, but a complex picture is emerging. While plenty of literature is available from pedagogical fields which describe teaching concepts, there are rarely examples of guidance for social behavior at the resolution required by social roboticists for designing robot behavior. The importance of social behavior in teaching and learning has been demonstrated between humans (Goldin-Meadow et al., 1992, 2001), but not enough is known for implementation in human–robot interaction (HRI) scenarios. This has led researchers to start exploring precisely how a robot should behave socially when information needs to be communicated to, and retained by, human learners (Huang and Mutlu, 2013; Kennedy et al., 2015d).

In this article, we seek to establish what constitutes appropriate social behavior for a robot with the aim of maximizing learning in educational interactions, as well as how such social behavior might be characterized across varied contexts. First, we review work conducted in the field of HRI between robots and children in learning environments, finding that the results are somewhat mixed and that it is difficult to draw comparisons between studies (Section 2.1). Following this, we consider how social behavior could be characterized, allowing for a better comparison between studies and highlighting *immediacy* as one potentially useful metric (Section 2). Immediacy literature is then used to generate a hypothesis for educational interactions between robots and children. In an evaluation to test this hypothesis, nonverbal immediacy scores are gathered for a variety of robot behaviors from the same context (Section 3). While the data broadly agrees with the predictions from the literature, there are important differences that are left unaccounted for. We discuss these differences and draw on the literature to hypothesize a possible model for the relationship between robot social cues and child learning (Section 2.5). The work contributes to the field by furthering our understanding of the impact of robot nonverbal social behavior on task outcomes, such as learning, and by proposing a model that generates predictions that can be objectively assessed through further empirical investigation.

2. RELATED WORK

2.1. Robot Social Behavior and Child Learning in HRI

There are many examples of compelling results, which support the notion that the physical presence of a robot can have a positive impact on task performance and learning. Leyzberg et al. (2012) found that adults who were tutored by a physical robot significantly outperformed those who interacted with a virtual character when completing a logic puzzle. A controlled classroom-based study by Alemi et al. (2014) employed a robot to support learning English from a standard textbook over 5 weeks with a (human) teacher. In one condition, normal delivery was provided, and in the other, this delivery was augmented with a robot that was preprogrammed to explain words through speech and actions. It was found that using a robot to supplement teaching over this period led to significant child learning increases when compared to the same material being covered by the human teacher without a robot. This is strong evidence for the positive impact that robots can have in education, which has been supported in other scenarios. Tanaka and Matsuzoe (2012) also found that children learn significantly more when a robot is added to traditional teaching, both immediately after the experiment and after a delayed period (3–5 weeks later). Combined, these findings suggest that the use of a physically embodied robot can positively contribute to child learning.

Aspects of a robot's nonverbal social behavior have been investigated in one-on-one tutoring scenarios with mixed results. Two studies in the same context by Kennedy et al. (2015c) and Kennedy et al. (2015d) have found that the nonverbal behavior of a robot does have an impact on learning, but that the effect is not always in agreement with predictions from the human–human interaction (HHI) literature. These studies will be considered in more detail in Section 3. Similarly, Herberg et al. (2015) found that the HHI literature would predict an increase in learning performance with increased gaze of a robot toward a pupil, but the opposite was observed: an Aldebaran NAO would look either toward or away from a child while they completed a worksheet based on material they had learnt from the robot, but this was not found to be the case. However, Saerbeck et al. (2010) varied socially supportive behaviors of a robot in a novel second language learning scenario. These behaviors included gestures, verbal utterances, and emotional expressions. Children learnt significantly more when the robot displayed these socially supportive behaviors.

The impact on child learning of verbal aspects of robot behavior has also been investigated. Gordon et al. (2015) developed robot behaviors to promote curiosity in children with the ultimate aim of increased learning. While the children were reciprocal in their curiosity, their learning did not increase as the HHI literature would predict. Kanda et al. (2012) compared a "social" robot to a "non-social" robot, operationalized through verbal utterances to children when they are completing a task. Children showed a preference for the social robot, but no learning differences were found.

Ultimately, it is a difficult task to present a coherent overview of the effect of robot social behavior on child learning, with many results appearing to contradict one another or not being comparable due to the difference in learning task or behavioral context. More researchers are now using the same robotic platforms and peripheral hardware than before (quite commonly the Aldebaran NAO with a large touchscreen, e.g., Baxter et al. (2012)), but there remain few other similarities between studies. Behavior of various elements of the system is reported alongside learning outcomes, but it is difficult to translate from these descriptions to something that can be compared between studies. As such, it becomes almost impossible to determine if differing results between studies (and discrepancies with HHI predictions) are due to differences in robot behavior, the study population, other contextual factors, or indeed a combination of all three. It is apparent that a characterization of the robot social behavior would help to clarify the differences between studies and provide a means by which certain factors could be accounted for in analysis; this will be explored in the following section.

2.2. Characterizing Social Behavior through Nonverbal Immediacy

To allow researchers to make clearer comparisons between studies and across contexts, a metric to characterize the social behavior of a robot is desirable. Various metrics have been used before in HRI. Retrospective video coding has been used in several HRI studies as a means of measuring differences in human behavioral responses to robots, for example, the studies by Tanaka and Matsuzoe (2012); Moshkina et al. (2014); Kennedy et al. (2015b). However, this method of characterizing social behavior is incredibly time consuming, particularly when the coding of multiple social cues is required. Furthermore, it provides data for social cues in isolation and does not easily provide a holistic characterization of the behavior. It is unclear what it means if the robot gazes for a certain number of seconds at the child in the interaction and also performs a certain number of gestures; this problem is exacerbated when a task context changes. The perception of the human directly interacting with the robot is also not accounted for. It is suggested that the direct perception of the human within the interaction is an important one, as they are the one being influenced by the robot behavior *in the moment*. This cannot be captured through *post hoc* video coding.

The Godspeed questionnaire series developed by Bartneck et al. (2009b) has been used in many HRI studies to measure users' perception of robots (Bartneck et al., 2009a; Ham et al., 2011). The animacy and anthropomorphism elements of the scale in particular consider the social behavior and perception of the robot. However, it is not particularly suited to use with children due to the language level (i.e., use of words such as "stagnant," "organic," and "apathetic"). It may also be that the questionnaire would measure aspects of the robot not directly related to social behavior as it is asking about more general perceptions. While this could be of use in many studies, for the aim of characterizing social behavior in the case here, these aspects prevent suitable application.

Nonverbal immediacy (NVI) was introduced in the 1960s by Mehrabian (1968) and is defined as the "psychological availability" of an interaction partner. Immediacy is further introduced as being a measure that indicates "the attitude of a communicator toward his addressee" and in a general form "the extent to which communication behaviors enhance closeness to and nonverbal interaction with another" (Mehrabian, 1968). A number of specific social behaviors are listed (touching, distance, forward lean, eye contact, and body orientation) to form a part of this measure, which were later utilized by researchers that sought to create and validate measuring instruments for NVI. However, it is also this feature that makes NVI a particularly enticing prospect for designers of robot behavior, as the social cues used in the measure are explicit (which is often not the case in other measures of perception commonly used in the field, e.g., Bartneck et al. (2009b)). A reasonable volume of data also already exists for studies considering immediacy, with over 80 studies (and *N* nearly 25,000) from its inception to 2001 (Witt et al., 2004) and more since. This provides a context for NVI findings in HRI scenarios and a firm grounding in the human–human literature from which roboticists can draw.

Several versions of surveys have been developed and validated for measuring the nonverbal immediacy of adults (Richmond et al., 2003). Surveys have also been developed for verbal immediacy (Gorham, 1988), but their ability to measure precisely the concept of verbal immediacy remains the subject of debate (Robinson and Richmond, 1995). Both verbal and nonverbal measures consider observed overt behavior more than, but not excluding, perceptions. Immediacy has recently been used in HRI as a means of motivating robot behavior manipulations (Szafir and Mutlu, 2012) and characterizing social behavior (Kennedy et al., 2017).

There is a consensus on the instruments used to measure nonverbal immediacy (whereas this is less clear for verbal immediacy), and it is also transparent in terms of how participants are judging the robot. The Godspeed questionnaire is a useful tool for gathering perceptions, but nonverbal immediacy is clearly measuring overt social behavior, and so it is ideal given our scope of trying to characterize social behavior (often with children). Use of the NVI metric brings several other advantages to researchers in HRI and for robot behavior designers. The NVI metric can be used as a guideline for an explicit list of social cues available for manipulation as a part of robot behavior. Characterization of robot social behavior at this relatively low level is not readily available in other metrics. This provides a useful first step in designing robot behavior but also a means of evaluating and modifying future social behaviors. NVI constitutes part of an overall social behavior; hence NVI is treated as a *characterization* of the overall behavior, not a complete description or definition. Not all aspects of sociality or interaction are addressed through the measure, but to the knowledge of the authors, nor are these aspects fully covered by any other validated metric.

The NVI metric can be used with either the subjects themselves or with observers (during or after the interaction). This permits flexibility depending on the needs of the researcher. It is not always practical to collect such data from participants (for example, when they are young children or following an already lengthy interaction), so having the flexibility to gather these data *post hoc* is advantageous. Due to this mixture of practical and theoretical benefits, nonverbal immediacy (NVI) will be adopted as a social behavior characterization metric for this article.

Immediacy has been validated through physical manipulation of some of the social cues, specifically eye gaze and proximity, to ensure that the phenomenon indeed works in practice and is not a product of affect or bias in survey responses (Kelley and Gorham, 1988). It was indeed found that the physical manipulations that were made which would lead to a higher immediacy score (standing closer and providing more eye gaze) did lead to increased short-term recall of information. While there is clearly a difference between recall and learning, recall of information is a promising first step to acquiring new understanding and skills. These results were hypothesized to exist in the other immediacy behaviors (such as gestures) as well. Overall, the link between teacher immediacy and student learning is hypothesized to be a positive one, as reflected in the meta-review by Witt et al. (2004) and many studies (Comstock et al., 1995; McCroskey et al., 1996; Christensen and Menzel, 1998). Thus, this prediction can be tested in human–robot interaction, where the robot takes the role of the tutor. As a result, we generate the following hypothesis:

H1. A robot tutor perceived to have higher immediacy leads to greater learning than a robot perceived to have lower immediacy.

3. APPLYING NONVERBAL IMMEDIACY TO HRI

In this section, an evaluation of nonverbal immediacy (NVI) in the context of cHRI is described. The aim is to explore whether the characterization that it provides can account for the differences between robot behaviors and learning outcomes of children. The wealth of literature that explores NVI in educational scenarios is generally in agreement that higher NVI of an instructor is positively correlated with learning outcomes of students. We evaluate 4 differently motivated robot behaviors and a human in a one-to-one maths-based educational interaction with children. The aim is to use these data to provide a comparison between behavioral manipulations to test predictions from the HHI immediacy literature regarding social behavior.

3.1. Task Design and Measures

All five behaviors under consideration use the same context and broader methodology. Children aged 8–9 years are taught how to identify prime numbers between 10 and 100 using a variation on the Sieve of Eratosthenes method. They interact with a tutor: in 4 conditions, this is an Aldebaran NAO robot, and in 1 condition, this is a human (**Figure 1**). Children complete pretests and posttests in prime number identification, as well as pretests and posttests for division by 2, 3, 5, and 7 (skills required by the Sieve of Eratosthenes method for numbers in the range used) on a large touchscreen. The tutor provides lessons on primes and dividing by 2, 3, 5, and 7 (**Figure 2**). In all cases, an experimenter briefs the child and introduces the child to the tutor. The experimenter remains in the room throughout the interaction, but out of view of the child. Two cameras record the interactions; one is directed toward the child and one toward the tutor. Interactions with the tutor would last for around 10–15 min, with an additional 5 min required afterward in conditions where nonverbal immediacy surveys were completed (details to follow).

At the start of the interaction, the children complete a pretest in prime numbers on the touchscreen without any feedback from the screen or the tutor. A posttest is completed by the children at the end of the interaction; again no feedback is provided to the child so as not to influence their categorizations. Two tests are used in a cross-testing strategy, so children have a different pretest and posttest, and the tests are varied as to whether they are used as a pretest or posttest. The tests require the children to categorize numbers as “prime” or “not prime” by dragging and dropping numbers on screen into the category labels. Each test has 12 numbers, so by chance, a score of 6 would be expected (given 2 possible categories 50% is chance). Learning is measured through the improvement in child score from the prime number pretest to posttest. By considering the improvement, any prior knowledge (correct or otherwise) or deviation in division skill is factored in to the learning measure. The mean and *SD* score (of 12) for the pretests are compared to those of the posttest to calculate the learning effect size (Cohen’s *d*) for each condition.

The prime number task was selected in consultation with education professionals to ensure that it was appropriate for the capabilities of children of this age. Children of this age have not yet learnt prime number concepts in school, but do have sufficient (but imperfect) skills for dividing by 2, 3, 5, and 7 as required by the technique for calculating whether numbers are prime. During the division sections of the interaction, the tutor provides feedback on child categorizations.

Nonverbal immediacy (NVI) scores are collected through questionnaires. For children, this was done after the interaction with the tutor had been completed, for adults, this was online (details in Section 3.4). A standard nonverbal immediacy questionnaire was adapted for use with children by modifying some of the language; the original and modified versions alongside the score formula can be seen online.[1] Both the Robot Nonverbal Immediacy Questionnaire (RNIQ) and Child-Friendly Nonverbal Immediacy Questionnaire (CNIQ) were used depending on condition for children. Adults had the same questionnaire but with “the child” in place of “you” as they were observing the interaction, rather than participating in it. The questionnaire consists of 16 questions about overt nonverbal behavior of the tutor. Each question is answered on a 5-point Likert scale, and a final immediacy score is calculated by combining these answers. Some count positively toward the nonverbal immediacy score, whereas some count negatively, depending on the wording of the question. The version in the Appendix shows the questionnaire used for this study when a robot (as opposed to a human) tutor was used as this has been validated for use in HRI (Kennedy et al., 2017) and corresponds to the validated version from prior human-based literature (Witt et al., 2004).

Existing immediacy literature extensively uses adults (often students) as subjects; studies with children are rare. Prior work

[1] http://goo.gl/UoL5QM, also included as an Appendix.

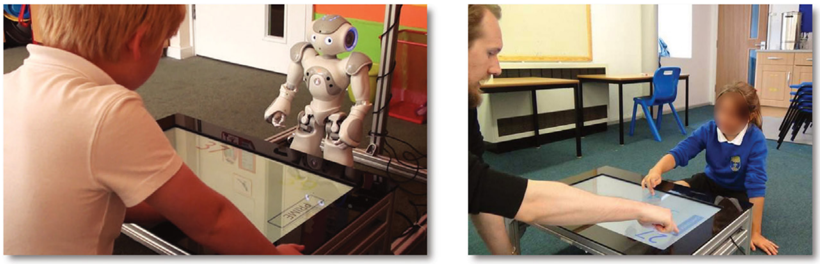


FIGURE 1 | (Left) Still image from a human–robot interaction (specifically, the “social” condition), and (right) still image from the human–human condition. The tutor (either robot or human) teaches children how to identify prime numbers using the Sieve of Eratosthenes method using a large horizontal touchscreen as a shared workspace. The robot can “virtually” move numbers on screen (numbers move in correspondence with robot arm movements, but physical contact is not made with the screen).

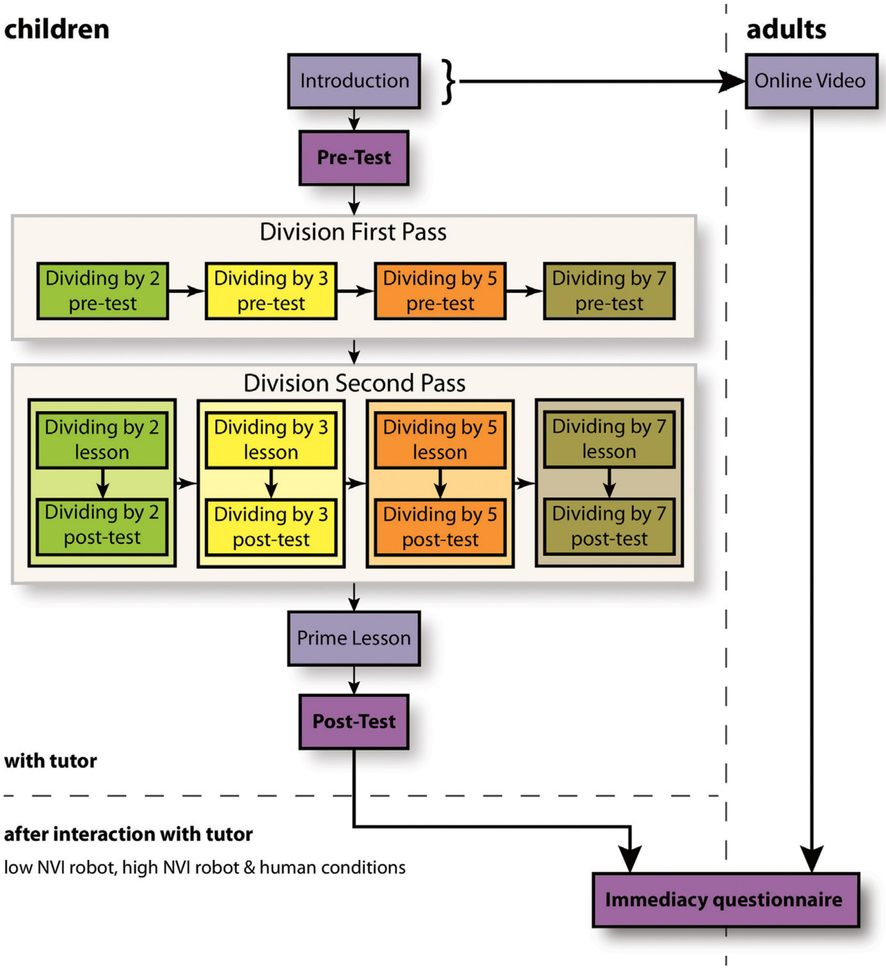


FIGURE 2 | **Task structure—the top section is led by the tutor and is aimed at teaching children how to calculate whether a number is prime**. The bottom section consists of completing the nonverbal immediacy questionnaire—this is done after the interaction for 3 of the child conditions and *via* online videos to get adult responses. Dark purple boxes (pretest, posttest, and immediacy questionnaire) are the metrics under consideration in this article.

has been conducted with the adapted nonverbal immediacy scale for use with robots and children (Kennedy et al., 2017); however, the task in this article is novel in this context (one-to-one interactions instead of group instruction). Children present unique challenges when using questionnaire scales, such as providing different answers for negatively worded questions to positively worded ones (Borgers et al., 2004) or trying to please experimenters (Belpaeme et al., 2013), which can consequently make it difficult to detect differences in responses (Kennedy et al., 2017). As children are not well represented in immediacy literature, using adults for NVI scores more tightly grounds our hypotheses and assumptions to the existing literature. However, NVI ratings are collected from children in robot conditions in which NVI is intentionally manipulated. As the nonverbal immediacy was intentionally manipulated between these conditions, and the adult results can provide some context, we can observe whether children do perceive the manipulation on this scale, potentially broadening the applicability of our findings.

3.2. Conditions

A total of 5 conditions are used in this evaluation.[2] As described in the introduction, an often adopted approach to social behavioral design is to consider how a human behaves and reproduce that (insofar as is possible) on the robot. As such, we use 2 conditions, seeking to follow and also invert this approach. We additionally use 2 conditions derived from the NVI literature, again seeking to maximize and minimize the behaviors along this scale. The final condition is a human benchmark. Further details for each can be seen in **Table 1** and below:

1. "Social" robot (SR)—this condition is derived from observations of an expert human–human tutor completing this task with 6 different children. This condition reflects a human

[2] Please note that while some data have previously been published for all of these conditions (Kennedy et al., 2015c,d, 2016), this article presents both novel data collection and different analysis perspectives in a new context to the prior work.

TABLE 1 | Operationalization of the differences in nonverbal behavior between the conditions considered in the study presented in this article.

Condition	Motivation	Nonverbal behavior	Other manipulations
"Social" robot (SR)	Based on a human model of the task	Seeks mutual gaze with child, frequent arm gestures	Uses child name, personalizes number of items in division posttests, "positive" feedback, variable feedback
"Asocial" robot (AR)	"Inverse" of the above human model	Avoids child gaze, frequent but mistimed arm gestures	Blunt feedback, repetitive feedback
High NVI robot (HNVI)	Intended to maximize the nonverbal immediacy	Seeks mutual gaze with child, frequent head/gaze movement, frequent arm gestures, lean forwards, continuous small upper body movements	
Low NVI robot (LNVI)	Intended to minimize the nonverbal immediacy	Avoids child gaze, infrequent head/gaze movement, no arm gestures, TTS parameters modified to give "dull" voice, lean backward, rigid/no upper body movements	
Human (HU)	Human benchmark	No instructions given for nonverbal behavior	

Further notes are provided about any other manipulations made besides nonverbal behavior.

model-based approach to designing the behavior. The social behavior of the tutor was analyzed through video coding, and these behaviors were implemented on the robot where possible.

2. "Asocial" robot (AR)—this condition considers the behavior generated for the SR condition and seeks to "invert" it. That is, the behavior is intentionally manipulated such that an opposite implementation is produced, for example, the SR condition seeks to maximize mutual gaze, whereas this condition actively minimizes mutual gaze. The quantity of social cues used in this condition is exactly the same as the SR condition above; however, the placement of these cues is varied (for example, a wave would occur during the greeting in SR, but during an explanation in AR).
3. High NVI robot (HNVI)—this condition uses the literature to drive the behavioral design. The behavior is derived from considering how the social cues within the nonverbal immediacy scale can be maximized. For example, the robot will seek to maximize gaze toward the child and make frequent gestures.
4. Low NVI robot (LNVI)—this condition is intended to be the opposite to the HNVI condition. Again, the nonverbal immediacy literature is used to drive the design, but in this case, all of the social cues are minimized. For example, the robot avoids gazing at the child and makes no gestures.
5. Human (HU)—this is a human benchmark. The human follows the same script for the lessons as the robot, but they are not constrained in their social behavior. The intention here is that we can then acquire data for a "natural", non-robot interaction where the social behavior is not being manipulated; this can then be used to provide context for the robot conditions.

A summary of the motivations for the conditions and the operationalization of the differences between conditions can be seen in **Table 1**. Further implementation details can be seen in "Robot Behavior." While the Aldebaran NAO platform cannot be manipulated for some of the cues involved in the nonverbal immediacy measure given the physical setup and modalities of the robot (i.e., smiling and touching), it has been manipulated on all of the other cues possible. This leaves only 4 of the 16 questions (2 of 8 cues) not manipulated in the metric. Specifically, these are questions 4, 8, 9, and 13, as seen in the Appendix, pertaining to frowning/smiling and touching.

3.2.1. Robot Behavior

Throughout the division sections of the interaction, the tutor (human or robot) would provide feedback on child categorizations and could also suggest numbers for the child to look at next. This was done through moving a number to the center of the screen and making a comment such as "why don't you try this one next?" The tutor would also provide some prescripted lessons (**Figure 2**) that would include 2 example categorizations on screen. These aspects are central to the delivery of the learning content, so are maintained across all conditions to prevent a confound in learning content.

All robot behavior was autonomous, apart from the experimenter clicking a button to start the system once the child was sat in front of the touchscreen. The touchscreen and a Microsoft Kinect were used to provide input for the robot to act in an autonomous manner. The touchscreen would provide information to the robot about the images being displayed and the child moves on screen, the Kinect would provide the vector of head gaze for the child and whether this was toward the robot. Through these inputs, the robot behavior could be made contingent on child actions, for example, by providing verbal feedback after child moves (in all conditions), or manipulating mutual gaze. In all robot conditions, the robot gaze was contingent on the child's gaze, but with differing strategies depending on the motivation of the condition. The AR and LNVI conditions would actively minimize mutual gaze by intentionally avoiding looking at the child, whereas the SR and HNVI conditions would actively maximize mutual gaze by looking at the child when data from the Kinect indicated that the child was looking at the robot. Robot speech manipulation executed in the LNVI condition to make the robot voice "dull" was achieved through lowering the vocal shaping parameter of the TTS engine (provided by Acapela).

Due to the human model-based approach, some personalization aspects such as use of child name were included as part of the social behavior in the SR condition. This was not done in the NVI conditions as these manipulations are not motivated through the NVI metric. The HNVI condition also addresses more of the NVI questionnaire items (leaning forward and continuous "relaxed" upper body movements) than the SR condition due to this difference in motivation. The AR condition has the same quantity of behavior as the SR condition, whereas the LNVI has a *lack* of behavior. As a concrete example, the AR condition includes inappropriately placed gestures, whereas the LNVI condition includes no gestures. Consequently, the LNVI and HNVI conditions provide useful comparisons both to one another and to the SR and AR conditions.

3.3. Participants

To provide NVI scores for all 5 conditions, video clips of the conditions were rated by adults. Nonverbal immediacy scores were also acquired at the time of running the experiments for 3 of the 5 conditions (high and low NVI robot and human) from children through paper questionnaires (**Table 2**). These scores allow a check that the NVI manipulation between the robot conditions could be perceived by the children, with the adult data provided context for these ratings. Written informed consent from parents/guardians was received for the children to take part in the study, and they additionally provided verbal assent themselves, in accordance with the Declaration of Helsinki. Written informed consent from parents/guardians and verbal assent from children were also received for the publication of identifiable images. The protocol was reviewed and approved by the Plymouth University ethics board. **Table 2** shows numbers of participants per condition and average ages for the adult conditions; all children were aged 8 or 9 years old and were recruited through a visit to their school, where the experiment took place.

3.4. Adult Nonverbal Immediacy Score Procedure

Videos shown to adults to acquire nonverbal immediacy scores were each 47 s long. The videos contained both the interaction video (42 s) and a verification code (5 s; details in the following paragraph). The length of video was selected to be 42 s as the literature suggests that at least around 6 s are required to form a judgment of social behavior (Ambady and Rosenthal, 1993), and there was a natural pause at 42 s in the speech in all conditions so that it would not cut part-way through a sentence. The interaction clips were all from the start of an interaction, so the same information was being provided by the tutor to the child in the clip.

To provide sufficient subject numbers for all of the conditions, an online crowdsourcing service[3] was used. The participants were restricted to the USA and could only take part if they had a reliable record within the crowdsourcing platform. A test question was put in place whereby participants had to enter a 4 digit number into a text box. This number was shown at the end of the video for 5 s (the video controls were disabled so it could not be paused and the number would disappear after the video had finished). A different number was used for each video. If the participants did not enter this number correctly, then their response was discarded. The crowdsourcing platform did not allow the prevention of users completing multiple conditions, so any duplicates were removed, i.e., only those seeing a video for the first time were kept as valid responses. A total of 366 responses were collected, but 209 were discarded as they did not answer the test question correctly, the user had completed another condition,[4] or the response was clearly spam (for example, all answers were "1"). This left 157 responses across 5 conditions; 90M/67F (**Table 2**).

4. RESULTS

When performing a one-way ANOVA, a significant effect is found for condition seen, showing that the robot behavior influences perceived nonverbal immediacy; $F(4,152) = 14.057$, $p < 0.001$. *Post hoc* pairwise comparisons with Bonferroni correction reveal that the adult-judged NVI of the LNVI condition is significantly different to all other conditions ($p < 0.001$ in all cases), but no other pairwise comparisons are statistically significant at $p < 0.05$. The nonverbal immediacy score means and learning effect sizes for each condition can be seen in **Table 3**. Children learning occurs in all conditions. Generally, it can be seen that the conditions with higher rated nonverbal immediacy lead to greater child improvement in identifying prime numbers.

While significance testing provides an indication that most of the conditions are similar (at least statistically) in terms of NVI, additional information for addressing the hypothesis can be gleaned by considering the trend that these data suggest (**Figure 3**). A strong positive correlation is found between the (adult) NVI score of the conditions and the learning effect sizes (Cohen's d) of children who interacted in those conditions ($r(3) = 0.70$, $p = 0.188$). This correlation is not significant, likely due to the small number of conditions under consideration, but the strength of the correlation suggests that a relationship could be present.

[3] http://www.crowdflower.com/.

[4] The majority of exclusions were due to users having completed another condition, thereby impairing the independence of the results.

TABLE 2 | Subject numbers by condition and average ages for adult participants by condition.

Condition	Child *N*	Adult *N*	Adult *M* age, *SD* in brackets	Child immediacy scores collected?
Low NVI robot	12	33	31.5 (*12.2*)	Yes
High NVI robot	11	31	35.6 (*11.7*)	Yes
Social robot	12	33	29.0 (*10.4*)	No
Asocial robot	11	30	39.0 (*12.2*)	No
Human	11	30	32.9 (*12.3*)	Yes

TABLE 3 | Adult and child nonverbal immediacy ratings and child learning (as measured through effect size between pretests and posttests for prime numbers) by tutor condition.

Condition	Adult *M* NVI rating *[95% CI]*	Child *M* NVI rating *[95% CI]*	Child learning (*d*)
Low NVI robot	40.2 *[38.1, 42.2]*	51.0 *[47.6, 54.4]*	0.30
High NVI robot	48.4 *[46.9, 50.0]*	55.1 *[52.3, 57.6]*	0.67
Social robot	49.0 *[47.6, 50.4]*	N/A	0.51
Asocial robot	48.5 *[46.1, 50.8]*	N/A	0.89
Human	47.7 *[45.3, 50.1]*	54.4 *[52.9, 55.9]*	0.89

When the immediacy scores provided by the children who interacted with the robot are also considered, a similar pattern can be seen (**Figure 4**). The adult and child immediacy ratings correlate well, with a strong positive correlation ($r(1) = 1.00$, $p < 0.001$). There is also a strong positive correlation for the children between immediacy score of the conditions and the learning effect sizes (Cohen's *d*) in those conditions ($r(1) = 0.86$, $p = 0.341$). Again, significance is not observed, but the power of the test is low due to the number of data points available for comparison. The strong positive correlations between child immediacy scores and learning and adult immediacy scores and learning provide some support for hypothesis H1 (that higher tutor NVI leads to greater learning), but further data points would be desired to explore this relationship further. It should be noted that we consider the results of 57 children and 157 adults across 5 conditions; acquiring further data points for more behaviors (and deciding what these behaviors should be) would be a time-consuming task.

5. DISCUSSION

There is a clear trend in support of hypothesis H1: that a tutor perceived to have higher immediacy leads to greater learning. As such, increasing the nonverbal immediacy behaviors used by a social robot would likely be an effective way of improving child learning in educational interactions. However, nonverbal immediacy does not account for all of the differences in learning. Three of the conditions have near identical NVI scores as judged by adults, but quite varied learning results (high NVI robot: $M = 48.4$ NVI score/$d = 0.67$ pre–post test improvement, asocial robot: NVI $M = 48.5/d = 0.89$, social robot: NVI $M = 49.0/d = 0.51$). This partially reflects the slightly mixed

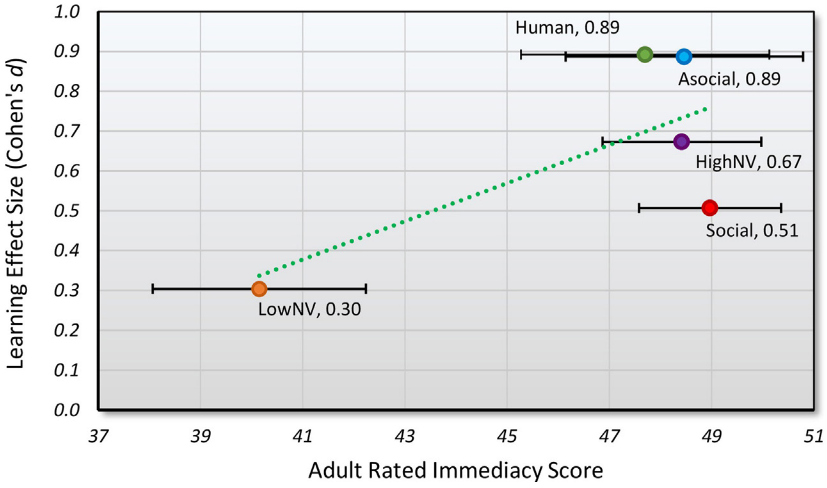


FIGURE 3 | **Nonverbal immediacy scores as judged by adults and learning effect sizes for the prime number task**. The dotted green line indicates a trend toward greater nonverbal immediacy of the tutor leading to increased learning. Error bars show 95% confidence interval.

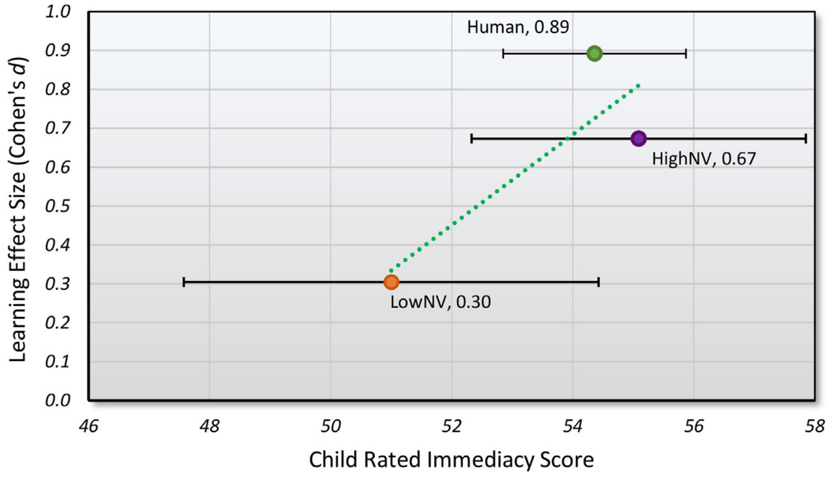


FIGURE 4 | **Nonverbal immediacy scores as judged by the children in the interaction and learning effect sizes for the prime number task**. The dotted green line indicates a trend toward greater perceived nonverbal immediacy of the tutor leading to increased learning. Error bars show 95% confidence interval.

picture of immediacy that the pedagogy literature presents; for example, the disagreement as to whether NVI has a linear (Christensen and Menzel, 1998) or curvilinear (Comstock et al., 1995) relationship with learning. Nonetheless, there are further factors that may be introduced by the use of a robot that may have had an influence on the results. Nonverbal immediacy only considers overt observed social behaviors, so by design does not cover all possible aspects of effective social behavior for teaching. While this seems to be enough in HHI (Witt et al., 2004), it may not be for HRI since various inherent facets of human behavior cannot be assumed for robots. Several possible explanations as to why this learning variation is present will now be discussed. From this, a possible model (suggested to be more accurate) of the relationship between social behavior and learning is proposed. Such a model may be useful in describing (and testing) the relationship between social behavior and child learning for future research.

5.1. Timing of Social Cues

The quantity of social cues used in both the social robot and the asocial robot conditions is exactly the same; however, the timing is varied. Timing is not considered as part of the nonverbal immediacy metric—the scale measures whether cues have, or have not, been used, rather than whether their timing was appropriate. The cues used in the asocial robot condition were intentionally placed at inappropriate times (for example, waving part-way through the introduction, instead of when saying hello). This is not factored into the nonverbal immediacy measure, but could impact the learning (Nussbaum, 1992).

The timing of social cues in the human condition may also explain why the learning in this condition was higher than the others. The robot conditions are contingent on aspects of child behavior, such as gaze and touchscreen moves, but are not adapted to individual children (for example, the number of feedback instances the robot provides would not be based on how well the child was performing). However, the human is presumably adaptive in both the number of social cues used and the timing of these cues. Again, this would not be directly revealed by the immediacy metric, but could account for some of the learning difference. Indeed, the nonverbal immediacy metric comes from HHI studies and has been validated in such environments. In HHI, there is a reasonable assumption that the timing of social cues will be appropriate, and so it may not be necessary to include it as part of a behavioral metric for HHI. However, when applied to social robotics, the assumption of appropriate timing no longer applies, and so to fully account for learning differences in HRI, timing may need more explicit incorporation into characterizations of social behavior. This constitutes a limitation of the NVI metric, but also an opportunity for expansion in future work to capture timing aspects.

5.2. Relative Importance of Social Cues

One substantial difference between the robot conditions and the human condition is the possibility of using facial expressions. The robotic platform used for the studies was the Aldebaran NAO. This platform has limited ability to generate facial expressions as none of the elements of the face can move, only the eye color can be changed. On the other hand, the human has a rich set of facial expressions to draw upon.

While the overall nonverbal immediacy scores for the asocial, social, and human conditions are tightly bunched, the make-up of the scores is not. For example, the robot scores (asocial and social combined) are higher for gesturing, averaging $M = 4.3$ (95% CI 4.1, 4.5) out of 5 for the nonverbal immediacy question about gesturing (the robot uses its hands and arms to gesture while talking to you), compared to $M = 3.1$ (95% CI 2.7, 3.5) for the human. However, the human is perceived to smile more ($M = 2.5$, 95% CI 2.1, 2.8) than the robot ($M = 1.8$, 95% CI 1.5, 2.0). Through principle component analysis, Wilson and Locker (2007) found that different elements of nonverbal behavior do not contribute equally to either the nonverbal immediacy construct or instructor effectiveness. Facial expressions (specifically smiles) have a large impact on both the nonverbal immediacy construct and the instructor effectiveness, whereas gestures do not have such a large effect (although still a meaningful contribution; smiles: 0.54, gestures: 0.30 component contribution from Wilson and Locker (2007)).

In the nonverbal immediacy metric, all social cues are given equal weighting. However, this may not always be the most appropriate method for combining the cues given the evidence, which suggests that some cues may contribute more than others to various outcomes (McCroskey et al., 1996; Wilson and Locker, 2007). This could be a further explanation as to why several of the conditions in the study conducted here have near identical overall nonverbal immediacy scores, but very different learning outcomes.

5.3. Novelty of Character and Behavior

The novelty of both the character (i.e., robot or human) and the behavior itself could have had an impact on the learning results found in the study. Novelty is often highlighted as a potential issue in experiments conducted in the field (Kanda et al., 2004; Sung et al., 2009). The novelty of the robot behavior could override the differences between the conditions and subsequently influence the learning of the child. In the social robot condition here, novel behavior (such as new gestures) was often introduced when providing lessons to the child. Between humans, this would likely result in a positive effect (Goldin-Meadow et al., 2001), but when done by a robot, the novelty of the behavior may counteract the intended positive effect.

There may also be a difference in the novelty effect for the children seeing the robot when compared to the human. Although the human is not one that they are familiar with, they are still “just” a human, whereas the robot is likely to be more exciting and novel as child interaction with robots is more limited than with humans. The additional novelty of the robot could have been a distraction from the learning, explaining why the learning in the human condition is higher.

Finally, the novelty may have impacted the nonverbal immediacy scores themselves. It is possible that observers (be they children or adults) score immediacy on a relative scale. It is reasonable to suggest that the immediacy of the characters is judged not as a standalone piece of behavior, but in the context of an observer’s prior experience, or expectations for what that character may be capable of. Clear expectations will likely exist for human behavior, but not for robot behavior, which may lead to an overestimation of robot immediacy. This would impact on the ability of considering the human and robots on the same nonverbal immediacy scale and drawing correlations with learning and cannot be ruled out as a factor in the results.

5.4. (In)Congruency of Social Cues

As previously discussed, the robot is limited in the social cues that it can produce (for example, it cannot produce facial expressions). This meant that the conditions all manipulated the available robot social cues, but if social cues are interpreted as a single percept by the human (as suggested by the literature (Zaki, 2013)), then this could lead to complications.

In the case of the social robot, many social cues are used to try and maximize the "sociality" of the robot. This means that there is a lot of gaze from the robot to the child, and the robot uses a lot of gestures. However, it still cannot produce facial expressions. This incongruency between the social cues could produce an adverse effect in terms of perception on the part of the child and subsequently diminish the learning outcome. There are clear parallels here with the concept of the Uncanny Valley (Mori et al., 2012), with models for the Uncanny Valley based on category boundaries in perception indicating issues arising from these mismatches (Moore, 2012).

The expectation the child has for the robot social behavior is suggested to be of great importance (Kennedy et al., 2015a). If their expectations are formed early on through high quantities of gaze and gestures, then there would be a discrepancy when facial expressions do not match this expectation. Again, this expectation discrepancy may lead to adverse effects on learning outcomes, as in the case of perceptual issues due to cue incongruence. These issues may become exacerbated as the overall level of sociality of behavior of the robot increases as any incongruencies then become more pronounced. As stated in the study by Richmond et al. (1987), higher immediacy generally leads to more communication, which can create misperceptions (of liking, or expected behavior).

As the nonverbal immediacy scale has been rigorously validated (McCroskey et al., 1996; Richmond et al., 2003), it is known that it does indeed provide a reliable metric for immediacy in humans (Cronbach's alpha is typically between 0.70 and 0.85 (McCroskey et al., 1996)). Typically, internal consistency measures of a scale would be used to evaluate the ability of items in a scale to measure a unidimensional construct, i.e., how congruent the items are with one another. As such, a consistency measure could be used as an indicator of the congruency between the cues. The robot lacks a number of capabilities when compared to humans, and there are several scale items that are known to be impaired on the robot, such as smiling/frowning. Using an internal consistency measure across all NVI questionnaire items (with the negatively worded question responses reversed) can reveal cases in which the cues are relatively more or less congruent. Greater internal consistency indicates lower variability between questionnaire items (the social cues) and, therefore, more congruence between the social cues. Lower internal consistency indicates larger variability between scale items and thus greater incongruency between the cues.

Guttman's λ_6 (or G6) for each condition has been calculated,[5] revealing that indeed there are differences in how congruent the cues could be considered to be (**Table 4**; **Figure 5**). All of the NVI questionnaire items are included in the λ_6 calculation. The behavioral conditions used here are restricted in such a way that a lower reliability would be expected (as several cues of the scale are not utilized) for some conditions. Indeed, these values fall in line with predictions that could be made based on the social behavior in each of the conditions. The human reliability score provides a "sanity check" as it is assumed that human behavior would have a certain degree of internal consistency between social cues, which is reflected by it having the highest value. In addition, the LNVI robot condition has intentionally low NVI behavior, so the lack of smiling or touching (high NVI behaviors) does not cause incongruency (signified by a lower λ_6 score), whereas the HNVI robot condition has intentionally high NVI behavior where possible on the robot, so the lack of smiling and touching cause greater overall incongruency, resulting in a considerably lower λ_6 score.

5.5. A Hypothesis: Social Cue Congruency and Learning

Taking Guttman's λ_6 to provide an indication of the congruency of social cues, then it is clear that this alone would not provide a strong predictor of learning (**Figure 5**). However, these data can be combined with the social behavior (as measured through immediacy) to be compared to learning outcomes. In the resulting space, both congruency and social behavior could have an impact on learning, as hypothesized in the previous section (**Figure 6**).

Our data show that learning is best with human behavior, which is shown to be highly social and reasonably congruent. When the social behavior used is congruent, but not highly social, then the learning drops to a low level. The general trend of our data shows that when the congruency of the cues increases

TABLE 4 | Guttman's λ_6 and learning effect size by condition.

Condition	Learning effect size (Cohen's *d*)	Guttman's λ_6 (G6)
Asocial robot	0.89	0.84
Social robot	0.51	0.83
High NVI robot	0.67	0.69
Low NVI robot	0.30	0.78
Human	0.89	0.87

λ_6 is used as an indicator of social cue congruency, with a higher value indicating greater congruency between cues.

[5] Cronbach's alpha tends to be the de facto standard for evaluating internal consistency and reliability; however, its use as such a measure has been called into question (Revelle and Zinbarg, 2009)—including by its own creator (Cronbach and Shavelson, 2004). Instead λ_6 is used, which considers the amount of variance in each item that can be accounted for by the linear regression of all other items (the squared multiple correlation) (Guttman, 1945). This provides a lower bound for item communality, becoming a better estimate with increased numbers of items. This would appear to provide a logical (but likely imperfect) indicator for the congruency of cues as required here.

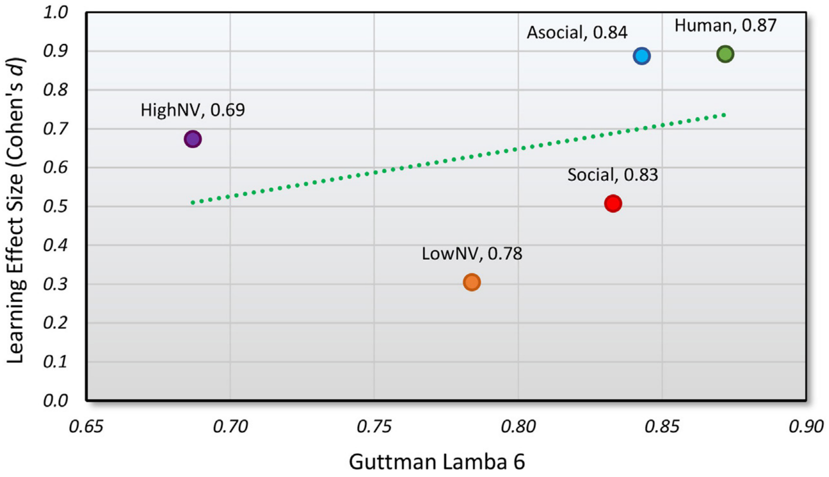


FIGURE 5 | Guttman's λ_6 against learning effect size for each of the prime tutoring conditions. The dotted line indicates a trend toward greater internal consistency (measured through λ_6) leading to greater learning.

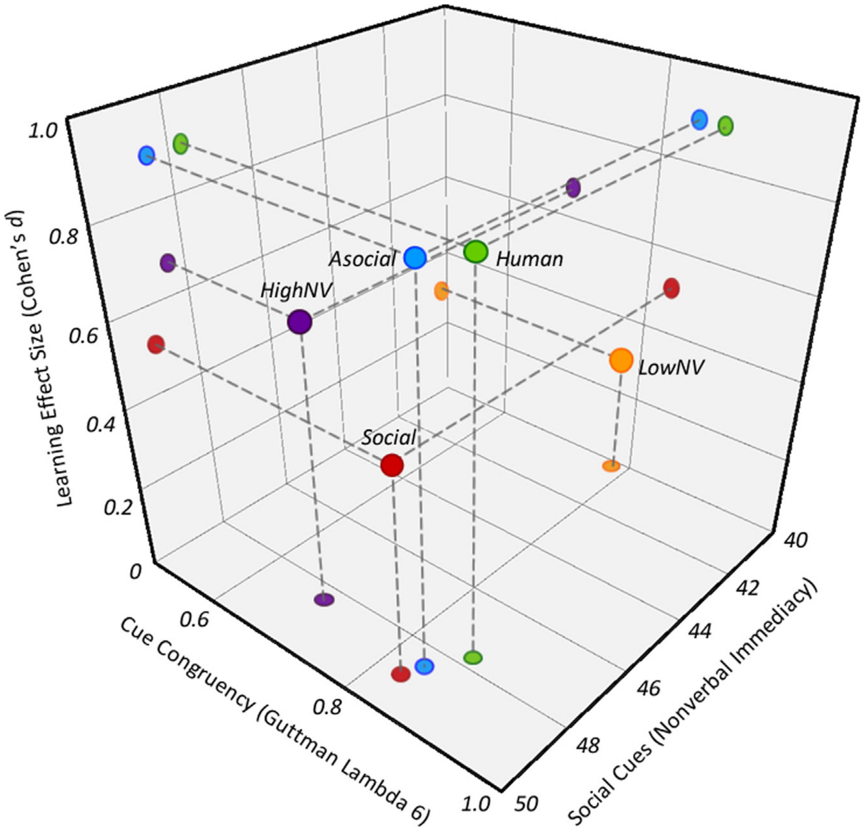


FIGURE 6 | Learning, congruency, and social behavior for each of the 5 conditions. Learning is measured in effect size between pretest and posttest for children. Congruency is indicated through Guttman's λ_6 of the adult nonverbal immediacy scores. Social behavior is characterized through nonverbal immediacy ratings from adults. An interactive version of this figure is available online to provide different perspectives of the space: https://goo.gl/ZNPxc8.

(indicated by Guttman's λ_6), learning also increases, and the same is true for social cues. The combination of congruency and social behavior as characterized by nonverbal immediacy provides a basis for learning predictions, where the combination of high social behavior and social cue congruency is necessary to maximize potential learning.

Such a hypothesis is supported by the view of social cues being perceived as a single percept, as suggested by Zaki (2013).

Experimental evidence with perception of emotions would seem to provide additional weight to such a perspective (Nook et al., 2015). This has clear implications for designers of social robot behavior when human perceptions or outcomes are of any degree of importance. The combination of all social cues in context must be considered alongside the expectations of the human to generate appropriate behavior. Not only does this give rise to a number of challenges, such as identifying combinatorial contextual expectations for social cues, but it could also have implications for how social cues should be examined experimentally. The isolation of specific social cues in experimental scenarios would not describe the role of that social cue, but the role of that social cue, *given the context of all other cues*. This is an important distinction that leads to a great deal more complexity in "solving" behavioral design for social robots, but that would also contribute to explanations of why a complex picture is emerging in terms of the effect of robot behavior on learning, as discussed in Section 2.1. The NVI metric and the predictions (that can be objectively examined) we put forward below provide a means through which robot behavior designers can iteratively implement and evaluate holistic social behaviors in an efficient manner, contributing to a more coherent framework in this regard. In particular, three predictions can be derived from the extremities of the space that is presented:

P1. Highly social behavior of a tutor robot (as characterized by nonverbal immediacy) with high congruency will lead to maximum potential learning.
P2. Low social behavior of a tutor robot with low congruency will lead to minimal potential learning.
P3. A mismatch in the social behavior of a tutor robot and the social cue congruency will lead to less than maximum potential learning.

Guttman's Lambda, as providing a measure of consistency, is used here as a proxy for the congruency of cues as observed by the study participants. We argue that this provides the necessary insight into cue congruency; however, the mapping between this metric and overtly judged congruency remains to be characterized. This would not necessarily be something that would be straightforward to achieve due to the potentially complex interactions between large numbers of social cues. For these predictions, use of the NVI metric as the characterization of social behavior would still suffer from some of the issues outlined earlier in this discussion: lack of timing information, relative cue importance, and novelty of behavior. The predictions are based on the general trends observed here, and it is noted that NVI is not a comprehensive measure of social behavior; indeed the SR condition in particular would not be fully explained using this means alone when compared to other results such as the AR condition. In addition, the data used for the learning axis were collected with relatively few samples (just over 10 per condition) in a specific experimental setup. Ideally, many further samples would be collected in both short and long term. The data collected here are over the short term and with children unfamiliar with robots. As longer term interactions take place, or as robots become more commonplace in society, expectations may change.

6. CONCLUSION

In this article, we have considered the use of nonverbal immediacy as a means of characterizing nonverbal social behavior in human–robot interactions. In a one-to-one maths tutoring task with humans and robots, it was shown that children and adults provide strong positively correlated ratings of tutor nonverbal immediacy. In addition, in agreement with the human–human literature, a positive correlation between tutor nonverbal immediacy and child learning was found. However, nonverbal immediacy alone could not account for all of the learning differences between tutoring conditions. This discrepancy led to the consideration of social cue congruency as an additional factor to social behavior in learning outcomes. Guttman's λ_6 was used to provide an indication of congruency between social cues. The combination of social behavior (as measured through nonverbal immediacy) and cue congruency (as indicated by Guttman's λ_6) provided an explanation of the learning data. It is suggested that if we are to achieve desirable outcomes with, and reactions to, social robots, greater consideration must be given to all cues in the context of multimodal social behavior and their possible perception as a unified construct. The hypotheses we have generated predict that the combination of high social behavior, and social cue congruency is necessary to maximize learning. The Robot Nonverbal Immediacy Questionnaire (RNIQ) developed for use here is offered as a means of gathering data for such characterizations.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of Plymouth University ethics board with written informed consent from all adult subjects. Child subjects gave verbal informed consent themselves, and written informed consent was provided by a parent or guardian. All subjects gave informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Plymouth University ethics board.

AUTHOR CONTRIBUTIONS

Conception and design of the work, interpretation and analysis of the data, and draft and critical revisions of the work: JK, PB, and TB. Acquisition of the data: JK.

FUNDING

This work is partially funded by the EU H2020 L2TOR project (grant 688014), the EU FP7 DREAM project (grant 611391), and the School of Computing, Electronics and Maths, Plymouth University, UK.

REFERENCES

Alemi, M., Meghdari, A., and Ghazisaedy, M. (2014). Employing humanoid robots for teaching English language in Iranian junior high-schools. *Int. J. Hum. Robot.* 11, 1450022-1–1450022-25. doi:10.1142/S0219843614500224

Ambady, N., and Rosenthal, R. (1993). Half a minute: predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *J. Pers. Soc. Psychol.* 64, 431. doi:10.1037/0022-3514.64.3.431

Bartneck, C., Kanda, T., Mubin, O., and Al Mahmud, A. (2009a). Does the design of a robot influence its animacy and perceived intelligence? *Int. J. Soc. Robot.* 1, 195–204. doi:10.1007/s12369-009-0013-7

Bartneck, C., Kuli, D., Croft, E., and Zoghbi, S. (2009b). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int. J. Soc. Robot.* 1, 71–81. doi:10.1007/s12369-008-0001-3

Baxter, P., Wood, R., and Belpaeme, T. (2012). "A touchscreen-based 'sandtray' to facilitate, mediate and contextualise human-robot social interaction," in *Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY: ACM), 105–106.

Belpaeme, T., Baxter, P., De Greeff, J., Kennedy, J., Read, R., Looije, R., et al. (2013). "Child-robot interaction: perspectives and challenges," in *Proceedings of the 5th International Conference on Social Robotics ICSR'*, Vol. 13 (Cham, Switzerland: Springer), 452–459.

Borgers, N., Sikkel, D., and Hox, J. (2004). Response effects in surveys on children and adolescents: the effect of number of response options, negative wording, and neutral mid-point. *Qual. Quant.* 38, 17–33. doi:10.1023/B:QUQU.0000013236.29205.a6

Christensen, L. J., and Menzel, K. E. (1998). The linear relationship between student reports of teacher immediacy behaviors and perceptions of state motivation, and of cognitive, affective, and behavioral learning. *Commun. Educ.* 47, 82–90. doi:10.1080/03634529809379112

Comstock, J., Rowell, E., and Bowers, J. W. (1995). Food for thought: teacher nonverbal immediacy, student learning, and curvilinearity. *Commun. Educ.* 44, 251–266. doi:10.1080/03634529509379015

Cronbach, L. J., and Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educ. Psychol. Meas.* 64, 391–418. doi:10.1177/0013164404266386

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., and Wagner, S. (2001). Explaining math: gesturing lightens the load. *Psychol. Sci.* 12, 516–522. doi:10.1111/1467-9280.00395

Goldin-Meadow, S., Wein, D., and Chang, C. (1992). Assessing knowledge through gesture: using children's hands to read their minds. *Cogn. Instr.* 9, 201–219. doi:10.1207/s1532690xci0903_2

Gordon, G., Breazeal, C., and Engel, S. (2015). "Can children catch curiosity from a social robot?" in *Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY: ACM), 91–98.

Gorham, J. (1988). The relationship between verbal teacher immediacy behaviors and student learning. *Commun. Educ.* 37, 40–53. doi:10.1080/03634528809378702

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika* 10, 255–282. doi:10.1007/BF02288892

Ham, J., Bokhorst, R., and Cabibihan, J. (2011). "The influence of gazing and gestures of a storytelling robot on its persuasive power," in *International Conference on Social Robotics* (Cham, Switzerland).

Han, J., Jo, M., Park, S., and Kim, S. (2005). "The educational use of home robots for children," in *Proceedings of the IEEE International Symposium on Robots and Human Interactive Communications RO-MAN*, Vol. 2005 (Piscataway, NJ: IEEE), 378–383.

Herberg, J., Feller, S., Yengin, I., and Saerbeck, M. (2015). "Robot watchfulness hinders learning performance," in *Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN*, Vol. 2015 (Piscataway, NJ), 153–160.

Huang, C.-M., and Mutlu, B. (2013). "Modeling and evaluating narrative gestures for humanlike robots," in *Proceedings of the Robotics: Science and Systems Conference, RSS'* (Berlin), 13.

Jung, Y., and Lee, K. M. (2004). "Effects of physical embodiment on social presence of social robots," in *Proceedings of the 7th Annual International Workshop on Presence* (Valencia, Spain), 80–87.

Kanda, T., Hirano, T., Eaton, D., and Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: a field trial. *Hum. Comput. Interact.* 19, 61–84. doi:10.1207/s15327051hci1901&2_4

Kanda, T., Shimada, M., and Koizumi, S. (2012). "Children learning with a social robot," in *Proceedings of the 7th ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY: ACM), 351–358.

Kelley, D. H., and Gorham, J. (1988). Effects of immediacy on recall of information. *Commun. Educ.* 37, 198–207. doi:10.1080/03634528809378719

Kennedy, J., Baxter, P., and Belpaeme, T. (2015a). "Can less be more? The impact of robot social behaviour on human learning," in *Proceedings of the 4th International Symposium on New Frontiers in HRI at AISB 2015* (Canterbury).

Kennedy, J., Baxter, P., and Belpaeme, T. (2015b). Comparing robot embodiments in a guided discovery learning interaction with children. *Int. J. Soc. Robot.* 7, 293–308. doi:10.1007/s12369-014-0277-4

Kennedy, J., Baxter, P., and Belpaeme, T. (2015c). "The robot who tried too hard: social behaviour of a robot tutor can negatively affect child learning," in *Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY: ACM), 67–74.

Kennedy, J., Baxter, P., Senft, E., and Belpaeme, T. (2015d). "Higher nonverbal immediacy leads to greater learning gains in child-robot tutoring interactions," in *Proceedings of the International Conference on Social Robotics* (Cham, Switzerland).

Kennedy, J., Baxter, P., and Belpaeme, T. (2017). Nonverbal immediacy as a characterisation of social behaviour for human-robot interaction. *Int. J. Soc. Robot.* 9, 109–128. doi:10.1007/s12369-016-0378-3

Kennedy, J., Baxter, P., Senft, E., and Belpaeme, T. (2016). "Heart vs hard drive: children learn more from a human tutor than a social robot," in *Proceedings of the 11th ACM/IEEE Conference on Human-Robot Interaction* (Piscataway, NJ: ACM), 451–452.

Leyzberg, D., Spaulding, S., Toneva, M., and Scassellati, B. (2012). "The physical presence of a robot tutor increases cognitive learning gains," in *Proceedings of the 34th Annual Conference of the Cognitive Science Society, CogSci*, Vol. 2012 (Austin, TX), 1882–1887.

McCroskey, J. C., Sallinen, A., Fayer, J. M., Richmond, V. P., and Barraclough, R. A. (1996). Nonverbal immediacy and cognitive learning: a cross-cultural investigation. *Commun. Educ.* 45, 200–211. doi:10.1080/03634529609379049

Mehrabian, A. (1968). Some referents and measures of nonverbal behavior. *Behav. Res. Methods Instrum.* 1, 203–207. doi:10.3758/BF03208096

Moore, R. K. (2012). A Bayesian explanation of the 'Uncanny Valley' effect and related psychological phenomena. *Nat. Sci. Rep.* 2:864. doi:10.1038/srep00864

Mori, M., MacDorman, K. F., and Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robot. Autom. Mag.* 19, 98–100. doi:10.1109/MRA.2012.2192811

Moshkina, L., Trickett, S., and Trafton, J. G. (2014). "Social engagement in public places: a tale of one robot," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY: ACM), 382–389.

Nook, E. C., Lindquist, K. A., and Zaki, J. (2015). A new look at emotion perception: concepts speed and shape facial emotion recognition. *Emotion* 15, 569–578. doi:10.1037/a0039166

Nussbaum, J. F. (1992). Effective teacher behaviors. *Commun. Educ.* 41, 167–180. doi:10.1080/03634529209378878

Reeves, B., and Nass, C. (1996). *How People Treat Computers, Television, and New Media like Real People and Places*. New York, NY: CSLI Publications, Cambridge University press.

Revelle, W., and Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: comments on Sijtsma. *Psychometrika* 74, 145–154. doi:10.1007/s11336-008-9102-z

Richmond, V., McCroskey, J., and Payne, S. (1987). *Nonverbal Behavior in Interpersonal Relations*. Englewood Cliffs, NJ: Prentice-Hall.

Richmond, V. P., McCroskey, J. C., and Johnson, A. D. (2003). Development of the Nonverbal Immediacy Scale (NIS): measures of self- and other-perceived nonverbal immediacy. *Commun. Q.* 51, 504–517. doi:10.1080/01463370309370170

Robinson, R. Y., and Richmond, V. P. (1995). Validity of the verbal immediacy scale. *Commun. Res. Rep.* 12, 80–84. doi:10.1080/08824099509362042

Saerbeck, M., Schut, T., Bartneck, C., and Janse, M. D. (2010). "Expressive robots in education: varying the degree of social supportive behavior of a robotic tutor," in

Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI'10 (New York, NY: ACM), 1613–1622.

Sung, J., Christensen, H. I., and Grinter, R. E. (2009). "Robots in the wild: understanding long-term use," in *Prcoeedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY: IEEE), 45–52.

Szafir, D., and Mutlu, B. (2012). "Pay attention! Designing adaptive agents that monitor and improve user engagement," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI'12* (New York, NY: ACM), 11–20.

Tanaka, F., and Matsuzoe, S. (2012). Children teach a care-receiving robot to promote their learning: field experiments in a classroom for vocabulary learning. *J. Hum. Robot Interact.* 1, 78–95. doi:10.5898/JHRI.1.1.Tanaka

Wainer, J., Feil-Seifer, D. J., Shell, D. A., and Mataric, M. J. (2007). "Embodiment and human-robot interaction: a task-based perspective," in *Proceedings of the 16th IEEE International Symposium on Robot and Human interactive Communication (IEEE), RO-MAN*, Vol. 2007 (Piscataway, NJ), 872–877.

Wilson, J. H., and Locker, L. Jr. (2007). Immediacy scale represents four factors: nonverbal and verbal components predict student outcomes. *J. Classroom Interact.* 42, 4–10.

Witt, P. L., Wheeless, L. R., and Allen, M. (2004). A meta-analytical review of the relationship between teacher immediacy and student learning. *Commun. Monogr.* 71, 184–207. doi:10.1080/036452042000228054

Zajonc, R. B. (1965). Social facilitation. *Science* 149, 269–274. doi:10.1126/science.149.3681.269

Zaki, J. (2013). Cue integration a common framework for social cognition and physical perception. *Perspect. Psychol. Sci.* 8, 296–312. doi:10.1177/1745691613475454

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Kennedy, Baxter and Belpaeme. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

A. Robot Nonverbal Immediacy Questionnaire (RNIQ)

The following is the questionnaire used by participants in the evaluation to rate the nonverbal immediacy of the robot, based on the short-form nonverbal immediacy scale-observer report. Options are provided in equally sized boxes below each question (or equally spaced radio buttons in the online version). The options are: **1 = Never; 2 = Rarely; 3 = Sometimes; 4 = Often; 5 = Very Often**. The questions are as follows:

1. The robot uses its hands and arms to gesture while talking to you
2. The robot uses a dull voice while talking to you
3. The robot looks at you while talking to you
4. The robot frowns while talking to you
5. The robot has a very tense body position while talking to you
6. The robot moves away from you while talking to you
7. The robot changes how it speaks while talking to you
8. The robot touches you on the shoulder or arm while talking to you
9. The robot smiles while talking to you
10. The robot looks away from you while talking to you
11. The robot has a relaxed body position while talking to you
12. The robot stays still while talking to you
13. The robot avoids touching you while talking to you
14. The robot moves closer to you while talking to you
15. The robot looks keen while talking to you
16. The robot is bored while talking to you

Scoring:

Step 1. Add the scores from the following items:
1, 3, 7, 8, 9, 11, 14, and 15.

Step 2. Add the scores from the following items:
2, 4, 5, 6, 10, 12, 13, and 16.

Total Score = 48 plus Step 1 minus Step 2.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 31 May 2017
doi: 10.3389/frobt.2017.00021

To Err Is Robot: How Humans Assess and Act toward an Erroneous Social Robot

Nicole Mirnig[1], Gerald Stollnberger[1], Markus Miksch[1], Susanne Stadler[1], Manuel Giuliani[2] and Manfred Tscheligi[1,3]*

[1] Center for Human-Computer Interaction, University of Salzburg, Salzburg, Austria, [2] Bristol Robotics Laboratory, University of the West of England, Bristol, United Kingdom, [3] Center for Technology Experience, Austrian Institute of Technology, Vienna, Austria

We conducted a user study for which we purposefully programmed faulty behavior into a robot's routine. It was our aim to explore if participants rate the faulty robot different from an error-free robot and which reactions people show in interaction with a faulty robot. The study was based on our previous research on robot errors where we detected typical error situations and the resulting social signals of our participants during social human–robot interaction. In contrast to our previous work, where we studied video material in which robot errors occurred unintentionally, in the herein reported user study, we purposefully elicited robot errors to further explore the human interaction partners' social signals following a robot error. Our participants interacted with a human-like NAO, and the robot either performed faulty or free from error. First, the robot asked the participants a set of predefined questions and then it asked them to complete a couple of LEGO building tasks. After the interaction, we asked the participants to rate the robot's anthropomorphism, likability, and perceived intelligence. We also interviewed the participants on their opinion about the interaction. Additionally, we video-coded the social signals the participants showed during their interaction with the robot as well as the answers they provided the robot with. Our results show that participants liked the faulty robot significantly better than the robot that interacted flawlessly. We did not find significant differences in people's ratings of the robot's anthropomorphism and perceived intelligence. The qualitative data confirmed the questionnaire results in showing that although the participants recognized the robot's mistakes, they did not necessarily reject the erroneous robot. The annotations of the video data further showed that gaze shifts (e.g., from an object to the robot or vice versa) and laughter are typical reactions to unexpected robot behavior. In contrast to existing research, we assess dimensions of user experience that have not been considered so far and we analyze the reactions users express when a robot makes a mistake. Our results show that decoding a human's social signals can help the robot understand that there is an error and subsequently react accordingly.

Keywords: social human–robot interaction, robot errors, user experience, social signals, likeability, faulty robots, error situations, *Pratfall Effect*

OPEN ACCESS

Edited by:
Ginevra Castellano,
Uppsala University, Sweden

Reviewed by:
Rodolphe Gelin,
Aldebaran Robotics, France
Patrícia Alves-Oliveira,
ISCTE – University Institute of Lisbon, Portugal

****Correspondence:***
Nicole Mirnig
nicole.mirnig@sbg.ac.at

Specialty section:
This article was submitted to Humanoid Robotics, a section of the journal Frontiers in Robotics and AI

Received: *30 December 2016*
Accepted: *10 May 2017*
Published: *31 May 2017*

Citation:
Mirnig N, Stollnberger G, Miksch M, Stadler S, Giuliani M and Tscheligi M (2017) To Err Is Robot: How Humans Assess and Act toward an Erroneous Social Robot. Front. Robot. AI 4:21. doi: 10.3389/frobt.2017.00021

1. INTRODUCTION

Social robots are not yet in a technical state where they operate free from errors. Nevertheless, most research approaches act on the assumption of robots performing faultlessly. This results in a confined standpoint, in which the created scenarios are considered as gold standard. Alternatives resulting from unforeseeable conditions that develop during an experiment are often not further regarded or simply excluded. It lies within the nature of thorough scientific research to pursue a strict code of conduct. However, we suppose that faulty instances of human–robot interaction (HRI) are nevertheless full with knowledge that can help us further improve the interactional quality in new dimensions. We think that because most research focuses on perfect interaction, many potentially crucial aspects are overlooked.

Research that is specifically directed at exploring erroneous instances of interaction could be useful to further refine the quality of HRI. For example, a robot that understands that there is a problem in the interaction by correctly interpreting the user's social signals could let the user know that it understands the problem and actively apply error recovery strategies. Knowing the severity of an error could further be helpful for the robot in finding the adequate corrective action.

Since robots in HRI are social actors, they elicit mental models and expectations known from human–human interaction (HHI) (Lohse, 2011). One aspect we know from HHI is that imperfections make human social actors more likeable and more believable. The psychological phenomenon *Pratfall Effect* states that people's attractiveness increases when they commit a mistake. Aronson et al. (1966) suggest that superior people may be viewed as superhuman and distant while a mistake would make them seem more human. Similarly, one could argue that robots are often seen as impeccable, since this is how they are presented in the media (Bruckenberger et al., 2013). Especially, people who have not interacted with robots themselves build their mental models and expectations about robots from those media. Moreover, experience with technology in general is mostly based on interaction with consumer products, such as smartphones or TVs. Those products are very common and need to work more or less error-free in order to get accepted on the market. For example, a TV which has problems in sound will not survive long on the market. People expect technology they paid for to work without errors. What makes the interaction with social robots different is that a TV is not seen as a social actor, in contrast to a social robot. This might result in people assuming robots to be without fail, which makes them likewise seem distant (*Pratfall Effect*). Robots that commit errors, on the other hand, could then be viewed as more human-like and, in subsequence, more likeable. With their study on an erroneous robot in a competitive game-play scenario Ragni et al. (2016) provided additional evidence that people consider robots in general as competent, functional, and intelligent.

In our effort to embrace the imperfections of social robots and create more believable robot characters, we propose to specifically explore faulty robot behavior and the social signals humans show when a robot commits a mistake. The term social signal is used to describe verbal and non-verbal signals that humans use in a conversation to communicate their intentions. Vinciarelli et al. (2009) argued that the ability to recognize social signals is crucial to mastering social intelligence. It is our long-term goal to enable robots to communicate about their errors and deploy recovery strategies. To achieve this ambitious goal, more general knowledge about robot errors is required. We report on a user study where we purposefully elicited faulty robot behavior.

Our user study is based on our previous research where we analyzed an extensive pool of video data showing social HRI instances where the robot made an error. The videos covered a variety of scenarios in different contexts, different robots, and a multitude of social signals. The robot errors happened unintentionally and, thus, the data created a sound basis for studying the nature of error situations. We found that there are two different kinds of robot errors, i.e., *social norm violations* (SNV) and *technical failures* (TF) (Giuliani et al., 2015), for which human interaction partners respond with typical social signals (Mirnig et al., 2015). A social norm violation means that the robot's actions deviate from the underlying social script, that is, the commonly known interaction steps a certain situation is expected to take. For example, a participant orders a drink from a bartender robot, the robot signals it has understood but then asks again for the participant's order. A technical failure means that the robot experiences a technical disruption that is perceived as such by the user. For example, a robot picks up an object but then loses it while grasping. From an expert perspective all robot errors might be considered as technical failures. Since, we are interested in the human perception of robot errors, we distinguish error types from how a human most likely perceives error events.

With the user study presented in this paper, we expand our previous research in purposefully eliciting robot errors and researching the resulting social signals of the human interaction partners. We measured how users perceive a robot that makes errors during interaction (social norm violations and technical failures) as compared to a robot operating free from errors.

The directed exploration of robot errors in social interaction is a new and upcoming topic. The HRI research community has reported first results on exploratory user studies. For example, Salem et al. (2015) conducted an experiment with an erroneous robot. The researchers measured how the robot's behavior influenced how the participants rated its trustworthiness and reliability. They also measured if robot errors affect the task performance. The researchers found that while participants rated the correctly behaving robot as significantly more trustworthy and reliable, the fact that a robot performs correctly or faulty did not influence the objective task performance.

In an earlier work, Salem et al. (2013) researched the effect of speech and gesture congruence on perceived anthropomorphism, likability, and task performance. In their experiment, a robot either spoke only, spoke while making congruent coverbal gestures, or spoke while making incongruent coverbal gestures. The researchers found that congruent coverbal gesturing makes a robot appear more anthropomorphic and more likeable. This effect was even stronger for incongruent coverbal gesturing. However, incongruent coverbal gesturing resulted in a lower task performance. Following our line of argumentation, such incongruent behavior violates the human social script, as humans do not expect incongruent messages from different modalities in everyday interactions (Schank and Abelson, 1977). Therefore, incongruent multimodal robot behavior results in a *social norm violation*. Ragni et al. (2016) reported similar effects. The researchers performed a study in which a human and a robot competed against each other in a reasoning task and a memory task. During the interaction, the robot either performed with or without errors. While participants rated the faulty robot as less competent, less reliable, less intelligent, and less superior than the error-free robot, participants reported having enjoyed the interaction more when the robot made errors. However, the task performance was significantly lower in the faulty robot condition.

Gompei and Umemuro (2015) investigated how a robot's speech errors influenced how familiar and sincere it was rated. The researchers found that speech errors made early in an interaction might lower the robot's sincerity rating. However, speech errors that are introduced later in the interaction might increase the robot's familiarity. Short et al. (2010) investigated people's perception when playing rock–paper–scissors with a robot that either played fair, cheated verbally by announcing a different hand gesture, or cheated with its actions by changing the hand gesture. The researchers found that a cheating robot resulted in a bigger social engagement, in comparison to one which plays fair. They stated that the results suggest that participants showed more verbal social signals to the robot that cheated. Participants were surprised by the cheating behavior of the robot, although verbal cheating was perceived as malfunction, while cheating through action was perceived as deliberate cheating behavior. These findings support our assumption that through unexpected behavior, people see a robot as a more social actor and that unexpected behavior might be interpreted as erroneous behavior.

In an online survey, Lee et al. (2010) found that when a service robot made a mistake, this has a strong negative impact on people's rating of the service quality and the robot itself. However, when the robot deployed a recovery strategy, both the rating of the service and the rating of the robot improved. The researchers deployed different recovery strategies and found that all of them increased the ratings of the robot's politeness. A robot which apologized for its mistake was seen more competent, people liked it more and felt closer to it, and a robot offering compensation for its mistake (such as a refund) was rated to be of more satisfying service quality but participants were hesitant to use the robot again. Whereas, an apology and a recovery strategy of offering options was perceived to foster reuse likelihood. In a related online survey, Brooks et al. (2016) explored people's reactions to the failure of an autonomous robot. In the survey, participants were asked to assess situations where an autonomous robot experienced different kinds of failures that affected a human interacting with it. They found that people who saw an erroneous robot rated it rather negatively on a series of items (i.e., How satisfying, pleasing, disappointing, reliably, dependable, competent, responsible, trustworthy, risky to use is the robot?), while people who experienced a robot without failure rated it positively. When the erroneous robot deployed mitigation strategies to overcome the error either by prompting human intervention or by deploying a different approach, people's ratings toward the erroneous robot became less negative. However, the amount the strategy influenced peoples reaction depended on the kind of task, the severity of the failure, and the risk of the failure.

To enable a robot to generate help requests in case of an error situation, Knepper et al. (2015) developed their inverse semantics algorithm. It allows the robot to phrase precise requests that specify the kind of help that is needed. The researchers evaluated their algorithm in a user study and found that participants preferred the precise request over high level, general phrasings. While in their approach errors are recognized through the robot's internal state and the environment (e.g., the robot is supposed to pick up an object which it can visually detect, but the object is out of its reach), we envision an approach where the robot can additionally detect an error through its human interaction partner's social signals. For example, Gehle et al. (2015) explored gaze patterns of human groups upon unexpected robot behavior in a museum guide scenario. They found that groups of visitors responded to unexpected robot behavior with stepwise gaze coordination, applying different modes of gaze constellation. Unexpected robot behavior is likely to conflict with the user expectations about the adequate social script in a certain situation. Therefore, unexpected robot behavior can lead to a social norm violation. A deviation from the social script resulted in a different strategy in the human gaze coordination (social signals). Hayes et al. (2016) performed a user study in which participants were instructed to teach a dance to a robot. They explored how humans implicitly responded when the robot made a mistake. The authors used a very small sample in their explorative study and did not provide a statistical analysis of their descriptive results.

Our approach extends the existing findings in several dimensions. While the errors in the study of Ragni et al. (2016) were based on errors from HHI, the errors we used were modeled based on data from HRI. Our work and that of Ragni et al. (2016) further cover different aspects: (a) their errors were task-related, ours non-task-related; (b) they covered the cognitive ability of the robot and we dealt with socially (in) appropriate robot behavior and more general soft- and hardware problems; and (c) they assessed the overall enjoyment of the interaction and users' task performance, while we looked into the interconnectedness of likability, anthropomorphism, and intelligence. We chose to examine these factors since they are commonly used and accepted measures in the HRI domain. We were especially interested in likability as it contributes to the overall user experience and it may foster technology acceptance. Since erroneous behavior potentially compromises intelligence ratings, we were also interested in exploring if our robot's mistakes make it seem less intelligent. In the light of the *Pratfall Effect*, we wanted to see if the robot's anthropomorphism level is influenced by the fact that it makes or does not make mistakes.

The related literature shows that the importance of exploring robot errors has been recognized. We extend the state of the art with our data-driven approach by systematically analyzing specific kinds of errors and their effects on the interaction experience, as well as the users' reactions to those errors (i.e., social signals).

2. MATERIALS AND METHODS

We set up a Wizard of Oz (WOz) user study to specifically explore robot errors. A human and a robot interacted with each other in two verbal sessions. The first session was a verbal interview where the robot asked a few questions to the participant. The second session was a LEGO task, where the robot invited the participant to build a few simple objects. We chose this setup in order to reenact the verbal context of the related work (Giuliani et al., 2015; Mirnig et al., 2015). In addition, the interview session enabled us to collect qualitative data on the participants' opinions, which we included in our data analysis.

The user study was performed between subjects, with each participant taking part in one of the following two conditions: (a) *no error* (baseline—the robot performs error-free) and (b) *error* (experimental condition—the robot commits eight errors over the entire interaction). To base the user study on the previous findings from Giuliani et al. (2015) and Mirnig et al. (2015), we programmed the robot to commit two social norm violations and two technical failures in each session. Based on our previous research, we defined these two types of error as the typical mistakes robots make in HRI. Therefore, we suppose that an interaction including these error types would be perceived as plausible. The complexity, severity, and risk level of the induced errors were chosen in alignment with our scenario. Naturally, different scenarios will entail other errors, different severity and risk levels. For example, Robinette et al. (2014) investigated faulty behavior of robots in safety critical situations. They simulated erroneous behavior of an emergency guiding robot that helps people to escape from a dangerous zone. They found that after the first error of the robot, people's attitude toward the robot decreased significantly. However, the decision to follow the robot in a follow-up interaction was not affected by their decreased attitude.

2.1. Hypotheses

As discussed in the previous sections, it is known that humans often base their expectations about robots on how robots are portrayed in the media. Since the media present robots frequently as perfect entities, we assume that social robots making errors negatively influence how their human interaction partners perceive them. Based on the findings on faulty robot actions in HRI as discussed so far and in light of the *Pratfall Effect*, we have postulated the following hypotheses for our user study:

H1: A robot that *commits errors* during its interaction with humans is perceived as *more likeable* than a robot that performs flawlessly.

H2: A robot that *commits errors* during its interaction with humans is perceived as *more anthropomorphic* than a robot that performs flawlessly.

H3: A robot that *commits errors* during its interaction with humans is perceived as *less intelligent* than a robot that performs flawlessly.

2.2. User Study Design

For the WOz user study, the participants were asked to interact with a NAO robot.[1] We set the interaction up in two sessions. During the first session, the robot asked a set of predefined questions to the participant in order to restrict the thematic dimension of the conversation. During the second session, the robot invited the participant to perform a couple of tasks using LEGO bricks.

In the interview session, the robot asked ten questions to the participant. The first three questions were meant to make the participant familiar with the situation and to create a comfortable atmosphere. For this reason, they were always presented in the same order and they never contained an error. The subsequent seven questions were asked in random order and four out of seven questions contained errors in the *error* condition.

In the LEGO session, the participant had to (dis-)assemble LEGO bricks according to the robot's instructions. The first two tasks were assigned in the same order for all participants and they did not contain errors. The subsequent eight tasks were assigned in random order and four out of eight tasks contained errors in the *error* condition.

The interview session lasted for an average of 3 min and 37 s (SD = 59 s) and the LEGO session 8 min and 14 s (SD = 1 min and 54 s). We decided for this two-part setup to keep the participants entertained with a diversified scenario. The two-part setup provided us also with the possibility to introduce a greater variety of errors and to achieve a higher number of errors in total.

The user study was performed in the User Experience and Interaction Experimentation Lab at the Center for Human-Computer Interaction at the University of Salzburg. The robot was wizarded from a researcher seated behind a bookshelf so that the wizarding was not obvious to the participant. A second researcher, likewise seated behind the bookshelf, controlled the video recording. During the entire interaction the participants stood adverse to the NAO robot at a distance of approximately 1.5 m. NAO was standing on a desk (see **Figure 1** for the setup). The transition between the two sessions was immediate with no break in between. Both sessions happened in the same setting. The only change was that the researcher placed a wooden box (80 cm × 50 cm × 50 cm) on the table in front of the robot right before the LEGO session started. The box was used to provide the participants with a comfortable height to complete the building tasks. Together with the box, the participants were given a set of LEGO blocks (prebuilt shapes) with which they were to perform the tasks (see **Figure 2**).

The between-subjects design required each person participating in either one of the two conditions. In the baseline condition, the robot performed free from errors. In the experimental condition, the robot committed two social norm violations and two technical failures each in both sessions. After each robot error, the researchers waited for the situation to unravel without them interfering. In many cases, the participants showed a reaction that confirmed that they had noticed the error (e.g., some participants laughed or frowned) and then moved on. The researchers only intervened in the rare cases where the interaction was severely

[1] https://www.ald.softbankrobotics.com/en/cool-robots/nao

interrupted, for example, when the participant directly addressed the researchers and commented on the error. In this case, the researcher simply asked the participant to continue interacting with the robot, in order to limit the interference as much as possible.

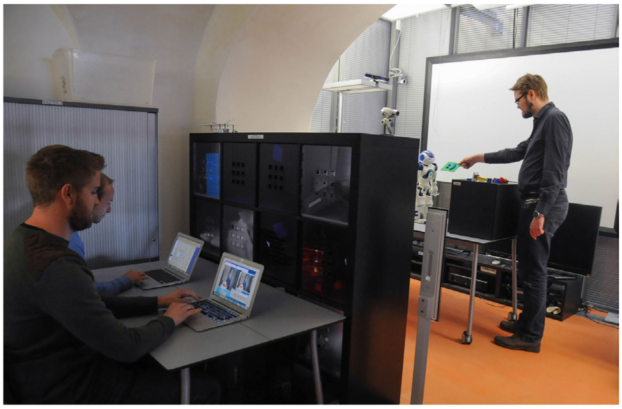


FIGURE 1 | Study setup with the participant interacting with the robot and two researchers seated behind a bookshelf who supervised the technology.

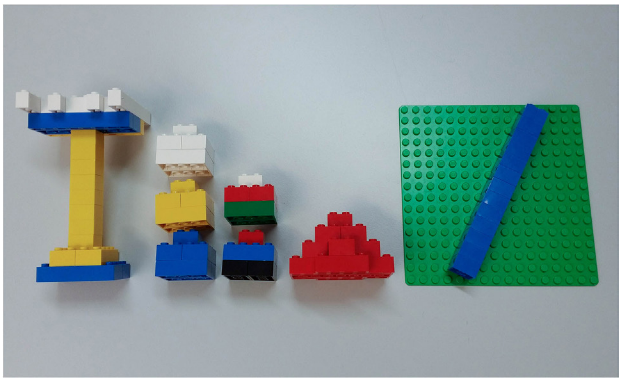


FIGURE 2 | LEGO blocks that were provided to the participants.

The three starting questions in the interview session and the first two building tasks were meant as an introduction and were not varied in order. Therefore, the robot errors occurred in the randomized questions/tasks only. **Tables 1** and **2** give an overview on the questions and tasks and which errors occurred together with which question or task. The questions were similar in both conditions. The difference between the baseline and the experimental condition was achieved by the presence or absence of the robot errors.

The induced errors were mainly modeled based on our previous findings on typical robot errors as reported in the studies of Giuliani et al. (2015) and Mirnig et al. (2015). Only LEGO task number 7 in the *error* condition was inspired by unusual requests as reported in the study of Salem et al. (2015).

The setup of our user study is based on real-life HRI. It is data-driven in representing actual error situations and corresponding robot errors that occur when humans interact with state-of-the-art social robots, which makes our setup ecologically valid.

2.3. User Study Procedure

The participants were welcomed to the laboratory. After a short briefing, they were asked to sign an informed consent. Next, the participants were asked to complete questionnaires to assess their demographics, personality traits, and attitude toward robots. The participants were introduced to the robot and they were given an overview on the process of the user study. As soon as the participants took their position opposite the robot, the user study began. First, the participants answered a set of questions the robot asked them (Session 1). Second, the robot instructed the participants to complete a set of building tasks with LEGO blocks (Session 2). After the interaction with the robot, the participants were again asked to complete the questionnaire assessing their attitude toward robots. They were further asked to complete a questionnaire rating the robot's likability, anthropomorphism, and perceived intelligence. The study was finalized with a closing interview where the researcher asked the participants four open-ended questions, which were followed by a short debriefing in

TABLE 1 | Interview session.

	#	Question	Error type	Error
Fixed order	1	What do you think is a robot?	–	None
	2	Which three properties come to your mind when you think about robots?	–	None
	3	Which robots do you know?	–	None
Randomized order	4	Would you like a robot that assists you with household chores?	SNV	The robot waits 15 s until it speaks again
	5	Why do you think some people are afraid of robots?	SNV	The robot starts speaking after 2.5 s, cutting off the participant
	6	Which skills would you like for a robot to have?	–	None
	7	In which areas could humanoid robots be helpful?	–	None
	8	Have you interacted with a robot before?	TF	The robot starts speaking but cuts the sentence off after "interac"
	9	Is hard- or software more important to you?	TF	The robot repeats the sentence 6 times
	10	Which tasks would you never entrust a robot with?	–	None

The questions comprised two Social Norm Violations (SNV) and two Technical Failures (TF).

TABLE 2 | LEGO session.

	#	Task	Error type	Error
Fixed order	1	Place all single-color blocks on top of each other. The order does not matter [participant performs task]. Unfortunately, the colors do not match how I imagined. Please take the blocks apart again.	–	None
	2	What animal comes to your mind? Please draw it with the blue blocks onto the green board and show it to me.	–	None
Randomized order	3	Pick the multicolor block you like least. Disassemble it and build something new.	–	None
	4	Build a tower from all blocks that have red pieces in them.	–	None
	5	Build a bridge from four blocks that gets as long as possible [participant performs task]. Wonderful! Please disassemble the bridge into the four original blocks.	–	None
	6	Count how many parts the red pyramid is made of. If you need to disassemble the pyramid to count the bricks put it back together in the end. Tell me the number.	–	None
	7	Place all single-color blocks on the right side and the remaining blocks on the left (*no error* condition)/Throw three blocks on the floor at once! (*error* condition).	SNV	In the *error* condition, instead of giving the sorting task to the participant, the robot instructs the participant to throw three blocks on the floor at once
	8	Place all blocks in a row sorting them by size. Begin with the smallest.	SNV	The robot waits 15 s until it speaks again
	9	Build something creative from the yellow and the blue block.	TF	The robot repeats the word yellow as if stuck in a loop ("Build something creative from the yellow, yellow, yellow, …")
	10	Which facial expression depicts your current emotional state? Please draw the expression with the blue blocks onto the green board [participant performs task]. Please place the picture in my hands. With the command "grasp!" I close my hands.	TF	The robot tries closing its hands but repeatedly fails to grasp the piece

The tasks comprised two Social Norm Violations (SNV) and two Technical Failures (TF).

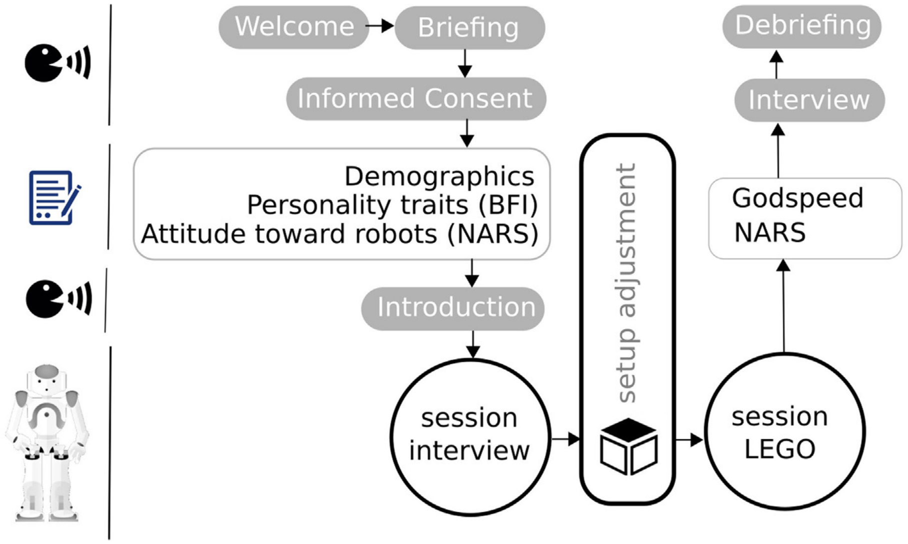


FIGURE 3 | **Study procedure**.

which the purpose of the study was explained to the participants. The study procedure is depicted in **Figure 3**.

2.4. Dependent Measures

Before the interaction, we asked our participants to fill in the Big Five Inventory (BFI) questionnaire by John et al. (2008). We used this questionnaire to analyze if people's personality influences how they perceive the robot. The BFI consists of 44 items (5-point Likert-scaled), constructing five subscales (extraversion, agreeableness, conscientiousness, neuroticism, and openness). This questionnaire is a well-accepted instrument among psychologists to assess the personality of humans. Therefore, we chose to use it for exploring potential connections between personality and how a social robot is perceived.

We used the Negative Attitude Toward Robots Scale (NARS) (Nomura et al., 2004) to assess participants' general attitude toward robots. The NARS consists of 14 items (5-point Likert-scaled) that account for three scales: people's negative attitude toward (S1) interaction with robots, (S2) social influence of robots, and (S3) emotions in interaction with robots. We asked the participants to complete the questionnaire before and after their interaction with the robot in order to measure if the interaction changed people's attitude. The NARS is a widely used questionnaire in the HRI community and it provides researchers with a comprehensive understanding of human fears around social robots.

To explore how our participants rate the robot, we used three subscales from the Godspeed Questionnaire Series by Bartneck et al. (2009), i.e., anthropomorphism, likability, and perceived intelligence. Each of the scales consists of five 5-point Likert-scaled items. The scales were developed in the HRI community to specifically assess users' perception of social robots. We chose the questionnaires since they are frequently used and widely accepted among the HRI community. The concepts the questionnaires cover are very relevant to social HRI and they represent the concepts we explore with our research. This questionnaire was administered once, after our participants' interaction with the robot.

2.5. Interview Data

We used two sources to gain qualitative data from the participants regarding their attitude toward robots. First, the robot asked the participants about their opinion on robots in the interview session (see **Table 1**). Second, in the concluding interview after the interaction and after all the other questionnaires were filled in, we asked the following questions:

1. Did you notice anything special during your interaction with the robot that you would like to tell us?
2. Did your attitude toward robots change during the interaction?
3. What would you change about the interaction with the robot?
4. What did you think when the robot made a mistake? (This question was only asked for participants who took part in the *error* condition.)

2.6. Participants

A total of 45 participants took part in our user study (25 males and 20 females). The participants were recruited over a university mailing list and social media. They were primarily university students and they had no previous experience with robots. Their age ranged from 16 to 76 years, with a mean age of 25.91 years (SD = 10.82). As regards conditions, 21 participants completed the *error* condition and 24 the *no error* condition. The participants' technology affinity was rated on average with a mean of 3.09 (SD = 1.49; 5-point Likert-scaled ranging from 1—"not technical" to 5—"technical") and their preexperience with robots was below average with a mean of 1.96 (SD = 0.82; 5-point Likert-scaled ranging from 1—"never seen" to 5—"frequent usage").

2.7. Manipulation Check

In order to verify that the manipulation programmed into the robot's behavior was effective, we analyzed the videos of the interactions. Out of the 21 participants of the *error* condition, 18 exhibited clearly noticeable reactions upon the robot's faults (e.g., laughing, looking up from the LEGO at the robot, annoyed facial expression). During the closing interview with the researcher, 15 of the 21 participants stated that they noticed the robot making errors. All three persons who had not shown reactions upon the robot's errors in the video mentioned them in the interview. We, therefore, conclude that our manipulation was effective.

3. RESULTS

We used non-parametric statistical test procedures for data analysis, since our data were mostly not normally distributed (Kolmogorov–Smirnov test). Mann–Whitney-*U* tests were used to compare between two independent samples (between the two conditions and between the genders). Wilcoxon rank-sum tests were used to compare paired samples (ratings of the same scales before and after the interaction).

We coded the qualitative data from both interviews thematically (the one the robot conducted and the concluding interview after the interaction). We further annotated the video recordings from the participants' interaction to investigate their social signals when experiencing an error situation with the robot. **Figure 4** shows a participant interacting with the robot during the LEGO building session. The coding was performed from one of the authors since we coded objectively visible events only.

3.1. Questionnaire Data

The gender distribution across conditions was roughly balanced. While 24 participants (15 males and 9 females) interacted with a flawless robot in the *no error* baseline condition, 21 participants (10 males and 11 females) were interviewed by an error-prone robot in the *error* experimental condition.

3.1.1. Participants' Personality

We explored if our participants' personality influenced their rating of the robot by measuring five major personality traits. The scales of the BFI are constructed with semantic differential items that measure the participants' position between two poles

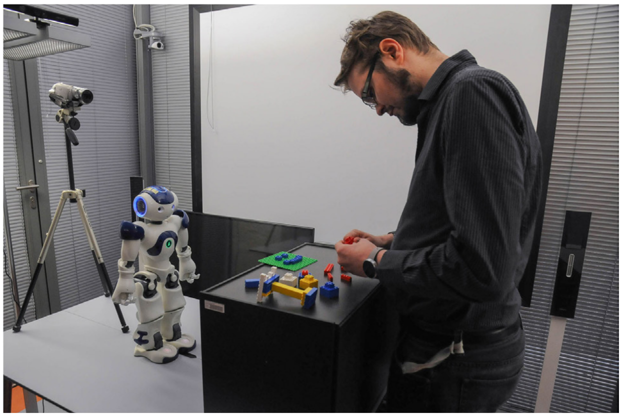


FIGURE 4 | **Participant interacting with the robot during the LEGO building session**.

(e.g., 1—introvert to 5—extravert). The arithmetic mean of these items with no emphasis on either one of the poles is 3.

3.1.1.1. Scale Reliability

The subscales extraversion, neuroticism, and openness resulted in high reliability (Cronbach's $\alpha = 0.82$, 0.81, and 0.85). The reliability for the conscientiousness scale was acceptable ($\alpha = 0.71$) and the one for agreeableness borderline acceptable ($\alpha = 0.61$).

3.1.1.2. Participants' Overall Personality

The results showed that the participants were slightly more extroverted (mean = 3.34, SD = 0.72), conscientious (mean = 3.42, SD = 0.57), and open (mean = 3.38, SD = 0.79) than the arithmetic mean. They were rather agreeable (mean = 3.79, SD = 0.47) and slightly less neurotic than average (mean = 2.91, SD = 0.73).

3.1.1.3. Participants' Personality Compared between Conditions

We performed Mann–Whitney-*U* tests to explore if participants' personality profile differed between conditions. The tests for all three subscales were non-significant, showing that participants' personality profile did not differ between people who completed the *error* condition and people who completed the *no error* condition ($U \geq 235$, $z \geq -0.388$, $p \geq 0.553$, $r \geq 0.03$).

3.1.2. Participants' Negative Attitude toward Robots

We measured people's negative attitude toward robots for two reasons. First, we wanted to assess our participants' general attitude. Therefore, we administered the NARS questionnaire before the participants' interaction with the robot. Second, we assumed that participants' attitude would be affected through the high number of errors. Therefore, we administered the questionnaire a second time, following the interaction. The individual NARS items range from 1—"I strongly disagree" to 5—"I strongly agree."[2] This means that low-scale values indicate that people have a more positive attitude toward robots and high-scale values denote a rather negative attitude.

3.1.2.1. Scale Reliability

We checked the reliability for all three subscales, before and after the interaction. The reliability for S1 before interaction resulted in borderline acceptable reliability (Cronbach's $\alpha = 0.64$), S1 after interaction in acceptable reliability ($\alpha = 0.74$). The reliability for S2 before interaction was too low ($\alpha = 0.51$). To increase reliability, we excluded item 2 (I feel that in the future society will be dominated by robots), and we recalculated the scale which resulted in borderline acceptable reliability ($\alpha = 62$). S2 after interaction was recalculated accordingly after excluding item 2 ($\alpha = 0.77$). S3 resulted in borderline acceptable reliability both before and after interaction (α before interaction = 0.62, after interaction = 0.67).

3.1.2.2. Participants' Overall Negative Attitude toward Robots

While our participants' rating for S2 and S3 resulted in a neutral standpoint, the rating for S1 showed that participants have a rather positive to neutral attitude toward interacting with robots (mean values before interaction are presented in **Table 3**).

3.1.2.3. Participants' Negative Attitude toward Robots Compared between Before and After Interaction

We were interested in investigating if our participants' negative attitude toward robots was influenced by their interaction with the robot. We conducted Wilcoxon rank-sum tests to evaluate if the ratings differed significantly before and after the interaction. The results showed that there was no significant difference in NARS ratings before and after the interaction with the robot (S1: $W = 248.00$, $z = -0.59$, $p = 0.558$, $r = -0.06$; S2: $W = 460.00$, $z = 1.66$, $p = 0.097$, $r = -0.18$; and S3: $W = 234.50$, $z = -1.81$, $p = 0.071$, $r = -0.19$). The mean values for the three scales before and after the participants' interaction with the robot are provided in **Table 3**.

3.1.2.4. Participants' Negative Attitude toward Robots Compared between Conditions

We explored if participants' rating after their interaction with the robot differed between the *error* and *no error* condition. We conducted Mann–Whitney-*U* tests for the scales completed after interaction. However, none of the scales resulted in significant differences between the conditions (S1: $U = 277.50$, $z = 0.85$, $p = 0.395$, $r = 0.13$; S2: $U = 324.50$, $z = 1.66$, $p = 0.098$, $r = 0.25$; and S3: $U = 277.00$, $z = 0.58$, $p = 0.564$, $r = 0.09$).

3.1.2.5. Participants' Negative Attitude toward Robots Compared between the Genders

We performed Mann–Whitney-*U* tests to assess if the NARS ratings differed between male and female participants. The ratings for S2 and S3 (both before and after interaction) did not differ significantly. However, both ratings for S1 differed significantly between men and women (S1 before interaction: $U = 419.50$, $z = 3.89$, $p = 0.000$, $r = 0.58$ and S1 after interaction: $U = 341.50$, $z = 2.41$, $p = 0.016$, $r = 0.36$). This result yielded in a large (before) and medium (after) effect size. For an overview on the mean

TABLE 3 | Mean values (SD) of the NARS questionnaire before and after the interaction (*error* and *no error* combined).

NARS scale	Before interaction	After interaction
S1: negative attitude toward situations of interaction with robots	Mean = 2.07 (SD = 0.59)	Mean = 2.09 (SD = 0.67)
S2: negative attitude toward social influence of robots	Mean = 2.94 (SD = 0.77)	Mean = 3.11 (SD = 0.89)
S3: negative attitude toward emotions in interaction with robots	Mean = 2.99 (SD = 0.87)	Mean = 2.79 (SD = 0.77)

[2] [15] recommend calculating the NARS scales by summing up the item values. Since the scales are constructed of a varying number of items, the scale scores are in that case not comparable at first sight (Scale 1 would range from 6-30, Scale 2 from 5-25, Scale 3 from 3-15). Therefore, we calculated the scale values by averaging the scale items. With this, the values of the three scales become comparable more quickly and they also correlate with the range of the individual items.

values refer to **Table 4**. Even though males and females rated their potential interaction with a robot as rather positive, male ratings are significantly more positive than those of the female participants.

3.1.3. Participants' Rating of the Robot

We measured how people rated the likability, anthropomorphism, and perceived intelligence of the robot after interacting with it. To do so, we used the three corresponding subscales of the Godspeed questionnaire, each of which consists of five semantic differential items. These items measure the participants' position between two poles. Therefore, the arithmetic mean of these items with no emphasis on either one of the poles is 3. The calculated likability score ranges from 1—"dislike" to 5—"like," anthropomorphism from 1—"fake" to 5—"natural," and perceived intelligence from 1—"incompetent" to 5—"competent".

3.1.3.1. Scale Reliability

The anthropomorphism and perceived intelligence scales resulted in acceptable reliability (Cronbach's $\alpha = 0.78$ and 0.79) and likability in high reliability ($\alpha = 0.83$).

3.1.3.2. Participants' Overall Rating of the Robot

Our participants rated the robot less anthropomorphic than the arithmetic mean (mean = 2.16, SD = 0.74), slightly more intelligent (mean = 3.28, SD = 0.69), and considerably more likeable (mean = 4.10, SD = 0.63).

3.1.3.3. Participants' Rating of the Robot Compared between Conditions

In order to explore if people who experienced erroneous robot behavior rated the robot differently from those participants who had interacted with a flawless robot, we conducted Mann–Whitney-*U* tests (see **Table 5**). While the mean ratings for anthropomorphism and perceived intelligence did not differ significantly between conditions, participants' rating of the robot's likability differed significantly between conditions. People who interacted with an erroneous robot liked the robot significantly more than people who interacted with a flawless robot. This difference yielded in a medium effect size.

TABLE 4 | NARS S1 mean values (SD) before and after interaction for male and female participants.

NARS S1	Males	Females
Before	Mean = 1.77, SD = 0.54	Mean = 2.46, SD = 0.42
After	Mean = 1.87, SD = 0.55	Mean = 2.35, SD = 0.73

3.1.3.4. Participants' Rating of the Robot Compared between the Genders

We conducted further Mann–Whitney-*U* tests to detect potential differences in robot ratings between the genders. The tests showed that none of the three scales resulted in different ratings for male and female participants (anthropomorphism: $U = 290.50$, $z = 0.93$, $p = 0.352$, $r = 0.14$; likability: $U = 317.50$, $z = 1.55$, $p = 0.121$, $r = 0.23$; perceived intelligence: $U = 323.00$, $z = 1.68$, $p = 0.094$, $r = 0.25$). We further checked if our participants' age, their preexperience with robots, and their technological affinity influenced how the robot was rated. None of these attributes resulted in significant differences.

Given our results, we can infer the following for our previously postulated hypotheses. Our participants liked the robot that made errors significantly more than the flawless robot which confirms our hypothesis 1. The hypotheses 2 and 3 have to be rejected since the robot committing errors did neither result in significantly higher anthropomorphism nor in significantly lower perceived intelligence ratings.

3.2. Qualitative Data

For the qualitative data analysis, we annotated the video recordings of the interview and LEGO sessions from the *error* condition. We hand coded the social signals the participants showed toward the robot, not toward the researcher, and which were objectively countable. Ambiguous events were discarded. For two of the participants, there was no video data due to technical problems from the recording equipment. The video data reported are based on the remaining 19 participants that completed the error condition. The data from the concluding interview were coded thematically in order to support our findings.

In this results section, we will report those findings from the qualitative data that are related to our research topic of robot errors.

3.2.1. Interview and LEGO Session

3.2.1.1. Interview Session

NAO began the interview with asking the participants to state their definition of a robot. The majority of people provided a very technical definition: 17 people used the word machine, 10 the word device, and 10 referred to a robot as some other technical object. While 2 people directly referred to NAO as being a robot ("NAO, you are a robot."), 4 participants used an "organic" noun (i.e., human, life form, and creature). However, they still used a technical adjective to further specify that noun (i.e., mechanical, artificial, electronic, and technical). Two participants provided unrelated answers.

TABLE 5 | Godspeed mean values (SD) compared between conditions.

Godspeed scale	Error	No error	Mann–Whitney-*U*
Anthropomorphism	Mean = 1.97, SD = 0.66	Mean = 2.33, SD = 0.78	$U = 182.00$, $z = -1.60$, $p = 0.109$, $r = 0.24$
Likability[a]	Mean = 4.30, SD = 0.49	Mean = 3.93, SD = 0.70	$U = 340.00$, $z = 2.02$, $p = 0.044$, $r = 0.30$
Perceived intelligence	Mean = 3.33, SD = 0.62	Mean = 3.23, SD = 0.76	$U = 267.50$, $z = 0.35$, $p = 0.723$, $r = 0.05$

[a] *Significant differences.*

We had the above question included in the robot's questionnaire to gather people's general standpoint on robots. Since most of the participants regarded a robot as a technical object, we assumed that they would want it to work reliably. In order to back our assumption up, the robot's next question targeted the three most prominent qualities people attribute with a robot. Again, many participants listed technical terms ($N = 24$; e.g., mechanical, electronic, and programmed). While 11 participants attributed a practical quality to robots (e.g., helpful, efficient, and diligent), 3 people said robots were intelligent, and 6 people pointed out that robots are controlled by humans (e.g., there is human intelligence in the background, not very intelligent, no free will). As regards performance, 3 people referred to robots as precise/reliable, 1 participant said that robots would do what they are meant to, given they are programmed correctly, and only one person said that robots often make errors. This confirms our previous assumption that people assume robots to perform error-free.

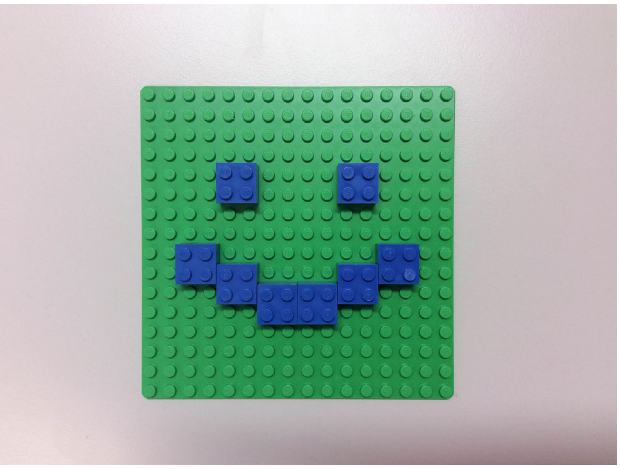


FIGURE 5 | **An example of how the participants showed their current emotion to NAO during the LEGO session**.

The questions reported above were asked at the beginning of the interview. In order to make the participant familiar with the situation, no errors were included in here, irregardless of the condition (for a complete description of the user study procedure refer to Section 2.2). Therefore, the answers were not influenced by the fact that the robot made or did not make mistakes. The following questions, however, contained robot errors in the *error* condition.

Upon asking the participants which skills they would want a robot to have, 8 participants referred to robots as error-free (e.g., should do what people tell it to do, work reliably, and make no mistakes). Other skills included that the robot should be helpful and take on work that is too difficult/tedious/dangerous for humans ($N = 13$), it should be communicative and understand the human ($N = 5$), it should be easy to handle ($N = 3$), and it should be witty ($N = 2$).

3.2.1.2. LEGO Session

The robot asked the participants to express their current emotional state with LEGO bricks. The emotional state declarations were classified through lip and/or eyebrow shape (for an example see **Figure 5**). Most of the emotional state declarations were closely modeled to emoticons that are widely used in social media. Depictions that could not clearly be matched to an emotion were excluded (no data entries in **Figure 6**). No apparent difference in participants' emotional state could be detected between the conditions. While the majority of participants was happy, only a few indicated a neutral expression. In the baseline condition, one participant reported a puzzled feeling and one felt silly. In the experimental condition, one participant indicated to be sad, one surprised. For an overview on all emotions refer to **Figure 6**.

In the *error* condition, the robot failed to grasp the LEGO board that the participants were supposed to hand over. Since the participants were instructed to tell the robot to grasp, we wanted to know how often participants were willing to repeat their instructions. The number of expressed instructions ("grasp!")

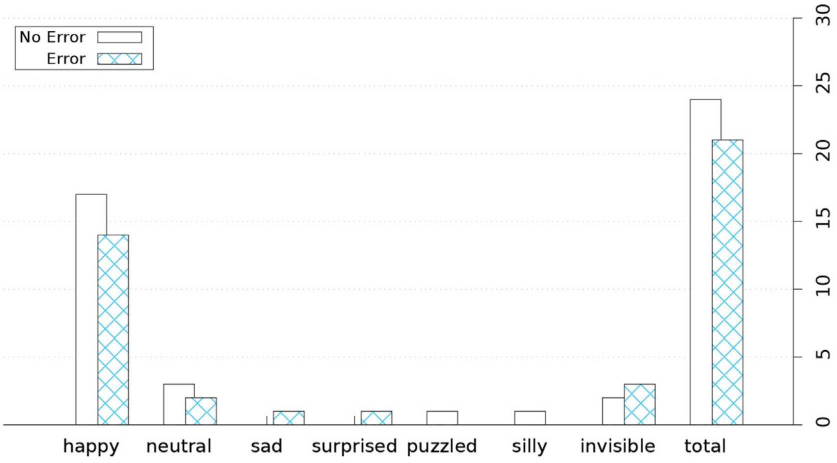


FIGURE 6 | **Emotions the participants expressed during the LEGO session**.

ranged from 2 to 7 (mean = 4.16, SD = 1.21). This result lets us assume that people are to some extent patient with a faulty robot.

Upon placing an unusual request in the *error* condition, the participants' willingness to comply was striking. A total of 17 participants threw LEGO blocks to the floor when asked to do so and 2 participants bent down and placed them on the floor, but no one refused to carry out the robot's request. The fact that the participants complied with the robot's unusual request links up with the research of Salem et al. (2015). The authors report that although people seemed to know that the robot's request was not right (the researchers made the robot ask a number of unusual things of the participants, such as throwing someone's personal mail in a garbage can), people complied as long as the action was not fatal and could be undone.

3.2.1.3. Social Signals

As we intended, the participants correctly interpreted the majority of social norm violations (SNV) and technical failures (TF) as error situations. The effectiveness manifests in the circumstance that most participants produced social signals when the robot made an error. Only the error where the robot waited for 15 s until it spoke was not recognized in 3 cases in the interview and in 7 cases in the LEGO session. The video footage showed that during the LEGO session, the participants were simply preoccupied with the previous task. This means that they were still dealing with the LEGO bricks (e.g., disassembling, counting, assembling, etc.) and, thus, did not pay attention to the robot's long silence. During the interview session, three participants were more patient than the rest of our sample and just waited for the robot to continue. The SNV in the interview session where the robot cut the participant off did not work in one case. This participant provided such a short but coherent answer that he was finished by the time the robot started speaking.

Each of the 19 participants experienced 8 error situations, which results in 152 error situations. From those, 11 were not recognized as error (see above) and in 19 cases, the participants did not show a reaction toward the robot. This leaves us with 122 error situations in which the participants showed 1 or more social signals (maximum 5). See **Table 6** for an overview on the mean number of social signals per error situation.

TABLE 6 | Mean number of social signals and standard deviation (SD) per error situation.

Error situation	Mean	SD
Interview—robot waits 15 s (SNV)	1.69	0.946
Interview—robot cuts participant off (SNV)	1.44	0.784
Interview—robot stops mid-word (TF)	0.95	0.911
Interview—speech loop (TF)	1.63	1.065
LEGO—throw block on the floor (SNV)	1.16	0.765
LEGO—robot waits 15 s (SNV)	1.00	0.953
LEGO—speech loop (TF)	2.00	1.106
LEGO—robot fails to grasp (TF)	2.63	1.26

The mean number of social signals expressed during a SNV is 1.36 (SD = 0.56) and during a TF 1.53 (SD = 0.72). A Kolmogorov–Smirnov test for normality over the differences of the variable scores indicated that the data are normally distributed ($D(19) = 0.131, p = 0.200$). We performed a paired-samples t-test and found that the amount of social signals the participants produced did not differ significantly between SNV and TF ($t(18) = -1.112, p = 0.281, d = 0.27$). **Table 7** gives an overview on how many social signals were made for each category in each type of error situation. The table also shows which kinds of social signals were grouped in the categories. Our analysis contains only social signals that were made toward the robot. Signals toward the present experimenters were not included in our analysis (e.g., verbal statements to the experimenter, head turns in the direction of the experimenter). We hand coded the data by counting the objectively perceivable events. Thereby, we distinguished a head tilt (head moves sideways with gaze staying in place) from a shift in gaze (the participant's gaze shifts visibly from, e.g., the robot to the LEGO parts). Head turns (head movements with the gaze leaving the scene) were all directed toward the present experiment and, thus, disregarded.

A Kolmogorov–Smirnov test for normality over the frequency differences of the variable scores for the speech category indicated that the data deviate from normal distribution ($D(19) = 0.250$, $p = 0.003$). Therefore, we performed Wilcoxon signed-rank tests to assess the differences in frequencies for each category. **Table 8** provides an overview on the mean number of social signal of each category per error situation type. The results show that during technical failures people made significantly more facial expressions, head movements, body movements, and gaze shifts.

3.2.2. Concluding Interview by the Researcher

After the participants finished interacting with the robot and after they completed the postinteraction questionnaires (NARS after interaction and Godspeed), they were asked four open-ended

TABLE 7 | Overview on social signal categories and frequencies per error type.

Category	Social signals	Frequencies in SNV	Frequencies in TF
Speech	Statements, questions	13	16
Smile/laughter	Smiles, laughs, giggle	29	30
Facial expressions	Frown, raised eyebrows, corners of the mouth lowered, eyes wide open	6	17
Head movements	Tilted head, nodding	5	12
Body movements	Lean forward, step back, touch face, adjust glasses, put hands on hip, put hands behind back, take hands out from pockets, raise arm and dance, sway, snap fingers, move LEGO parts around in front of the robot	8	19
Gaze shift	Shift gaze to or away from robot, wandering gaze	26	43
Total number of social signals		87	137

TABLE 8 | Social signals shown during social norm violations and technical failures.

Social signal	Social norm violation	Technical failure	Wilcoxon signed rank		
	Mean (SD)	Mean (SD)	*Z*	*p*-Value	*r*-Value
Speech	0.68 (0.820)	0.84 (0.958)	0.758	0.448	0.12
Smile/laughter	1.53 (1.219)	1.58 (0.902)	−0.074	0.941	−0.01
Facial expressions	0.32 (0.582)	0.89 (0.809)	−2.147	0.032	−0.35
Head movements	0.26 (0.562)	0.63 (1.165)	−2.121	0.034	−0.34
Body movements	0.42 (0.607)	1.00 (0.816)	−2.484	0.013	0.40
Gaze shift	1.37 (0.831)	2.26 (1.098)	−3.090	0.002	0.50

questions in the final interview. While the questions 1–3 asked about some general aspects of the participants' impression of the interaction and the robot, question 4 specifically targeted the robot's errors (see Section 2.5 for the specific questions). Therefore, question 4 was only asked for participants in the *error* condition. The resulting data were analyzed through an affinity diagram (Holtzblatt et al., 2004). An affinity diagram is a method for organizing ideas, challenges, and solutions into a wall-sized hierarchical diagram.

In question 1, participants were asked to report anything particular they had noticed during their interaction with the robot. Here, 12 participants reported that the robot had made some mistakes (e.g., *it went in a loop; it cut my word*). The participants' answers to question 2 did not include any mentions about the robot's mistakes. In question 3, 7 participants reported that they would like to change the faulty robot behavior (e.g., *fix the technical bugs; it does not leave time for you to respond; loops*).

With the final question in the interview, we specifically targeted the robot's errors, in asking what the participants thought of the robot making mistakes. While 7 participants uttered specifically negative aspects (e.g., *unpleasant; confusing; that's just what one would expect from technology; I was unsure if the interaction had stopped; I thought I had made a mistake*), 10 participants uttered positive feelings when asked about the fact that the robot made mistakes (e.g., *funny; friendly; it was great that the robot did not make it look like I made a mistake; I do not like it less because of the mistakes; it would be scary if all went smooth because that would be too human-like*).

4. DISCUSSION

Our results showed that the participants liked the faulty robot significantly more than the flawless one. This finding confirms the *Pratfall Effect*, which states that people's attractiveness increases when they make a mistake as shown by Aronson et al. (1966). Therefore, the psychological concept can successfully be transferred from interpersonal interaction to HRI. Upon the attempt of including socially acting robots into this concept, we can extend it to: "*Imperfections and mistakes carry the potential of increasing the likability of any social actor (human or robotic).*" The same effect was previously researched by Salem et al. (2013), where incongruent behavior of a robot can be seen as a social norm violation as such behavior violates participants' expectations from a *social script*. To overcome this error situation, participants changed their social signals, but on the other hand they rated the likability of the robot higher. Similarly, Ragni et al. (2016) showed that the participants in their study enjoyed the interaction with the faulty robot significantly more, than the participants who had interacted with a flawless robot. On the other hand, their participants who had interacted with the faulty robot, rated it less intelligent, less competent, and less superior, which again confirms the *Pratfall Effect*.

The repeated evidence of this phenomenon existing in HRI strengthens our argument to create robots that do not lead to believe they perform free from errors. We recommend that robot creators design social robots with their potential imperfections in mind. We see two sources for these imperfections that link back to the two error types found in HRI. On one hand, creators of social robots should follow the notions of interpersonal interaction to meet the expectations humans have about social actors and with it socially interacting robots. On the other hand, it is advisable to embrace the imperfections of robot technology. Technology that is created with potential shortcomings in mind can be designed to include methods for error recovery. Therefore, one way to go here would be to make robots understand they made an error by correctly interpreting the human's social signals and indicate their understanding to the human user. Both of these sources of imperfections will lead to more believable robot characters and more natural interaction. Of course, this applies to social robots operating in non-critical environments. Safety-relevant applications and scenarios must under all circumstances operate at zero-defect level.

Interestingly, we could not find a comparable effect for anthropomorphism in our data. The robot's anthropomorphism level was rated similar, irregardless of the fact if the robot made errors or not. Our result is different from the findings of Salem et al. (2013), who also used a human-like robot, and where the participants rated the faulty robot more anthropomorphic as the flawless one. The researchers used coverbal gestures, while we programmed the robot to provide mostly random gestures to make it appear more life-like. This might have in general diminished the effect of anthropomorphism in our setup (which is indicated by the low overall anthropomorphism level). However, more research is required to further explore the role of anthropomorphism in faulty robot behavior.

Contrary to our assumption, the faulty robot was not rated as less intelligent than the flawless one. This seems striking since the robot made several errors over a relatively short interaction time. Furthermore, most participants had noticed the robot making errors, while, at the same time, they had indicated to regard a robot as something very technical that should perform reliably. One potential explanation could be the fact that the induced errors were non-task-related. Follow-up research is required to further explore the perceived intelligence of erroneous robot behavior.

Upon asking the participants about their current emotional state, the majority of participants showed the robot that they were happy. The participants were also quite patient and tried handing the object several times, when the robot failed grasping it. All of these observations point toward the notion that a faulty social robot is a more natural social robot. In our future research on this topic, we will extend our approach to include more user experience measures to get a more profound understanding on the users' perception of the robot. For example, it will be interesting to further investigate possible impacts on subjective performance and acceptance.

Our data showed that when people interacted with a social robot that made an error, they were likely to show social signals in response to that error. In our previous research, we performed an analysis of video material in which robot errors occurred unintentionally and we found that users showed social signals in about half the interactions (Mirnig et al., 2015). In the herein reported study, however, most participants showed at least one social signal per error situation. We explain this difference in part with the high error rate (8 errors in an average total interaction time of about 12 min). Users seem to anticipate the robot making more errors once they experienced it is not flawless and responded more frequently with social signals. The reason for the increased number of social signals could also be based on the size of the robot. While the majority of interactions from the previous study were with a human-sized robot at eye level, the robot in our case was small and placed slightly below participants' eye level. This aspect remains to be studied further.

With our results we show again that humans respond to a robot's error with social signals. Therefore, recognizing social signals might help a robot to understand that an error happened. According to the frequencies of occurrence, gaze shifts and smile/laughter carry most potential for error detection, which is in line with our previous findings in the study of Giuliani et al. (2015). Upon a detailed analysis on the categories of social signals we found that people make significantly more gaze shifts during technical failures. This result is in contrast to our previous findings where significantly more gaze shifts were made during social norm violations. We take from this that gaze shifts are a potential indicator for robot errors, but it remains to be studied if they can be used to distinguish between the two error types.

We also found that people made significantly more facial expressions, head-, and body movements during technical failures. The increase in social signals during technical failures may be rooted in the circumstance that the technical failures were more obvious in the present user study. For example, in the video material from the previous study the robot failed to grasp an object that was placed in front of it. In our setup, the robot failed to grasp an object that the participant handed to it, which made the participant more actively perceive the robot's error.

Contrary to our previous findings, we did not detect significant differences in spoken social signals. This could be grounded in the fact that due to the setup, the robot had in general a much larger share in spoken utterances.

In response to the robot's unusual request, most users showed social signals. The kind of signals (gaze shifts and laughter) displayed the users' slight discomfort and provided evidence that they knew the robot's request implied a deviation from the social script of the situation. However, most users nevertheless followed the robot's order and threw the LEGO blocks to the floor. In addition to the previous results as reported in the study of Mirnig et al. (2015), this result provides further evidence that users show specific social signals in response to robot errors. Future research should be targeted at making a robot understand the signals and make sense of them. A robot that can understand its human interaction partner's social signals will be a better interaction partner itself and the overall user experience will improve.

Since most of our participants had not interacted with a robot before, a potential novelty denotes a certain limitation to our results. Some participants were probably captivated with the technology, which made them remain patient. It remains to be studied how such novelty wears off over time and how this influences people's willingness to interact. It will, furthermore, be interesting to assess the dimensions of faults. That is, how extensive can an error become until it becomes a deal-breaker. Ragni et al. (2016) already provided evidence that erroneous robot behavior decreases performance of a human interacting with the robot. It could also be interesting to explore how users react in case of the robot giving ambiguous information. Further aspects of robot errors that are worthwhile exploring are, for example, the following. What kinds of errors are forgivable and which ones are not? What is the threshold for error rate or number of errors until the participants' patience is over or performance drops considerably? A lot more specific research is required to understand and make use of the effects of errors in social HRI.

5. CONCLUSION

With our user study we explored how people rated a robot making errors in comparison to a perfectly performing robot. We measured the robot's likability, anthropomorphism, and perceived intelligence. We found that the faulty robot was rated as more likeable, but neither more anthropomorphic nor less intelligent. We recommend robots to be designed with their possible shortcomings in mind as we believe that this will result in more likeable social robots. Similar to interpersonal interaction, imperfections might even have a positive influence in terms of likability. We expect social HRI that embraces the imperfectness of today's robots to result in more natural interaction and more believable robot characters.

Our results confirm existing HRI research on robot likability such as the studies of Salem et al. (2013) and Ragni et al. (2016), hinting at error-prone robots supposedly resulting in more believable robots. Our work successfully proves the existence of the psychological concept *Pratfall Effect* in HRI and suggests that it should be our community's aim to bear potential shortcomings of social robots in mind when creating them. The nature and extent of errors that can be handled through the interactional design remain yet to be studied.

With our results we could again show that humans respond to faulty robot behavior with social signals. A robot that can recognize these social signals can, in subsequence, understand that an error happened. We detected gaze shifts and laughter/smiling as the most frequently shown social signals, which is in line with our previous research.

We see the following next steps to the ambitious goal of creating social robots that are able to overcome an error situation. First, it needs to be studied how we can let robots understand that an error occurred. Second, robots must be enabled to communicate about such errors. Third, robots need to know how to behave in an error situation in order to effectively apply error recovery strategies.

ETHICS STATEMENT

This study was carried out in accordance with the ethical regulations of conducting user studies at the University of Salzburg. The entire process was supervised, and the protocol was approved by the department director, Prof. Manfred Tscheligi. Each of our participants was given information about the study process beforehand, including the information that it was possible to quit participating at every point in time. Every participant gave their written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

NM is the main author of this article who provided the storyline and most of the text is written by her. She was the responsible supervisor of the user study and she performed the data analysis and statistics. GS contributed to the setup of the user study, he assisted with writing and the storyline, and he contributed to data analysis. MM recruited participants and performed the user study. SS provided input on the related work. She provided **Figure 3** and she helped with formatting the tables. MG provided related work and he contributed to the overall storyline. MT was supervising the user study and writing processes.

ACKNOWLEDGMENTS

The authors of this paper would like to thank Michael Miksch for his contribution in performing the user study.

FUNDING

We gratefully acknowledge the financial support by the Austrian Federal Ministry of Economy, Family and Youth and the National Foundation for Research, Technology and Development (Christian Doppler Laboratory for “Contextual Interfaces”). This work was additionally funded in part by the European Commission in the project ReMeDi (Grant No. 610902). We acknowledge financial support by the Open Access Publication Fund of the University of Salzburg.

REFERENCES

Aronson, E., Willerman, B., and Floyd, J. (1966). The effect of a pratfall on increasing interpersonal attractiveness. *Psychon. Sci.* 4, 227–228. doi:10.3758/BF03342263

Bartneck, C., Kulić, D., Croft, E., and Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int. J. Soc. Robot.* 1, 71–81. doi:10.1007/s12369-008-0001-3

Brooks, D. J., Begum, M., and Yanco, H. A. (2016). “Analysis of reactions towards failures and recovery strategies for autonomous robots,” in *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2016)* (New York, NY: IEEE), 487–492.

Bruckenberger, U., Weiss, A., Mirnig, N., Strasser, E., Stadler, S., and Tscheligi, M. (2013). “The good, the bad, the weird: audience evaluation of a “real” robot in relation to science fiction and mass media,” in *Proceedings of the International Conference on Social Robotics* (Bristol, UK: Springer), 301–310.

Gehle, R., Pitsch, K., Dankert, T., and Wrede, S. (2015). “Trouble-based group dynamics in real-world HRI—reactions on unexpected next moves of a museum guide robot,” in *Proceedings of the International Symposium on Robot and Human Interactive Communication* (Kobe: IEEE), 407–412.

Giuliani, M., Mirnig, N., Stollnberger, G., Stadler, S., Buchner, R., and Tscheligi, M. (2015). Systematic analysis of video data from different human-robot interaction studies: a categorisation of social signals during error situations. *Front. Psychol.* 6:931. doi:10.3389/fpsyg.2015.00931

Gompei, T., and Umemuro, H. (2015). “A robot’s slip of the tongue: effect of speech error on the familiarity of a humanoid robot,” in *Proceedings of the International Symposium on Robot and Human Interactive Communication* (Kobe: IEEE), 331–336.

Hayes, C. J., Maryam, M., and Riek, L. D. (2016). “Exploring implicit human responses to robot mistakes in a learning from demonstration task,” in *Proceedings of the International Symposium on Robot and Human Interactive Communication* (New York, NY: IEEE), 246–252.

Holtzblatt, K., Wendell, J. B., and Wood, S. (2004). *Rapid Contextual Design: A How-to Guide to Key Techniques for User-Centered Design*. San Francisco, CA: Elsevier.

John, O. P., Naumann, L. P., and Soto, C. J. (2008). “Paradigm shift to the integrative big-five trait taxonomy: history, measurement, and conceptual issues,” in *Handbook of Personality: Theory and Research*, eds O. P. John, R. W. Robins, and L. A. Pervin (New York, NY: Guilford Press), 114–158.

Knepper, R. A., Tellex, S., Li, A., Roy, N., and Rus, D. (2015). Recovering from failure by asking for help. *Auton. Robots* 39, 347–362. doi:10.1007/s10514-015-9460-1

Lee, M. K., Kielser, S., Forlizzi, J., Srinivasa, S., and Rybski, P. (2010). “Gracefully mitigating breakdowns in robotic services,” in *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction* (Osaka: IEEE Press), 203–210.

Lohse, M. (2011). The role of expectations and situations in human-robot interaction. *New Front. Hum. Robot Interact.* 2, 35–56. doi:10.1075/ais.2.04loh

Mirnig, N., Giuliani, M., Stollnberger, G., Stadler, S., Buchner, R., and Tscheligi, M. (2015). “Impact of robot actions on social signals and reaction times in HRI error situations,” in *Proceedings of the International Conference on Social Robotics* (Paris: Springer), 461–471.

Nomura, T., Kanda, T., Suzuki, T., and Kato, K. (2004). “Psychology in human-robot communication: an attempt through investigation of negative attitudes and anxiety toward robots,” in *Proceedings of the International Symposium on Robot and Human Interactive Communication* (Kurashiki: IEEE), 35–40.

Ragni, M., Rudenko, A., Kuhnert, B., and Arras, K. O. (2016). “Errare humanum est: erroneous robots in human-robot interaction,” in *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2016)* (New York, NY: IEEE), 501–506.

Robinette, P., Wagner, A. R., and Howard, A. M. (2014). “Assessment of robot guidance modalities conveying instructions to humans in emergency situations,” in *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on*, (Edinburgh, UK: IEEE), 1043–1049.

Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., and Joublin, F. (2013). To err is human (-like): effects of robot gesture on perceived anthropomorphism and likability. *Int. J. Soc. Robot.* 5, 313–323. doi:10.1007/s12369-013-0196-9

Salem, M., Lakatos, G., Amirabdollahian, F., and Dautenhahn, K. (2015). “Would you trust a (faulty) robot? Effects of error, task type and personality on human-robot cooperation and trust,” in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (Portland, OR: ACM), 141–148.

Schank, R., and Abelson, R. (1977). *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*, Vol. 2. Hillsdale, NJ: Lawrence Erlbaum Associates.

Short, E., Hart, J., Vu, M., and Scassellati, B. (2010). “No fair!!: an interaction with a cheating robot,” in *Proceedings of the 5th ACM/IEEE International Conference on Human-robot Interaction, HRI’10* (Piscataway, NJ: IEEE Press), 219–226.

Vinciarelli, A., Pantic, M., and Bourlard, H. (2009). Social signal processing: survey of an emerging domain. *Image Vis. Comput.* 27, 1743–1759. doi:10.1016/j.imavis.2008.11.007

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Mirnig, Stollnberger, Miksch, Stadler, Giuliani and Tscheligi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

ORIGINAL RESEARCH
published: 23 May 2017
doi: 10.3389/frobt.2017.00016

Personality Perception of Robot Avatar Teleoperators in Solo and Dyadic Tasks

Paul Adam Bremner[1*†], *Oya Celiktutan*[2†] *and Hatice Gunes*[2]

[1] Bristol Robotics Laboratory, University of West England, Bristol, UK, [2] Computer Laboratory, University of Cambridge, Cambridge, UK

Humanoid robot avatars are a potential new telecommunication tool, whereby a user is remotely represented by a robot that replicates their arm, head, and possible face movements. They have been shown to have a number of benefits over more traditional media such as phones or video calls. However, using a teleoperated humanoid as a communication medium inherently changes the appearance of the operator, and appearance-based stereotypes are used in interpersonal judgments (whether consciously or unconsciously). One such judgment that plays a key role in how people interact is personality. Hence, we have been motivated to investigate if and how using a robot avatar alters the perceived personality of teleoperators. To do so, we carried out two studies where participants performed 3 communication tasks, solo in study one and dyadic in study two, and were recorded on video both with and without robot mediation. Judges recruited using online crowdsourcing services then made personality judgments of the participants in the video clips. We observed that judges were able to make internally consistent trait judgments in both communication conditions. However, judge agreement was affected by robot mediation, although which traits were affected was highly task dependent. Our most important finding was that in dyadic tasks personality trait perception was shifted to incorporate cues relating to the robot's appearance when it was used to communicate. Our findings have important implications for telepresence robot design and personality expression in autonomous robots.

Keywords: telepresence, Big Five personality traits, personality perception

OPEN ACCESS

Edited by:
Giuseppe Carbone,
University of Cassino, Italy

Reviewed by:
Fulvio Mastrogiovanni,
University of Genoa, Italy
Paolo Boscariol,
University of Udine, Italy

****Correspondence:***
Paul Adam Bremner
paul.bremner@brl.ac.uk

[†] These authors have contributed equally to this work.

Specialty section:
This article was submitted to Humanoid Robotics, a section of the journal Frontiers in Robotics and AI

Received: *20 December 2016*
Accepted: *27 April 2017*
Published: *23 May 2017*

Citation:
Bremner PA, Celiktutan O and Gunes H (2017) Personality Perception of Robot Avatar Teleoperators in Solo and Dyadic Tasks.
Front. Robot. AI 4:16.
doi: 10.3389/frobt.2017.00016

1. INTRODUCTION

Telecommunication is omnipresent in today's society, with people desiring to be able to communicate with one another, regardless of distance, for a variety of social and practical reasons. While video-enabled communication offers a number of benefits over voice-only communication, it is still lacking compared to face-to-face interactions (Daly-Jones et al., 1998). For example, remotely located team members are less included in cooperative activities than colocated team members (Daly-Jones et al., 1998) and have fewer conversational turns and speaking time in group conversations (O'Conaill et al., 1993). Suggested reasons for these disparities are a lack of social presence of these remote group members, reduced engagement, and reduced awareness of actions (Tang et al., 2004). A suggested underlying cause for the disparities found in traditional telecommunication is a lack of physical presence. An alternative is the use of teleoperated robots as communication media. A common approach to such embodied telecommunication is the use of mobile remote presence (MRP) devices: a screen displaying the operators face mounted on a stalk attached to a wheeled base (Kristoffersson et al., 2013). Though studies examining the utility of MRPs have found that there are some improvements in social presence, different social norms are observed when people use them to interact, and there are impacts on trust and rapport (Lee and Takayama, 2011; Rae et al., 2013). Further, such systems are not able to effectively transmit non-verbal communication cues, a key element of human communication not only for information conveyance but also in maintaining engagement and building rapport (Salam et al., 2016).

A proposed method for further improving social presence and effectively transmitting body language is to use a humanoid robot as a communication medium. In such a system, the operator's body language is duplicated on a humanoid robot such that it is comprehensible and highly salient (Bremner and Leonards, 2016; Bremner et al., 2016b). Using a humanoid robot as a communications avatar has benefits with regard to engagement of conversational partners (Hossen Mamode et al., 2013), social presence (Adalgeirsson and Breazeal, 2010), group interaction (Hossen Mamode et al., 2013), and trust (Bevan and Stanton Fraser, 2015).

However, when using a robot as a remote proxy for communication, the operator is represented with a different physical appearance, much as computer generated avatars do in virtual environments. Appearance has been observed to be utilized in making interpersonal judgments (Naumann et al., 2009), and this can extend to virtual avatars (Wang et al., 2013; Fong and Mar, 2015). It was observed that judges made relatively consistent inferences based on avatar appearance alone (Wang et al., 2013; Fong and Mar, 2015), and more attractive avatars were rated more highly in an interview scenario (Behrend et al., 2012). How this might manifest with robot avatars, in particular in the interaction between a robot appearance and human voice communication, remains unclear and is yet to be explored.

Here, the particular judgment we are concerned with is that of personality perception, an important facet of communication. Researchers in psychology have shown that personality plays a key role in forming interpersonal relationships, and predicting future behaviors (Borkenau et al., 2004). These findings have motivated a significant body of work for how people judge others' personalities based on their observable behaviors. A key component of these social cues for personality are non-verbal behaviors. We aim to investigate if such non-verbal personality cues transmitted by a teleoperated humanoid robot continue to be utilized in personality judgments, and how they interact with verbal cues. Non-verbal cues can be transmitted as our robot teleoperation system utilizes a motion capture-based approach so that arm and head movements the operator performs while talking are recreated with minimal delay on a NAO humanoid robot (Bremner and Leonards, 2016). The control system is intuitive and immersive, and we observe people behaving similarly to how they do face-to-face (Bremner et al., 2016b).

We designed two experiments which follow an experimental methodology common in the personality analysis literature, i.e., videos of participants performing different communication tasks are shown to external observers (judges) for personality assessment (e.g., Borkenau et al. (2004)). Personality judgments are made on the so-called big five traits, *extroversion*, *conscientiousness*, *agreeableness*, *neuroticism*, and *openness* (multiple questions relate to each trait). We varied communication media between judges, either video only or robot mediated (also recorded on video). Two main measures are used to see whether there was an effect of communication condition on personality judgments: (1) judge consistency in how they evaluate a given trait, both within and between judge (low consistency indicates lack of cues or conflicting cues); and (2) personality shifts between high and low classification for each trait between the video and robot conditions.

Hence we address the following research questions:

- **RQ1:** Are there differences in judges' consistency in assessing personality traits (within-judge consistency)?
- **RQ2:** Are there differences in how much judges agree with one another on personality judgments (between-judge consistency)?
- **RQ3:** Are personality judgments less accurate compared to self-ratings (self-other agreement)?
- **RQ4:** Are perceived personalities systematically shifted to incorporate characteristics associated with the robot's appearance (personality shifts)?

This paper is an extended version of our work published by Bremner et al. (2016a). We extended our previous work by adding a second experiment that refined our experimental procedure and used dyadic rather than solo tasks. Our discussions and conclusions are extended to include both experiments, evaluating all our results to give a clearer picture.

In the first experiment, three tasks are performed direct to camera, i.e., solo tasks. In the second experiment, participants performed three tasks that involved interaction with a confederate, i.e., dyadic. The first experiment provided some limited evidence for shifts in personality perception. Further, by adding an audio-only communication condition, we were able to show that the robot was not simply ignored, and gesture cues performed on the robot were utilized. An important finding from the first experiment was that effects were very task dependent, as the literature suggested. Borkenau et al. (2004) found that *openness* is better inferred in more ability-demanding tasks such as pantomime task. Hence, the second experiment used additional tasks, which by being dyadic will engender personality cues differently; it is also a refinement of our experimental procedure, improving the reliability of our results. It produced compelling evidence that cues related to the robot's appearance were incorporated in personality judgments, causing consistent shifts in perceived personality.

2. RELATED WORK

A common approach to investigating personality judgments is first impression or thin slice personality analysis. It is a body of research that studies the accuracy with which people are able to make personality judgments of others based only on short behavioral episodes (termed thin slices). This approach is taken as it is believed that these judgments provide insight into the assessments people make in everyday interactions (Funder and Sneed, 1993; Borkenau et al., 2004). In such studies, targets are typically asked to perform a range of communication tasks, either solo performances to camera or dyadic with confederates, and are filmed while doing so. *Judges* then observe the video clips and complete personality assessment questionnaires. Ratings of judges are compared with target self-ratings, acquaintance ratings, and for inter-judge agreement. For many traits, there is sufficient inter-judge agreement for the method to be useful in assessing the impressions a person creates on those they interact with (Borkenau et al., 2004); however, the accuracy of judge ratings to self/acquaintance ratings is typically a lot lower, as self/acquaintance ratings are error prone, and use different sources to make their judgments (Vinciarelli and Mohammadi, 2014).

Often analyzed in thin slice personality studies are the cues that appear to be utilized in people making their judgments. Appearance, speaking style, gaze, head movements, and hand gestures have been frequently reported to be significant predictors of personality (Riggio and Friedman, 1986; Borkenau and Liebler, 1992; Borkenau et al., 2004). Indeed, this sort of analysis forms the basis for automated personality analysis systems. Aran and Gatica-Perez (2013) focused on personality perception in a small group meeting scenario. They extracted a set of multimodal features including speaking turn, pitch, energy, head and body activity, and social attention features. Thin slice analysis yielded the highest accuracy for *extroversion*, while *openness* was better modeled by longer time scales. With regard to the related work in personality computing, the closest approach was presented in the study by Batrinca et al. (2016). In order to analyze the Big Five personality traits, Batrinca et al. conducted a study where a set of participants were asked to interact with a computer, which was controlled by an experimenter, and then a different set of participants were asked to interact with the experimenter face-to-face to collaborate on completing a map task. In order to elicit the participants' personality traits, the experimenter exhibited four different levels of collaborative behaviors from fully collaborative to fully non-collaborative. Self-reported personality traits were used to study the manifestation of traits from audiovisual cues. In the human-machine interaction setting, their results showed that (1) extroversion and neuroticism can be predicted with a high level of accuracy, regardless of the collaboration modality; (2) prediction of the agreeableness and conscientiousness traits depends on the collaboration modality; and (3) openness was the only trait that cannot be modeled. In contrast to their findings in the human–machine interaction setting, they showed that openness was the trait that can be predicted with highest accuracy in the human–human interaction setting.

Applying such personality perception analysis to robot teleoperators has so far been limited. Perception of teleoperator's personality is important not only in social interactions but is also crucial where teleoperated robots are used in a service capacity such as for elderly care (Yamazaki et al., 2012), and search and rescue (Martins and Ventura, 2009). In these settings, perception of the operator will effect system utility for carrying out the desired service and achieving the desired outcome. In the study by Celiktutan et al. (2016), we showed that many of the aforementioned personality cues can be transmitted by a telepresence robot. We trained support vector machine classifiers with a set of features extracted from participants' voice and body movements. We found that the use of a robot avatar helps to discriminate between different personality types (e.g., extroverted vs.introverted) better than audio-only mediated communication for extroversion (65%) and conscientiousness (60%).

Studies with Mobile Remote Presence devices (MRPs) have briefly mentioned perceiving the operator's personality (Lee and Takayama, 2011), but it has not been deliberately studied as we do here. There are two studies that look directly at personality perception of teleoperators. Kuwamura et al. (2012) examined an effect that they term *personality distortion*, demonstrated by reduction in internal consistency of the personality questionnaire they used, for two different robot platforms and communication using video. They use 3 tasks: (1) an experimenter talks freely with the participant, (2) a different experimenter introduces and talks about themselves, and (3) a third experimenter interviews the participant. They only observed *personality distortion* for one of the robot platforms, for *extroversion* in the interview task, and for *agreeableness* in the introduction task. Using a single fixed person for each task, particularly members of the experimental team who are aware of the goals of the study, greatly reduces the ecological validity of their results. In contrast, here we use a large number of naïve targets performing naturalistic communication, and conduct far more in-depth data analysis.

In a study with a teleoperated, highly humanlike robot, Straub et al. (2010) examined both how participant teleoperators incorporate the fact that they are operating a robot into their presented identity, and how interlocutors at the robot's location blend operator and robot identities. They used language analysis to make their assessments. They observed that many operators pretended they themselves were a robot, and interlocutors often referred to the operator as a robot. These behaviors are different from what we typically observe with our teleoperation system, where most operators appeared to act naturally as themselves (Bremner et al., 2016b).

3. MATERIALS AND METHODS

We designed a two-stage experimental method for assessing changes in perceived personality that we used in two studies. First, a set of participants (targets) were recorded performing three communication tasks in two conditions, directly visible on video camera (audiovisual condition) and communicating using the teleoperated robot (teleoperated robot condition, also recorded on camera). This ensures that we have a large set of natural communication behaviors, and hence personality cues, for a range of personality types, that can be viewed directly or when mediated by a robot.

In the second stage of the study, the recorded data were used to create a set of video clips for each target in each communication condition. The video clips were pseudorandomly assigned to a set of surveys in such a way as to have one of each task and communication condition combinations present, with a given target only appearing once in a given survey (i.e., communication condition was varied between surveys). Each survey was viewed by a set of 10 judges, who after watching each clip assessed the personality of that target. We used an online crowdsourcing service to have the clips assessed. Employing judges *via* online crowdsourcing services has recently gained popularity due to its efficiency and practicality as it enables collecting responses from a large group of people within a short period of time (Biel and Gatica-Perez, 2013; Salam et al., 2016).

Personality was assessed by a questionnaire that aims to gather an assessment along the widely known Big Five personality traits (Vinciarelli and Mohammadi, 2014). These five personality traits are *extroversion* (EX—assertive, outgoing, energetic, friendly, socially active), *agreeableness* (AG—cooperative, compliant, trustworthy), *conscientiousness* (CO—self-disciplined, organized, reliable, consistent), *neuroticism* (NE—having tendency to negative emotions such as anxiety, depression, or anger), and *openness* (OP—having tendency to changing experience, adventure, new ideas). Each trait is measured using a set of items (the BFI-10 (Rammstedt and John, 2007) with 2 per trait in the Solo Tasks Study, and the IPIP-BFM-20 (Topolewska et al., 2014) with 4 per trait in the Dyadic Tasks Study) scored on 10-point Likert scales. As well as being assessed by external observers, each target completed the personality questionnaire for self-assessment.

3.1. Teleoperation System

In order to reproduce the gestures of targets on the NAO humanoid robot platform from Softbank Robotics (Gouaillier et al., 2009), we used a motion capture-based teleoperation system. Previously we have demonstrated the system to be capable of producing comprehensible gestures (Bremner and Leonards, 2015, 2016). The arm motion of the targets is recorded using a Microsoft Kinect and Polhemus Patriot,[1] and used to produce equivalent motion on the robot. Arm link end points at the wrist, elbow, and shoulder are tracked and were used to calculate joint angles for the robot so that its upper and lower arm links reproduce human arm link positions and motion. This method ensures that joint coordination, and hand trajectories are as similar as possible between the human and the robot within the constraints of the NAO robot platform. **Figure 1** shows a gesture produced by one of the targets, and the equivalent gesture on the NAO.

3.2. Solo Tasks Study

3.2.1. Tasks

In the first study, the three tasks performed by participants involved them performing directly to the camera, i.e., solo, and were based upon a subset of tasks used by Borkenau et al. (2004). Each of the tasks was framed as an interaction with the experimenter who stood beside the video camera used in the recordings, and provided non-verbal feedback and prompt questions to ensure as natural communicative behaviors as possible. Targets were instructed to speak for as long as they felt able, with a maximum time of 2 min for each task. The majority of the targets talked for 30–60 s on each task, with occasional prompts for missing information. Prior to performing tasks, we asked the targets to introduce themselves and give some information about themselves, e.g., where they work, what they do, their family, etc. This stage was purely to help naturalize the target to the experimental setting. It was not used to produce clips for judge rating.

3.2.1.1. Task 1 (Hobby)

This task asked targets to describe one of their hobbies, providing as much detail as possible. Suggested detail included what their hobby involves, why they like it, how long have they been doing it for, etc. Example personality cues we anticipated from this task include what targets have as their hobby, and what detail and the depth of detail they provide while describing their hobby.

[1] Product of http://polhemus.com/.

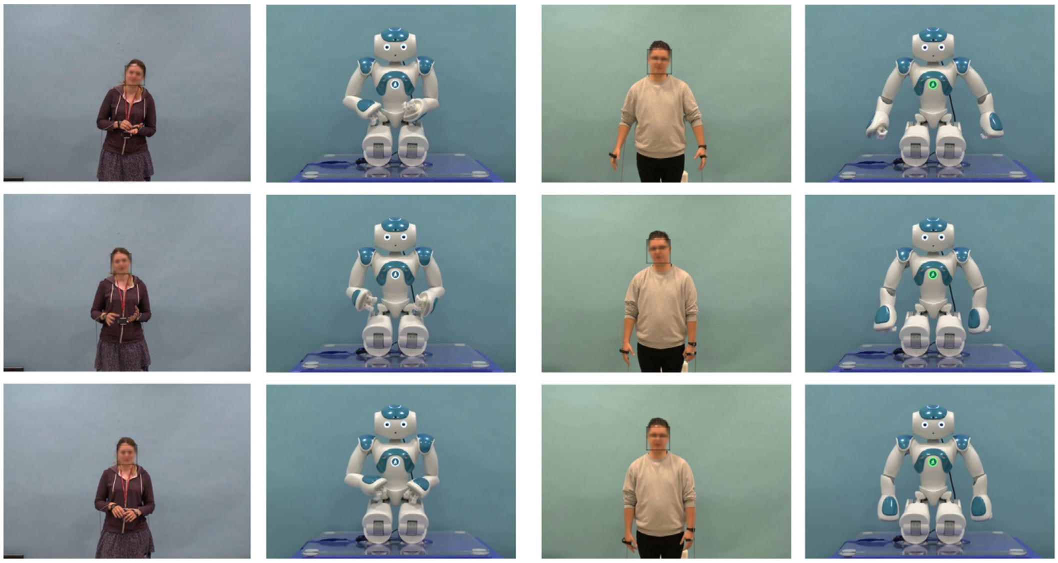


FIGURE 1 | Snapshots from the Solo Tasks Study. Left hand side: a target perceived to be *extroverted* by judges. Right hand side: a target perceived to be *introverted* by judges.

3.2.1.2. Task 2 (Story)

This task is based on Murray's thematic apperception test (TAT), where the target is shown a picture and is asked to tell a dramatic story based on a picture (Murray, 1943). They are asked what is happening in the picture,[2] what are the characters thinking and feeling, what happens before the events in the picture and what happens after. The picture is purposely designed to be ambiguous so that the target has the scope to interpret the picture as they see fit, and has to be creative in their story telling. It is a projective test, where the details given by the target, and how they relate the actions of the characters, provide cues about their personality.

3.2.1.3. Task 3 (Mime)

This task required the targets to mime preparing and cooking a meal of their choice. This was different from the mime task used by Borkenau et al. (2004), where targets had to mime alternative uses for a brick. Our pretests showed little variability between targets for that task. Instead, the chosen task gave the desired variability, and the gestures were better suited to performance on the NAO robot. Which meal was selected, and the complexity of the mime, are example personality cues we anticipated from this task.

3.2.2. Participants

Twenty-six participants were recorded as targets (16 female, mean age = 30.85, SD = 7.58) and gave written informed consent for their participation, they were reimbursed with a £5 gift voucher for their time. Recordings for 20 of the targets were used to create the clips used for judgments (6 targets were omitted due to recording problems). The study was approved by the ethics committee of the Faculty of Environment and Technology of The University of the West of England.

Clip ratings were undertaken by 143 judges recruited through the CrowdFlower online crowdsourcing platform.[3] Judges were compensated 50 cents for annotating a total of four clips.

3.2.3. Recordings

All tasks were recorded by one RGB video camera and the motion capture system used for teleoperation. The recorded motion capture data were then used to produce robot-mediated versions of the targets' performances on the NAO robot using the aforementioned teleoperation system, which were also recorded on video.

In addition to the audiovisual and teleoperated robot conditions, an audio-only condition was created using the audio from hobby and story tasks. Hence, each target had a total of 8 clips split over 3 communication conditions: 3 clips for the audiovisual condition, 2 clips for the audio-only condition, and 3 clips for the teleoperated robot condition. This resulted in a total of 158 clips (two clips became corrupted).

To avoid confusion, prompt questions were edited out of the clips. Further, for the few tasks where performance exceeded 60 s, clips were edited to be close to this length as pretests showed a decrease in the reliability of judgments with overly long clips. Mean clip duration was 50 s (SD = 20 s).

The clips were split up into surveys each containing four clips: one of each task and one of the audio-only clips, each of a unique target. Communication condition was pseudo-randomized across the three tasks in each survey, but always contained at least one of each communication condition.

3.3. Dyadic Tasks Study

3.3.1. The Extended Teleoperation System

The teleoperation system was extended to enable interactive multimodal communication. The first addition made was a stereo camera helmet on the NAO robot, the images from which are displayed in an Oculus Rift head-mounted display (HMD). Coupled with using the Rift's inertial measurement unit to drive the robot's head, meant the operator could see from the robots point of view, and their gaze direction and head motion could be observed on the robot. Secondly we used a voice over IP communication system to allow full duplex audio communication. Finally, due to feedback from participants in the Solo Tasks Study, we did not use the Polhemus Patriot in the Dyadic Tasks Study to make behaviors more natural; importantly, wrist rotation was only really needed for the mime task in the Solo Tasks Study, and is less important for normal gesturing. **Figure 2** shows the teleoperation system and the setup during performance of dyadic tasks in the teleoperation (TO) condition.

3.3.2. Tasks

In the second study, the three tasks performed by participants involved interacting with a confederate, i.e., dyadic. A confederate was used to ensure that each participant had the same interactive partner, giving us a measure of control over the interactions, while still seeming natural to the participants. The three selected tasks were based on the suggestions by Funder et al. (2000) of having an informative task, a competitive task, and a cooperative task. The intention of these task types is that they each engender personality cues in different ways.

The three tasks were briefly explained to the participant and the confederate together, and more detailed written instructions were provided to be used during the experimental session. This was done to ensure that the experimenters could leave the room for the participant and confederate to converse alone. The two communication conditions (audiovisual and teleoperated robot) were performed sequentially, in a pseudorandomized order, in the same room. The audiovisual condition was recorded face-to-face, i.e., with both participant and confederate seated across a table from one another. In the teleoperated robot condition, the participant moved to an adjoining room where the teleoperation controls were located, while the confederate sat at a table across from the robot.

3.3.2.1. Task 1 (Informative)

Participants watched a clip from a Sylvester and Tweety cartoon, which they then had to describe to the confederate. This is a task commonly used to examine gesturing (Alibali, 2001), as describing the action filled cartoon often engenders gestures, which may be useful personality cues that can be produced by the robot. Another key reason for this task choice was that all

[2] Image used was https://www.flickr.com/photos/bassclarinetist/, used under creative commons licence.

[3] CrowdFlower, a data enrichment, data mining and crowdsourcing company, http://www.crowdflower.com/.

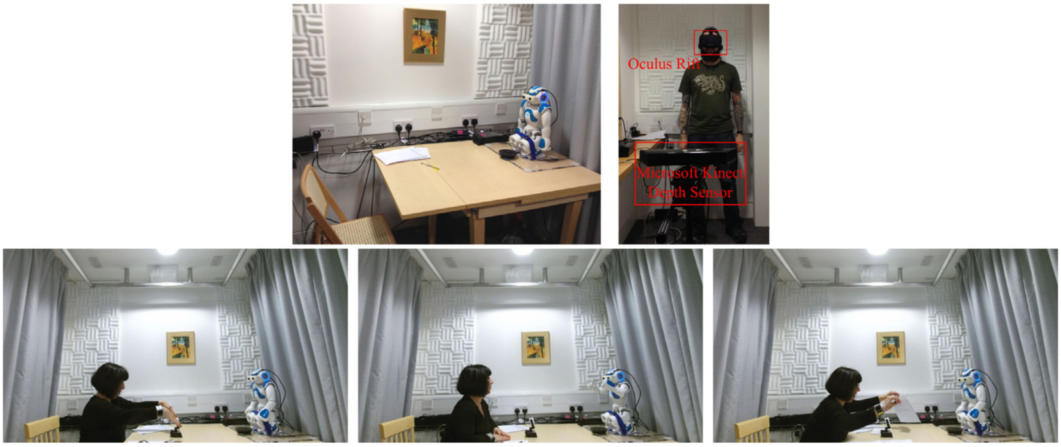


FIGURE 2 | Snapshots from the Dyadic Tasks Study. Upper row: illustration of teleoperation (TO) room and interaction room. Lower row: snapshots from the dyadic interaction sequences.

participants have the same things to talk about: in the previously used hobby task several participants struggled to find much to say without significant prompting. Two different Sylvester and Tweety cartoons were used, one for each communication condition; cartoon assignment was randomized between conditions. We expected there to be an abundance of gestural cues, as well as cues related to the participants' verbal behavior (such as how detailed the description was).

3.3.2.2. Task 2 (Competitive)

The participants and the confederate played a memory-based word game adapted from the traditional *Grandmothers Trunk* game. The first player says "My Grandmother went on holiday and she…" and adds something she did, accompanied by a gesture, the other player then repeats what the first said and their gesture, and adds something else she did. Play continues alternating between players who repeat the whole list of things and perform the gestures, adding a new thing each time, until one player forgets something and that player loses. How they approach the competitive nature of the task, and the actions they select are personality cues we expected from this task.

3.3.2.3. Task 3 (Cooperative)

The participants and the confederate cooperated to put a set of 5 items into utility order for surviving in a given scenario. There were two scenarios each with its own set of items, surviving a ship wreck, and surviving a crash landing on the moon. One scenario was presented per communication condition and was randomly assigned. How agreement is reached, and how the task is approached are the main cues we expect from this task.

3.3.3. Participants

Thirty participants were recorded as targets (13 female, mean age = 25.01, SD = 4.2), and gave written informed consent for their participation, they were reimbursed with a £5 gift voucher for their time. Recordings for 25 of the targets were used to create the clips used for judgments (5 targets were omitted due to recording problems). The study was approved by the ethics committee of the University of Cambridge.

Clip ratings were undertaken by 250 judges recruited through the Prolific Academic online crowdsourcing platform.[4] Each judge rated 6 clips and was compensated £2 for their time.

3.3.4. Recordings

In all tasks, both the confederate and the participant were recorded by separate RGB video cameras. The confederate was only recorded to obscure the fact that she was a confederate. In the teleoperated robot condition, a video camera recorded the robot instead of the participant. In order to produce videos of identical length for all targets and tasks, the video clips were further edited to select a 60 s segment from the beginning of the Informative task and from the end of Competitive and Cooperative tasks. This is in line with suggestions by Carney et al. (2007b) for using clips of this length of a task to maximize consistent judgment conditions for each target. Thus, each target had a set of three 60 s clips for each of the two communication conditions. One survey consisted of a pseudo-randomized set of 6 clips, 1 example of each task in each communication condition, with unique targets in each clip. Additionally a practice clip of the confederate was added to the start of all surveys to use as a measure of judge reliability, it also served to demonstrate her voice such that it could be ignored when she spoke during the target clips.

In **Table 1**, we summarized both studies in terms of number of participants, tasks, communication conditions, and communicated cues.

4. RESULTS AND ANALYSIS

To address the research questions introduced in Section 1, we analyzed the level of agreement and the extent of shifts with respect to different communication conditions (e.g., audiovisual/

[4] Prolific Academic online crowd sourcing platform, https://www.prolific.ac/.

TABLE 1 | Summary of the conducted studies.

Study	Number of participants	Tasks	Communication conditions	Communicated cues
Solo	26	Hobby, story, mime	AO, AV, TO	Wrist, elbow, shoulder motion, wrist orientation
Dyadic	30	Informative, competitive, cooperative	AV, TO	Wrist, elbow, shoulder motion; head motion; gaze direction

AO, audio-only; AV, audiovisual; TO, teleoperation.

AV, audio-only/AO, teleoperation/TO) and different tasks for each personality trait. We evaluated personality judgments to measure intra-/inter-agreement, self-other agreement, and personality shifts as below.

- *Intra-judge Agreement:* Intra-judge agreement (also known as internal consistency) evaluates the quality of personality judgments based on correlations between different questionnaire items that contribute to measuring the same personality trait by each judge. We measured intra-judge agreement in terms of standardized Cronbach's α: $\alpha = \frac{K\bar{r}}{(1+(K-1)\bar{r})}$ where K is the number of the items ($K = 2$ in the Solo Tasks Study, and $K = 4$ in the Dyadic Tasks Study) and $\bar{r}$ is the mean of pairwise correlations between values assigned. The resulting α coefficient ranges from 0 to 1; higher values are associated with higher internal consistency and values less than 0.5 are usually unacceptable (McKeown et al., 2012).
- *Inter-judge Agreement:* Inter-judge agreement refers to the level of consensus among judges. We computed the inter-judge agreement in terms of intraclass correlation (ICC) (Shrout and Fleiss, 1979). ICC assesses the reliability of the judges by comparing the variability of different ratings of the same target to the total variation across all ratings and all targets. We used ICC(1,k) as in our experiments each target subject was rated by a different set of k judges, randomly sampled from a larger population of judges. ICC(1,k) measures the degree of agreement for ratings that are averages of k independent ratings on the target subjects.
- *Self-other Agreement:* Self-other agreement measures the similarity between the personality judgments made by self and others. We computed self-other agreement in terms of Pearson correlation and tested the significance of correlations using Student's t distribution. Pearson correlation was computed between the target's self-reported responses and the mean of the others' scores per trait.
- *Personality Shifts:* Personality shift refers to the extent to which people shifted from one personality class to another, in judges' perception, between AV and TO conditions. In order to measure shifts, we first classified each target into low or high (e.g., *introverted* or *extroverted*) for each trait according to if their average judge rating for each task was above or below the mean for all targets in AV. For each trait, each target was grouped according to their classification in both conditions, creating 4 groups (i.e., AV: high and TO: high, AV: high and TO: low, etc.). We presented these results in terms of contingency tables and tested the significance using McNemar's test with Edwards's correction (Edwards, 1948).

In the following subsections, we present these results for each study (solo and dyadic) separately.

4.1. Solo Tasks Study

4.1.1. Elimination of Low-Quality Judges

Although crowdsourcing techniques have many advantages, identifying annotators who assign labels without looking at the content (low-quality judges or spammers) is necessary to get informative results. As a first measure, we eliminated judges who incorrectly answered a test question about the content of the clips. After this elimination mean-judges-per-clip was 7.9 (SD = 1.5), with minimum judges-per-clip being 5.

To assess whether there remained further low-quality judges we calculated within-judge consistency for the AV clips using Cronbach's α, which measures whether the values assigned to the items that contribute to the same trait are correlated. The average value across all tasks was lower than we expected (less than 0.5), indicating some judges answer randomly. With no low-quality judges, we would expect values for the AV clips greater than 0.5, i.e., in line with values reported in the literature for the BFI-10 with video clips assessed by online judges (Credé et al., 2012). We therefore used a judge selection method to remove these additional low-quality judges. We used a ranking-based method based on pairwise correlations instead of standard methods for outlier detection. For each clip, we calculated an average correlation score for each judge from pairwise correlations (using all 10 questions in the BFI-10) with the remaining judges. Judges with low correlation scores are deemed to be spammers. The judges were then ranked in order of correlation score and the k highest ranked selected.

To evaluate the efficacy of this ranking procedure we calculated within-judge consistency results for the AV clips for different judge numbers ranging from $k = 10$ (without elimination) to $k = 3$. These values averaged over all tasks are presented in **Figure 3A**. We further validated this by computing ICC with varying number of judges, **Figure 3C**. Selecting 5 judges per clip (based on pairwise comparisons) was found to be sufficient to increase reliability to acceptable levels for the AV clips (greater than 0.5) for all traits except for *openness*. We use 5 judges as it allows us to exclude all judges who failed the test question while having the same number of judges for all clips [5 judges is common in this type of study, e.g., Borkenau and Liebler (1992)].

4.1.2. Within-Judge Consistency

Within-judge consistency was measured in terms of Cronbach's α. For the selected 5 judges per clip, the detailed results with respect to different communication conditions and tasks are presented in **Table 2**(a), where α values that indicate sufficient reliability for the BFI-10 (greater than 0.5, in line with values reported in the literature (Credé et al., 2012)) are highlighted in bold. To compare α values between communication conditions we follow the method suggested by Feldt et al. (1987): 95% confidence intervals are calculated for each α value, and if the value from

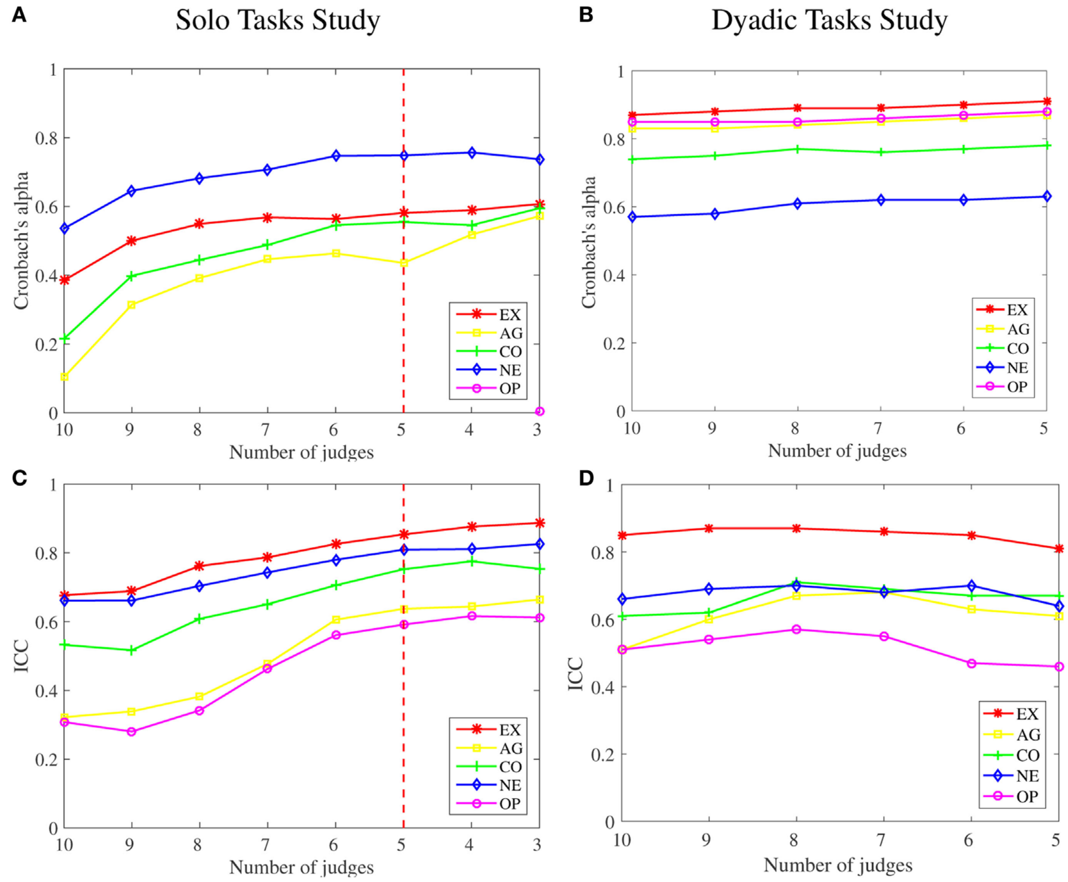


FIGURE 3 | **Changes in Cronbach's α values (A,B) and ICC values (C,D) as a function of number selected judges (k) for different traits in the AV communication condition for Solo Tasks Study (A–C) and Dyadic Tasks Study (B–D)**.

TABLE 2 | Analysis of personality judgments across 3 communication conditions and 3 tasks.

	Audiovisual (AV)				Audio-only (AO)			Teleoperation (TO)			
	Hobby	**Story**	**Mime**	**All**	**Hobby**	**Story**	**All**	**Hobby**	**Story**	**Mime**	**All**
(a) Within-judge											
EX	**0.64**	**0.56**	**0.63**	**0.62**	**0.57**	−0.15	0.34	**0.61**	0.39	0.19	0.47
AG	**0.54**	0.41	**0.60**	**0.52**	**0.61**	0.33	**0.52**	0.40	**0.56**	0.37	0.44
CO	0.47	**0.60**	**0.54**	**0.55**	**0.50**	0.21	0.39	**0.54**	**0.56**	**0.57**	**0.55**
NE	**0.76**	**0.76**	**0.78**	**0.78**	**0.75**	0.42	**0.63**	**0.66**	**0.54**	0.30	**0.50**
OP	−0.6	0.05	0.22	−0.04	−0.14	0.12	0.05	0.17	−0.24	−0.14	−0.07
(b) Between-judge											
EX	0.84***	0.81***	0.74***	0.81***	0.72***	0.51*	0.70***	0.72***	0.63**	−0.12	0.66***
AG	0.46*	0.61**	0.40	0.55***	0.25	−0.15	0.32	0.21	0.54**	−0.95	0.39**
CO	0.78***	0.67***	0.71***	0.72***	0.37	−0.10	0.22	0.32	0.65***	−0.35	0.36*
NE	0.80***	0.71***	0.55**	0.75***	0.57**	0.12	0.55***	0.70***	0.36	−0.56	0.44**
OP	0.12	0.67***	0.40	0.52***	0.49	0.40	0.55***	0.34	0.17	0.04	0.36*
(c) Self-other											
EX	0.34***	0.32**	0.26*	0.30***	0.44***	0.01	0.24***	0.12	−0.02	0.04	0.05
AG	0.04	0.13	0.04	0.07	0.28**	−0.05	0.12	0.08	−0.01	0.10	0.06
CO	−0.17	0.09	0.16	0.03	0.13	−0.13	0.01	0.05	0.16	−0.16	0.01
NE	0.00	−0.07	0.05	−0.01	0.07	0.09	0.07	0.02	−0.08	0.04	0.00
OP	0.06	0.03	0.00	0.03	0.10	0.04	0.07	0.16	0.07	0.03	0.09

*(a) Within-judge consistency in terms of Cronbach's α (good reliability > 0.80 is highlighted in bold); (b) Between-judge consistency in terms of ICC(1,k) (at a significance level of *p < 0.05, **p < 0.01, ***p < 0.001); (c) Self-other agreement in terms of Pearson correlation (at a significance level of *p < 0.05, **p < 0.01, and ***p < 0.001).*

one condition falls outside the confidence intervals from a condition it is being compared to, this suggests it is significantly less consistent. Comparing AO with AV for the hobby task, values for all traits, except for *agreeableness*, fall outside the 95% confidence intervals of the AV values. Comparing TO with AV for the mime task, values for all traits, except for *conscientiousness*, fall outside the 95% confidence intervals of the AV values. This indicates AV is found to be more consistent as compared to AO for the hobby task (except for *agreeableness*) and TO for the mime task (except for *conscientiousness*). No other comparisons indicate significant differences.

4.1.3. Between-Judge Consistency

We computed between-judge consistency in terms of intraclass correlation, ICC(1,k) proposed by Shrout and Fleiss (1979), where $k = 5$. Our judge selection method uses the k most correlated judges so might bias the ICC results (see Section 4.1.1). To evaluate this, we calculated ICC for $k = (10, \ldots 3)$ for the AV condition. **Figure 3B** shows that, for *extroversion*, *conscientiousness*, and *neuroticism*, ICC does not change meaningfully as the number of judges varies, while selecting the 5 most correlated judges slightly biases the results for *agreeableness* and *openness*.

The detailed results for the selected 5 judges per clip are presented in **Table 2**(b). We obtained significant correlations for most traits in the AV condition, with values in the same range ($0.40 < ICC(1, k) < 0.81$) as reported in the literature for online judges using a 10-item test ($0.42 < ICC(1, k) < 0.76$) (Biel and Gatica-Perez, 2013). Fewer significant correlations were observed in the other communication conditions, particularly in the story task for AO and the mime task for TO. *Extroversion* was the only trait that consistently maintained correlation across conditions.

4.1.4. Self-Other Agreement

We examined the extent to which judges agree with the target's self-assessment. Pearson correlations between the self-ratings and the judge's ratings of conditions and tasks are reported in **Table 2**(c) for the selected 5 judges per clip. We observed that the judge's ratings bear a significant relation to the target's self-ratings for *extroversion* only ($r = 0.24 - 0.44$ and $p < 0.05$). However, we did not obtain any significant correlations in the TO condition (all $r < 0.2$ and $p > 0.05$).

4.1.5. Personality Shifts

We examined the extent to which people shifted from one personality class to another, in judges' perception, between AV and TO conditions, in the hobby and story tasks for the selected 5 judges per clip. We did not examine shifts involving AO or Mime task as the ICC scores indicated that personality ratings in this condition would be too unreliable. These results are presented in **Table 3** as 2 × 2 contingency tables. To aid analysis we have also illustrated each shift as a proportional change (%) both from high to low (HIGH2LOW) and from low to high (LOW2HIGH) in **Figure 4** (see the figure on the left hand side).

We found a significant shift from high to low for *neuroticism* (70%). Note that the corrected McNemar's test is very conservative in estimating significance, particularly for small sample sizes. Although not statistically significant, we observed large shifts from low to high for *extroversion* (56%), *conscientiousness* (67%), and *openness* (57%).

4.2. Dyadic Tasks Study

As in the Solo Tasks Study, we assessed whether there existed low-quality judges (spammers) in the judge pool used for the Dyadic Tasks Study. To do so, we repeated the same method that

TABLE 3 | Contingency tables for each trait (at a significance level of *$p < 0.05$).

EX	TO: high	TO: low	AG	TO: high	TO: low	CO	TO: high	TO: low
AV: high	16	**6**	AV: high	16	**11**	AV: high	13	**9**
AV: low	**10**	8	AV: low	**5**	8	AV: low	**12**	6
NE	**TO: high**	**TO: low**	**OP**	**TO: high**	**TO: low**			
AV: high	6	**14***	AV: high	13	**6**			
AV: low	**1***	19	AV: low	**12**	9			

Shift between two classes (from high to low or vice versa) are highlighted in bold.

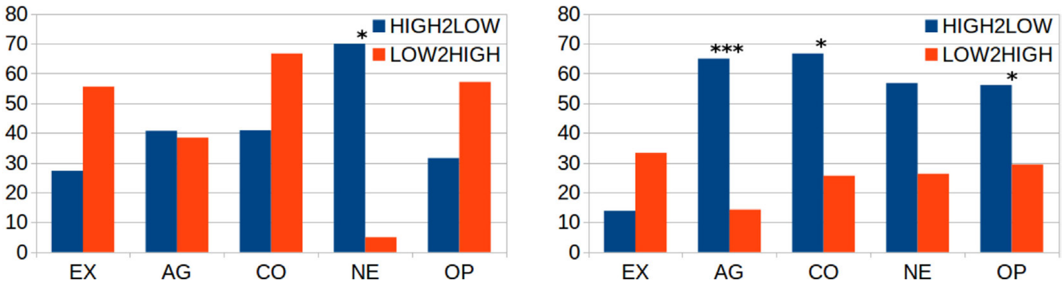


FIGURE 4 | Amount of shifts (%) from high to low (HIGH2LOW) and from low to high (LOW2HIGH) (*$p < 0.05$, ***$p < 0.001$) between AV and TO: solo tasks (left hand side) versus dyadic tasks (right hand side).

we used for the Solo Tasks Study, where we evaluated ICC values, and used judge rating techniques to selectively remove judges. These results are presented in **Figures 3B,D**. As we observed ICC values for the AV condition in line with expectation with all judges included, and cannot observe large changes in the Cronbach's α values and the ICC values, by excluding judges, we concluded that the judges were reliable. Hence, we present the results for the Dyadic Tasks Study without eliminating any judges.

4.2.1. Within-Judge Consistency

Within-judge consistency was measured in terms of Cronbach's α. The detailed results with respect to different communication conditions and tasks are presented in **Table 4**(a), where α values that indicate sufficient reliability for the IPIP-BFM-20 (greater than 0.75, in line with values reported in the literature (Credé et al., 2012)) are highlighted in bold. Values are above or close to good reliability (>0.7) for all traits except for *neuroticism*. Comparing values across communication conditions, we observe little difference, hence judges were able to make consistent trait evaluations when the robot is used for communication.

4.2.2. Between-Judge Consistency

We computed between-judge consistency in terms of intraclass correlation, ICC(1,k), where $k = 10$ (Shrout and Fleiss, 1979). The detailed results for the 10 judges per clip are presented in **Table 4**(b). *Extroversion* and *openness* are the only traits with significant agreement across most tasks and both conditions ($0.47 \leq ICC(1, k) \leq 0.85$ at a significance level of $p < 0.01$). Other traits vary between tasks and conditions as to where significant agreement is achieved. A clearer picture can be gained from the all task results, where it can be seen that agreement on *conscientiousness* deteriorates in the TO condition relative to AV (a drastic drop from 0.61 to −0.26 over all tasks).

4.2.3. Self-Other Agreement

We examined the extent to which judges agree with the target's self-assessment. Pearson correlations between the self-ratings and the judge's ratings of conditions and tasks are reported in **Table 4**(c). Significant agreement was found for *agreeableness* and *openness* across most tasks and both conditions ($r_{ag} = 0.75$ and $r_{op} = 0.71$ over all tasks), although agreement is much lower in the TO condition ($r_{ag} = 0.63$ and $r_{op} = 0.46$ over all tasks). For *extroversion* and *neuroticism*, agreement is much lower than for other traits, and this is fairly consistent across conditions. Again we observe the larger difference across conditions for *conscientiousness* ($r_{co} = 0.17$), with almost no significant agreement in the TO condition compared to significant agreement across all tasks in the AV condition ($r_{co} = 0.31$).

4.2.4. Personality Shifts

We examined the extent to which people shifted from one personality trait classification to another, in judges' perception, between AV and TO conditions for each task. These results are presented in **Tables 3** and **5** as 2 × 2 contingency tables. To aid analysis, we have also illustrated each shift as a proportional change (%) both from high to low (HIGH2LOW) and from low to high (LOW2HIGH) in **Figure 4** (see the figure on the right hand side). We found a significant shift from high to low for *agreeableness* (65%), *conscientiousness* (67%) and *openness* (56%). Although not statistically significant, we observed a large shift from high to low for *neuroticism* (57%).

5. DISCUSSION

In this section, we discuss our results, including comparisons with related work introduced in Section 2. We present in-depth discussion of meta-data (i.e., judge ratings, self-ratings) in terms

TABLE 4 | Analysis of personality judgments across 2 communication conditions and 3 tasks.

	Audiovisual (AV)				Teleoperation (TO)			
	Informative	Competitive	Cooperative	All	Informative	Competitive	Cooperative	All
(a) Within-judge								
EX	**0.85**	**0.87**	**0.85**	**0.87**	**0.84**	**0.85**	**0.84**	**0.86**
AG	**0.77**	**0.80**	**0.84**	**0.83**	**0.86**	**0.84**	**0.81**	**0.84**
CO	0.71	**0.75**	0.**77**	0.74	**0.76**	0.70	0.72	0.73
NE	0.57	0.60	0.54	0.57	0.54	0.64	0.60	0.59
OP	**0.78**	**0.82**	**0.87**	**0.85**	**0.75**	**0.79**	**0.85**	**0.81**
(b) Between-judge								
EX	0.83***	0.84***	0.70***	0.85***	0.61***	0.78***	0.78***	0.82***
AG	0.18	0.21	0.58***	0.51**	0.08	0.35	0.37*	0.41*
CO	0.27	0.28	0.48**	0.61***	−0.24	−0.11	0.24	−0.26
NE	0.52**	0.53**	0.22	0.66***	0.38*	0.13	−0.35	0.46**
OP	0.21	0.67***	0.57***	0.51**	0.55**	0.47**	0.29	0.52**
(c) Self-other								
EX	0.29**	−0.12	−0.29**	−0.06	0.32**	0.21*	−0.15	0.18
AG	0.74***	0.73***	0.44***	0.75***	0.57***	0.65***	0.27**	0.63***
CO	0.22*	0.28**	0.31**	0.31**	−0.01	0.27**	0.14	0.17
NE	0.16	0.18	0.28**	0.24*	0.24*	0.19	0.07	0.23*
OP	0.68***	0.61***	0.17	0.71***	0.51***	0.37***	0.04	0.46***

*(a) Intra-judge consistency in terms of Cronbach's α (good reliability > 0.80 is highlighted in bold); (b) Inter-judge consistency in terms of ICC(1,k) (at a significance level of *p < 0.05, **p < 0.01, ***p < 0.001); (c) Self-other agreement in terms of Pearson correlation (at a significance level of *p < 0.05, **p < 0.01, and ***p < 0.001).*

TABLE 5 | Contingency tables for each trait (at a significance level of *$p < 0.05$ and ***$p < 0.001$).

EX	TO: high	TO: low	AG	TO: high	TO: low	CO	TO: high	TO: low
AV: high	31	**5**	AV: high	14	**26*****	AV: high	12	**24***
AV: low	**13**	26	AV: low	**5*****	30	AV: low	**10***	29
NE	**TO: high**	**TO: low**	**OP**	**TO: high**	**TO: low**			
AV: high	16	**21**	AV: high	18	**23***			
AV: low	**10**	28	AV: low	**10***	24			

Shift between two classes (from high to low or vice versa) are highlighted in bold.

of intra/inter-judge agreement, accuracy of judgments and personality shifts, with regard to different communication conditions (i.e., AO: audio-only, AV: audiovisual, and TO: teleoperation) and different tasks (i.e., solo and dyadic tasks). Note that in the majority of related works results were not directly comparable as personality recognition accuracy is typically the reported metric, as opposed to agreement as used here; accuracy as measured by comparing human responses with machine learning systems (e.g., Aran and Gatica-Perez (2013), Batrinca et al. (2016)), or between self-ratings and judge ratings (e.g., Funder (1995), Borkenau et al. (2004)). Nevertheless, for which traits this reported accuracy is high or low helps provide some explanation for our findings.

5.1. Intra-Judge Agreement

Consistency within judges for how each trait is judged (**Table 2**(a) and **Table 4**(a)) is used to address RQ1. In both studies, judges were sufficiently consistent in their trait ratings in the audiovisual condition (AV), with the exception of *openness* in the Solo Tasks Study, and to a lesser extent *neuroticism* in the Dyadic Tasks Study for us to conclude that the tasks and judges' behaviors were reliable. Batrinca et al. (2016) also reported a similar finding that openness was not modeled successfully in the human-machine interaction, whereas, in the human–human interaction setting, it was the only trait that could be predicted with a high accuracy over all collaboration tasks. In our case, the difference between the two studies with regard to consistent judgment of the *openness* trait indicates that cues for this trait may be more evident in dyadic tasks. Some researchers have suggested that one aspect of *openness* is intellect, where intellect incorporates the facets of intelligence, intellectual engagement, and creativity (DeYoung, 2011), and the tasks in the Dyadic Tasks Study are more conducive to displaying these facets.

In the Solo Tasks Study, there were some notable differences between the audio-only (AO) and the teleoperated robot (TO) conditions. For the hobby task, judges remained consistent in both the AO and TO conditions, indicating they were able to use audio cues to make judgments for this task, and robot appearance had no effect on consistency. However, for the story task, judges were much less consistent in the AO than in the AV condition, for all traits except for *agreeableness*. This is in contrast to the teleoperated robot condition (TO), where they remained as consistent as in the AV condition. The only additional cues available with the robot compared to audio only are gestures and appearance. The results indicate that such cues are used to aid judgments in the same way that they do in the AV condition, though their utility appears to be task dependent (only of apparent benefit in the story task). Importantly, the fact that they are utilized provides good evidence that the robot is not simply ignored when making judgments. Hence, the findings of high levels of agreement across both conditions in all tasks in the Dyadic Tasks Study indicate that in dyadic tasks the robot transmits sufficient cues to make judgments as consistently as observing the target directly.

The use of gesture to aid personality judgments appears to be dependent on it accompanying speech, as in the Solo Tasks Study ratings in the TO condition are far less consistent than in the AV condition for the mime task. That is to say, gestures alone do not provide sufficient information for judging personality. This was in contrast to what was reported by Aran and Gatica-Perez (2013), where the best results were achieved when they used visual cues only for predicting personality traits, and using audio cues or combining them with visual cues resulted in lower accuracy. This showed that either other behavior cues not transmitted by the robot are needed, or appearance cues are used which conflict with gesture cues in the TO condition.

Taking the results from both studies together, it is apparent that judges are able to remain consistent in their judgments of a given trait whether they are observing someone directly or their communication relayed through a teleoperated robot. Indeed, where there are slight shifts in consistency between AV and TO conditions, they are not large; the one exception being for the mime task in the Solo Tasks Study. Hence, each judge appears to formulate a relatively consistent evaluation of a given targets' personality traits based on speech, gesture, and appearance, combining them to assess each trait facet. This finding is in contrast to the study by Kuwamura et al. (2012) where they suggested small shifts in intra-judge consistency provided evidence of robot appearance effects on personality perception. While in subsequent sections we do observe evidence for effects of robot mediation on perception, we do not find such small shifts in intra-judge consistency convincing in this regard.

5.2. Inter-Judge Agreement

Looking at inter-judge agreement results to address RQ2 (**Table 2**(b) and **Table 4**(b)), *extroversion* was the only trait on which judges reached consensus in both studies, regardless of the communication condition, and task (the mime task in the Solo Tasks Study being the one exception). This result is in line with the widely accepted idea that *extroversion* is the easiest trait to infer upon others (Barrick et al., 2000). Hence, the strength of the available cues was sufficient to overcome any conflict between appearance, vocal, and gesture based cues. Indeed it indicates that judges had a common set of interpretations for the available cues.

On the other hand, where agreement was reached on *agreeableness*, *conscientiousness*, and *neuroticism* for some tasks in the AV condition in each study, it had mostly deteriorated in the TO condition, and the AO condition in the Solo Tasks Study. The clearest example of this is for *conscientiousness* taking all three tasks together in the Dyadic Tasks Study (and to some extent in the Solo Tasks Study as well), where agreement drastically deteriorated in the TO condition as compared to the AV condition. As explained in the study by Macrae et al. (1996), physical appearance based impressions (facial and vocal features) are often used in the judgment of *conscientiousness*. In particular, low *conscientiousness* is conveyed by a childlike face (Macrae et al., 1996), which the face of the NAO robot can be considered to have, and this may conflict with the vocal cues of the operator. *Neuroticism* is mainly related to emotions, and *agreeableness* is related to trust, cooperation and sympathy (Zillig et al., 2002), both of which it seems reasonable to suggest judges might perceive as being low for a robot (particularly NAO with its lack of facial expressions), again creating conflicts. It would appear that judges do not have a consistent manner with which to resolve such conflicts.

Task-based analyzes in the Solo Tasks Study show that for *agreeableness* and *conscientiousness* the story task provides sufficient cues for agreement to be maintained in the TO condition, whereas the hobby task does so for *neuroticism*. As agreement being maintained in the TO condition indicates sufficient cues to overcome appearance/behavior conflicts, it is instructive to consider how those tasks might relate to the traits. In telling the story, targets might demonstrate their morality, and relation to others, components of *agreeableness* (Zillig et al., 2002). How well structured and clear the story is could relate to facets of the *conscientiousness* trait. The hobby task on the other hand might demonstrate how self-conscious a person is about their hobby, a facet of *neuroticism* (Zillig et al., 2002). While these two tasks might provide some cues for facets of the traits for which consistency was not maintained, they appear to do so in a way that conflicts with cues related to the robot.

We also compared differences in agreement between the TO and AO conditions in the Solo Tasks Study. Where there is agreement in TO for *agreeableness*, *conscientiousness*, and *neuroticism*, we found it was greatly reduced for *agreeableness* and *conscientiousness*, and to a lesser extent for *neuroticism*. This provides further evidence that physical cues, be they behavioral or appearance based, are utilized in the TO condition. Again, this appears to be dependent on the presence of speech: in the mime task for the Solo Tasks Study, judges were unable to provide a consistent rating for any trait in the TO condition, in contrast to the consistent ratings for *extroversion*, *conscientiousness*, and *neuroticism* in the AV condition. A likely reason for this observation is that without vocal cues there is an increased reliance on appearance based cues, often based on stereotypes (Kenny et al., 1994), and judges do not have consistent stereotypes relating to robot appearance.

Batrinca et al. (2016) showed that the prediction of agreeableness and conscientiousness in the human-machine interaction setting and the prediction of conscientiousness and neuroticism were highly dependent on the collaboration task, where the extroversion trait was the only trait yielding consistent results over all tasks in both settings. Similarly, our task-based analyses in the Dyadic Tasks Study show that in the AV condition, while the cooperative task provided a higher level of agreement for *agreeableness* and *conscientiousness*, the competitive task yielded better results for *neuroticism* and *openness*. Indeed, the results are somewhat expected given the nature of the tasks: the cooperative task was to agree upon how to order five items in a survival scenario, in which participants were expected to exhibit the *agreeableness* facet of personality; the competitive task was more related to creativity and intelligence, that are strongly associated with *openness* (Zillig et al., 2002). Though agreement is lower, it is still maintained for *agreeableness* in the cooperative task and *openness* in the competitive task in the TO condition. This indicates that in these cases, for at least some of the judges, either the vocal cues override the visual cues, or movement cues are utilized (with the vocal cues).

Taken together, the findings from both studies indicate that the ability of judges to make judgments based on a common interpretation of cues is affected not only by communication condition but is also dependent on the task. While in some cases it is apparent that a particular task is conducive to providing more verbal cues than another for a particular trait (as indicated by higher agreement, and inferred from the literature), whether these override the physical cues in the TO condition is hard to predict. Indeed, whether clear cues in the AV condition translate into agreement in the TO condition vary a great deal between all tasks. Hence, it seems reasonable to suggest that whether inter-judge consistency is observed also depends on how much appearance cues are utilized for a given task and trait, and thus how all the cues interact. This complex interaction effect provides strong evidence that personality perception is likely to be altered when communicating *via* a robot, and this depends on what cues are produced.

5.3. Accuracy of Judgments

In order to assess RQ3, we analyzed the extent to which judge ratings correlated with self-ratings provided by target participants (**Table 2**(c) and **Table 4**(c)). In general in the Solo Tasks Study, there was very little correlation between self and other ratings. This is in contrast to previous findings where they found low, but significant, self-other correlation ($0.11 - 0.42$) (Carney et al., 2007a). The one exception to this was self-other correlation for *extroversion* in the AV condition. This suggests that participant targets did not present cues relating to their self-perception in the tasks we used, other than for *extroversion* which is commonly reported as the trait with the most available cues. Audio cues were sufficient for this correlation to be maintained in the hobby task in the AO condition, but not in the story task, or in either task in the TO condition.

In contrast to the tasks used in the Solo Tasks Study, the tasks of the Dyadic Tasks Study resulted in self-other agreement for *extroversion*, *agreeableness*, *conscientiousness*, and *openness* in the majority of tasks for the AV condition. This indicates that the tasks we used in the Dyadic Tasks Study were better at engendering more naturalistic behavior, and hence personality cues than the tasks in the Solo Tasks Study. Indeed, an important factor in thin slice personality analysis is how easy a person is to judge

(Funder, 1995), and people behaving more naturally produce better cues. However, despite these apparently better cues, there was a large reduction in agreement for *conscientiousness*, *neuroticism*, and *openness* (and to a lesser extent *agreeableness*) in the TO condition relative to the AV condition. This finding combined with those of the Solo Tasks Study suggests that there is a shift in the way personality cues are interpreted caused by their interaction with the appearance of the robot, and the way non-verbal communication cues are reproduced on it.

5.4. Personality Shifts

In order to address RQ4, we analyzed the difference in perceived personality in terms of the occurrences of personality shifts. We principally consider the results from the Dyadic Tasks Study as it provides the more compelling evidence. The main reason for this assertion is that more naturalistic cues appeared to be produced in the Dyadic Tasks Study (see previous section), and we consider such cues and their interaction with the TO condition more ecologically valid. In addition, by being able to consider three tasks rather than the two considered in the Solo Tasks Study we have increased statistical power. The shifts we observed (**Figure 4**) provide evidence that cues related to the robots appearance are incorporated into, or even override personality judgments based on speech. Indeed, this is somewhat to be expected given that (Behrend et al., 2012) observed that, in judgments of suitability, attractiveness of a graphical avatar superseded qualities perceived in an interviewees words.

There are two likely causal factors in the perceived personalities being shifted, first human-based physical appearance stereotypes (inferred from humanlike characteristics of the robot) might be applied, second characteristics related to robots might be applied. Here, we will discuss possible underlying causes for the shifts observed in the Dyadic Tasks Study. In the case of *conscientiousness* and *neuroticism* a childlike face, as the NAO might be considered to have, conveys low ratings for both traits (Borkenau and Liebler, 1992; Macrae et al., 1996). Further, *conscientiousness* and *neuroticism* were also observed to be influenced by face shape in graphical avatars (Fong and Mar, 2015), and as the NAO has a face shape that differs from a human, hence this could lead to distortions in perceptions of these traits. Additionally, *neuroticism* is mainly related to emotions (Zillig et al., 2002), something which robots are rarely considered to have. Also linked to emotions is *openness*, which combined with its other facets of imagination and creativity, might also be reasonably expected to be low for a robot, which could also be considered to have *hard facial linaments*, also linked to low *openness* (Borkenau and Liebler, 1992). The NAO robot could also be considered male in appearance, and male avatars have been found to cue for lower *conscientiousness* and *openness* (Fong and Mar, 2015). Low *agreeableness* is more difficult to rationalize, but one facet is trustworthiness (Zillig et al., 2002), and judges may have perceived using a robot to communicate as less trustworthy. The vocal cues for *extroversion* appeared to be very strong, and this might explain why little influence on this trait was observed.

An important thing to note from these findings is that people appear to be attributing personality stereotypes to NAO for characteristics other than the *extroversion* trait, which has been previously examined (Park et al., 2012; Aly and Tapus, 2013; Celiktutan and Gunes, 2015). Hence, in future work in which a desired personality is to be expressed by an autonomous robot, its appearance based cues must be considered alongside any behavioral cues expressed. We suggest that strong behavioral cues may be required to overcome such stereotypes.

5.5. Conclusion

In this paper, we have shown that judges are able to make personality trait judgments that are as consistent with a robot avatar as when the same people are viewed on video in contrast to past work (Kuwamura et al., 2012). One possible reason for this difference in findings is that our teleoperation system allows reproduction of some non-verbal communication cues on the robot which might improve the ease with which judges can assess personality. Hence, we suggest that it is important for telepresence systems to be able to transmit non-verbal communication cues, whether this be actuation of physical systems, or large enough screens on remote presence devices.

We have shown that the appearance of a teleoperated robot avatar influences how the personality of its controller is perceived, i.e., robot appearance based personality cues are utilized along with cues in the speech of the operators. Hence, the perceived personality of a teleoperator is shifted toward that related to the robot's appearance. In light of these findings, we suggest that robot avatar appearance and behavior be carefully considered relative to the person who will be controlling it, and this needs to be done on an individual basis. Training of operators to produce clear cues, or having some cues appropriate to the operator's personality autonomously generated, might allow some control of appearance effects.

Having the correct robot personality has been found to have a positive effect on interactions with people (Park et al., 2012; Aly and Tapus, 2013; Celiktutan and Gunes, 2015), and our findings also have implications for such autonomous robot personality expression. It is important to consider what appearance cues for personality a robot has, as we have observed humanlike personality inferences, and whether the planned behavioral cues might conflict with them. Cues that work on one platform may not be transferable to another. Additionally, we suggest that future experiments on robots expressing personality need to carefully consider tasks undertaken, as we observed that intra-judge agreement on personality perception was highly task dependent.

5.6. Limitations and Future Work

While this paper provides evidence for how personality perception is affected for people teleoperating a humanoid robot avatar, it has a number of limitations we hope to address in future work.

One area of limitation in our work relates to the movement capabilities of the NAO robot, and the inherent differences with human movement capabilities. Although our previous work showed reproduced gestures are comprehensible (Bremner and Leonards, 2015, 2016), there are clearly appreciable differences in the way some movements are reproduced. Indeed, while these differences have limited affect on perceived meaning, they likely contribute to the observed distortions in personality. The main limitations in this regard are in elbow flexion, movement speed, and wrist and hand motion: the NAO elbow can only bend to ~90°, the main effect of which being a reduction in vertical travel of the hand for some gestures; humans are capable of extremely rapid motions that the robot cannot match, consequently it will catch up as best it can, but the usual response will be to not express some motions due to the method of motion processing; wrist flexion and hand shape are clearly of utility in many gestures, and their absence (as well as wrist rotation in study 2) restricts the expression of components of some gestures. These movement restrictions are added to by limitations in the Kinect sensor and software processing: movements that result in hand occlusions can lead to imprecision, as well as noise in the sensor data can lead to some added jitter on the robot (though this is filtered as much as possible).

It is also important to note that robot operators had little to no awareness of the limitations of the robot as none of them had prior experience with NAO, and when in control of it they could not observe its motion. The only instruction given pertaining to system capabilities was to not to rest with the arms flat against the body or behind the back as tracking would be lost. While this resulted in some initial poses that were a bit unnatural (video of which was not used in the studies), participants soon reverted to "normal" behavior. Indeed, qualitative comparison of participants in the dyadic study in each condition (video of participants recorded while they were operating the robot allowed this) reveals little difference in gesturing behavior for the majority of participants. Exceptions were the two participants with prior experience working with robots who moved more than they did face-to-face. In further work, we aim to more closely examine the data for any differences (which may be subtle), and if present test how they contribute to the observed personality distortion effects.

In the study by Celiktutan et al. (2016), our AV condition results showed that face gestures and head activity play an important role in the recognition of the extroversion, agreeableness and conscientiousness traits. This implies another limitation of the robotic platform used in this study. To convey the teleoperators personality traits more accurately, the robot should portray head pose or facial activity together with audio and arm gestures.

A further limitation is that there are some differences between our two studies, the Dyadic Tasks Study has a slightly different design due to correcting issues we encountered in the Solo Tasks Study, making the study comparison slightly less fair. In particular, we addressed the issue with low-quality judges, by utilizing a different recruiting platform which allowed us to recruit better quality judges, and thus did not require a judge removal process. In the Solo Tasks Study, the issues with low-quality judges meant we used a judge selection method based on the gathered responses. The procedure we used had a slight biasing effect on the between-judge consistency (ICC) result for *agreeableness* and *openness*. This bias means that where ICC values are not significant it is strong evidence that there is either a lack of cues or conflicting cues, as even amongst the most agreeing judges consensus of opinion was not possible. Where there is significant agreement, it indicates there are cues for that trait in the particular task and condition and some judges are able to pick up on these cues. Indeed, Funder points out that there exists good and bad judges of personality (Funder, 1995), and we suggest our selection method allowed us to bias toward good judges. This limits the generalizability of our results to judges more adept at picking up on personality cues. By changing crowdsourcing platforms we were able to remove the need for this selection process in the Dyadic Tasks Study.

In addition to recruiting better quality judges, we also utilized a larger personality questionnaire, making our results more accurate, especially with regard to measuring intra-judge and inter-judge consistency.

In the work reported here, it is not clear how different cues are utilized in the aforementioned personality perception. Given that there was such high variability in affects of robot appearance dependent on the task, it seems likely this is due to differences in use of audio and visual cues. Hence, we intend to analyze in-depth the behaviors of targets relative to their judged personality for different tasks. To facilitate this, we aim to extend our work on automatic personality classification, which can extract and identify useful cues automatically (Celiktutan et al., 2016), and apply it to the recordings from the Dyadic Tasks Study. A comparative cue analysis could not only allow us to gain a better understanding of the causes of personality shifts, but could also be useful in synthesizing robot personality behavioral cues.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the ethics committees of the University of the West of England and the University of Cambridge with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the ethics committees of the University of the West of England and the University of Cambridge.

AUTHOR CONTRIBUTIONS

PB: substantial contributions to the conception and design of the work, the acquisition, analysis, and interpretation of data; drafting the work; final approval of the version to be published; and agreement to be accountable. OC: substantial contributions to the conception and design of the work, the acquisition, analysis, and interpretation of data; drafting the work; final approval of the version to be published; and agreement to be accountable. HG: substantial contributions to the design of the work, analysis, and interpretation of data; revising the work critically for important intellectual content; final approval of the version to be published; and agreement to be accountable.

FUNDING

This work was funded by the EPSRC under its IDEAS Factory Sandpits call on Digital Personhood (Grant Ref: EP/L00416X/1).

REFERENCES

Adalgeirsson, S. O., and Breazeal, C. (2010). "MeBot: a robotic platform for socially embodied telepresence," in *Proc. of Int. Conf. Human Robot Interaction* (Osaka: ACM/IEEE), 15–22.

Alibali, M. (2001). Effects of visibility between speaker and listener on gesture production: some gestures are meant to be seen. *J. Mem. Lang.* 44, 169–188. doi:10.1006/jmla.2000.2752

Aly, A., and Tapus, A. (2013). "A model for synthesizing a combined verbal and nonverbal behavior based on personality traits in human-robot interaction," in *Proc. of ACM/IEEE Int. Conf. on Human-Robot Interaction*, Tokyo.

Aran, O., and Gatica-Perez, D. (2013). "One of a kind: inferring personality impressions in meetings," in *Proc. of ACM Int. Conf. on Multimodal Interaction*, Sydney.

Barrick, M. R., Patton, G. K., and Haugland, S. N. (2000). Accuracy of interviewer judgments of job applicant personality traits. *Personnel Psychol.* 53, 925–951. doi:10.1111/j.1744-6570.2000.tb02424.x

Batrinca, L., Mana, N., Lepri, B., Sebe, N., and Pianesi, F. (2016). Multimodal personality recognition in collaborative goal-oriented tasks. *IEEE Trans. Multimedia* 18, 659–673. doi:10.1109/TMM.2016.2522763

Behrend, T., Toaddy, S., Thompson, L. F., and Sharek, D. J. (2012). The effects of avatar appearance on interviewer ratings in virtual employment interviews. *Comput. Human Behav.* 28, 2128–2133. doi:10.1016/j.chb.2012.06.017

Bevan, C., and Stanton Fraser, D. (2015). "Shaking hands and cooperation in tele-present human-robot negotiation," in *Proc. of Int. Conf. Human Robot Interaction* (Portland: ACM/IEEE), 247–254.

Biel, J., and Gatica-Perez, D. (2013). The YouTube lens: crowdsourced personality impressions and audiovisual analysis of Vlogs. *IEEE Trans. Multimedia* 15, 41–55. doi:10.1109/TMM.2012.2225032

Borkenau, P., and Liebler, A. (1992). Trait inferences: sources of validity at zero acquaintance. *J. Pers. Soc. Psychol.* 62, 645–657. doi:10.1037/0022-3514.62.4.645

Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., and Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *J. Pers. Soc. Psychol.* 86, 599–614. doi:10.1037/0022-3514.86.4.599

Bremner, P., Celiktutan, O., and Gunes, H. (2016a). "Personality perception of robot avatar tele-operators," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction, HRI '16* (Christchurch: IEEE), 141–148.

Bremner, P., Koschate, M., and Levine, M. (2016b). "Humanoid robot avatars: an 'in the wild' usability study," in *RO-MAN* (New Zealand: IEEE).

Bremner, P., and Leonards, U. (2015). "Efficiency of speech and iconic gesture integration for robotic and human communicators – a direct comparison," in *Proc. of IEEE Int. Conf. on Robotics and Automation* (Seattle: IEEE), 1999–2006.

Bremner, P., and Leonards, U. (2016). Iconic gestures for robot avatars, recognition and integration with speech. *Front. Psychol.* 7:183. doi:10.3389/fpsyg.2016.00183

Carney, D. R., Colvin, C. R., and Hall, J. A. (2007a). A thin slice perspective on the accuracy of first impressions. *J. Res. Pers.* 41, 1054–1072. doi:10.1016/j.jrp.2007.01.004

Celiktutan, O., Bremner, P., and Gunes, H. (2016). "Personality classification from robot-mediated communication cues," in *25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, New York.

Celiktutan, O., and Gunes, H. (2015). "Computational analysis of human-robot interactions through first-person vision: personality and interaction experience," in *24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (Kobe: IEEE), 815–820.

Credé, M., Harms, P., Niehorster, S., and Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the big five personality traits. *J. Pers. Soc. Psychol.* 102, 874–888. doi:10.1037/a0027403

Daly-Jones, O., Monk, A., and Watts, L. (1998). Some advantages of video conferencing over high-quality audio conferencing: fluency and awareness of attentional focus. *Int. J. Human Comput. Stud.* 49, 21–58. doi:10.1006/ijhc.1998.0195

DeYoung, C. D. (2011). "Intelligence and personality," in *The Cambridge Handbook of Intelligence*, eds R. J. Sternberg and S. B. Kaufman (New York, NY: Cambridge University Press), 711–737.

Edwards, A. L. (1948). Note on the correction for continuity in testing the significance of the difference between correlated proportions. *Psychometrika* 13, 185–187. doi:10.1007/BF02289261

Feldt, L. S., Woodruff, D. J., and Salih, F. A. (1987). Statistical inference for coefficient alpha. *Appl. Psychol. Measure.* 11, 93–103. doi:10.1177/014662168701100107

Fong, K., and Mar, R. A. (2015). What does my avatar say about me? Inferring personality from avatars. *Pers. Soc. Psychol. Bull.* 41, 237–249. doi:10.1177/0146167214562761

Funder, D. C. (1995). On the accuracy of personality judgment: a realistic approach. *Psychol. Rev.* 102, 652–670. doi:10.1037/0033-295X.102.4.652

Funder, D. C., Furr, R. M., and Colvin, C. R. (2000). The riverside behavioral q-sort: a tool for the description of social behavior. *J. Pers.* 68, 451–489. doi:10.1111/1467-6494.00103

Funder, D. C., and Sneed, C. D. (1993). Behavioral manifestations of personality: an ecological approach to judgmental accuracy. *J. Pers. Soc. Psychol.* 64, 479–490. doi:10.1037/0022-3514.64.3.479

Gouaillier, D., Hugel, V., Blazevic, P., Kilner, C., Monceaux, J., Lafourcade, P., et al. (2009). "Mechatronic design of NAO humanoid," in *Proc of IEEE Int. Conf. on Robotics and Automation* (Kobe: IEEE), 769–774.

Hossen Mamode, H. Z., Bremner, P., Pipe, A. G., and Carse, B. (2013). "Cooperative tabletop working for humans and humanoid robots: group interaction with an avatar," in *IEEE Int. Conf. on Robotics and Automation* (Karlsruhe: IEEE), 184–190.

Kenny, D. A., Albright, L., Malloy, T. E., and Kashy, D. A. (1994). Consensus in interpersonal perception: acquaintance and the big five. *Psychol. Bull.* 116, 245–258. doi:10.1037/0033-2909.116.2.245

Kristoffersson, A., Coradeschi, S., and Loutfi, A. (2013). A review of mobile robotic telepresence. *Adv. Human-Comput. Interact.* 2013, 17. doi:10.1155/2013/902316

Kuwamura, K., Minato, T., Nishio, S., and Ishiguro, H. (2012). "Personality distortion in communication through teleoperated robots," in *Proc of IEEE Int. Symp. on Robot and Human Interactive Communication* (Paris: IEEE), 49–54.

Lee, M. K., and Takayama, L. (2011). "Now, I have a body," in *Proc. of the Conf. on Human Factors in Computing Systems* (Vancouver, BC: ACM Press), 33.

Macrae, C. N., Stangor, C., and Hewstone, M. (1996). *Stereotypes and Stereotyping* (New York, NY: The Guilford Press).

Martins, H., and Ventura, R. (2009). "Immersive 3-d teleoperation of a search and rescue robot using a head-mounted display," in *IEEE Conf. on Emerging Technologies Factory Automation (ETFA)* (Mallorca: IEEE), 1–8.

McKeown, G., Valstar, M., Cowie, R., Pantic, M., and Schroder, M. (2012). The semaine database: annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE Trans. Affect. Comput.* 3, 5–17. doi:10.1109/T-AFFC.2011.20

Murray, H. A. (1943). *Thematic Apperception Test*. Cambridge, MA: Harvard University Press.

Naumann, L. P., Vazire, S., Rentfrow, P. J., and Gosling, S. D. (2009). Personality judgments based on physical appearance. *Pers. Soc. Psychol. Bull.* 35, 1661–1671. doi:10.1177/0146167209346309

O'Conaill, B., Whittaker, S., and Wilbur, S. (1993). Conversations over video conferences: an evaluation of the spoken aspects of video-mediated communication. *Human Comput. Interact.* 8, 389–428. doi:10.1207/s15327051hci0804_4

Park, E., Jin, D., and del Pobil, A. P. (2012). The law of attraction in human-robot interaction. *Int. J. Adv. Rob. Syst.* 9, 1–7. doi:10.5772/50228

Rae, I., Takayama, L., and Mutlu, B. (2013). "In-body experiences," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems – CHI '13* (New York, NY: ACM Press), 1921–1930.

Rammstedt, B., and John, O. P. (2007). Measuring personality in one minute or less: a 10-item short version of the big five inventory in English and German. *J. Res. Pers.* 41, 203–212. doi:10.1016/j.jrp.2006.02.001

Riggio, R. E., and Friedman, H. S. (1986). Impression formation: the role of expressive behavior. *J. Pers. Soc. Psychol.* 50, 421–427. doi:10.1037/0022-3514.50.2.421

Salam, H., Celiktutan, O., Hupont, I., Gunes, H., and Chetouani, M. (2016). Fully automatic analysis of engagement and its relationship to personality in human-robot interactions. *IEEE Access* 5, 705–721. doi:10.1109/ACCESS.2016.2614525

Shrout, P., and Fleiss, J. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* 86, 420–428. doi:10.1037/0033-2909.86.2.420

Straub, I., Nishio, S., and Ishiguro, H. (2010). "Incorporated identity in interaction with a teleoperated android robot: a case study," in *Proc of Int. Symp. in Robot and Human Interactive Communication* (Viareggio: IEEE), 119–124.

Tang, A., Boyle, M., and Greenberg, S. (2004). "Display and presence disparity in mixed presence groupware," in *Proc. of Australasian User Interface Conf* (Dunedin: Australian Computer Society, Inc.), 73–82.

Topolewska, E., Skiminia, E., Strus, W., Cieciuch, J., and Rowinski, T. (2014). The short ipip-bfm-20 questionnaire for measuring the big five. *Ann. Psychol.* 2, 385–402.

Vinciarelli, A., and Mohammadi, G. (2014). A survey of personality computing. *IEEE Trans. Affect. Comput.* 5, 273–291. doi:10.1109/TAFFC.2014.2330816

Wang, Y., Geigel, J., and Herbert, A. (2013). "Reading personality: avatar vs. human faces," in *Proc. of HAC Conf. on Affective Computing and Intelligent Interaction* (Geneva: IEEE), 479–484.

Yamazaki, R., Nishio, S., Ogawa, K., and Ishigur, H. (2012). "Teleoperated android as an embodied communication medium: a case study with demented elderlies in a care facility," in *RO-MAN* (Paris: IEEE), 1066–1071.

Zillig, L. M. P., Hemenover, S. H., and Dienstbier, R. A. (2002). What do we assess when we assess a big 5 trait? A content analysis of the affective, behavioral, and cognitive processes represented in big 5 personality inventories. *Pers. Soc. Psychol. Bull.* 28, 847–858. doi:10.1177/0146167202289013

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Bremner, Celiktutan and Gunes. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 30 August 2017
doi: 10.3389/fpsyg.2017.01366

Understanding the Uncanny: Both Atypical Features and Category Ambiguity Provoke Aversion toward Humanlike Robots

Megan K. Strait[1,2]*, Victoria A. Floerke[2], Wendy Ju[3], Keith Maddox[4], Jessica D. Remedios[5], Malte F. Jung[6] and Heather L. Urry[2]

[1] Social Systems Laboratory, Computer Science, University of Texas Rio Grande Valley, Edinburg, TX, United States, [2] Emotion, Brain, and Behavior Laboratory, Psychology, Tufts University, Medford, MA, United States, [3] Center for Design Research, Mechanical Engineering, Stanford University, Stanford, CA, United States, [4] Social Cognition Laboratory, Psychology, Tufts University, Medford, MA, United States, [5] Social Identity and Stigma Laboratory, Psychology, Tufts University, Medford, MA, United States, [6] Robots in Groups Laboratory, Information Science, Cornell University, Ithaca, NY, United States

OPEN ACCESS

Edited by:
Bilge Mutlu,
University of Wisconsin-Madison,
United States

Reviewed by:
Roope Raisamo,
University of Tampere, Finland
Silvia Gabrielli,
Fondazione Bruno Kessler, Italy

***Correspondence:**
Megan K. Strait
megan.strait@utrgv.edu

Specialty section:
This article was submitted to
Human-Media Interaction,
a section of the journal
Frontiers in Psychology

Received: 31 December 2016
Accepted: 27 July 2017
Published: 30 August 2017

Citation:
Strait MK, Floerke VA, Ju W, Maddox K, Remedios JD, Jung MF and Urry HL (2017) Understanding the Uncanny: Both Atypical Features and Category Ambiguity Provoke Aversion toward Humanlike Robots. Front. Psychol. 8:1366. doi: 10.3389/fpsyg.2017.01366

Robots intended for social contexts are often designed with explicit humanlike attributes in order to facilitate their reception by (and communication with) people. However, observation of an "uncanny valley"—a phenomenon in which highly humanlike entities provoke *aversion* in human observers—has lead some to caution against this practice. Both of these contrasting perspectives on the anthropomorphic design of social robots find some support in empirical investigations to date. Yet, owing to outstanding empirical limitations and theoretical disputes, the uncanny valley and its implications for human-robot interaction remains poorly understood. We thus explored the relationship between *human similarity* and people's aversion toward humanlike robots via manipulation of the agents' appearances. To that end, we employed a picture-viewing task ($N_{agents} = 60$) to conduct an experimental test ($N_{participants} = 72$) of the uncanny valley's existence and the visual features that cause certain humanlike robots to be unnerving. Across the levels of human similarity, we further manipulated agent appearance on two dimensions, *typicality* (prototypic, atypical, and ambiguous) and *agent identity* (robot, person), and measured participants' aversion using both subjective and behavioral indices. Our findings were as follows: (1) Further substantiating its existence, the data show a clear and consistent uncanny valley in the current design space of humanoid robots. (2) Both category ambiguity, and more so, atypicalities provoke aversive responding, thus shedding light on the visual factors that drive people's discomfort. (3) Use of the Negative Attitudes toward Robots Scale did not reveal any significant relationships between people's pre-existing attitudes toward humanlike robots and their aversive responding—suggesting positive exposure and/or additional experience with robots is unlikely to affect the occurrence of an uncanny valley effect in humanoid robotics. This work furthers our understanding of both the uncanny valley, as well as the visual factors that contribute to an agent's uncanniness.

Keywords: anthropomorphism, emotion regulation, humanoid robots, human-robot interaction, uncanny valley, social robotics

1. INTRODUCTION

By capitalizing on traits that are familiar and intuitive to people, robots designed with greater *human similarity*—both physically and behaviorally—can offer more natural and effective human-robot interactions (Duffy, 2003; Złotowski et al., 2015). For example, incorporating humanlike cues into a robot's design elicits feelings of empathy toward it (Riek et al., 2009) and causes attribution of greater agency (Gray and Wegner, 2012; Broadbent et al., 2013; Stafford et al., 2014a). In turn, this has significant prosocial outcomes such as increases in people's comfort around a robot (Sauppé and Mutlu, 2015) and their willingness to collaborate with it (Andrist et al., 2015).

With the emergence of increasingly humanlike robots, however, researchers have observed an unintended consequence: the *uncanny valley* effect (Mori et al., 2012). The valley effect, originally described by Masahiro Mori nearly a half-century ago, refers to the phenomenon wherein highly humanlike (but not prototypically human) entities provoke aversion in people (for a review, see Kätsyri et al., 2015). For example, highly humanlike robots are rated more negatively (MacDorman, 2006), avoided more frequently (Strait et al., 2015), and attributed less trustworthiness (Mathur and Reichling, 2016) than their less humanlike counterparts and humans. Moreover, such effects do not appear to be limited to adults, as valley-like effects have been observed in infants (Lewkowicz and Ghazanfar, 2012; Matsuda et al., 2012), children (Yamamoto et al., 2009), and even other primates (Steckenfinger and Ghazanfar, 2009), suggesting the general phenomenon is relatively pervasive.

Yet, the uncanny valley continues to be a poorly understood and even contentious topic in human-robot interaction (HRI) research, due to gaps in the current literature and various empirical inconsistencies. These issues stem, at least in part, from challenges inherent to conducting empirical HRI studies (in particular, the limited accessibility of robotic platforms that only partially represent the large design space). This has lead researchers to turn to more accessible alternatives, such as the use of computer-generated stimuli to make inferences about embodied counterparts (e.g., Inkpen and Sedlins, 2011) and careful case studies of only one or a few robotic platforms (e.g., Bartneck et al., 2009; Kupferberg et al., 2011; Saygin et al., 2012; Strait et al., 2014). But the small range of methodologies for investigating the valley, in turn, has lead to conflicting findings. For example, amongst studies utilizing few robots or non-embodied robot stimuli, there are both many studies which fail to find a valley effect (or find the opposite – more positive responding to the most humanlike stimuli; e.g., Bartneck et al., 2009; Kupferberg et al., 2011; Piwek et al., 2014) as well as many that confirm its existence (e.g., Saygin et al., 2012; Koschate et al., 2016; Strait et al., 2015).

Considering that the theoretical comparisons are being made across such dissimilar methodologies, it is unsurprising that inconsistencies have arisen and that gaps in the literature remain. Researchers have begun to address such shortcomings through systematic review of the literature (Kätsyri et al., 2015; Rosenthal-von der Pütten and Krämer, 2015; MacDorman and Chattopadhyay, 2016) and development of alternative methodologies. For example, two recent studies utilized picture-based stimuli (photographs depicting embodied robots) to evaluate a large portion[1] of the current design space in humanoid robotics (Strait et al., 2015; Mathur and Reichling, 2016). In combination, recent work paints a more consistent picture in which there exists a robust uncanny valley as a function of human similarity.

Despite perspectives on the valley's existence trending toward agreement, many critical questions remain. In particular, *when, why, and how do robots fall into the uncanny valley*? Researchers have long pointed to *human similarity* as the cause of the valley effect—wherein a robot with "too much" similarity is unnerving. However, several studies indicate that similarity alone is not sufficient to cause a humanoid robot to fall into the valley. For example, Rosenthal-von der Pütten and Krämer (2014, 2015) have repeatedly shown that people respond negatively toward some instances of highly humanlike robots but positively toward others. Moreover, an experiment by Schein and Gray (2015) showed that humans too can be perceived as unnerving, suggesting that humanness (and a biologically-human appearance) is not enough to avoid the valley.

Finding the answers to these questions has particular relevance to human-robot interaction and the design of social robots. Despite the superficial nature of a robot's appearance, its appearance nevertheless substantially impacts how people perceive it and whether they are willing to interact with it (e.g., Strait et al., 2015; Mathur and Reichling, 2016). Thus, to achieve effective robot designs (or, at least, avoid ineffective ones), it remains crucial to gain better understanding of the uncanny valley and the variables (both visual and behavioral) that drive it.

1.1. Present Work

Here, we aimed to further examine the uncanny valley as it pertains to human-robot interaction. Our contributions are three-fold: in addition to providing another experimental test of the valley's existence, we investigated what design factors cause a robot to fall into the valley. In particular, we tested two theoretically-motivated factors – *atypicality* and *category ambiguity* – for their effects on perceptions of uncanniness and people's corresponding aversion. Finally, we aimed to address an outstanding shortcoming of the current literature, namely whether people's aversion can be explained by pre-existing negative attitudes toward robots.

Recent reviews of valley literature have pointed to two explanatory mechanisms underlying the effect: *atypicality* and *category ambiguity* (cf. Kätsyri et al., 2015; MacDorman and Chattopadhyay, 2016). *Atypicality* (also called "feature atypicality" and "realism inconsistency") refers to the presence of features unusual for an agent's category. For example, Albert Hubo is an *atypical* robot with its prototypically mechanical body combined with an atypical (highly humanlike) head. Derived from theories of perceptual mismatch, *atypicality* is proposed to underlie uncanniness via violation of expectations about how an agent should look/behave based on its category membership

[1] In contrast to the aforementioned studies (which involved 1-3 robots), both studies referenced here involved 45–80 robots.

(Groom et al., 2009; Saygin et al., 2012). Perceptual mismatch theories thus predict that *any* atypical agent (robot or human) will provoke aversion.

Category ambiguity, on the other hand, refers to a difficulty in determining the category to which an entity belongs (e.g., Burleigh et al., 2013; Yamada et al., 2013). For example, people have difficulty perceiving the Geminoid HI as being a robot because of its very humanlike design (Rosenthal-von der Pütten et al., 2014). Derived from theories of categorical perception, *category ambiguity* is proposed to underlie uncanniness via doubt about what an entity is (Jentsch, 1997). Contrary to the above, categorical perception theories predict that the valley effect is greatest at category boundaries (e.g., the robot-human boundary), with aversion decreasing outwards with increasing distance.

In the present study, we observed people's subjective and behavioral aversion toward 60 distinct robots and humans using the popular picture-viewing methodology used in emotion research (see Vujovic et al., 2013), as adapted for HRI research involving social signals (Strait et al., 2015; see **Figure 2**). Participants were presented with the 60 photographs sequentially and for 12 s each. For each viewing, participants had the option to press a button if they wished to terminate the encounter early (thereby engaging in behavioral avoidance). In total, we collected participants' subjective ratings of the agents' eeriness, the frequency at which they terminated encounters with the various agents, and their reasons for terminating.

Per Mori's uncanny valley theory, we hypothesized that people would be averse to *highly humanlike* – but not prototypic – agents (**H1: Valley Hypothesis**). Specifically, relative to people of prototypically human appearances and robots of low human similarity, we expected that the appearance of highly humanlike agents would be so discomforting (as evidenced by higher ratings of eeriness; H1a) that people would avoid their encounters more frequently (H1b), and that they would report doing so due to being unnerved (H1c).

In confirming the existence of a valley in the design space included, we looked at the governing mechanisms underlying uncanniness (when, why, and how an agent falls into the valley) with two further predictions following from the literature. Specifically, we hypothesized that people would be more averse to *atypical* agents than prototypic agents (**M1: Feature Atypicality**). We also hypothesized that people would be more averse to *ambiguous* agents than prototypic agents (**M2: Category Ambiguity**). In addition to the above predictions, we explored how the two proposed mechanisms – *atypicality* vs. *ambiguity*—interact with the agents' actual category membership (whether the agent in question is a robot or a person) in provoking aversion, and further, whether people's aversive responding can be explained by pre-existing negative attitudes toward robots.

2. MATERIALS AND METHODS

Based on Mori's valley hypothesis, we expected that highly humanlike (but not prototypic) agents may be so eerie (H1a) that people avoid their encounters because due to being unnerved (H1b–c). We further predicted, based on perceptually-oriented theories of categorization and processing, that salient atypicalities (M1) and/or high category ambiguity (M2) might underlie such discomfort.

2.1. Design

To test our predictions, we conducted a within-subjects experiment in which we presented participants with 60 distinct agents which spanned two ontological **categories** (*robot*, *person*) and were of appearances that varied semi-hierarchically across two overlapping dimensions – **human similarity** (three levels: *low*, *high*, and *prototypic*) and **typicality** (three levels: *prototypic*, *atypical*, and *ambiguous*)[2]. In total, the study involved six agent conditions (with 10 agents per condition):

- 10 agents of *low* human similarity (i.e., prototypic robots such as the mechanomorphic REEM-C);
- 40 agents of *high* (but not prototypically human) human similarity:
  - 10 robots with *atypical* features (e.g., Albert Hubo),
  - 10 robots of *ambiguous* category membership (e.g., the Geminoid DK),
  - 10 people with *atypical* features (e.g., persons with bionic prostheses), and
  - 10 people of *ambiguous* category membership (persons wearing black, full-sclera contacts);
- 10 agents of *prototypic* human similarity (i.e., people of typical appearances).

Table 1 shows exemplars of each agent condition, as well as the semi-hierarchical mapping between the three manipulations (the agent's approximate human similarity and their typicality relative to their respective category membership).

2.1.1. Valley Hypothesis

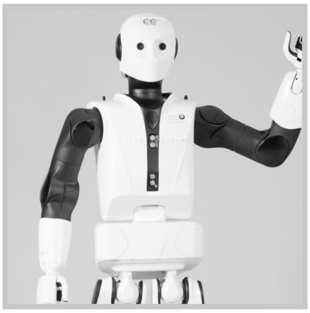
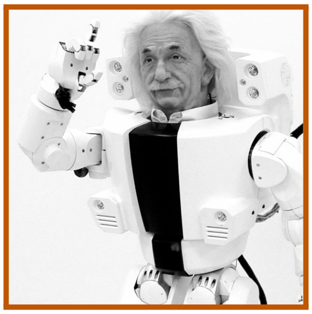
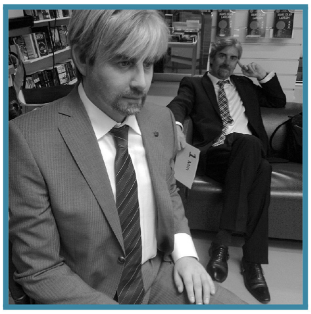
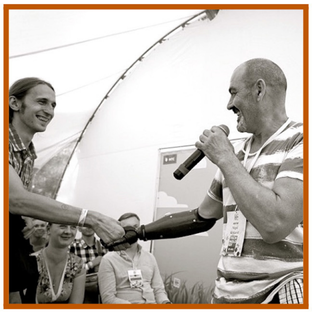
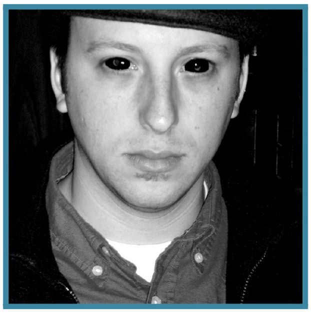
The manipulation of the agents' human similarity was used to test whether or not there exists an uncanny valley within the current design space of humanoids and range of human appearances (H1). Note that, in testing the valley hypothesis, we collapse across the four sets of robots and people of *atypic* and *ambiguous* designations as their normalized ratings of human similarity constitute *high*—but not prototypic—human similarity. That is, they are rated as significantly more humanlike than mechanomorphic humanoids and significantly less humanlike than people of prototypic appearances.

2.1.2. Mechanisms

Via the typicality manipulation, we further tested whether two mechanisms (M1: feature atypicality; M2: category ambiguity) drive the valley's effects by drilling down within the set of highly humanlike agents. Specifically, via the explicit inclusion and clustering of highly humanlike agents by those with appearances atypic for their respective category and those of

[2] Due to the current design space of humanoid robotics and range of human appearances (e.g., there do no exist stimuli depicting people of "low human similarity"), the study did not involve a factorial design.

TABLE 1 | Exemplars of the six agent conditions, with: the agent's **category** membership reflected across the dotted y-axis (top row: *robot*; bottom: *person*); the **human similarity** manipulation shown along the x-axis (increasing left to right: from *low* similarity to *high*—inclusive of atypic and ambiguous typicality levels – to *prototypically human*), and the corresponding **typicality** levels indicated via color-coding (gray: *prototypic* for a given ontology; orange: *atypical*; and blue: *ambiguous*). **Robots** (top; from left to right): a prototypic robot (PAL ROBOTICS' REEM-C); a robot with a salient atypicality (KAIST's Albert Hubo); and a robot of ambiguous ontology (the Geminoid DK; shown, for comparison, in front of Henrik Scharfe – the person after which it was modeled). **People**: a person of prototypically human human similarity; a person with a prosthetic arm; and a person of "*ambiguous*" humanness (a person wearing black sclera contacts; face enlarged for emphasis).

low **(prototypic robot)**	**high** **(inclusive of both atypical and ambiguous agents)**		**prototypic** **(prototypic person)**
			
			

Attribution (from top left to bottom right): shown are adaptations of photographs by JosepPAL, Dayofid, and Eirik Newth; SalganikEA and Matthew Batchelder. Original photos (https://goo.gl/38yUr1, https://goo.gl/00o07k, https://goo.gl/T7Ym4O; https://goo.gl/YfndfO, https://goo.gl/UB62Ac) available under Creative Commons Attribution-Share Alike 3.0 Unported, Attribution 2.5 Generic, or Attribution-NonCommercial-ShareAlike 2.0 Generic licenses[3].

ambiguous category membership, we contrasted the role of each of the two mechanisms (against prototypicality) in eliciting discomfort. Here, we additionally included the manipulation of ontological category (robot vs. person), as both the feature atypicality and category ambiguity hypotheses require that the valley effect be evident regardless of the agent's actual category membership. Thus, in testing the two hypothesis, the three typicality levels (*prototypic*, *atypical*, and *ambiguous*) are robot-human inclusive (e.g., *prototypic* included mechanomorphic humanoids and people of prototypically human appearances).

2.2. Materials

To construct a final set of high quality and relatively comparable photos, the stimuli used in this experiment were selected from an initial superset of 120 photos. The 120 photos were obtained from various academic and online sources based on strict inclusion/exclusion criteria and pretested for their fit within the six intended agent categories to reduce within-category variability.

2.2.1. Set Construction

We constructed our initial stimulus set via a systematic search using stringent inclusion criteria based on that developed by Mathur and Reichling (2016). The purpose of the criteria was to reduce any researcher bias that may be present in image selection (e.g., agent expression, pose, etc.). The criteria were as follows:

- Visibility: the agent's face/torso and eyes are fully visible (shown from top of head to waist; face is shown in frontal to 3/4 aspect).
- Embodiment: the agent is capable of interacting socially with humans (e.g., if a robot, the agent has been built and is capable of physical movement).
- Affect: the agent is expressionless/affect-neutral.
- Familiarity: the agent is not a replica of a well-known character or a famous person (e.g., Albert Hubo).
- Image characteristics: the resolution of the image is sufficient to yield a final cropped image of 6x6" with a resolution of 100 DPI.

We performed ten Google image searches on a single day using the following search phrases: "humanoid robot," "humanlike robot," "robot with humanlike face," "android robot," "highly

[3] https://goo.gl/eTRg2B

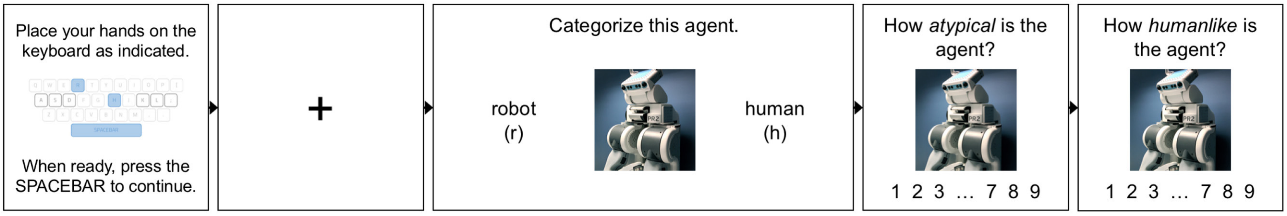

Place your hands on the keyboard as indicated.
When ready, press the SPACEBAR to continue.
+
Categorize this agent.
robot (r)
human (h)
How *atypical* is the agent?
1 2 3 … 7 8 9
How *humanlike* is the agent?
1 2 3 … 7 8 9

FIGURE 1 | Structure of Pretesting Trials/Manipulation Checks. Each trial began with a prompt to the participant to place their hands on the keyboard as shown (with the left and right index fingers on the "r" and "h" keys respectively). When the participant was ready to continue, they pressed the spacebar to start the categorization task. After a response was entered for the categorization, participants completed two prompts for explicit ratings of the agent's atypicality[5] and human similarity. *Pretesting only*: each trial was preceded by a 1 s fixation point and followed by a 2 s rest period.

humanlike robot," "robot that looks human"; "black sclera contacts," "people wearing sclera contacts," "people with bionic prostheses," "person candid photograph." In collating a set of 20 *atypical* humanoids, we intentionally searched for humanoid robots with a salient mismatch in the realism of their head/torso due to greater availability of robots with this particular design. As the closest human analog (in appearance) to the set of atypical humanoids (robots with features that are atypical in terms of frequency of appearance within the humanoid design space), we specifically searched for people with a bionic prosthetic. To collate an analogous set of 20 people of "ambiguous" ontology (i.e., questionable membership in the *person* category), we intentionally searched for people wearing black, full-sclera contacts as it is a visual modification often used in media to convey different category membership (see for example: the Supernatural TV series, 2005–) and prior literature suggests that people perceive such stimuli as uncanny (Schein and Gray, 2015).

When a search returned multiple images of a particular agent, we included only the first image encountered. For each of the intended agent categories, we included the first 20 photographs satisfying inclusion criteria and depicting distinct agents. However, we note that our resulting set of *ambiguous* robots was comprised of robots that were predominately female (15 of 20) and Asian (13 of 20) in appearance.[4] For comparability between conditions, we thus adjusted the composition of our human stimuli to reflect similar demographics. Specifically, we manually searched for replacements (per the above criteria) for the initially-selected images to adjust the gender and racial composition of the three sets of human stimuli.

2.2.2. Pretesting

To confirm that perception of the agents was as expected (e.g., atypical agents rated as high in atypicality, etc.), we first pretested these 120 photographs (20 agents per each of the six intended design conditions) with 30 participants (recruited from Tufts university and granted course credit in exchange for their participation). Participants were shown the 120 photographs sequentially and in an order randomized by participant. For each image, we measured the agent's "category ambiguity" (indexed by participants' accuracy in a categorization task and their latency to respond), atypicality, and human similarity (see **Figure 1**). Then, to concentrate atypicality within the set of atypical agents and category ambiguity within the set of ambiguous agents, we reduced the pretested set of 120 photographs down to 60 (with 10 instances per agent category) by selecting for category-ambiguous agents with lowest atypicality and atypical agents with lowest category ambiguity.

2.2.3. Manipulation Checks

To confirm that this final set of 60 images reflected our design assumptions (that agents labeled as atypical were perceived as most atypical, agents labeled as ambiguous were most ambiguous, etc.), analyses of variance (ANOVA) were conducted on the dependent variables indexing ambiguity (categorization error rate, response time) and atypicality with *typicality* as the independent variable. Each ANOVA revealing significant effects was followed by t-tests examining the planned, pairwise contrasts (*atypical*, *ambiguous* vs. *prototypic*)[6].

ANOVAs on categorization error rate and response time confirmed a significant main effect of *typicality* on perceptions of agent ambiguity [F_{error} (1.15, 33.42) = 86.94, $p < 0.01$, $\eta_p^2 = 0.75$; F_{RT} (2, 58) = 10.89, $p < 0.01$, $\eta_p^2 = 0.27$], in which *ambiguous* agents elicited the greatest difficulty ($p < 0.01$) in categorization [$M_{error} = 0.32$, $SD = 0.18$; M_{RT} (2, 58) = 1.92s, $SD = 1.08$ s] relative to both agents with prototypic appearances ($M_{error} = 0.01$, $SD = 0.03$; $M_{RT} = 1.11$ s, $SD = 0.40$ s) and those categorized as atypical ($M_{error} = 0.04$, $SD = 0.06$; $M_{RT} = 1.78$ s, $SD = 1.23$ s). Similarly, an ANOVA on atypicality

[4] The current design space of highly humanlike robots is largely comprised of robots that are gendered/racialized (designed with physical features that convey gender/race) and skewed toward appearances that are female and Asian

[5] The atypicality prompt in **Figure 1** is shortened from: "How *mismatched* are this agent's features relative to its overall appearance?" due to space constraints.

[6] All analyses were run in R (Version 3.3.1), with statistical significance defined as $\alpha = 0.05$. For each ANOVA, the assumption of equal variance was confirmed using Mauchly's test of sphericity. In cases of violation, the reported degrees of freedom and corresponding p-value reflect a Greenhouse-Geisser adjustment as per Girden (1992). For the pairwise contrasts, two-tailed (rather than one-tailed) t-tests were used to reveal if/when a contrast went in the direction opposite to that which was predicted and reduce the overall rate of false positive results. Additionally, all pairwise contrasts reflect a Bonferroni correction for multiple comparisons. Lastly, note that while we defined statistical significance at $\alpha = 0.05$, all significant results (including both tests of the hypotheses and Bonferroni-corrected contrasts) have a p-value of $\leq$ .01 except where explicitly stated otherwise.

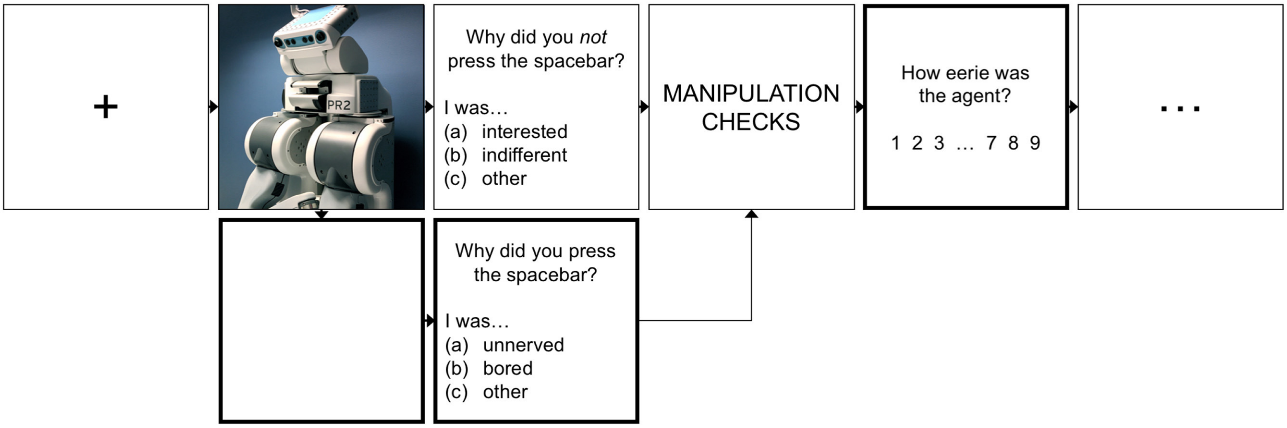


FIGURE 2 | Trial Structure. Each trial began with a 1 s fixation point, followed by a picture viewing of up to 12 s. During the viewing, participants had the option to terminate the encounter early by pressing the spacebar. In doing so, the image was removed from display, leaving a blank screen for the remainder of the 12 s period. After the viewing, participants (1) were prompted for their reasons as to why they did or did not press the spacebar, (2) completed a series of checks to confirm whether the manipulations of agent appearance had the intended effects, and (3) provided an explicit rating of the agent's eeriness.

ratings confirmed a main effect of *typicality* [$F_{(2, 58)} = 276.46$, $p < 0.01$, $\eta_p^2 = 0.91$], in which the set of *atypical* agents received significantly higher ratings of atypicality ($M = 4.59$, $SD = 1.00$) relative to prototypic agents ($M = 1.78$, $SD = 0.60$; $p < 0.01$).

In addition, an ANOVA on ratings of human likeness confirmed a main effect of *human similarity* [$F_{(1.58, 45.80)} = 512.97$, $p < 0.01$, $\eta_p^2 = 0.95$], in which the highly humanlike (but not prototypically human) agents received significantly higher ratings ($M = 7.02, SD = 0.84$) than prototypic robots ($M = 2.66, SD = 1.26$; $p < 0.01$) and significantly lower ratings than prototypic persons ($M = 8.91, SD = 0.22$; $p < 0.01$).

2.3. Experiment

2.3.1. Participants

Seventy-five new participants (participants who took part in pretesting were excluded from participating here) were recruited from Tufts University and the surrounding community (the Greater Boston Area), and received either course credit ($n = 45$) or monetary compensation ($n = 30$) at a rate of $10/h for their participation. Data were unavailable for three participants due to software crashes ($n = 2$) and termination of a session due to failure to follow instructions ($n = 1$). Thus, a total of 72 participants (26 male) with ages ranging from 18 to 49 years ($M = 19.73, SD = 4.00$) were included in our final sample.

2.3.2. Procedure

The final set of 60 photographs were shown using Processing 3.2.1 (©The Processing Foundation) in random order. Each trial began with a 1 s fixation point followed by the image presentation, and ended with a 2 s rest period (see **Figure 2**). During the viewing period, an image was presented for up to 12 s during which time participants had the option to press a button (the spacebar) to remove the image from the screen. If the participant did not press the spacebar, the image was shown for the full viewing duration (12 s). Otherwise, the image was removed as soon as participants pressed the spacebar, leaving a blank screen for the remainder of the viewing period[7]. After the viewing period, participants were prompted for their rationale as to why they terminated or did not terminate the encounter, followed by several manipulation checks (see **Figure 1**) and prompt for participants' explicit perceptions of the agent's eeriness. At the end of the picture-viewing protocol, participants were given a brief questionnaire to assess their attitudes toward robots.

2.3.3. Measures

To index participants' aversion, we employed three primary measures derived from those developed in Strait et al. (2015):

- **Eeriness**: participants' subjective ratings of the agents' appearances. As we used a fully within-subjects design, ratings were averaged (by participant) across trials within each of the six agent categories.
- **Termination frequency**: the frequency at which participants elected to end their encounters with the various agents (computed within each of the six agent conditions as the proportion of trials in which participants pressed the spacebar to terminate the trial).
- **Terminations due to discomfort**: the proportion of terminated trials in which participants reported terminating due to being unnerved by the shown agent.

Finally, to index participants' attitudes toward robots, we used the Negative Attitudes Toward Robots Scale (NARS; Nomura et al., 2006). The scale is comprised of 14 questionnaire items and produces an overall NARS score (Cronbach's $\alpha = 0.87$), as well as three subscores: negative attitude toward situations concerning interaction with robots (6 items; $\alpha = 0.78$), negative attitude toward the social influence of robots (5 items; $\alpha = 0.70$),

[7] Replacement with a blank screen was done to ensure that the button press could not be used as a strategy to finish the experiment more quickly.

TABLE 2 | Main effects of the *human similarity* manipulation (within-subjects; three levels: low, high[10], and prototypic) and corresponding descriptive statistics (means and standard deviation for each of the three levels).

	n	DF_n	DF_d	F	p	η_p^2	Low	High	Prototypic
MANIPULATION CHECKS									
Human Similarity Rating	72	1.55	110.26	1188.44	< 0.01	0.94	2.99 (1.16)	7.06 (0.81)	8.83 (0.33)
HYPOTHESIS TESTING									
Eeriness Rating	72	1.73	122.68	250.92	< 0.01	0.78	2.50 (1.37)	5.05 (1.13)	1.26 (0.69)
Termination Frequency	72	2	142	250.92	< 0.01	0.09	0.30 (0.37)	0.38 (0.35)	0.37 (0.38)
Rationale for Terminating:									
Unnerved	39	2	76	39.84	< 0.01	0.51	0.11 (0.29)	0.48 (0.33)	0.03 (0.11)
Bored	39	2	76	33.44	< 0.01	0.47	0.74 (0.39)	0.36 (0.33)	0.83 (0.32)
Other	39	2	76	0.18	0.83	0.00	0.14 (0.30)	0.16 (0.27)	0.14 (0.28)
Rationale for Viewing:									
Interested	58	2	114	25.38	< 0.01	0.31	0.52 (0.33)	0.62 (0.29)	0.32 (0.36)
Indifferent	58	1.79	102.12	30.01	< 0.01	0.34	0.45 (0.32)	0.32 (0.29)	0.64 (0.36)
Other	58	–	–	–	–	–	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)

Note that inferential statistics are unavailable for the "other" response rationale (for viewing), as the variance of the data was zero.

and negative attitude toward emotions in interacting with robots (3 items; $\alpha = 0.77$).

3. RESULTS

3.1. Valley Hypothesis (H1)

Based on Mori's uncanny valley theory, we hypothesized that—relative to robots of low human similarity and persons of prototypically human appearances—highly humanlike (but not prototypic) agents can be so discomforting that people would be averse to interacting with them. To test our hypotheses, a repeated-measures ANOVA was conducted on each of the three dependent variables and relevant manipulation check.[6, 8] All statistics (descriptive and inferential) are reported in **Table 2**, with effect sizes[9] for significant contrasts reported in the discussion below.

3.1.1. Manipulation Check

We assumed that the three similarity designations—robots of low human similarity, highly humanlike agents, and people of prototypically human appearances—would be perceived as having monotonically increasing human similarity (from *low* to *prototypic*). To first confirm this assumption, we conducted an ANOVA on participants' ratings of the agents' human similarity with *human similarity* (*low*, *high*, *prototypic*) as the independent variable. As expected, the results showed a main effect of similarity ($\eta_p^2 = 0.94$). All pairwise contrasts were significant, with ratings increasing from robots designated as low in human similarity to highly humanlike agents (Cohen's $d_z = 3.50$) to people of prototypic similarity ($d_z = 2.50$).

3.1.2. Hypothesis Testing

We expected that, relative to both robots of low human similarity and persons of prototypically human appearances, participants would be averse to highly humanlike agents, evidenced by higher ratings of eeriness (**H1a**), more frequent termination of their encounters (**H1b**), and a greater proportion of terminated encounters terminated due to being unnerved (**H1c**). As expected, there was a main effect of *human similarity* on all three indices of aversion—eeriness ratings ($\eta_p^2 = 0.78$), termination frequency ($\eta_p^2 = 0.09$), and the frequency of terminations due to being unnerved ($\eta_p^2 = 0.51$) (see **Figure 3**).

Consistent with the valley theory, participants rated highly humanlike agents as eerier than both robots of low human similarity ($d_z = 1.47$) and prototypic persons ($d_z = 2.93$). In addition, they terminated encounters with highly humanlike agents more frequently than those with robots of low human similarity ($d_z = 0.45$). Lastly, although there was no significant difference in participants' termination frequency between encounters with highly humanlike agents vs. prototypic persons, significant differences did manifest in their rationale for terminating. Specifically, participants reported terminating encounters due to being unnerved more frequently in response to highly humanlike agents than they did in response to robots of less human similarity ($d_z = 1.03$) and prototypic persons ($d_z = 1.29$). For comparison, when participants terminated encounters with

[8] Only participants who provided data in all conditions relevant to each particular test were included (e.g., only participants who terminated at least one encounter with each of the six agent types were included in analysis of termination frequencies). Thus, due to listwise deletion of participants with missing data, the number of observations (and consequently the degrees of freedom) vary across tests. In addition, while the proportion of encounters terminated due to being unnerved is the only rationale item central to our hypotheses, all data for participants' rationale is included (including rationale for electing to view photographs in full) for completeness of reporting.

[9] Cohen's d_z, corrected for the within-subjects design per Morris and DeShon (2002), is reported for all significant contrasts.

[10] In testing the valley hypothesis (H1), the set of agents of *high* human similarity (40) includes both robots and people of the *atypical* and *ambiguous* designations. Note also that the set of agents of *prototypic* human similarity (10) refers only to the set of people of prototypically human appearances.

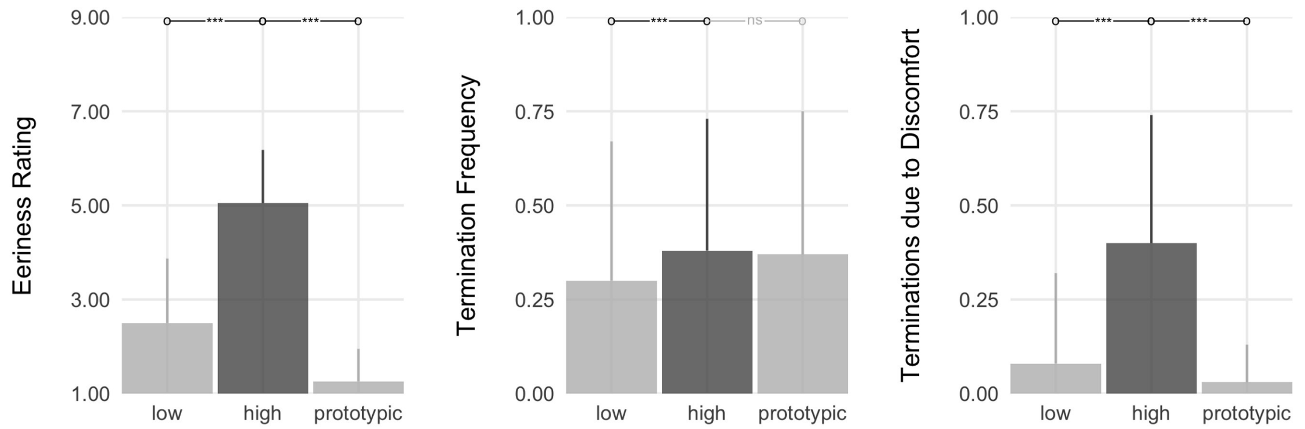


FIGURE 3 | Test of the valley hypothesis (H1). Shown are the three indices of aversion by *human similarity* level: *low*, *high* (inclusive of both atypical and ambiguous robots and people), and *prototypic*. The bars show the planned contrasts, with asterisks denoting the significance level (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$; or *ns*: nonsignificant).

prototypic persons or with robots of low human similarity, their rationale for doing so stemmed largely from boredom (see **Table 2**).

Taken together, the results show strong support of Mori's valley hypothesis. Specifically, relative to robots of low human similarity and persons of prototypically human appearances, participants exhibited greater aversion (as evidenced by their eeriness ratings and avoidance rationale) toward highly humanlike—but not prototypically human—agents.

3.2. Mechanisms Underlying Uncanniness (M1–M2)

In identifying an uncanny valley in the current design space of humanoid robots and range of human appearances, we moved to testing the mechanisms underlying uncanniness. Here, we had hypothesized that both atypicality (M1: Feature Atypicality) and ambiguity (M2: Category Ambiguity) drive people's aversion toward highly humanlike (but not prototypically human) agents. Specifically, to understand *when/why/how* certain agents fall into the uncanny valley, we investigated two visual variables (atypicality, ambiguity) for their impact on people's perceptions of highly humanlike agents relative to agents of prototypic appearances.

In testing these hypotheses and corresponding assumptions, we ran 2 × 3 within-subjects ANOVAs with the IVs—*category* (two levels: *robot* and *person*) and *typicality* (three levels: *prototypic*, *atypical*, and *ambiguous*)—on each of the three indices of aversion.[6,8] Note that, while we included *category* as an IV (due to its inclusion in the experimental design), both of the two mechanisms require that the valley effect is evident regardless of the agent's category membership. Thus, the testing of the two mechanisms relies on the main effect of *typicality*, not the *category* × *typicality* (which we explore later). To test the two mechanisms, we examined two *a priori* contrasts of interest as follows: *prototypic* vs. *atypical* (M1) and *prototypic* vs. *ambiguous* (M2). All statistics (descriptive and inferential) are reported in **Table 3**, with effect sizes[9] for significant contrasts reported in the discussion below.

3.2.1. Manipulation Checks

Here we made two additional assumptions in our experimental design. First, we expected that the agents categorized as *atypical* would be perceived as more atypical than the other typicality conditions (prototypic, ambiguous). As expected, an ANOVA on atypicality ratings showed a main effect of *typicality* condition ($\eta_p^2 = 0.86$). Contrary to our expectations, however, the post hoc contrasts showed *ambiguous* agents to prompt the highest ratings of atypicality, followed by atypical agents ($d_z = -1.23$), and lastly, prototypic agents ($d_z = 2.79$). A significant interaction with *ontological category* ($\eta_p^2 = 0.79$) confirmed our assumption with respect to the robotic agents. Specifically, atypical robots elicited the highest ratings of atypicality relative to both prototypic ($d_z = 2.77$) and ambiguous robots ($d_z = 0.91$). Whereas, amongst human agents, persons of ambiguous category membership elicited higher ratings than persons with atypical features ($d_z = -2.25$)[11]. Nevertheless, participants rated persons with atypical features as more atypical than prototypic persons ($d_z = 0.97$).

Second, we assumed the *ambiguous* agents (agents proximate to a nonhuman–human category boundary) would elicit difficulty in deciding their category membership (robot or person) on a categorization task. Furthermore, we assumed categorization difficulty would be reflected by participants' error in categorizing and latency to respond (RT). As expected, there was a main effect of *typicality* on both categorization error ($\eta_p^2 = 0.61$) and RT ($\eta_p^2 = 0.36$). Specifically, ambiguous agents

[11] We speculate that this asymmetry in perception of atypicality stems from two potential sources. First, participants may have been uncomfortable in making explicit atypicality ratings of persons with prostheses, and thus the ratings may not be representative of participants' actual perception. Second, participants exposure to persons wearing black, full-sclera contacts is likely lower than that of exposure to persons with prostheses. Thus, by definition, black eyes are likely perceived as more atypical of humans than prostheses.

TABLE 3 | Main effects of the *typicality* manipulation (within-subjects; three levels: prototypic[12], atypical, ambiguous) and corresponding descriptive statistics (means and standard deviation for each of the three levels).

	n	DF_n	DF_d	F	p	η_p^2	Prototypic	Atypical	Ambiguous
MANIPULATION CHECKS									
Atypicality Rating	72	1.56	110.70	419.60	< 0.01	0.86	1.80 (.68)	4.25 (1.13)	5.75 (1.52)
Categorization Error (%)	72	1.09	77.64	111.68	< 0.01	0.61	0.00 (.01)	0.02 (0.04)	0.20 (0.15)
RT (s)	72	1.23	87.60	39.74	< 0.01	0.36	0.76 (0.51)	0.93 (0.65)	1.29 (0.98)
HYPOTHESIS TESTING									
Eeriness Rating	72	2	142	216.52	< 0.01	0.75	1.88 (0.85)	5.30 (1.33)	4.81 (1.34)
Termination Frequency	72	1.84	130.89	4.57	0.01	0.06	0.33 (0.36)	0.36 (0.35)	0.40 (0.36)
Rationale for Terminating:									
Unnerved	23	1.51	33.25	21.76	< 0.01	0.50	0.04 (0.12)	0.37 (0.34)	0.44 (0.37)
Bored	23	1.02	22.37	21.85	< 0.01	0.50	0.78 (0.33)	0.45 (0.37)	0.38 (0.38)
Other	23	2	44	1.52	0.22	0.06	0.17 (0.29)	0.21 (0.31)	0.17 (0.28)
Rationale for Viewing:									
Interested	52	1.52	77.30	29.61	< 0.01	0.37	0.42 (0.28)	0.67 (0.28)	0.55 (0.32)
Indifferent	52	1.68	85.45	38.37	< 0.01	0.43	0.55 (0.29)	0.28 (0.29)	0.39 (0.32)
Other	52	–	–	–	–	–	0.00 (.00)	0.00 (0.00)	0.00 (0.00)

Inferential statistics are unavailable for the "other" response rationale (for viewing), as the variance of the data was again zero.

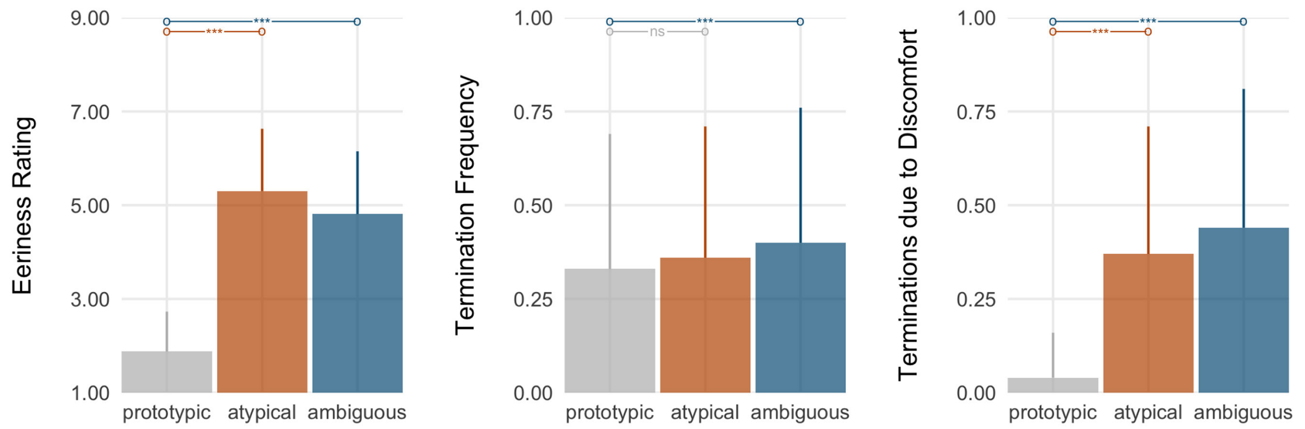


FIGURE 4 | Test of the underlying mechanisms (M1–M2). Shown are the main effects of *typicality* (prototypic, atypical, and ambiguous) on the three indices of aversion. Bars show the planned contrasts, with asterisks denoting significance.

elicited greater categorization error and longer response times in categorizing relative to both prototypic (d_{error} = 1.29; d_{RT} = 0.81) and atypical agents ($d_{error} = 1.22$; $d_{RT} = 0.65$).

3.2.2. Hypothesis Testing

We had hypothesized that, relative to agents of prototypic appearances, agents with *feature atypicality* (**M1**) and *category ambiguity* (**M2**) would elicit aversion in participants. Consistent with our predictions, the results show a main effect of *typicality* on the three indices of aversion: eeriness ratings (η_p^2 = 0.75), termination frequency (η_p^2 = 0.06), and the frequency of terminations due to being unnerved (η_p^2 = 0.50) (see **Figure 4**).

Specifically, the planned contrasts show that participants rated both atypical and ambiguous agents as significantly eerier than prototypic agents ($d_{atypical}$ = 2.07; $d_{ambiguous}$ = 2.02). In addition, when participants terminated encounters with atypical and ambiguous agents, they did so more frequently due to being unnerved ($d_{atypical}$ = 1.01; $d_{ambiguous}$ = 1.13) than they did in response to prototypic agents. For comparison, when participants terminated encounters with prototypic agents, their rationale for doing so stemmed largely from boredom (see **Table 3**).

However, only agents with ambiguous appearances prompted more frequent avoidance. Specifically, participants terminated encounters with ambiguous agents more frequently than they did with prototypic agents ($d_z = 0.31$).

[12] In testing the underlying mechanisms (M1–M2), the set of agents of *prototypic* typicality (20) includes both mechanomorphic robots and prototypic persons. Note also that the set of agents of *atypical* (20) and *ambiguous* (20) agents is inclusive of both robots and people.

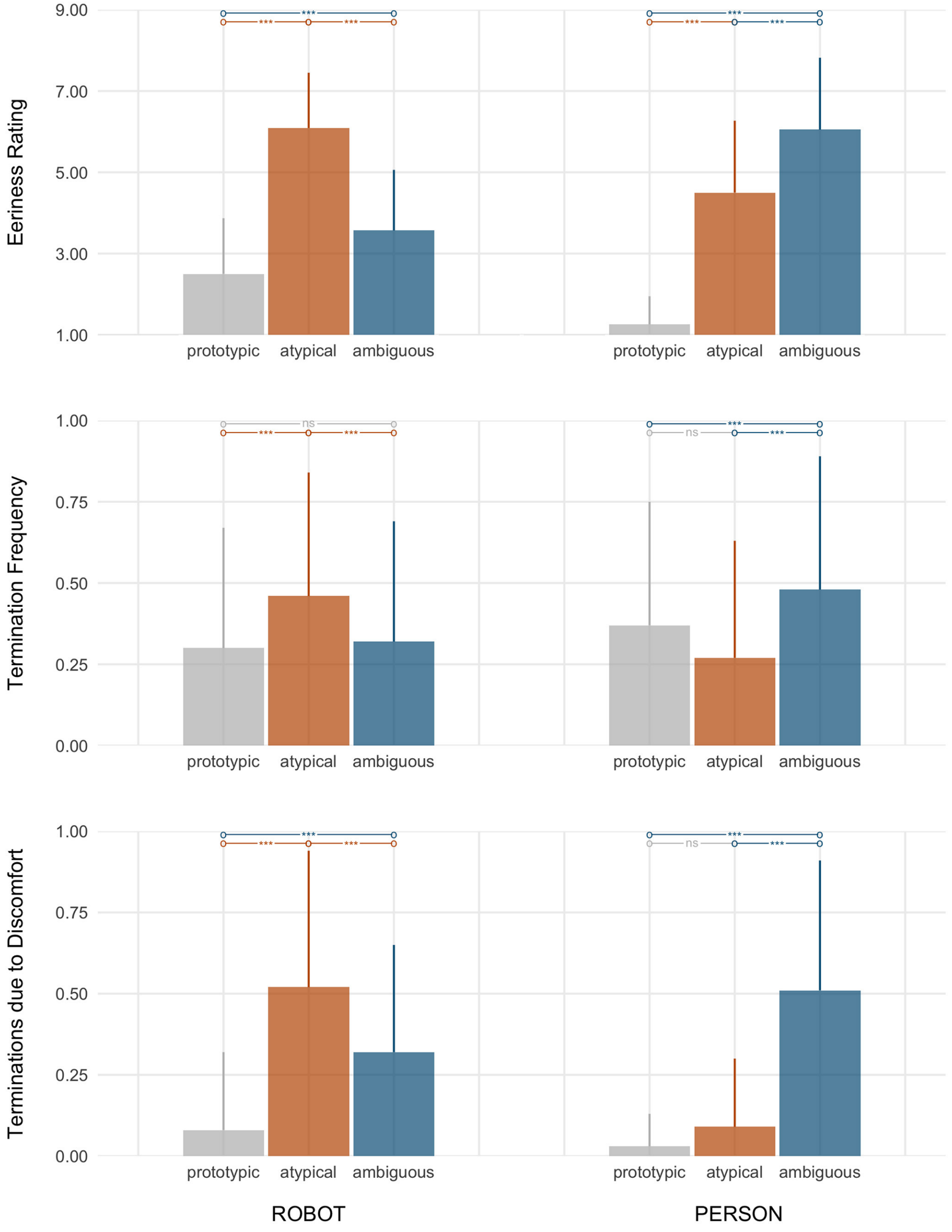


FIGURE 5 | Participants' aversive responding (top: eeriness; middle: termination frequency; bottom: proportion of terminations due to being unnerved) by *typicality* (prototypic, atypical, and ambiguous) and within agent *category* (left grouping: robot; right grouping: person).

In sum, the results here show support for both theoretical accounts (feature atypicality and category ambiguity). Specifically, consistent with **M1** (*feature atypicality*), participants rated atypical agents as eerier than prototypic agents and avoided them more frequently due to being unnerved. Similarly, consistent with **M2** (*category ambiguity*), participants rated ambiguous agents as eerier than prototypic agents, avoided them more frequently, specifically due to being unnerved.

3.2.3. Secondary Analyses

While we found support for both theoretical mechanisms, we also observed a significant interaction between the agents' ontological *category* and *typicality* on eeriness ratings ($\eta_p^2 = 0.71$), termination frequency ($\eta_p^2 = 0.34$), and the frequency of terminations due to being unnerved ($\eta_p^2 = 0.59$), thus indicating that the effect of typicality manifests differently depending on whether the agent in question is a robot or a human (see **Figure 5**). Hence, we proceeded to explore the

TABLE 4 | Interaction between the *typicality* × *ontology* manipulations, as well as the corresponding descriptive statistics (means and standard deviation) for each of the three typicality conditions (prototypic, atypical, and ambiguous) by category membership (top: robot; bottom: person).

	n	DF_n	DF_d	*F*	*p*	η_p^2	Prototypic	Atypical	Ambiguous
MANIPULATION CHECKS									
Atypicality Rating	72	1.80	127.53	260.84	< 0.01	0.79	2.37 (1.22)	6.30 (1.66)	5.01 (1.72)
							1.22 (0.34)	2.20 (1.05)	6.49 (1.80)
Categorization Error (%)	72	1.04	74.11	4.71	0.03	0.06	0.00 (0.01)	0.03 (0.06)	0.26 (0.27)
							0.00 (0.02)	0.01 (0.04)	0.14 (0.27)
RT (s)	72	1.35	95.63	1.42	0.24	0.02	0.76 (0.53)	0.90 (0.63)	1.35 (1.07)
							0.77 (0.52)	0.95 (0.72)	1.24 (1.09)
HYPOTHESIS TESTING									
Eeriness Rating	72	2	142	177.61	< 0.01	0.71	2.50 (1.37)	6.09 (1.36)	3.57 (1.49)
							1.26 (.69)	4.50 (1.77)	6.05 (1.77)
Termination Frequency	72	1.68	119.18	37.06	< 0.01	0.34	0.30 (0.37)	0.46 (0.38)	0.32 (0.37)
							0.37 (0.38)	0.27 (0.36)	0.48 (0.41)
Rationale for Terminating:									
Unnerved	23	2	44	32.04	< 0.01	0.59	0.08 (0.24)	0.52 (0.42)	0.32 (0.33)
							0.03 (0.10)	0.09 (0.21)	0.51 (0.40)
Bored	23	2	44	14.39	< 0.01	0.40	0.75 (0.38)	0.31 (0.40)	0.44 (0.38)
							0.81 (0.32)	0.66 (0.39)	0.37 (0.40)
Other	23	2	44	5.41	< 0.01	0.20	0.17 (0.31)	0.18 (0.31)	0.24 (0.31)
							0.16 (0.28)	0.25 (0.36)	0.12 (0.28)
Rationale for Viewing:									
Interested	52	2	102	12.05	< 0.01	0.19	0.50 (0.33)	0.61 (0.35)	0.65 (0.32)
							0.31 (0.35)	0.69 (0.29)	0.42 (0.41)
Indifferent	52	2	102	7.83	< 0.01	0.13	0.47 (0.32)	0.30 (0.35)	0.29 (0.31)
							0.64 (0.35)	0.28 (0.29)	0.49 (0.40)
Other	52	–	–	–	–	–	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
							0.00 (0.00)	0.00 (0.00)	0.00 (0.00)

pairwise contrasts between typicality levels and agent category.

Note that the exploration, however, was limited to *within* the respective agent category. This follows from the theoretical motivations for investigating feature atypicality (M1) and category ambiguity (M2) for their role in the valley effect, which rely on contrasts between agents of prototypic appearances relative to those of atypic and ambiguous appearances. Here, we also explored the contrast between agents of atypic vs. ambiguous appearances toward understanding whether one vs. the other more strongly provokes discomfort.

3.2.3.1. Responding toward robots

Across all three measures of interest (eeriness ratings, termination frequency, and terminations due to being unnerved), the within-category pairwise contrasts suggest that *atypicality* drove participants' aversion toward robots (see **Figure 5**, top).

Prototypic robots, as expected, were rated as the least eerie of all robot stimuli. In addition, participants terminated their encounters less frequently, and when they did so, it was rarely due to being unnerved (see **Table 4**). On the other end, atypical robots—relative to *both* prototypic and ambiguous robots—were rated as most eerie ($d_z = 1.94$; $d_z = 1.48$). Participants also terminated encounters with atypical robots at the highest frequencies ($d_z = 0.59$; $d_z = 0.63$), and did so most frequently due to being unnerved ($d_z = 1.07$; $d_z = 0.90$). Participants did also exhibit aversion to interacting with ambiguous robots (though less so than their aversive responding toward the set of atypical robots). Specifically, participants rated ambiguous robots as more eerie ($d_z = 0.63$) and terminated their encounters more frequently due to being unnerved ($d_z = 0.73$) relative to prototypic robots. Surprisingly, however, participants were not any more avoidant (evidenced by the frequency at which participants' terminated their encounters) of ambiguous robots than they were of prototypic robots.

3.2.3.2. Responding toward people

Similar to prototypic robots, persons of prototypic appearances were rated as the least eerie of all persons depicted. Furthermore, though participants terminated approximately a third of their encounters with prototypic persons, when they did so, it was again rarely due to being unnerved (see **Table 4**). In contrast, however, to participant responding toward non-prototypic robots (in which atypicality provoked the greatest aversion), category ambiguity appeared to drive participants' aversion toward the human stimuli (see **Figure 5**, bottom). Specifically, participants rated persons of ambiguous category membership

TABLE 5 | Correlation matrix between the NARS scales (overall score and by subscales: negative attitude toward situations concerning *interactions with robots*; negative attitude toward the *social influence of robots*; and negative attitude toward *emotions* in interacting with robots) and participants' responding toward robots (overall and by category – prototypic, atypical, and ambiguous) on the three indices of aversion (eeriness rating, termination frequency, and proportion of terminations terminated due to being unnerved).

	NARS	Interactions with robots	Social influence of Robots	Emotions in interactions
Eeriness Rating	0.14	0.13	0.17	0.04
Prototypic	0.17	0.13	0.16	0.15
Atypical	0.01	−0.05	0.02	0.09
Ambiguous	0.10	0.18	0.15	−0.14
Termination Frequency	0.13	0.12	0.11	0.08
Prototypic	0.16	0.12	0.13	0.16
Atypical	0.08	0.14	0.05	−0.02
Ambiguous	0.11	0.07	0.13	0.09
Unnerved Rationale	0.09	0.20	0.00	−0.06
Prototypic	0.09	0.21	−0.04	−0.01
Atypical	0.04	0.12	−0.04	−0.03
Ambiguous	0.10	0.25	−0.02	−0.07

No significant correlations exist.

as most eerie, relative to both prototypic persons ($d_z = 2.54$) and persons with atypical features ($d_z = -0.84$). They also terminated their encounters with ambiguous persons at the highest frequencies ($d_z = 0.38$; $d_z = -0.70$), and did so most frequently due to being unnerved ($d_z = 1.21$; $d_z = -1.22$). In fact, participants terminated their encounters with persons with atypical features significantly *less* frequently than their encounters with prototypic persons ($d_z = -0.43$) and there was no significant difference between atypical and prototypic persons in the proportion of encounters that they terminated due to being unnerved.

Overall, the secondary analyses reveal that the data reflect greater support for the feature atypicality hypothesis with respect to robotic agents. With respect to human agents, the results are suggestive of greater support for the category ambiguity hypothesis, but uncertainty arising from the study's manipulation checks warrants further investigation of this finding[13].

3.3. Negative Attitudes Toward Robots

Lastly, we explored whether participants' aversive responding toward our stimuli could be explained by pre-existing negative attitudes about robots. Using the Negative Attitudes toward Robots Scale, we tested participants' overall NARS score and scores on the three NARS subscales – negative attitude toward situations concerning interactions with robots (S1), negative attitude toward the social influence of robots (S2), and negative attitude toward emotions in interacting with robots (S3) – for any relationship to their subjective and behavioral responding on the three indices of aversion (eeriness ratings, termination frequency, terminations due to being unnerved). For each aversion index, we computed participants' average response toward all robotic stimuli and by category (prototypic, atypical, ambiguous). In total, we computed 48 correlations (three NARS subscales, plus an overall NARS score; three agent categories, plus an overall response; three aversion indices) using Pearson's *r* test (see **Table 5**). However, no significant relationships were found.

4. DISCUSSION

In the nearly 50 years since Mori's formalization of the uncanny valley (Mori et al., 2012), substantial empirical support has been found for the hypothesis that agents with highly humanlike (but not prototypically human) appearances provoke aversive responding in observers (Kätsyri et al., 2015; Rosenthal-von der Pütten and Krämer, 2015; MacDorman and Chattopadhyay, 2016). Yet, the mechanisms that lead to such feelings of discomfort are largely unknown. Moreover, many still question whether a valley even exists (e.g., Brenton et al., 2005; Hanson et al., 2005; Bartneck et al., 2009; Burleigh et al., 2013; Zlotowski et al., 2013; Złotowski et al., 2015).

Those questioning uncanny valley theory are not wrong: evidence of the valley effect is not in overabundance and the evidence which does exist varies widely in methodologies used (Kätsyri et al., 2015), leaving numerous gaps in the literature. In particular, much of the valley literature is based on (1) stimuli that represent a small subset of a large design space (humanoid robots) and (2) measures that do not capture behavioral implications (relying instead on explicit perception). Thus, the questions of whether the valley effect is robust (i.e., does it generalize to the broader design space) and relevant to human-robot interaction remain.

Two recent studies, using the largest stimulus sets to date (45-80 robots), suggest that the valley effect is both robust and profoundly impactful (Strait et al., 2015; Mathur and Reichling, 2016). Specifically, using picture-based methodologies and behavioral measures to supplement the traditional metrics, the two studies evaluated the impact of a robot's appearance on people's behavior toward a broad range of humanoid robots. In particular, Mathur and Reichling (2016) found that the valley reduces people's trust in highly humanlike robots and we (Strait et al., 2015) found that, not only do people dislike highly humanlike robots, but people actively avoid interacting with them.

As a test of its replicability and extension to this recent work, we adapted the methodologies of Mathur and Reichling (2016) and Strait et al. (2015) for another experimental investigation

[13] We note, however, that the results of the manipulation checks leave us unable to assert this implication definitively. Specifically, participants rated the set of "ambiguous" human stimuli as *more* atypical than those intended to comprise the "atypical" set (comparable to their robotic counterparts). Though we suggested that the atypicality ratings of the atypical human stimuli may have been reduced (due to participants' discomfort at making explicit ratings), without resolution of the manipulation check outcome (divergence from what was expected), we are unable to assume that the asymmetry in responding to human stimuli – that is, the more aversive responding to stimuli of ambiguous humanness – is driven by category ambiguity alone.

of the valley's existence and the design factors that underlie uncanniness. In particular, we tested two theoretically-motivated factors—atypicality and category ambiguity—for their effects on perceptions of uncanniness and resulting avoidant behaviors. Furthermore, we tested an outstanding and common critique of the valley—namely, whether people's aversive responding can be alternatively explained by pre-existing negative attitudes toward robots.

4.1. Summary of Findings

4.1.1. Replication of the Valley Effect (H1)

Consistent with our expectations and previous literature (e.g., MacDorman, 2006; Kätsyri et al., 2015; Strait et al., 2015; Mathur and Reichling, 2016), participants exhibited clear aversion toward agents of high human similarity (highly humanlike robots and humans with non-prototypic appearances), as evidenced by higher ratings of eeriness, more frequent avoidance (early termination of their encounters[14]), and more frequent termination due to being *unnerved*.

While there was not a significant difference in termination frequencies between agents of high similarity and (prototypic) humans, participants' endorsed different rationales for terminating these encounters. Specifically, participants terminated encounters with human stimuli largely due to boredom. By contrast, participants terminated over a third of their encounters with highly humanlike agents due to being unnerved. In particular, it is worth noting that, while the stimuli used in the present study were both innocuous and fleeting, participants nevertheless exhibited significant aversion in their encounters with the highly humanlike agents. That is, the appearances of the highly humanlike agents was discomforting enough that participants often preferred to look at a blank screen, rather than the agents themselves.

Beyond the confirmation of our first hypothesis, the data here fully replicate and thus validate the findings of Strait et al. (2015), demonstrating empirically that the uncanny valley—as a function of human similarity—provokes robust, emotionally-motivated responses to humanlike robots. Our results also lend further support to the findings by Mathur and Reichling (2016) that robots with highly humanlike appearances profoundly (and negatively) impact people's behavioral responding.

4.1.2. Understanding the Uncanny (M1, M2)

As hypothesized and consistent with prior indications (Mitchell et al., 2011; Chattopadhyay and MacDorman, 2016; MacDorman and Chattopadhyay, 2016), *atypicality* provoked aversive responding relative to agents with more typical appearances as evidenced by participants' ratings of the agents' eeriness and the proportion of encounters terminated early due to being unnerved (M1). Support was also found for the hypothesized effect of category *ambiguity* (M2). Specifically, similar to participants' responding toward atypical agents, participants exhibited significant aversion toward agents of ambiguous category membership relative to prototypic agents as evidenced by all three indices of aversion (respective eeriness ratings, termination frequency, and proportion of encounters terminated due to being unnerved).

Exploration of the *typicality* × *category* interaction, however, suggests that the mechanisms have differential impact on responding depending on whether the agent in question is robot or human. Specifically, within the set of robotic stimuli, atypicality provoked the greatest aversion (highest ratings of eeriness, more frequent termination of encounters, and greatest proportion of encounters terminated due to being unnerved). In fact, while the set of ambiguous robots – relative to prototypic robots – prompted higher eeriness ratings and more encounters to be terminated due to being unnerved, they did not elicit greater avoidance (there was no significant difference in the termination frequency from that in response to prototypic robots). Moreover, the ambiguous stimuli were neither the eeriest nor the most discomforting.

In contrast, within the set of human agents, ambiguity provoked the greatest aversion in participants (higher ratings of eeriness, more frequent termination of encounters, and greater proportion of encounters terminated due to being unnerved). Surprisingly, while participants rated atypical stimuli as eerier than persons of prototypically human similarity, participants terminated their encounters with atypical stimuli *less* frequently than with ambiguous and prototypic stimuli.

4.1.3. Negative Attitudes Toward Robots

Exploration of alternative explanations of the above findings did *not* yield support for the suggestion that people's behavior may be explained by pre-existing and negative attitudes toward robots (rather than as the result of an uncanny valley phenomenon). Specifically, no significant relationships were found in 48 correlational tests between participants' aversion and their attitudes toward robots, as indexed by the NARS scales. These findings suggest that positive exposure and/or additional experience with robots is unlikely to affect the occurrence of an uncanny valley effect in humanoid robotics.

4.2. Implications

The present research has three primary theoretical and practical implications.

4.2.1. Methodological Practices

We validated a simple – but effective – laboratory procedure for assessment of people's aversion to social robots. In particular, we adapted a standard procedure from psychology research for the measurement of social signals (particularly, the experience and regulation of negative emotion) in laboratory-based human-robot interactions. The protocol contributes both instrumentation (the measurement of emotion-related social signals in HRI contexts), as well as an effective work-around for a longstanding methodological limitation (accessibility of physical robotic platforms).

Consistency across the multiple measures (of participants' emotion experience and emotionally-motivated responding) and between studies (Strait et al., 2015 and here) demonstrates the reliability of this approach. Whether and how these results

[14] Relative only to robots of low human similarity.

transfer to more ecologically valid contexts (e.g., actual human-robot interaction in the wild) remains to be investigated. However, at a minimum, the protocol provides a means of making systematic probes of the various visual variables present in an agent's appearance.

4.2.2. Uncanny Valley Theory

In providing another experimental test of the uncanny valley hypothesis, our study reveals a robust uncanny valley in the design space of social robots in terms of people's attribution of eeriness to highly humanlike (but not prototypically human) agents. More importantly, it validates the previously suggested (cf. Strait et al., 2015) link between avoidant behavior (early termination of encounters due to being unnerved) and highly humanlike robots. Furthermore, this work extends Mori's initial postulations to consider specific visual aspects that lead to uncanniness. Specifically, the findings point to both atypicality and category ambiguity as driving forces in people's discomfort. The two visual variables (atypicality and category ambiguity) resulted in higher ratings of eeriness, more frequent terminations of encounters, and a greater proportion of terminations terminated due to being unnerved.

Of particular note, our exploratory analyses showed that the atypical robots (which were atypical in the combination of a highly humanlike head atop a mechanomorphic body) and ambiguous humans (which were dehumanized via the use of black, full-sclera contacts, thus occluding the iris) elicited the greatest aversion. These findings are consistent with prior literature evaluating *mind*-related (features related to the head, and in particular, the eyes) atypicalities (Gray and Wegner, 2012; Schein and Gray, 2015; Appel et al., 2016). In addition, the findings support the (relatively common) use of certain visual effects in media and film to instill a sense of unease in observers. Consider for example: Pixar's Babyface (see **Figure 6**) who was an unnerving (albeit eventually sympathetic) character in *Toy Story* (1995); Joshu Kasei, an ultimately terrifying character in *Psycho-Pass* (2012–); and the generally unsettling Ava in *Ex Machina* (2015), amongst others.

4.2.3. Design Considerations

Correspondingly, the findings here provide soft guidelines for the design of future humanoid systems. Participants' strong negative responding—particularly their frequent avoidance of encounters due to discomfort—establishes a shortcoming of the current design space. Moreover, the lack of any predictive relationship between participants' preexisting attitudes toward robots (as indexed by NARS) and their aversive responding suggests that the valley effect is not learned (e.g., via negative portrayals of robots in media) and furthermore, unlikely to dissipate with time/exposure. Thus, there is a clear need to consider alternatives to blanket anthropomorphization.

Broadly, participants' consistent aversion to highly humanlike robots demonstrates a significant cost to designing robots with high human similarity in their appearance. Our results do show evidence of increased interest in the robots corresponding to increased human similarity (consistent with the empirical motivations for increasingly anthropomorphized robot designs; e.g., Riek et al., 2009). However, the increase in interest we've observed pales in magnitude relative to the corresponding increase in avoidance due to discomfort. Moreover, despite significantly increased interest and stimuli that were both innocuous and fleeting, we have consistently observed participants' *avoidance* of encounters with (*photographs* of) highly humanlike robots. In considering that such aversion can be elicited in these settings and in spite of increased interest, we suggest that designing robots with *less* human similarity (at least in their appearance) is a practical and fast solution to the issues underscored by the present findings.

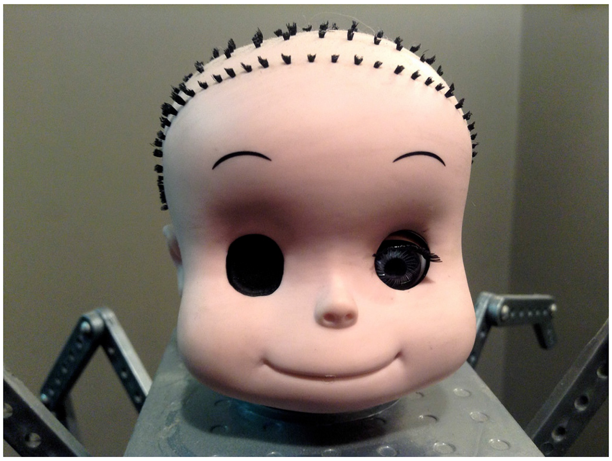


FIGURE 6 | "Babyface" from Pixar's Toy Story (1995). Attribution: photograph (https://goo.gl/GkuBLQ) by Mike Mozart, available under a Creative Commons Attribution 2.0 Generic license1.

That being said, our results do not suggest that efforts to design humanlike robots are futile. Rather, they hint that attention to certain attributes when designing humanlike robots may mitigate aversive responding. Specifically, we note that the set of atypical robots provoked the greatest aversion in participants, more so than the set of "ambiguous" robots (androids). This finding is consistent with prior indications that androids do not necessarily elicit the most negative reactions (e.g., Rosenthal-von der Pütten and Krämer, 2014), and further, suggests that the valley effect can be attenuated, if not overcome. Thus, when designing humanlike robots, our data indicate that greater consistency amongst features may avoid the elicitation of aversion. For example, a prototypically mechanical body should be accompanied by a prototypically mechanical head, even if it means forgoing more humanlike features. Conversely, a highly humanlike head should be accompanied by a highly humanlike body.

4.3. Limitations and Future Directions

The present study contributes a replication and extension of prior research on the uncanny valley in the domain of social robotics and human-robot interaction. In particular, it demonstrates the use of a simple laboratory procedure to evaluate aversive responding with a large portion of the current design space of humanoid robots. While we are confident that the present study was well-suited to address our primary goals, the work also has its limitations that underscore important avenues for future research.

4.3.1. Demographics

One potentially significant limitation in particular concerns the demographics of both our participants and of the humanlike stimuli employed in this study. Specifically, our sampling – despite attempts to recruit broader participation via public advertisement within the local metropolitan area – drew a largely homogenous (predominately white, well-educated, American, and young) participant population. While these demographics reflect those of the local university and to some extent, the geographical region in which the study was conducted, it nevertheless constrains the interpretation of our results. In particular, it remains unknown as to whether the observed valley effects extend to the general population as variations in participant demographics (e.g., age, culture, etc.) have been found to affect people's general perceptions of social robots (e.g., Bartneck et al., 2005; Kuo et al., 2009; Li et al., 2010; Lee and Sabanović, 2014; Stafford et al., 2014b; Sundar et al., 2016). Though these variations have not been studied directly in relation to the uncanny valley, still there may be a multitude of sociocultural factors relevant to understanding the valley phenomenon and its effects on the perception of and emotionally-motivated responding to robots.

In addition, it is important to note the simultaneous imbalance in the race/gender of our stimuli. Specifically, the set of highly humanlike robots is primarily composed of robots that are female-gendered and phenotypically Asian, while robots with lesser degrees of human similarity lack explicit race and gender cues. This imbalance stems from the "demographics" of the current design space of android robots, in which a majority of platforms have been modeled after women (who are predominately Asian) and white men. Though we balanced our set of human stimuli to reflect the demographics of the highly humanlike robots, the skewed demographics of both our stimuli and the participants evaluating it leave the potential for differential responding on the basis of the agents' gender/race (e.g., Fiske et al., 2007; Zebrowitz and Montepare, 2008). This thus poses a methodological consideration that warrants further investigation.

4.3.2. Instrumentation

In addition to the above considerations, we also note a potential limitation with respect to the measurement of negative attitudes toward robots. Specifically, we employed the NARS scales (Nomura et al., 2006) for indexing participants' attitudes in order to address a longstanding critique of valley theory, namely whether people's aversion stems from pre-existing negative attitudes. Though no significant relationship was observed between the NARS and aversion indices, it is possible that the NARS scales do not capture negative attitudes that are relevant to the uncanny valley. Specifically, the content of the NARS questionnaire items range from context-related (e.g., "I would feel nervous just standing in front of a robot") to highly philosophical in nature (e.g., "I would feel uneasy if robots really had emotions," "I am concerned that robots would be a bad influence on children," "I feel that in the future, society will be dominated by robots."). Thus, the scale may align more with attitudes pertaining to human identity and replacement by robots (e.g., MacDorman, 2006; Rosenthal-von der Pütten and Krämer, 2015), which may not drive the behavioral valley effects observed here.

4.3.3. Development

Finally, the majority of literature probing the valley and its effects is limited to young adults. Thus, it remains to be determined as to when/how the uncanny valley emerges over development. Specifically, are the indices of aversion that we observed here present in infants/children in a qualitatively similar way? Or is the valley limited to adults? While there is evidence of valley effects in infants (Lewkowicz and Ghazanfar, 2012; Matsuda et al., 2012), it is methodologically limited. In particular, the valley effects in infants are evidenced only by their gaze behavior and only in response to a very small set of agents. Additional studies evaluating valley effects in children would be useful both theoretically and practically. Theoretically, observation of a valley before young adulthood would lend support to the notion that the valley stems from more intrinsic perceptual mechanisms (e.g., the category uncertainty hypothesis and categorization theory). Practically, regardless of its innateness, understanding how younger populations perceive social robots would determine whether their design needs to be modified as a function of age of the population for which the robot is designed.

5. CONCLUSIONS

Our results both replicated and extended prior research, providing further empirical support for Mori's uncanny valley hypothesis and its relevance to human-robot interaction. Specifically, we demonstrated a robust valley effect within the current design space of humanoid robotics, wherein people showed significant behavioral aversion to highly humanlike robots. Moreover, we found no relationship between people's aversion and any pre-existing attitudes toward robots, suggesting that time and/or exposure to robots is unlikely to mitigate the valley effect. These findings underscore both a need for careful attention to the appearance of humanoid robots and the importance of measuring people's emotional responses to robots during the design phase.

At present, the findings serve to provide general guidance in the design of future social robots. In particular, our exploration points to two visual factors that should be considered, namely atypicality and category ambiguity. Our results suggest, for example, that it would be wise to design new robots with greater consistency between features and greater distance from the robot-human boundary (in either direction). Doing so may help to mitigate aversive reactions and, thus, maximize the utility of robots in contexts requiring interaction with humans.

ETHICS STATEMENT

All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Tufts University Institutional Review Board.

AUTHOR CONTRIBUTIONS

MS and HU conceived and designed the study, with significant input from VF, MJ, WJ, KM, and JR. MS conducted the data acquisition. MS analyzed the data, with significant input from HU, VF, MJ, WJ, KM, and JR. MS and HU wrote the manuscript, with significant input from VF, MJ, WJ, KM, and JR.

FUNDING

This work was supported by Tufts University and The University of Texas Rio Grande Valley.

REFERENCES

Andrist, S., Mutlu, B., and Tapus, A. (2015). "Look like me: matching robot personality via gaze to increase motivation," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul: ACM), 3603–3612.

Appel, M., Weber, S., Krause, S., and Mara, M. (2016). "On the eeriness of service robots with emotional capabilities," in *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Christchurch), 411–412.

Bartneck, C., Kanda, T., Ishiguro, H., and Hagita, N. (2009). "My robotic doppelgänger-a critical look at the uncanny valley," in *Proceedings of the 18th IEEE Symposium on Robot and Human Interactive Communication (RO-MAN)* (Toyama), 269–276.

Bartneck, C., Nomura, T., Kanda, T., Suzuki, T., and Kato, K. (2005). "Cultural differences in attitudes towards robots," in *Proceedings of Symposium on Robot Companions (SSAISB 2005 Convention)* (Hatfield), 1–4.

Brenton, H., Gillies, M., Ballin, D., and Chatting, D. (2005). The Uncanny Valley: does it exist and is it related to presence. *Presence Connect*.

Broadbent, E., Kumar, V., Li, X., Sollers, J. III. Stafford, R. Q., MacDonald, B. A., et al. (2013). Robots with display screens: a robot with a more humanlike face display is perceived to have more mind and a better personality. *PLoS ONE* 8:e72589. doi: 10.1371/journal.pone.0072589

Burleigh, T. J., Schoenherr, J. R., and Lacroix, G. L. (2013). Does the uncanny valley exist? an empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Comput. Hum. Behav.* 29, 759–771. doi: 10.1016/j.chb.2012.11.021

Chattopadhyay, D., and MacDorman, K. F. (2016). Familiar faces rendered strange: why inconsistent realism drives characters into the uncanny valley. *J. Vis.* 16, 7–7. doi: 10.1167/16.11.7

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robot. Auton. Sys.* 42, 177–190. doi: 10.1016/S0921-8890(02)00374-3

Fiske, S. T., Cuddy, A. J., and Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends Cogn. Sci.* 11, 77–83. doi: 10.1016/j.tics.2006.11.005

Girden, E. R. (1992). *ANOVA: Repeated Measures*. Number 84.

Gray, K., and Wegner, D. M. (2012). Feeling robots and human zombies: mind perception and the uncanny valley. *Cognition* 125, 125–130. doi: 10.1016/j.cognition.2012.06.007

Groom, V., Nass, C., Chen, T., Nielsen, A., Scarborough, J. K., and Robles, E. (2009). Evaluating the effects of behavioral realism in embodied agents. *Int. J. Hum. Comput. Stud.* 67, 842–849. doi: 10.1016/j.ijhcs.2009.07.001

Hanson, D., Olney, A., Prilliman, S., Mathews, E., Zielke, M., Hammons, D., et al. (2005). "Upending the uncanny valley," in *Proceedings of the National Conference on Artificial Intelligence*, Vol. 20 (Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press), 1728.

Inkpen, K. M., and Sedlins, M. (2011). "Me and my avatar: exploring users' comfort with avatars for workplace communication," in *Proceedings of the ACM International Conference on Computer Supported Cooperative Work* (Hangzhou: ACM), 383–386.

Jentsch, E. (1997). On the psychology of the uncanny. *J. Theor. Human.* 2, 7–16.

Kätsyri, J., Förger, K., Mäkäräinen, M., and Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Front. Psychol.* 6:390. doi: 10.3389/fpsyg.2015.00390

Koschate, M., Potter, R., Bremner, P., and Levine, M. (2016). "Overcoming the uncanny valley: displays of emotions reduce the uncanniness of humanlike robots," in *Proceedings of the 11th ACM/IEEE International Conference on Human Robot Interaction (HRI)*, 359–365.

Kuo, I. H., Rabindran, J. M., Broadbent, E., Lee, Y. I., Kerse, N., Stafford, R., et al. (2009). "Age and gender factors in user acceptance of healthcare robots," in *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, (IEEE) (Toyama), 214–219.

Kupferberg, A., Glasauer, S., Huber, M., Rickert, M., Knoll, A., and Brandt, T. (2011). Biological movement increases acceptance of humanoid robots as human partners in motor interaction. *AI Soc.* 26, 339–345. doi: 10.1007/s00146-010-0314-2

Lee, H. R., and Sabanović, S. (2014). "Culturally variable preferences for robot design and use in south korea, turkey, and the united states," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction* (Bielefeld: ACM), 17–24.

Lewkowicz, D. J., and Ghazanfar, A. A. (2012). The development of the uncanny valley in infants. *Dev. Psychobiol.* 54, 124–132. doi: 10.1002/dev.20583

Li, D., Rau, P. P., and Li, Y. (2010). A cross-cultural study: effect of robot appearance and task. *Int. J. Soc. Robot.* 2, 175–186. doi: 10.1007/s12369-010-0056-9

MacDorman, K. F. (2006). "Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley," in *ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science*, 26–29.

MacDorman, K. F., and Chattopadhyay, D. (2016). Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. *Cognition* 146, 190–205. doi: 10.1016/j.cognition.2015.09.019

Mathur, M. B., and Reichling, D. B. (2016). Navigating a social world with robot partners: a quantitative cartography of the uncanny valley. *Cognition* 146, 22–32. doi: 10.1016/j.cognition.2015.09.008

Matsuda, Y.-T., Okamoto, Y., Ida, M., Okanoya, K., and Myowa-Yamakoshi, M. (2012). Infants prefer the faces of strangers or mothers to morphed faces: an uncanny valley between social novelty and familiarity. *Biol. Lett.* 8, 725–728. doi: 10.1098/rsbl.2012.0346

Mitchell, W. J., Szerszen, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., and MacDorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception* 2, 10–12. doi: 10.1068/i0415

Mori, M., MacDorman, K. F., and Kageki, N. (1970/2012). The uncanny valley [from the field]. *IEEE Robot. Automat. Magazine* 19, 98–100. doi: 10.1109/MRA.2012.2192811

Morris, S. B., and DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychol. Methods* 7:105. doi: 10.1037/1082-989X.7.1.105

Nomura, T., Kanda, T., and Suzuki, T. (2006). Experimental investigation into influence of negative attitudes toward robots on human–robot interaction. *AI Soc.* 20, 138–150. doi: 10.1007/s00146-005-0012-7

Piwek, L., McKay, L. S., and Pollick, F. E. (2014). Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion. *Cognition* 130, 271–277. doi: 10.1016/j.cognition.2013.11.001

Riek, L. D., Rabinowitch, T.-C., Chakrabarti, B., and Robinson, P. (2009). "Empathizing with robots: Fellow feeling along the anthropomorphic spectrum," in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops* (Amsterdam: IEEE), 1–6.

Rosenthal-von der Pütten, A. M., and Krämer, N. C. (2014). How design characteristics of robots determine evaluation and uncanny valley related responses. *Comput. Hum. Behav.* 36, 422–439. doi: 10.1016/j.chb.2014.03.066

Rosenthal-von der Pütten, A. M., and Krämer, N. C. (2015). Individuals' evaluations of and attitudes towards potentially uncanny robots. *Int. J. Soc. Robot.* 7, 799–824. doi: 10.1007/s12369-015-0321-z

Rosenthal-von der Pütten, A. M., Krämer, N. C., Becker-Asano, C., Ogawa, K., Nishio, S., and Ishiguro, H. (2014). The uncanny in the wild. analysis of unscripted human–android interaction in the field. *Int. J. Soc. Robot.* 6, 67–83. doi: 10.1007/s12369-013-0198-7

Sauppé, A., and Mutlu, B. (2015). "The social impact of a robot co-worker in industrial settings," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (ACM), 3613–3622.

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., and Frith, C. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Soc. Cogn. Affect. Neurosci.* 7, 413–422. doi: 10.1093/scan/nsr025

Schein, C., and Gray, K. (2015). The eyes are the window to the uncanny valley: mind perception, autism and missing souls. *Inter. Stud.* 16, 173–179. doi: 10.1075/is.16.2.02sch

Stafford, R. Q., MacDonald, B. A., Jayawardena, C., Wegner, D. M., and Broadbent, E. (2014a). Does the robot have a mind? mind perception and attitudes towards robots predict use of an eldercare robot. *Int. J. Soc. Robot.* 6, 17–32. doi: 10.1007/s12369-013-0186-y

Stafford, R. Q., MacDonald, B. A., Li, X., and Broadbent, E. (2014b). Older people's prior robot attitudes influence evaluations of a conversational robot. *Inter. J. Soc. Robot.* 6, 281–297. doi: 10.1007/s12369-013-0224-9

Steckenfinger, S. A., and Ghazanfar, A. A. (2009). Monkey visual behavior falls into the uncanny valley. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18362–18366. doi: 10.1073/pnas.0910063106

Strait, M., Canning, C., and Scheutz, M. (2014). "Let me tell you! investigating the effects of robot communication strategies in advice-giving situations based on robot appearance, interaction modality and distance," in *Proceedings of the 9th ACM/IEEE Conference on Human-Robot Interaction (HRI)* (Bielefeld), 479–486.

Strait, M., Vujovic, L., Floerke, V., Scheutz, M., and Urry, H. (2015). "Too much humanness for human-robot interaction: exposure to highly humanlike robots elicits aversive responding in observers," in *Proceedings of the 33rd ACM Conference on Human Factors in Computing Systems (CHI)* (Seoul), 3593–3602.

Sundar, S. S., Waddell, T. F., and Jung, E. H. (2016). "The hollywood robot syndrome: media effects on older adults' attitudes toward robots and adoption intentions," in *Proceedings of the 11th ACM/IEEE Conference on Human-Robot Interaction (HRI)* (Christchurch), 343–350.

Vujovic, L., Opitz, P. C., Birk, J. L., and Urry, H. L. (2013). Cut! that's a wrap: regulating negative emotion by ending emotion-eliciting situations. *Front. Psychol.* 5:165. doi: 10.3389/fpsyg.2014.00165

Yamada, Y., Kawabe, T., and Ihaya, K. (2013). Categorization difficulty is associated with negative evaluation in the "uncanny valley" phenomenon. *Jpn. Psychol. Res.* 55, 20–32. doi: 10.1111/j.1468-5884.2012.00538.x

Yamamoto, K., Tanaka, S., Kobayashi, H., Kozima, H., and Hashiya, K. (2009). A non-humanoid robot in the "uncanny valley": experimental analysis of the reaction to behavioral contingency in 2–3 year old children. *PLoS ONE* 4:e6974. doi: 10.1371/journal.pone.0006974

Zebrowitz, L. A., and Montepare, J. M. (2008). Social psychological face perception: why appearance matters. *Soc. Person. Psychol. Compass* 2, 1497–1517. doi: 10.1111/j.1751-9004.2008.00109.x

Zlotowski, J., Proudfoot, D., and Bartneck, C. (2013). "More human than human: does the uncanny curve really matter?," in *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction, Workshop on Design of Humanlikeness in HRI from Uncanny Valley to Minimal Design*, ed H. Kuzuoka (Piscataway, NJ: IEEE Press), 7–13.

Złotowski, J., Proudfoot, D., Yogeeswaran, K., and Bartneck, C. (2015). Anthropomorphism: opportunities and challenges in human–robot interaction. *Int. J. Soc. Robot.* 7, 347–360. doi: 10.1007/s12369-014-0267-6

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Strait, Floerke, Ju, Maddox, Remedios, Jung and Urry. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

ORIGINAL RESEARCH
published: 31 October 2017
doi: 10.3389/frobt.2017.00047

Role of Speaker Cues in Attention Inference

Jin Joo Lee[1]*, *Cynthia Breazeal*[1] *and David DeSteno*[2]

[1] Personal Robots Group, Media Lab, Massachusetts Institute of Technology, Cambridge, MA, United States,
[2] Social Emotions Group, Department of Psychology, Northeastern University, Boston, MA, United States

Current state-of-the-art approaches to emotion recognition primarily focus on modeling the nonverbal expressions of the sole individual without reference to contextual elements such as the co-presence of the partner. In this paper, we demonstrate that the accurate inference of listeners' social-emotional state of attention depends on accounting for the nonverbal behaviors of their storytelling partner, namely their speaker cues. To gain a deeper understanding of the role of speaker cues in attention inference, we conduct investigations into real-world interactions of children (5–6 years old) storytelling with their peers. Through in-depth analysis of human–human interaction data, we first identify nonverbal speaker cues (i.e., backchannel-inviting cues) and listener responses (i.e., backchannel feedback). We then demonstrate how speaker cues can modify the interpretation of attention-related backchannels as well as serve as a means to regulate the responsiveness of listeners. We discuss the design implications of our findings toward our primary goal of developing attention recognition models for storytelling robots, and we argue that social robots can proactively use speaker cues to form more accurate inferences about the attentive state of their human partners.

Keywords: attention and engagement, nonverbal behaviors, speaker cues, listener backchannels, emotion recognition, children and storytelling, human-robot interaction

OPEN ACCESS

Edited by:
Hatice Gunes,
University of Cambridge,
United Kingdom

Reviewed by:
Gerard Bailly,
UMR5216 Grenoble Images
Parole Signal Automatique
(GIPSA-lab), France
Amit Kumar Pandey,
Aldebaran Robotics, France

Correspondence:
Jin Joo Lee
jinjoo@media.mit.edu

Specialty section:
This article was submitted
to Humanoid Robotics,
a section of the journal
Frontiers in Robotics and AI

Received: *14 January 2017*
Accepted: *05 September 2017*
Published: *31 October 2017*

Citation:
Lee JJ, Breazeal C and DeSteno D
(2017) Role of Speaker Cues
in Attention Inference.
Front. Robot. AI 4:47.
doi: 10.3389/frobt.2017.00047

1. INTRODUCTION

Storytelling is an interaction form that is mutually regulated between storytellers and listeners where a key dynamic is the back-and-forth process of speaker cues and listener responses. Speaker cues, also called backchannel-*inviting* cues, are signaled nonverbally through changes in prosody, gaze patterns, and other behaviors. They serve as a mechanism for storytellers to elicit feedback from listeners (Ward and Tsukahara, 2000). Listeners contingently respond using backchannel feedback which is signaled linguistically (e.g., "I see"), para-linguistically (e.g., "mm-hmm"), and nonverbally (e.g., head nod).

To support human-robot interactions (HRI), prior approaches have typically treated speaker cues as timing mechanisms to predict upcoming backchannel opportunities. In contingently responding to a person's speaker cues, robot listeners are able to support more fluid interactions, engender feelings of rapport, and communicate attention (Gratch et al., 2007; Morency et al., 2010; Park et al., 2017). In this paper, we introduce additional functions speaker cues have in social interactions beyond this stimulus-response contingency. Our main contribution is demonstrating how:

1. speaker cues serve as a means to *regulate* the responsiveness of listeners.
2. speaker cues can modify the *interpretation* of backchannels when inferring listener's attention.

For our first claim, we begin by identifying backchannels that signal the attention and engagement of listeners as well as speaker cues capable of eliciting those backchannels. We examine multimodal speaker cues (prosody and gaze) and their emission as either singlets or combinations, and we find that compounded cues have a higher likelihood of eliciting a response from listeners.

We support our second claim through a two-part process. First, our video-based human-subjects experiment demonstrates that accurate inference about listeners' attentive state depends on observing not just the listeners but also their storytelling partner. Second, through a finer-grain analysis, we find that the interpretation of backchannels from a listener depends on the storyteller's cueing behaviors. This cue-response pair is necessary for an accurate understanding of listener's attention.

Our primary research goal is to develop contextually aware attention recognition models for social robots in storytelling applications. In this paper, we focus on the nonverbal behaviors of storytellers as key context in which we evaluate the attentive state of listeners. A storyteller's speaker cues play an important role in the attention inference about listeners. This social and interpersonal context to attention, or more broadly emotion, recognition is especially relevant for human–robot interactions. HRI researchers depend on emotion recognition technologies to better understand user experience. But a common approach in affective computing is to model only the expressions of the sole individual without reference to external context like the co-presence of a social agent. In using these technologies for storytelling robots, we miss out on the added value their cueing actions can bring to the inference process. In pursuit of our research goal, this paper's approach is to first deeply understand the interpersonal nature of attention inference from the human perspective. Based on our findings from human–human interaction studies, we extract design implications when developing attention recognition models for social robots.

Our paper is outlined as follows:

- **Section 2: Background:** We elaborate on how current emotion recognition technologies disagree with modern theories of human nonverbal communication. We review speaker cues and listener backchannels that have been studied among adult populations and highlight the limited findings surrounding young children in peer-to-peer interactions.
- **Section 3: Effect of Storyteller Context on Inferences about Listeners:** Through a video-based human-subjects experiment, we manipulate the presence, absence, or falseness of storytellers from original interactions with listeners. Although the listeners' nonverbal behaviors remain exactly the same, perceptions about their attentive state from a third-party observer are different across these contextual manipulations.
- **Section 4: Effect of Speaker Cues on Listener Response Interpretation and Regulation:** Through a data collection of peer-to-peer storytelling, we identify attention-related listener responses as well as speaker cues that children use amongst peers. We examine which speaker cues, taken singly or in combination, can elicit a contingent response from listeners, and we find that listeners are more likely to respond to stronger cueing contexts. Lastly, using a logistic regression model, we find that backchannels are interpreted differently if observed after a weak, moderate, or strong cue.
- **Section 5: General Discussion:** We summarize our findings based on our human–human interaction studies and draw implications when modeling attention recognition for HRI.

2. BACKGROUND

2.1. Context in Emotion Recognition—Humans vs Machines

Emotion recognition systems typically discretize emotional states as a basic set of anger, surprise, happiness, disgust, sadness, and fear, while states such as boredom, confusion, frustration, engagement, and curiosity are considered to be non-basic (D'Mello and Kory, 2015). In our work, we focus on the social-emotional state of engagement which we interchangeably use with the word attention. Note, this should not be confused with joint attention, which is a different research problem of inferring what people are attending to in a physical environment (Scassellati, 1999). The nonverbal behaviors that support joint attention serve more as a mechanism to attend to objects and events rather than ones associated with communicating emotional states.

Emotion recognition systems have primarily focused on detecting prototypical facial expressions through facial muscle action units (FACS) (Sariyanidi et al., 2015). Based on a recent survey, facial expressions are still the main modality used for affect detection but have also extended to include gaze behaviors, body movements, voice features, spoken language, and biosignals such as electrodermal activity (D'Mello and Kory, 2015). Of the 90 systems reported, 93% of approaches focus on these within-person features and exclude extrinsic factors such as the environment or interaction partners.

This representation follows a classical theory in human nonverbal communication of nonverbal leakage where emotional states are direct influencers of exhibited nonverbal behaviors (Knapp and Hall, 2010). Traditional emotion understanding models such as those utilized by Ekman (1984) focus on the nonverbal expressions of single individuals without reference to any contextual elements such as setting, cultural orientation, or other people. By contrast, modern theories emphasize the contextual nature of nonverbal inference where greater accuracy comes from decoding expressions with reference to the social context (Barrett et al., 2011; Hassin et al., 2013).

Toward this, a growing amount of work has started to model the behaviors of both interactants to recognize social-emotional states, such as trust (Lee et al., 2013), rapport (Yu et al., 2013), and bonding (Jaques et al., 2016). Although the behaviors of both interactants are now being considered, they are fundamentally represented as a pair of independent events or captured as joint or dyadic features (like the number of conversational turns) for non-temporal models. As such, these approaches do not consider the added information that comes from the interpersonal call–response dynamic of social interactions. Although this is a foundation when modeling other domains such as turn-taking

(Thórisson, 2002) or conversational structure (Otsuka et al., 2007), emotion recognition models for dyadic interactions currently do not consider the causal properties between the behaviors of dyads and how they can influence each other.

2.2. Speaker Cues and Listener Responses—Children vs Adults

A well-known dynamic in face-to-face communication is the call–response contingencies between speaker cues and listener backchannels, which we will also refer to more broadly as "listener responses." The role of listener responses in conversations have been comprehensively characterized as carrying different functions such as signaling understanding, support, empathy, and agreement (Maynard, 1997) as well as facilitating conversational flow (Dittmann, 1972; Duncan and Fiske, 1977). However, in this paper, we specifically focus on the role of backchannels as evidence of continued attention, interest, and engagement of listeners (Kendon, 1967; Schegloff, 1982). It is important to note that we will consistently use the words *cues* and *responses* to differentiate the source of the emitted nonverbal behavior as either from a speaker or listener, respectively.

Although there is extensive research on adult listening and speaking behaviors, limited prior work exists in investigating younger populations especially in the context of peer-to-peer storytelling. In adult–child conversations, prior works have focused on demonstrating the effect of age on the backchanneling behaviors of children. More specifically, 11-year-olds were found to provide significantly more listener responses to adults than 7- or 9-year-olds and with a threefold increase between 7-year-olds and 11-year-olds (Hess and Johnston, 1988). In a separate study investigating 2- to 5-year-olds, older preschool children were found to use more head nods and spent more time smiling and gazing at adult speakers, suggesting that older children better understand a listener's role in providing collaborative feedback (Miller et al., 1985).

Both children and adult listeners were found to respond more frequently to joint cues (e.g., co-occurring speaker cues like simultaneous eye-contact with long speech pauses) over single cues. Joint cues were found to quadratically increase the likelihood of eliciting a backchannel response (Hess and Johnston, 1988; Gravano and Hirschberg, 2009). For an organized collection of prior research into speaker cues and listener responses of adults and children, see Tables S1 and S2 in the Supplementary Materials. We extend these prior works by pioneering the identification of attention-related listener responses and speaker cues that children employ amongst peers (not with adults) in storytelling interactions.

3. EFFECT OF STORYTELLER CONTEXT ON INFERENCES ABOUT LISTENERS

3.1. Overview

Although modern theories of human nonverbal communication emphasize the contextual nature of emotion understanding, current state-of-the-art approaches to emotion recognition primarily focus on the sole individual without reference to contextual elements such as the co-presence of interaction partners. The goal of this section is to demonstrate how a similar expectation placed on human observers results in them forming less accurate inferences about the emotions of others. Through a video-based experiment, we manipulate the presence, absence, or falseness of storytellers from original interactions with listeners. Although the listeners' behaviors remain exactly the same, we expect that the perception about their attentive state from third-party observers will be different across these conditions. We hypothesize the following:

> **Main Hypothesis:** Inference performance about a listener's attentive state is best when observing both the storyteller's and listener's behaviors of a social interaction and worst when missing the storyteller context.

We quantify inference performance as a function of prediction speed and accuracy and aim to demonstrate that both measures improve when observing the true storyteller context to the listener's behaviors. We argue that accurate inference about listeners' attention depends on also observing the storyteller.

3.2. Method

Through a video-based human-subjects experiment, we study how the perception of listeners changes when observing their original behaviors in different storyteller contexts.

3.2.1. Participants

Participants were recruited online through Amazon Mechanical Turk. Turk Workers were from the United States to ensure cultural relevance. To limit the participation pool to high-quality workers, their qualification requirements met the following:

- Number of approved HITs (Human Intelligence Tasks) greater than 5000,
- Approval rating from former requesters greater than 98%.

From the 542 Turk workers that submitted to the HIT task, 36 individuals were rejected for not fully completing all parts of the task or for not properly following the task's instructions. The average age of the remaining 506 participants was 38-years-old (SD = 11). Nearly half (56%) were parents and gender was close to balanced (53% female). Below we detail two exclusion principles applied in removing participants from our analysis.

3.2.2. Study Procedure

The online survey-based experiment took an average 19 minutes (SD = 12) to complete the following three parts: Affect Recognition Assessment, Training Exercise, and Inference Task.

3.2.2.1. Affect Recognition Assessment

The Diagnostic Analysis of Nonverbal Behavior (DANVA2) is an assessment to measure an individual's nonverbal affect recognition ability (Nowicki and Duke, 1994). The evaluation consists of viewing a series of facial expressions as well as listening to paralinguistic expressions of children to identify the expressed emotion: happiness, sadness, anger, or fear. Individuals are scored based on the number of items incorrectly identified from 24 different pictures of children's faces and 24 different recordings of children's voices. Participants took this assessment through a web-based flash program that would present the stimuli and record their multiple choice response.

Overall, participants scored a mean error of 2.9 (SD = 2.0) in recognizing children's facial expression and 4.8 (SD = 2.6) in recognizing children's paralinguistic expressions. To ensure that our population consisted of individuals of average affect recognition ability, 23 participants that scored an error greater than two standard deviations from the population's average on either the DANVA face or voice subtests were excluded from our analysis below.[1]

3.2.2.2. Training Exercise

To familiarize participants with the procedure of the primary task, they first experienced the task procedure on a simple example video as a training exercise. Participants were asked to carefully watch the video and immediately pause it when they heard the word "bat." Then they were instructed to report the number in the upper-left hand corner of the video, which represented the video frame corresponding to the paused scene.

Overall, participants were on average 41 frames, or 1.4 seconds, away from the exact moment of the target event (SD = 147 frames or 4.9 seconds). Participants that were within two standard deviations from the population's average response frame passed this training exercise. As a measure of task adherence to filter out low-quality Turk workers, the 22 participants who failed to meet these criteria were excluded from our analysis below.

3.2.2.3. Inference Task

Participants were asked to watch a series of videos (each around 30 seconds in duration) of different children listening to a storyteller. Participants were told that in all the videos the listener is at first paying attention to the story, but we want to know when/if the listener stops being attentive to the narrator's story. Following the same procedure introduced in the training exercise, participants reported their paused frame, which represented the moment they perceived the listener transitioning from attentiveness to inattentiveness. They also had the option of reporting if they believed that the listener was paying attention the entire time.

3.2.3. Experiment Design

From an original interaction between a listener and their storytelling partner, we manipulate the presence, absence, or falseness of the storyteller through a video-based experiment. Although the listener's behaviors remain the same, we investigate how an observer's perception about the listener's attentive state changes across the different contextualizations. As a within-subject study design, a participant viewed a video from each of the three conditions but of three different listeners in a random order. In using three different listeners, we can generalize our results to be beyond a listener-specific phenomenon. Our three conditions are defined as the following:

1. **TRUE (control):** Participants viewed the original interaction between a storyteller and listener. With access to both the storyteller's and the listener's behaviors, they made an inference about the listener's attentive state.
2. **ABSENT:** Participants only viewed the listener. They made their inference based solely on the listener's nonverbal behaviors.
3. **FALSE:** Participants viewed an unmatched interaction where the original storyteller is replaced with one from a different storytelling episode.

From three different storytelling interaction videos collected in *Section 4.2*, we created a set for the TRUE condition with the audio and video (AV) of the original storyteller, a set for the ABSENT condition with the storyteller's AV removed, and a set for the FALSE condition with the AV of a different storyteller (see **Figure 1A**). It is important to note that although the audio recordings captured both of the storyteller's and listener's voices, in general only the storyteller is speaking and the listener is quiet. To preserve the illusion that the FALSE condition was showing real interactions, we avoided moments containing any dialog-related coordination. For example, we carefully selected video snippets that did not include when a storyteller asked a direct question or was interrupted by the listener.

All the videos were composed and edited to allow a viewer to easily see the facial expressions of both the storyteller and listener. We also preserved their gaze cues by arranging the images to mimic the original interaction geometry. As shown in **Figure 1B**, we ensured that a listener's behavior between each condition remained exactly the same. Please see Videos S1–S3 in the Supplementary Materials to watch an example set of videos used for this experiment.

3.2.4. Dependent Measures

The video snippets contain a single point where the listener transitions from attentiveness to inattentiveness as illustrated in **Figure 2**. This transition point is based on the hand-annotated attention labels from trained experts (see *Section 4.2.3*). From a participant's report on where he/she believed the transition point to be, we defined two dependent measures for inference performance.

1. **Accuracy:** A response frame after the transition point is marked as correct and elsewhere as incorrect, including the option of reporting the listener as attentive for the entire time. Accuracy is a dichotomous variable, where a value of 0 means incorrect and 1 means correct.
2. **Latency:** Latency is measured as the distance between the response frame from the target frame. This difference represents the participant's delay and is only calculated for correct inferences.

In accordance with our hypotheses, we expect an increasing trend (TRUE > FALSE > ABSENT) where participants achieve their best inference performance in both accuracy and latency

[1] There is a bit of irony in using a standard contextless test to exclude participants from a study that is investigating the influence of context on affect recognition. It is possible to make an inference (of lesser accuracy) in contextless situations, but we are investigating the added value of context. This exclusion is to ensure a population of typical development.

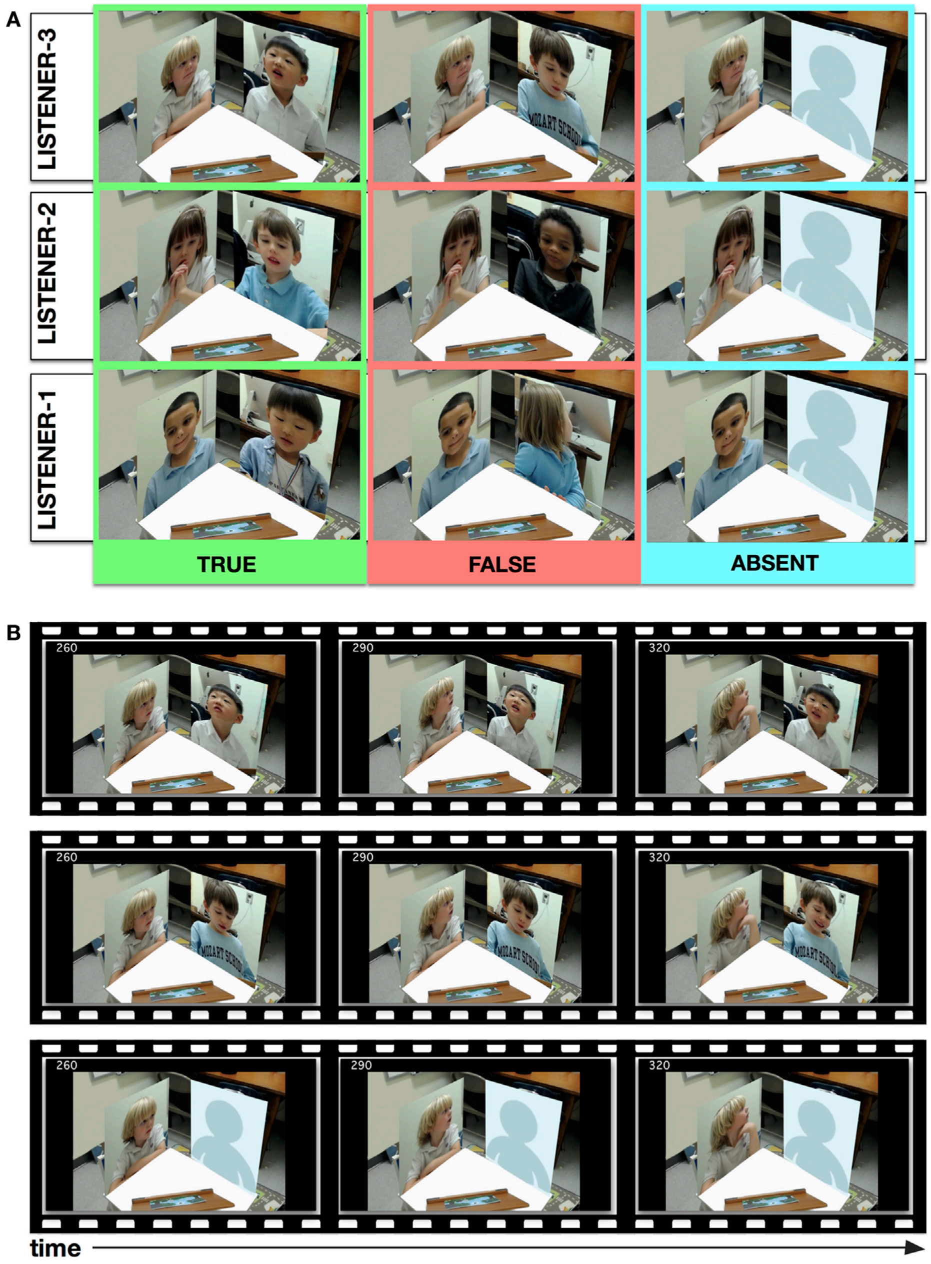


FIGURE 1 | Video-based human-subjects experiment. **(A)** From TRUE interactions between a storyteller and listener, we manipulate the absence and falseness of the storyteller context. For the FALSE condition, we replace the original storyteller with the audio and video of a different storyteller. The ABSENT condition removes all storyteller context (both audio and video). **(B)** We illustrate how for Listener-3, at frames 260, 290, and 320, we retain his exact behavior across the three conditions: TRUE (*top row*), FALSE (*middle row*), ABSENT (*bottom row*).

with the TRUE condition and their worst inference performance with the ABSENT condition since it lacks any storyteller context. We anticipate that participants will have a difficult time with the FALSE condition since the disjointed set of storyteller's cues to listener's responses will either delay or confuse their inference process. Although participants were informed, through a brief description, of the storytelling context of their upcoming videos, the FALSE condition at least visually presents the listeners' behaviors in an interpersonal context. As such, we hypothesize that a false/unmatched context is better than having no context.

3.3. Analysis of Inference Accuracy

We examine the ability of storyteller context to predict an increasing trend of inference accuracy using generalized linear models (GLM). A multilevel (i.e., mixed-model) logistic regression was performed to determine the effect of storyteller context on the likelihood of participants making a correct

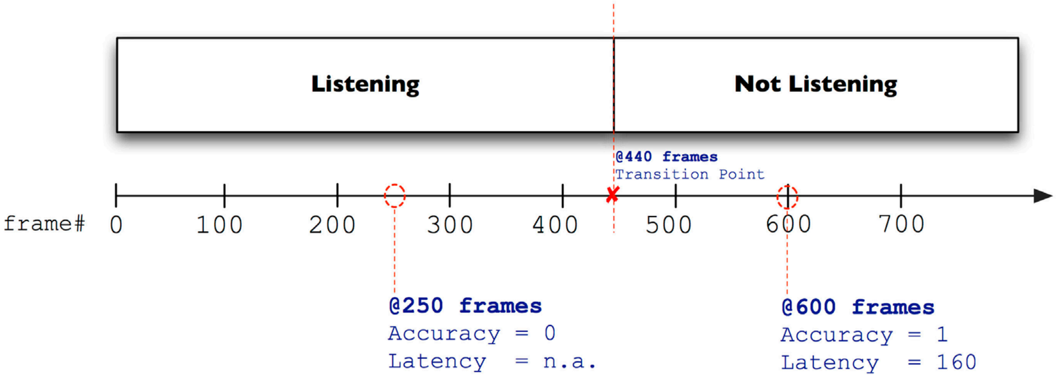


FIGURE 2 | Example of scoring accuracy and latency. At frame 440, a listener is annotated by experts as transitioning from an attentive to inattentive state. As such, a participant who predicted and reported the transition occurring at 250 frames is marked as being incorrect (accuracy of 0). A participant who made a prediction at 600 frames is accurate with a latency of 160 frames.

TABLE 1 | Effect of storyteller context on inference accuracy and latency.

	Conditions					
	True		**False**		**Absent**	
	Measures	**N**	**Measures**	**N**	**Measures**	**N**
Accuracy	58.4% correct	461	57.7% correct	461	51.0% correct	461
Latency	96 frames	269	107 frames	266	99 frames	235

Accuracy is reported as the percentage of participants correctly inferring attention transitions of a condition. Latency is reported as the median time-to-respond when a correct inference is made. N is the number of samples per condition (542 participants − 81 exclusions = 461 samples). Note, latency's N varies per condition since it only includes the samples with correct inferences.

inference about listeners' attentive state while controlling for within-subject dependencies from repeated measures (see Table S3 in Supplementary Material for model details). Based on our expectation that inference accuracy increases across treatment groups, the predictor variable is contrast coded as ordered values $[-1, 0, 1]$ to model a linear trend, where having access to the true context yields the highest likelihood of accurate inference while having access to no context yields the lowest.

Based on the Wald Chi-square statistic, the logistic regression model was statistically significant, $[\chi^2(1) = 4.15, p = 0.04]$, which indicates a linear relationship between our expected order of storyteller-context treatment and likelihood of correct inference. As shown in **Table 1**, an ascending trend in inference accuracy is observed with the TRUE condition obtaining the highest percentage of participants that correctly predict the attentive state of listeners and the ABSENT condition obtaining the lowest.

3.4. Analysis of Inference Latency

Similar to the trend analysis described for accuracy, we examine the ability of storyteller context to predict an increasing trend of inference latency values. The latency observations are positive whole numbers and have a skewed distribution since the highest density of observations are found closest to the target frame and then drop-off over time. Given the nature of the data, we used a gamma GLM (versus the typical normal distribution assumption) with storyteller context as the primary predictor of the log latency while again controlling for within-subject dependencies. We expected an increasing trend where participants experienced the greatest delays in the ABSENT condition, followed by the FALSE condition, and with the TRUE condition obtaining the lowest latencies, but no significant trend was found $[\chi^2(1) = 0.01, p = 0.94]$.

However, rather than looking for a trend, we instead looked for any differences between the conditions. By treating the predictor as a categorical variable, a statistically significant gamma GLM was found, $[\chi^2(2) = 6.35, p = 0.04]$[2] (see Table S4 in Supplementary Material for model details). More specifically, there is a significant difference between the TRUE and FALSE conditions, $t(767) = 2.21$, $p = 0.03$, with the TRUE condition obtaining lower latencies ($\tilde{x} = 96$ frames) than the FALSE condition ($\tilde{x} = 107$ frames). No significant difference was found between the TRUE and ABSENT conditions.

3.5. Discussion

Our main hypothesis was upheld regarding inference accuracy and partially upheld regarding inference latency. When

[2] There are two degrees-of-freedom since the three conditions are dummy coded as two categorical predictor variables.

predicting the attentive state of listeners, participants are most accurate when able to observe the true storyteller, less accurate with a false storyteller, and worst with no storyteller.

In regard to inference latency, we found that participants were faster in forming correct inferences with true storytellers over false ones. We had anticipated participants to be slowest in forming their predictions when missing the storyteller context, but they actually achieved similar speeds as when having it. If we view latency as an operationalization of confidence, we can interpret this result to mean that they felt similarly confident about their appraisals.

In sum, by changing the storyteller context in which listener behaviors are observed, we can delay or even cause incorrect inferences to be formed about the listener's attentive state. Participants are most accurate when observing *both* the storyteller's and listener's behaviors of a social interaction. They are least accurate when missing the interpersonal context of the storyteller. When presented with a false storyteller context, participants are again less accurate but also slower. This demonstrates the extent to which we can degrade an observer's perceptions about the social-emotional state of listeners.

4. EFFECT OF SPEAKER CUES ON LISTENER RESPONSE INTERPRETATION AND REGULATION

4.1. Overview

Our video-based human-subjects experiment demonstrated that the accurate interpretation of listeners' attentive state depends on also observing the storyteller. But what is it about the partner's behaviors that lead human observers to form more accurate inferences? In this section, our goal is to better understand the relationship between the storyteller's speaker cues and listener's backchannels as well as how their joint meaning impacts perceptions about listener's attention.

We conduct a series of analyses of human–human interactions. We begin by detailing our method for data collection and annotation of peer-to-peer storytelling interactions of young children in *Section 4.2: Data Collection*. Before we can start to investigate the relationship between cues and responses, we first identify the relevant nonverbal behaviors of our particular young population. As such, in *Section 4.3: Analysis of Listener Behavior*, we find backchanneling behaviors that communicate attention.

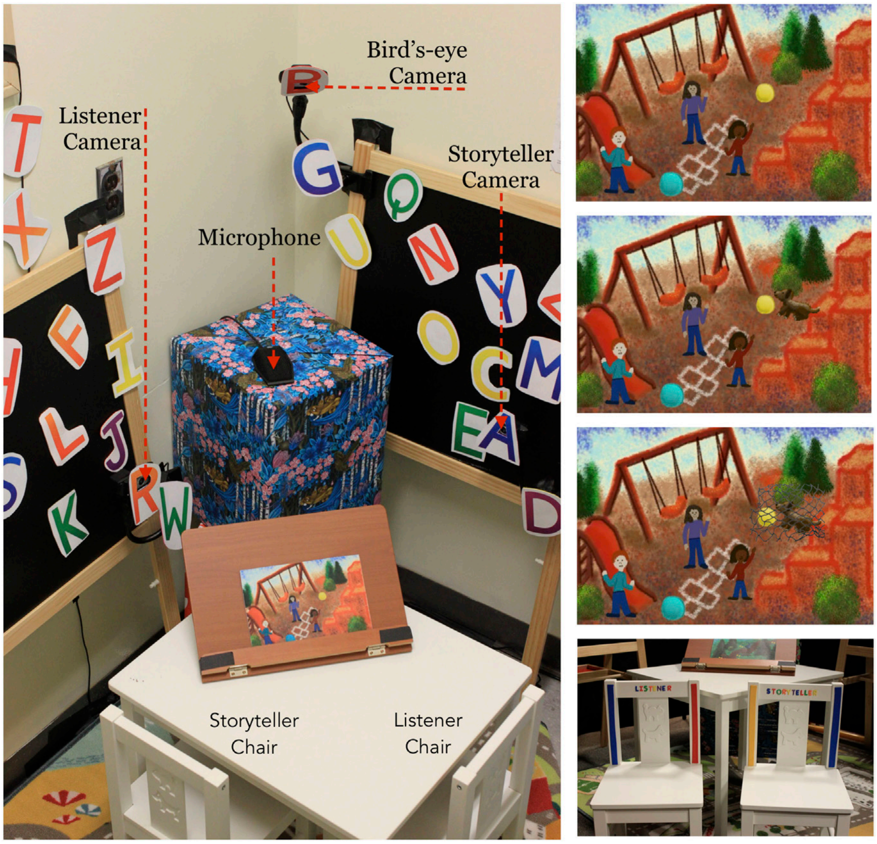


FIGURE 3 | Story space setup. Setup included three different camera angles, a high-quality microphone, listener and storyteller chairs, and a storybook with compounding story elements per page. The bottom-right photo shows how we labeled each chair to further emphasize to a child his/her role as either the storyteller or listener.

Furthermore, in *Section 4.4: Analysis of Speaker Cues*, we examine which of the coded multimodal speaker cues are observed to elicit contingent backchannels from listeners. Finally, in *Section 4.5: Analysis of Cues and Responses to Predict State*, we model the relationship between cues and responses and their effects on the perceived attentiveness of listeners.

4.2. Data Collection

4.2.1. Participants

Children of typical development were recruited from a Boston public elementary school whose curriculum already included an emphasis on storytelling. A total of 18 students from a single kindergarten (K2) classroom participated in the study. The average age was 5.22 years old (SD = 0.44) and 61% were male. Overall, 10 participants identified as White, 3 as Black or African American, 2 as Hispanic or Latino, 1 as Asian, 1 as Mixed, and 1 not specified.

4.2.2. Storytelling Task

Over a span of 5 weeks, each child completed at least three rounds of storytelling with different partners and storybooks. The storybooks were a series of colored pictures with illustrated characters and scenes that the children used to craft their own narratives (see **Figure 3** for an example storybook). In a dyad session, the pair of students took turns narrating a story to their partner with each turn generating a storytelling episode. Importantly, for each child participant, we had multiple examples of them being a storyteller and a listener. In sum, our data collection consisted of 58 storytelling episodes. The average length of a child's story was 1 minute and 17 seconds.

4.2.3. Video-Coded Annotations and Data Extraction

For each storytelling episode, the behaviors of both the listener and storyteller were manually annotated by multiple independent coders. We achieved moderate levels of agreement (Fleiss' $\kappa = 0.55$). For storytellers, we coded for gaze- and prosodic-based speaker cues. For listeners, we annotated for gaze direction, posture shifts, nods, eyebrow movement, smiles and frowns, short utterances, and perceived attentiveness. From the video recordings of the three time-synchronized cameras shown in **Figure 4**, coders used a video-annotation software called ELAN (Wittenburg et al., 2006) to mark the start and stop times for all the behaviors listed in **Table 2** except for the prosodic cues. For the attentive state annotation, a "listening" label meant that the participant was paying attention to the storyteller's story. It is important to note that our state annotation included when a listener took a "speaking-turn" as a mutually exclusive event. This enabled us to filter observations regarding conversational behaviors or turn-yielding cues, which has been demonstrated to be different from backchannel-inviting cues (Gravano and Hirschberg, 2009). Based on the "Task" annotation, we further excluded moments from our analyses when both children participants were off-task from the storytelling activity.

We developed a custom program to help coders easily annotate when and what type of prosodic cue was detected in speech. The program played back the audio recording of a storytelling episode, and coders were asked to simulate in real time being a listener and mark the moments when they wanted to backchannel by simply tapping the space bar. After this simulation, coders reviewed the audio snippets surrounding these moments to reflect on what

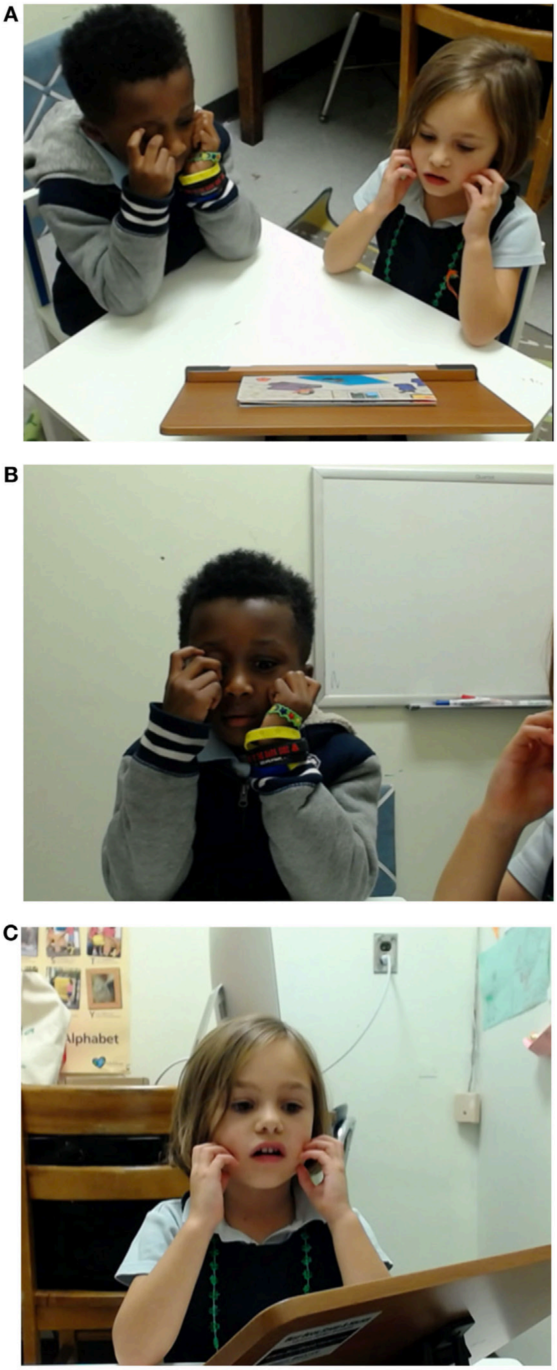


FIGURE 4 | Video-recordings of storytelling interactions. Three time-synchronized cameras captured the frontal view of each participant along with a bird's-eye-view **(A)**. Image **(B)** is the listener view. Image **(C)** is the storyteller view.

TABLE 2 | List of all annotated behaviors.

Category	Labels	S	L
Gaze	*book*, partner, away	X	X
Posture	*upright*, toward, away, other		X
Nod	*none*, nod		X
Eyebrows	*neutral*, raise, furrow		X
Mouth	*neutral*, smile, frown, other		X
Utterance	*none*, "ok," "oh," "so," "then," "yeah," "uh-huh," "ok then," "and then," "and they"		X
Voicing	*silence*, storyteller's voice, listener's voice, both	joint	
Task	*on-task*, off-task	joint	
Attentive State	listening, not listening, speaking-turn		X
Prosodic Cue	*none*, pitch, energy, pause, filled pause, long utterance, other	X	

The selected set of nonverbal behaviors were either found in prior works (see Tables S1 and S2 in Supplementary Material) or commonly observed in the storytelling interactions. Each annotation category has a set of mutually exclusive labels and was coded for storytellers (S) and/or listeners (L) or jointly evaluated (joint). An italicized label is the default behavior of an annotation category.

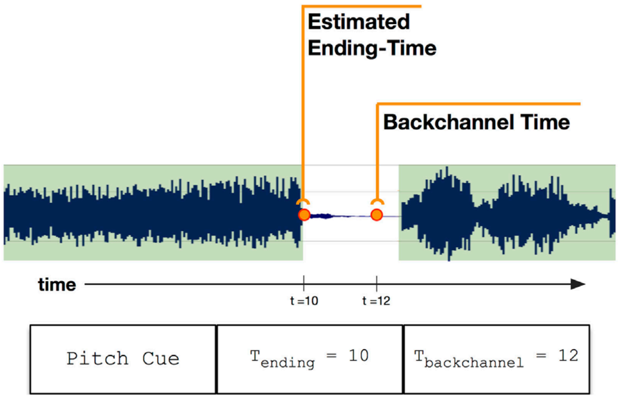


FIGURE 5 | Estimating the ending-time of a prosodic cue. Based on the backchannel time, we extract the last speaking-turn of the storyteller and estimate that its terminating edge is the ending-time of the prosodic cue. This is calculated for all prosodic-based cues except for the pause cue, which is roughly estimated to be halfway between the backchannel time and the terminating edge.

prompted their backchannel and categorize their reasoning into one or more of the following prosodic cues:

- pitch (intonation in voice, change in tone)
- energy (volume of voice, softness/loudness)
- pause (pause in speech, long silence)
- filled pause (e.g., "um," "uh," "so," "and")
- long utterance or wordy (a long contiguous speech segment)
- other

This stimulus-based coding was a method for annotators to identify *when* they wanted to backchannel (i.e., backchannel moment) in addition to categorizing their *why* (i.e., speaker cue(s) event). The null-space that was not marked had an implied default label of "none."

Three coders underwent this simulation, and we followed the Parasocial Consensus approach from Huang et al. (2010) to build consensus of when backchannel opportunities occurred. More specifically, each of our three coders' registered backchannel times were added as a "vote" on a consensus timeline with a duration of one second around the central moment. An area in the timeline with more than two total votes was counted as a valid backchannel moment.

From these backchannel moments and the voicing annotations, we estimated the emission time of prosodic cues (see **Figure 5** for more detail). To capture the complete cue context embodied by storytellers, we combined the prosodic cues with physical gaze cues to gain sets of multimodal cues.[3] In sum, for the proceeding set of analyses, we know when and which multimodal cues occurred throughout the storytelling episodes.

4.3. Analysis of Listener Behavior

A logistic regression analysis finds the best model to describe the relationship between the outcome and explanatory variables. Based on the fitted coefficients (and its significance levels), we can determine how much the explanatory variables can predict the outcome. Our goal is to identify nonverbal behaviors (explanatory variables) that can predict a listener's perceived attentive state (outcome). More specifically, for each annotated listener behavior listed in **Table 2**, a logistic regression analysis was performed to predict attention (0/1) based on the behavior's normalized duration and frequency rate. Normalized duration and frequency rates of behaviors were observed during a block period of either attentiveness or inattentiveness. Note, multiple block periods can exist in a single storytelling episode. For nonverbal behaviors that are quickly expressed (i.e., an average duration less than 90 seconds), the frequency rate was the only predictor.

Shown in **Table 3**, gazes, leans, brow-raises, smiles, nods, and utterances are nonverbal behaviors that significantly predict listeners' attention. Based on the sign of the coefficients (b) and significance (p) of the explanatory variables, we determine that frequent partner-gazes, frequent forward-leans, frequent brow-raises, prolonged smiles, frequent nods, and frequent utterances are positively associated with an attentive listener. By contrast, prolonged away-gazes from the partner, frequent away-leans, and prolonged brow-raises are negatively associated. Interestingly, brow-raises can hold opposite associations depending on their form of emission.

4.4. Analysis of Speaker Cues

In *Section 4.2.3*, adult-coders annotated when and what type of speaker cue was detected in the storytelling interactions. Therefore, the annotated speaker cues are based on adult

[3] Based on the prosodic cue ending-time and gaze onset-time, events were considered to be co-occurring and merged if they are within an empirically found 1.3 seconds of each other. This merging averages the times and reflects a collective moment of emission. When looking at the period between back-to-back gaze cues, we found a minimum time of separation of 1.5 seconds between cues. This establishes an upper bound of a merge window when trying to collect co-occurring cues. Beyond this window, we start encroaching on cues that could be a part of the next cueing instance.

TABLE 3 | Descriptive statistics and logistic regression models to estimate attention from listener behaviors.

					Logistic Regression Models		
Behavior	**Total**	**Mean Freq**	**Mean Dur**	**% Pop**	**Overall**	**Freq Term**	**Dur Term**
Gaze Partner	270	4.66	2.19	100	$\chi^2(2, 192) = 62.06$ $p^* = 3.34e^{-14}$	$b = 10.23$ $p^* = 8.25e^{-08}$	$b = 3.06$ $p = 0.22$
Gaze Away	698	12.03	4.43	100	$\chi^2(2, 192) = 152.34$ $p^* = 8.33e^{-34}$	$b = 0.10$ $p = 0.95$	$b = -8.56$ $p^* = 3.27e^{-13}$
Lean Toward	110	1.90	8.98	100	$\chi^2(2, 173) = 22.25$ $p^* = 1.48e^{-05}$	$b = 2.97$ $p^* = 7.57e^{-04}$	$b = 0.79$ $p = 0.18$
Lean Away	78	1.34	5.81	94	$\chi^2(2, 173) = 11.60$ $p^* = 3.02e^{-03}$	$b = -1.98$ $p^* = 4.57e^{-03}$	$b = 0.04$ $p = 0.96$
Brow–Raise	102	1.76	2.33	100	$\chi^2(2, 141) = 11.88$ $p^* = 2.63e^{-03}$	$b = 2.47$ $p^* = 0.01$	$b = -5.28$ $p^* = 0.02$
Brow-Furrow	17	0.29	3.23	44	$\chi^2(2, 141) = 2.06$ $p = 0.36$	$b = 1.96$ $p = 0.38$	$b = -3.43$ $p = 0.33$
Smile	173	2.98	7.23	94	$\chi^2(2, 173) = 12.35$ $p^* = 2.08e^{-03}$	$b = 0.88$ $p = 0.26$	$b = 1.69$ $p^* = 0.04$
Frown	9	0.16	2.55	28	$\chi^2(2, 173) = 1.50$ $p = 0.47$	$b = -3.83$ $p = 0.27$	$b = 2.17$ $p = 0.58$
Nod	18	0.31	1.13	39	$\chi^2(1, 34) = 7.61$ $p^* = 5.80e^{-03}$	$b = 5.28$ $p^* = 0.03$	–
Utter	18	0.31	0.94	50	$\chi^2(1, 42) = 4.24$ $p^* = 0.04$	$b = 6.72$ $p^* = 1.27e^{-03}$	–

Total is the collective frequency counts found in the dataset. The Mean Frequency is the average number of occurrences in a storytelling episode (i.e., Total/58). The Mean Duration is the average duration of an emitted behavior in seconds. % Pop refers to the proportion of the population (of the 18 participants) that demonstrated a single instance of the behavior across the repeated interactions. The logistic regression models predict the listener's attention based on the normalized duration and/or frequency rate of the nonverbal behavior. Note, the number of observations N for the chi-squared tests (i.e., χ^2 (DF, N)) are different for each annotation category since each analysis includes block periods only from storytelling episodes where at least one instance of the behavior type was observed.

perception. But which ones do children perceive, understand, and know to respond to? In our next set of analyses, we examine which speaker cues, taken singly or in combination, were observed to elicit a contingent backchannel from child listeners. We marked a backchannel as being contingent if the listener responded within [0.5–3.0] seconds[4] after the emitted cue with any of the previously found attentive behaviors. Those attentive behaviors were the onset of partner-gazes, forward-leans, brow-raises, nods, and utterances as well as prolonged smiles.

To further refine our proceeding analyses, we considered the situation where a speaker cue occurred but during a period when the listener was not paying attention to the storyteller. Their lack of a contingent response in this situation does not add relevant information to determining which cues children know to respond to. As such, our analyses only included data from moments when listeners were marked as attentive. This way, we can reason that when an attentive listener is unresponsive to a particular speaker cue, it is because the listener does not know to respond to this type of cueing signal.

4.4.1. As Individual Signals

A logistic regression analysis was performed to determine which speaker cues predict that an attentive listener will contingently backchannel. The overall logistic regression model was statistically significant [$\chi^2(6) = 45.9, p = 3.15e^{-08}$], and the speaker cues—gaze, pitch, filled pause, and long utterance taken singly—can elicit a response from young listeners (see **Table 4**). As expected, some of the speaker cues—energy and pause—do not offer significant predictive ability when examined in isolation. However, young children have been previously observed to respond more often in stronger cue contexts where two or more cues are co-occurring (Hess and Johnston, 1988).

4.4.2. As Co-Occurring Signals

Using the set of multimodal cues (extracted in *Section 4.2.3*), we examined the ability of cue combinations to predict that an attentive listener will contingently backchannel. The likelihood of observing a combination of cues is much smaller than individual cues, resulting in small sample sizes for each unique combination. Rather than performing a logistic regression analysis, we use the binomial exact test to determine whether the response rate of a cue combination is greater than an expected rate of 0.5. As shown in **Table 5**, the one-sided binomial test indicates that the response rates of the co-occurring cues Pitch-Energy, Gaze-Pause, Gaze-Pitch, Gaze-Pitch-Pause, and Gaze-Pitch-Energy are significantly higher than the expected rate. Interestingly, as the number of co-occurring cues increases (1 → 2→ 3), the likelihood of receiving a response also increases (0.68 → 0.82 → 0.93)[5]. Stronger the cue context, the more likely a listener will respond.

[4] We found that children positively respond on average 1.77 seconds (SD = 1.30) after an emitted cue. As such, we considered only the listener behaviors within a standard deviation from this average response time.

[5] Averaged response rates of only significantly predictive cues.

TABLE 4 | Descriptive statistics and the logistic regression model for individual speaker cues.

	Logistic Regression Model					
Predictors	**Gaze**	**Pitch**	**Energy**	**Pause**	**Filled pause**	**Long utterance**
b	1.89	0.65	0.08	0.09	1.33	1.05
t-stat	5.35	2.16	0.22	0.31	2.13	2.25
p-value	$p^* = 8.67e^{-08}$	$p^* = 0.03$	$p = 0.82$	$p = 0.76$	$p^* = 0.03$	$p^* = 0.02$
N	174	147	52	122	27	17
rate	0.76	0.59	0.58	0.51	0.59	0.76

The logistic regression model predicts the likelihood of a contingent response from an attentive listener based on the emitted speaker cues. N is the collective frequency counts found in the dataset. Rate is the likelihood of a response from listeners.

TABLE 5 | Descriptive statistics and the one-sided binomial exact test for co-occurring speaker cues.

2 Cues	N	Rate	p-value	3 Cues	N	Rate	p-value
..PE..	14	0.57	p = 0.40	.CPE..	10	0.70	p = 0.17
.CP...	64	0.56	p = 0.19	GCP...	14	0.93	$p^* = 9.16e^{-04}$
..P.F.	19	0.63	p = 0.18	GC.E..	15	0.93	$p^* = 4.88e^{-04}$
.C...W	12	0.75	p = 0.07				
.C.E..	44	0.66	p* = 0.02				
G.P...	18	0.89	$p^* = 6.56e^{-04}$				
GC....	39	0.90	$p^* = 1.68e^{-07}$				

We show the most frequently observed cue combinations in our dataset. Cue combinations are specified through the presence of the cue's symbolic letter. G: Gaze, C: Pitch, P: Pause, E: Energy, F: Filled Pause, W: Wordy (for long utterances). A dot represents the absence of that cue. N is the total occurrences of the cue combination found in the dataset. The one-sided binomial exact test determines whether the response rate of a cue combination is greater than an expected rate of 0.5.

4.5. Analysis of Cues and Responses to Predict State

Our primary goal is to better understand the relationship between speaker cues and listener responses and how their joint meaning can influence perceptions about listeners' attention. To fully model how the unique combinations of cue-response pairs effect this perception, we need much more data. Given our dataset, we instead create similarity heuristics to form groups that define a smaller range of possible behavior combinations.

Based on the relationship between cue-strength and response rate from our prior analysis (*Section 4.4.2*), we categorize multimodal cues based on their number of co-occurring cues as either weak, moderate, or strong cue contexts. For example, a Gaze-Pitch-Energy multimodal cue is represented with a value of 3, or a strong cue context.

Based on our analysis in *Section 4.3*, listener response combinations are grouped based on their overall valence score. Measured as a sum of individual valences, a forward-lean (+), prolonged smile (+), and an away-gaze (−) observed from the listener within [0.5−3.0] seconds after an emitted cue is represented as a total valence value of +1, an overall weak positive response. By accounting for both positive and negative behaviors, we roughly measure the magnitude and direction of listeners' overall response during this time window.

We also recorded whether annotators marked listeners as being attentive or inattentive by the end of this time period. This captured whether the perception about a listener changed (or remained the same) after witnessing the emoted response to the cue context.

In sum, for the following analyses, we use data tuples of <Cue, Response, State>:

- **Cue:** number of co-occurring speaker cues representing strength of weak, moderate, or strong [1 to 3].
- **Response:** measure of listener response to a cue as an overall valence rating [−3 to +4].
- **State:** perception of listener's attentive state sampled immediately after the response window [0 or 1].

We first examine the ability of cues and responses as independent predictors to explain state by themselves (*Section 4.5.1*: Only Main Effects) and then compare what happens when we add an interaction term that represents the relationship between cues and responses (*Section 4.5.2*: With Interaction Effects).

4.5.1. Only Main Effects

We examined the ability of cue-strength and response-valence to predict listeners' attention. The overall logistic regression model was statistically significant, [$\chi^2(2) = 71.4$, p = $3.2e^{-16}$], where response-valence was the primary predictor in estimating state (see **Table 6A**). One unit increase in the response-valence makes the listener 2.91 times more likely to be paying attention. This result is not surprising since listeners' behaviors are, of course, good predictors of their attentive state. But this analysis also serves as a means to validate our method of measuring listener response as an overall valence rating.

4.5.2. With Interaction Effects

In adding an interaction term to our previous logistic regression model, we find that the overall model is again statistically significant, [$\chi^2(3) = 78.4$, p = $6.74e^{-17}$], but can explain more of the variance $R^2 = 19.5\%$ compared to $R^2 = 17.8\%$ of the previous model. As shown in **Table 6B**, the interaction term is significant (p = 0.02), which indicates that the predictive power of listener response is modified by the cue context.

As shown in **Figure 6**, the strong-cue curve approaches areas of higher likelihood (i.e., y-axes limits) more quickly than the other curves, especially in comparison to the weak-cue curve which has less steep tails. This means, that for the same valence of listener response, stronger cues facilitate higher levels of certainty regarding listener's attentive state.

TABLE 6 | Logistic regression models predicting attention based on cues and responses.

(A) Main Effects Model

Predictor	Cue	Response
b	0.02	1.07
t-stat	0.09	7.20
p-value	$p = 0.93$	$p^* = 5.98e^{-13}$

(B) With Interaction Effects

Predictor	Cue	Response	Cue·Response
b	0.09	0.10	0.70
t-stat	0.38	0.24	2.41
p-value	$p = 0.71$	$p = 0.81$	$p^* = 0.02$

(A) The first logistic regression model considers only the main effects of cue-strength and response-valence to predict state. (B) The second model adds an interaction term which represents the relationship between cues and responses.

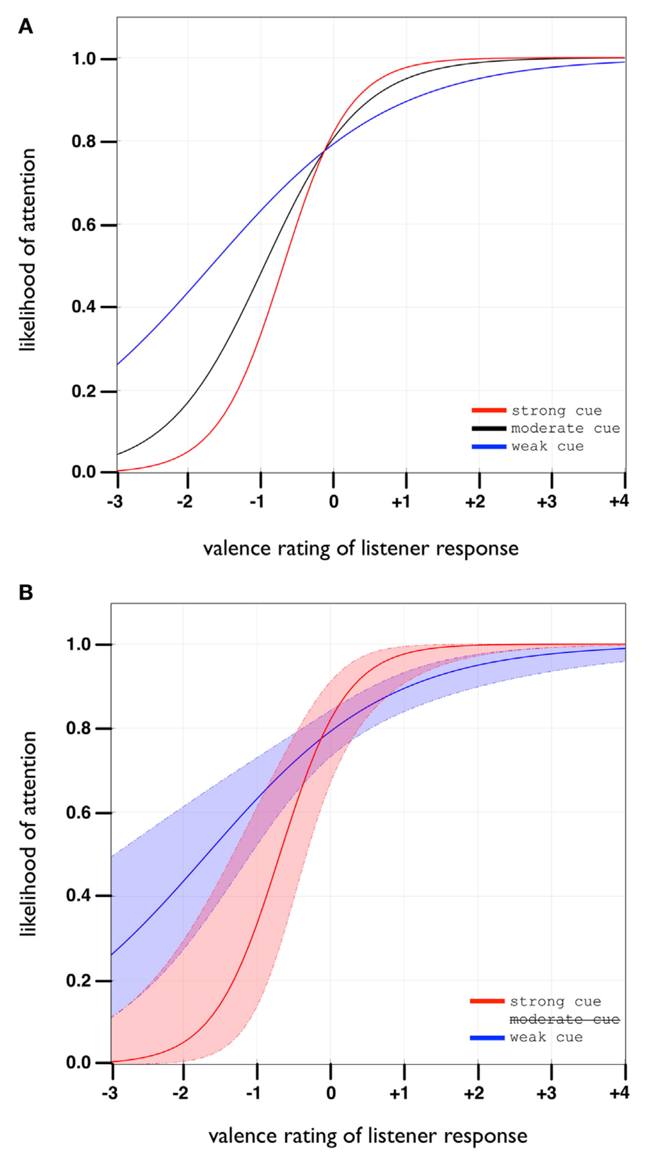


FIGURE 6 | Graphical representation of the logistic regression model from *Section 4.5.2*. **(A)** The model predicts the attentive state of listeners based on cue-strength and response-valence as well as their interaction. The x-axis represents overall listener's response as either very positive (+4) to very negative (−3). The y-axis represents the likelihood of attention, or inversely as inattention. **(B)** Shows the 95% confidence bounds of strong vs weak cue contexts. For the same valence of listener response (e.g., x = −2), there is a difference in interpretation if we observed it after a weak vs a strong cue. Strong cues buy us greater certainty that the listener is not paying attention (likelihood of 70–100% vs 50–70%).

4.6. Discussion

4.6.1. Attention-Related Backchannels of Young Listeners

We identified nonverbal behaviors that are indicative of a child either attentive or inattentive to their partner's storytelling (see summary **Table 7**). We determined the form in which these nuanced behaviors are emitted and differentiate the relevance of a behavior as either a prolonged expression or as a frequent occurrence. Of the behaviors identified, the most unexpected result was the opposing interpretations of frequent versus prolonged brow-raises. Prolonged brow-raises most often co-occurred when listeners were also looking away from storytellers (see Figure S1 in Supplementary Material for a correlation map); their joint emission can serve as a strong signal of a listener losing attention.

4.6.2. Response Rate of Multimodal Speaker Cues

By examining prosodic- and gaze- based cues, we identified multimodal speaker cues, taken singly or in combinations, that can elicit a response from listeners at different rates of success (see summary **Table 8**). Some prosodic cues such as pauses in speech or changes in energy seem to be too subtle for young children to perceive, but their cueing context can be strengthened by adding co-occurring behaviors such as a gaze cue. We confirm prior work by Hess and Johnston (1988) in demonstrating that children respond more often in stronger cue contexts. However, we differentiate our work by finding cues that young listeners know to respond to as well as employ themselves as storytellers.

4.6.3. Magnifying Certainty about Listeners' Attentive State

We found that speaker cues can modify the interpretation of backchannel responses. For the same valence and quality of listener response, there is a difference in interpretation if observed after a weak versus a strong speaker cue. We found that stronger cues buy us greater certainty that a listener is attentive or inattentive. We need both speaker cues and their associated listener responses for an accurate understanding of attention. Backchannels are more informative about the attentive state of listeners when we also know the manner in which they were elicited.

5. GENERAL DISCUSSION

Our primary contribution is introducing the role speaker cues can have in the process of attention inference. We found

TABLE 7 | Listener response summary.

Attentive Behaviors		Inattentive Behaviors	
Frequent	Partner-gazes	Prolonged	Away-gazes
Frequent	Forward-leans	Frequent	Away-leans
Frequent	Brow-raises	Prolonged	Brow-raises
Prolonged	Smiles		
Frequent	Nods		
Frequent	Utterances		

Summary of nonverbal behaviors, as prolonged expressions or frequent occurrences, that are indicative of an attentive or inattentive child listener.

TABLE 8 | Speaker cue summary.

Single Cue	rate	Dual Cues	rate	Tri Cues	rate
Pitch	0.59	Pitch-Energy	0.66	Gaze-Pitch-Energy	0.93
Filled Pause	0.59	Gaze-Pause	0.89	Gaze-Pitch-Pause	0.93
Long Utterance	0.76	Gaze-Pitch	0.90		
Gaze	0.76				

Summary of multimodal cues that children storytellers are observed to use and also can elicit a contingent response from children listeners with varying rates of success.

that speaker cues add interpretive value to attention-related backchannels and also serve as a means to regulate the responsiveness of listeners for those backchannels. Although these findings are based on human–human interaction studies, their implications are noteworthy toward our research goal of developing attention recognition models for social robots. We detail two major implications that will need further validation in an HRI context, which open promising directions for future research.

5.1. Design Implication 1: Modeling the Cueing Actions of Robots Can Increase Attention Recognition Accuracy

Since speaker cues and listener responses are both necessary for accurate attention inference, robot storytellers capable of accounting for their own nonverbal cueing behaviors in their attention models can form more accurate inferences about their human partners. Current approaches to attention recognition primarily focus on modeling the nonverbal behaviors of the sole individual, e.g., just the listener. As we saw in our video-based human-subjects experiment, this approach is akin to asking participants to form accurate inferences about listeners while removing the context of the storyteller. But, inference performance decreases when missing this interpersonal context.

Furthermore, we found that the interpretation of backchannels from listeners depends on whether it was observed after a weak, moderate, or strong speaker cue. A strong cue is a strong elicitation for a response. As such, the cue-response pair is more informative.

By including both the *robot* storyteller's cues and the *human* listener's backchannels, attention recognition models can achieve more accurate predictions especially when used in social situations.

5.2. Design Implication 2: Social Robots Can Pursue a Proactive Form of Attention Recognition in HRI

Since compounded cue contexts have a higher likelihood of eliciting a response from listeners, robot storytellers can manipulate their production of nonverbal speaker cues to deliberately gain more information. In moments of high uncertainty about the listener, a social robot can plan to emit an appropriate cue context to strongly elicit a response that can reduce state uncertainty. Through cueing actions, social robots can pursue a proactive form of inference to better understand their partner's emotional state. Toward this, an immediate extension of this work is to validate whether robot-generated speaker cues result in similar response rates from children listeners. To develop a robot capable of producing these nonverbal cues, we refer readers to our prior work in modeling prosodic-based cues through a rule-based method (Park et al., 2017).

5.3. Limitations

Admittedly, our work does not explicitly include a robot in the studies. But strong evidence exist in demonstrating the readiness of the human mind to respond to technology as social actors—capable of evoking the same social responses as they would with a human partner (Reeves and Nass, 1996; Desteno et al., 2012). We expect our finding from studying human–human interactions will carry over to human-robot interactions. However, further experimental validation is necessary to confirm the effectiveness of robot-generated speaker cues to boost attention recognition accuracies when incorporated into the model and evaluated in a human–robot interaction context.

CONCLUSION

Socially situated robots are not passive observers, but their own nonverbal behaviors contribute to the interaction context and can actively influence the inference process. We argue for a move away from the contextless approaches to emotion recognition, especially for human–robot interaction. A robot's awareness of the contextual effects of its own nonverbal behaviors has an important role in affective computing.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of MIT's Committee on the Use of Humans as Experimental Subjects (COUHES) with written informed consent from all subjects' legal guardian including the publication of subjects' photos. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by MIT's COUHES.

AUTHOR CONTRIBUTIONS

All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take responsibility for the content, design, analysis, interpretation, writing, and/or critical revision of the manuscript. More specifically, CB and JJL developed the concept of robot actions in affective computing and designed human-subjects studies with DD. JJL collected the storytelling demonstrations and performed the set of analyses under the guidance of DD. The manuscript was drafted by JJL and edited by CB and DD for important intellectual merit. JJL, CB, and DD give approval of the final version to be published.

ACKNOWLEDGMENTS

We thank Dr. Paul Harris for the crucial insights regarding children behaviors and Dr. Jesse Gray for advice on video-editing and data-processing.

REFERENCES

Barrett, L. F., Mesquita, B., and Gendron, M. (2011). Context in emotion perception. *Curr. Dir. Psychol. Sci.* 20, 286–290. doi:10.1177/0963721411422522

Desteno, D., Breazeal, C., Frank, R. H., Pizarro, D., Baumann, J., Dickens, L., et al. (2012). Detecting the trustworthiness of novel partners in economic exchange. *Psychol. Sci.* 23, 1549–1556. doi:10.1177/0956797612448793

Dittmann, A. (1972). Developmental factors in conversational behavior. *J. Commun.* 22, 404–423. doi:10.1111/j.1460-2466.1972.tb00165.x

D'Mello, S., and Kory, J. (2015). A review and meta-analysis of multimodal affect detection systems. *ACM Comput. Surveys* 47, 1–36. doi:10.1145/2682899

Duncan, S., and Fiske, D. W. (1977). *Face-to-Face Interaction: Research, Methods, and Theory*. Cambridge University Press.

Ekman, P. (1984). "Expression and the nature of emotion," in *Approaches to Emotion*, eds K. Scherer and P. Ekman (Hillsdale, NJ: Lawrence Erlbaum), 319–343.

Gratch, J., Wang, N., Gerten, J., Fast, E., and Duffy, R. (2007). "Creating Rapport with Virtual Agents," in *Proceedings of the International Conference on Intelligent Virtual Agents* (Paris, France), 125–138.

Gravano, A., and Hirschberg, J. (2009). "Backchannel-inviting cues in task-oriented dialogue," in *Proceedings of the International Conference of INTERSPEECH* (Brighton, UK), 1019–1022.

Hassin, R. R., Aviezer, H., and Bentin, S. (2013). Inherently ambiguous: facial expressions of emotions, in context. *Emot. Rev.* 5, 60–65. doi:10.1177/1754073912451331

Hess, L., and Johnston, J. (1988). Acquisition of back channel listener responses to adequate messages. *Discourse Process.* 11, 319–335. doi:10.1080/01638538809544706

Huang, L., Morency, L.-P., and Gratch, J. (2010). "Parasocial consensus sampling: combining multiple perspectives to learn virtual human behavior," in *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems* (Toronto, Canada), 1265–1272.

Jaques, N., McDuff, D., Kim, Y. L., and Picard, R. W. (2016). "Understanding and predicting bonding in conversations using thin slices of facial expressions and body language," in *Proceedings of the International Conference on Intelligent Virtual Agents* (Los Angeles, CA), 64–74.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychol.* 26, 22–63. doi:10.1016/0001-6918(67)90005-4

Knapp, M., and Hall, J. (2010). *Nonverbal Communication in Human Interaction*. Boston, MA: Wadsworth Publishing.

Lee, J. J., Knox, W. B., Wormwood, J. B., Breazeal, C., and DeSteno, D. (2013). Computationally modeling interpersonal trust. *Front. Psychol.* 4:893. doi:10.3389/fpsyg.2013.00893

Maynard, S. (1997). Analyzing interactional management in native/non-native English conversation: a case of listener response. *Int. Rev. Appl. Linguist. Lang. Teach.* 35, 37–60.

Miller, L., Lechner, R., and Rugs, D. (1985). Development of conversational responsiveness: preschoolers' use of responsive listener cues and relevant comments. *Dev. Psychol.* 21, 473–480. doi:10.1037/0012-1649.21.3.473

Morency, L.-P., de Kok, I., and Gratch, J. (2010). A probabilistic multimodal approach for predicting listener backchannels. *Auton. Agents Multi Agent Syst.* 20, 70–84. doi:10.1007/s10458-009-9092-y

FUNDING

We acknowledge the support of the National Science Foundation (grant IIS-1138986), Dolores Zohrab Liebmann Fellowship, and William Asbjornsen Albert Memorial Fellowship.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http://www.frontiersin.org/article/10.3389/frobt.2017.00047/full#supplementary-material.

VIDEO S1 | The FALSE condition for Listener-2 in **Figure 1A**.

VIDEO S2 | The ABSENT condition for Listener-2.

VIDEO S3 | The TRUE condition for Listener-2.

Nowicki, S., and Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: the diagnostic analysis of nonverbal accuracy scale. *J. Nonverbal Behav.* 18, 9–35. doi:10.1007/BF02169077

Otsuka, K., Sawada, H., and Yamato, J. (2007). "Automatic inference of cross-modal nonverbal interactions in multiparty conversations," in *Proceedings of the International Conference on Multimodal Interaction* (Nagoya, Aichi, Japan), 255–262.

Park, H. W., Gelsomini, M., Lee, J. J., and Breazeal, C. (2017). "Telling stories to robots: the effect of backchanneling on a child's storytelling," in *Proceedings of the International Conference on Human-Robot Interaction* (Vienna, Austria), 100–108.

Reeves, B., and Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. New York, NY: Cambridge University Press.

Sariyanidi, E., Cavallaro, A., and Gunes, H. (2015). Automatic analysis of facial affect: a survey of registration, representation, and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 37, 1113–1133. doi:10.1109/TPAMI.2014.2366127

Scassellati, B. (1999). "Imitation and mechanisms of joint attention: a developmental structure for building social skills on a humanoid robot," in *Computation for Metaphors, Analogy, and Agents*, ed. C. L. Nehaniv (Berlin, Heidelberg, Germany: Springer), 176–195.

Schegloff, E. (1982). "Discourse as an interactional achievement: some uses of 'uh huh' and other things that come between sentences," in *Analyzing Discourse: Text and Talk*, ed. D. Tannen (Washington, DC: Georgetown University Press), 71–93.

Thórisson, K. R. (2002). Natural turn-taking needs no manual: computational theory and model, from perception to action. *Multimodality Lang. Speech Syst.* 19, 173–207. doi:10.1007/978-94-017-2367-1_8

Ward, N., and Tsukahara, W. (2000). Prosodic features which cue back-channel responses in English and Japanese. *J. Pragmat.* 32, 1177–1207. doi:10.1016/S0378-2166(99)00109-5

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). "ELAN: a professional framework for multimodality research," in *Proceedings of the International Conference on Language Resources and Evaluation* (Genoa, Italy), 1556–1559.

Yu, Z., Gerritsen, D., Ogan, A., Black, A. W., and Cassell, J. (2013). "Automatic prediction of friendship via multi-model dyadic features," in *Proceedings of the SIGdial Meeting on Discourse and Dialogue* (Metz, France), 51–60.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Lee, Breazeal and DeSteno. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers
in Robotics and AI

ORIGINAL RESEARCH
published: 12 July 2017
doi: 10.3389/frobt.2017.00029

Narratives with Robots: The Impact of Interaction Context and Individual Differences on Story Recall and Emotional Understanding

Iolanda Leite[1], Marissa McCoy[2], Monika Lohani[3], Daniel Ullman[1], Nicole Salomons[1], Charlene Stokes[4], Susan Rivers[5] and Brian Scassellati[1]*

[1] Social Robotics Lab, Department of Computer Science, Yale University, New Haven, CT, United States, [2] Ball Aerospace, Boston, MA, United States, [3] Yale Center for Emotional Intelligence, Department of Psychology, Yale University, New Haven, CT, United States, [4] Air Force Research Laboratory, Dayton, OH, United States, [5] iThrive, Boston, MA, United States

OPEN ACCESS

Edited by:
Hatice Gunes,
University of Cambridge,
United Kingdom

Reviewed by:
Erol Sahin,
Middle East Technical
University, Turkey
Rodolphe Gelin,
Aldebaran Robotics, France

****Correspondence:***
Iolanda Leite
iolanda@kth.se

Specialty section:
This article was submitted to
Humanoid Robotics,
a section of the journal
Frontiers in Robotics and AI

Received: *30 December 2016*
Accepted: *12 June 2017*
Published: *12 July 2017*

Citation:
Leite I, McCoy M, Lohani M, Ullman D, Salomons N, Stokes C, Rivers S and Scassellati B (2017) Narratives with Robots: The Impact of Interaction Context and Individual Differences on Story Recall and Emotional Understanding. Front. Robot. AI 4:29. doi: 10.3389/frobt.2017.00029

Role-play scenarios have been considered a successful learning space for children to develop their social and emotional abilities. In this paper, we investigate whether socially assistive robots in role-playing settings are as effective with small groups of children as they are with a single child and whether individual factors such as gender, grade level (first vs. second), perception of the robots (peer vs. adult), and empathy level (low vs. high) play a role in these two interaction contexts. We conducted a three-week repeated exposure experiment where 40 children interacted with socially assistive robotic characters that acted out interactive stories around words that contribute to expanding children's emotional vocabulary. Our results showed that although participants who interacted alone with the robots recalled the stories better than participants in the group condition, no significant differences were found in children's emotional interpretation of the narratives. With regard to individual differences, we found that a single child setting appeared more appropriate to first graders than a group setting, empathy level is an important predictor for emotional understanding of the narratives, and children's performance varies depending on their perception of the robots (peer vs. adult) in the two conditions.

Keywords: socially assistive robotics, emotional intelligence, individual differences, multiparty interaction

1. INTRODUCTION

The typical use case for socially assistive robotics applications involves one robot and one user (Tapus et al., 2007). As assistive technology becomes more sophisticated, and as robots are being used more broadly in interventions, there arises a need to explore other types of interactions. Contrasting the typical "one robot to one user" and "one robot to many users" situations, there are cases where it is desirable to have multiple robots interacting with one user or multiple users. As an example, consider the case of role-playing activities in emotionally charged domains (e.g., bullying prevention, domestic violence, or hostage scenarios). In these cases, taking an active role in the interaction may bring about undesirable consequences, while observing the interaction might serve as a learning experience. Here, robots offer an inexpensive alternative to human actors, displaying controlled behavior across interventions with different trainees.

Our goal is to use socially assistive robots to help children build their emotional intelligence skills through interactive storytelling activities. Storytelling is an effective tool for creating a memorable and creative learning space where children can develop cognitive skills (e.g., structured oral summaries, listening, and verbal aptitude), while also bolstering social-emotional abilities (e.g., perspective-taking, mental state inference). Narrative recall, for example, prompts children to logically reconstruct a series of events, while explaining behavior and attributing mental and emotional states to story characters (Capps et al., 2000; John et al., 2003; McCabe et al., 2008). As this is a novel research direction, several questions can be posed. What is the effect of having multiple robots in the scene or, more importantly, what is the optimal context of interaction for these interventions? Should the interaction context focus on groups of children (as in traditional role-playing activities) or should we aim for single child interactions, following the current trend in socially assistive robotics?

In our previous work, we began addressing the question of whether socially assistive robots are as effective with small groups of children as they are with a single child (Leite et al., 2015a). While we found that the interaction context can impact children's learning, we anticipate that it might not be the only contributing factor. Previous research exploring the effects of single versus small group learning with technology points out that the two different contexts can be affected by learner characteristics such as gender, grade level, and ability level (Lou et al., 2001). A recent Human–Robot Interaction (HRI) study suggests that children behave differently when interacting alone or in dyads with a social robot (Baxter et al., 2013). However, it remains unknown whether interaction context and individual factors impact the effectiveness of robot interventions in terms of how much users can learn or recall from the interaction. In this paper, we extend previous work by investigating whether individual human factors play a role in different interaction contexts (single vs. group). For example, do individual differences such as gender or empathy level influence the optimal interaction context for socially assistive robotics interventions?

To address these questions, we developed an interactive narrative scenario where a pair of robotic characters played out stories centered around words that contribute to expanding children's emotional vocabulary (Rivers et al., 2013). To evaluate the effects of interaction context (single vs. group), we conducted a three-week repeated exposure study where children interacted with the robots either alone or in small groups, and then were individually asked questions on the interaction they had just witnessed. We analyzed interview responses in order to measure participants' story recall and emotional understanding abilities and looked into individual differences that might affect these measures. Our results show that although children interacting alone with the robots were able to recall the narrative more accurately, no significant differences were found in the understanding of the emotional context of the stories. Furthermore, we found that individual differences such as grade, empathy level, and perception of the robots are important predictors of the optimal interaction for students. We discuss these implications for the future design of robot technology in learning environments.

2. BACKGROUND

2.1. Learning Alone or in Small Groups

Educational research highlights the benefits of learning in small groups as compared to learning alone (Pai et al., 2015). These findings also apply to learning activities supported by computers (Dillenbourg, 1999; Lou et al., 2001). It has long been acknowledged that groups outperform individuals in a variety of learning tasks such as concept attainment, creativity, and problem solving (Hill, 1982). More recently, Schultze et al. (2012) conducted a controlled experiment to show that groups perform better than individuals in quantitative judgments. Interestingly, the authors attribute this finding to within-group interactions instead of weighting the individual judgment of each group member. A situation in which "two or more people learn or attempt to learn something together" is often referred to as *collaborative learning* (Dillenbourg, 1999).

It is important to note, however, that most of these findings were obtained with adults. Additionally, as previously stated, interaction context (i.e., whether students are alone or in small groups) is not the only factor that influences learning. In a meta-review focusing on computer-supported learning, Lou et al. (2001) enumerated several learning characteristics that can affect learning as much as interaction context. Factors such as ability level, gender, grade, or past experience with computers are among the most common individual differences that affect learning.

2.2. Individual Differences in Narrative Recall and Understanding

Research has shown that a child's ability to reconstruct a cohesive and nuanced narrative develops with age (Griffith et al., 1986; Crais and Lorch, 1994; Bliss et al., 1998; John et al., 2003). While three-year-olds tend to focus on an isolated event within a narrative, by the time a child is five, the capacity to create a more structured narrative with a logical sequence of story events is already developed (John et al., 2003). Among seven to eleven-year-olds, Griffith et al. (1986) found more story inaccuracies in older children's retellings as their narratives became longer. A central notion of the current study is to increase emotional understanding by allowing children to see the impact of their decisions play out in the story. As such, knowing at what age children begin to develop the capacity to attribute meaning to characters' behaviors is important. While the foundations of a story – story setting, opening scene, and story conclusion – are typically included in narratives of children aged four to six, the presence of a character's thoughts and intentions within a narrative takes longer to develop (Morrow, 1985). Identifying more overt story structure elements may be easier for children than attributing meaning, intentions, and emotions to a character's behavior (Renz et al., 2003). A study carried out by Camras and Allison (1985) found that when emotion-laden stories are given to children from kindergarten to second grade, the accuracy of children's emotion labeling improved with age.

Several authors have found gender differences in narrative recall and understanding. For example, research shows that females are more verbal than males (Smedler and Törestad, 1996; Buckner and Fivush, 1998; Crow et al., 1998) and excel on verbal tasks (Bolla-Wilson and Bleecker, 1986; Capitani et al., 1998), while males are more successful at spatial tasks (Maccoby and Jacklin, 1974; Linn and Petersen, 1985; Iaria et al., 2003). However, Andreano and Cahill (2009) found that gender differences are more nuanced and extensive, with females outperforming males in spatial, autobiographical abilities, and general episodic memory. In a test of verbal learning among children aged 5–16, females outperformed males in long-term memory recall and delayed recognition, while males produced more intrusion errors (Kramer et al., 1997). Yet, Maccoby and Jacklin (1974) concluded that "the two genders show a remarkable degree of similarity in the basic intellectual processes of perception, learning, and memory." Additionally, females tend to generate more accurate (Pohl et al., 2005), detailed (Ross and Holmberg, 1992), and exhaustive narratives that take social context and emotions into account (Buckner and Fivush, 1998; John et al., 2003). Females are also generally thought to be more emotive both verbally (Smedler and Törestad, 1996) and non-verbally (Briton and Hall, 1995). Gender differences in emotional dialog and understanding are broadly attributed to the view that, beginning in early childhood, girls are socialized to be more emotionally attuned and, therefore, more skilled at perspective-taking (Hoffman, 1977; Greif et al., 1981; Dunn et al., 1987).

2.3. Individual Differences in Emotional Intelligence

Emotions are functional and impact our attention, memory, and learning (Rivers et al., 2013). Emotional intelligence (EI) is defined as "the ability to monitor one's own and other's feelings and emotions, to discriminate among them and to use this information to guide one's thinking and action" (Salovey and Mayer, 1990). Previous research has determined that a child's emotional understanding advances with age (Pons et al., 2004; Harris, 2008). The ability to recognize basic emotions and understand that emotions is affected by external causes, which is generally established by the age of 3–4 (Yuill, 1984; Denham, 1986). Between 3 and 6 years, children begin to understand how emotions are impacted by desires, beliefs, and time (Harris, 1983; Yuill, 1984), while children aged 6–7 begin to explore strategies for emotion regulation (Harris, 1989).

In terms of gender, Petrides and Furnham (2000) concluded that females scored higher than males on the "social skills" factor of measured trait EI, and a cross-cultural study carried out by Collis (1996) found that females had higher empathy than males at the first-grade level. These results are reinforced with findings from a meta-analysis of 16 studies in which females scored higher on self-reported empathy (Eisenberg and Lennon, 1983).

3. RELATED WORK

In this section, we review previous research in the three main research thrusts that inform this work: robots for education, multiparty interactions, and individual differences in HRI.

3.1. Robots As Educational Tools

Kim and Baylor (2006) posit that the use of non-human pedagogical agents as learning companions creates the best possible environment for learning for a child. Virtual agents are designed to provide the user with the most interactive experience possible; however, research by Bainbridge et al. (2011) indicated that physical presence matters in addition to embodiment, with participants in a task rating an overall more positive interaction when the robot was physically embodied rather than virtually embodied. Furthermore, Leyzberg et al. (2012) found that the students who showed the greatest measurable learning gains in a cognitive skill learning task were those who interacted with a physically embodied robot tutor (a Keepon robot), as compared to a video-represented robot and a disembodied voice.

Research by Mercer (1996) supports talk as a social mode of thinking, with talk in the interaction between learners beneficial to educational activities. However, Mercer identifies the need for focused direction from a teaching figure for the interaction to be as effective as possible. In line with these findings, Saerbeck et al. (2010) showed the positive effects of social robots in language learning, especially when the robot was programmed with appropriate socially supportive behaviors.

A great deal of research has been conducted into the use of artificial characters in the context of educational interactive storytelling with children. Embodied conversational agents are structured using a foundation of human-human conversation, creating agents that appear on a screen and interact with a human user (Cassell, 2000). Interactive narratives, where users can influence the storyline through actions and interact with the characters, result in engaging experiences (Schoenau-Fog, 2011) and increase a user's desire to keep interacting with the system (Kelleher et al., 2007; Hoffman et al., 2008). *FearNot* is a virtual simulation with different bullying episodes where a child can take an active role in the story by advising the victim on possible coping strategies to handle the bullying situation. An extensive evaluation of this software in schools showed promising results on the use of such tools in bullying prevention (Vannini et al., 2011). Although some authors have explored the idea of robots as actors (Bruce et al., 2000; Breazeal et al., 2003; Hoffman et al., 2008; Lu and Smart, 2011), most of the interactive storytelling applications so far are designed for virtual environments.

3.2. Multiparty Interactions with Robots

Research on design and evaluation of robots that interact with groups of users has become very prominent in the past few years in several application domains such as education (Kanda et al., 2007; Al Moubayed et al., 2012; Foster et al., 2012; Gomez et al., 2012; Johansson et al., 2013; Bohus et al., 2014; Pereira et al., 2014). Despite this trend, few authors investigated differences between one single user and a group of users interacting with a robot in the same setting.

One of the exceptions is Baxter et al. (2013), who reported a preliminary analysis that consisted of a single child or a pair of children interacting with a robot in a sorting game. Their observations indicate differences between the two conditions:

when alone with the robot, children seem to treat it more as a social entity (e.g., engage in turn-taking and shared gaze with the robot), while these behaviors are less common when another peer is in the room.

Shahid et al. (2014) conducted a cross-cultural examination of variation between interactions in children who either played a game alone, with a robot, or with another child. They found that children both enjoyed playing more and were more expressive when they played with the robot, as compared to when they played alone; unsurprisingly, children who played with a friend showed the highest levels of enjoyment of all groups.

With this previous research serving as foundation, one of the goals of our work is to investigate whether interactions with robots in a group setting could benefit information retainment and emotional understanding. However, in addition to interaction context (single vs. group), there might be other individual factors contributing to these differences.

3.3. Individual Differences in Human–Robot Interaction

One of the underlying aims of studying individual differences in HRI is personalization. Understanding how different user groups perceive and react to robots, and adapting the robot's behavior accordingly, can result in more effective and natural interactions. Andrist et al. (2015) provided one of the first empirical validations on the positive effects of personalization. In a controlled study where a social robot matched each participant's extroversion personality dimension through gaze, they showed that introverted subjects had a marginally significant preference for the robot displaying introverted behaviors and that both introverts and extroverts showed higher compliance when interacting with the robot that matched their personality dimension.

Most of the research reporting individual differences in HRI so far has mainly focused on gender (Mutlu et al., 2006; Nomura et al., 2008; Schermerhorn et al., 2008; MacDorman and Entezari, 2015) and certain personality traits such as extroversion and agreeableness (Walters et al., 2005; Syrdal et al., 2007; Takayama and Pantofaru, 2009; Andrist et al., 2015), but there are also studies exploring other factors such as pet ownership (Takayama and Pantofaru, 2009) and perception of robots (Nomura et al., 2008; Schermerhorn et al., 2008; Mumm and Mutlu, 2011; MacDorman and Entezari, 2015).

Gender seems to play an important role in individual's perceptions and attitudes toward robots. In a study where a storytelling robot recited a fairy tale to two participants, Mutlu et al. (2006) manipulated the robot's gaze behavior by having the robot look at one of the participants 80% of the time. This manipulation had a significant interaction effect on gender, with males who were looked at more rating the robot more positively, and females who were looked at less rating the robot more positively. More recently, the same authors investigated gender differences (among other factors) in a scenario where the robot was able to monitor participants' attention using brain electrophysiology and adapt its behavior accordingly (Szafir and Mutlu, 2012). Females interacting with the adaptive robot gave higher ratings in rapport toward the robot and self-motivation, while no significant differences were found for males on the same measures. In the studies conducted to validate the Negative Attitudes Toward Robots Scale (NARS) and Robot Anxiety Scale (RAS), Nomura et al. (2008) found several gender effects. For instance, males with higher NARS and RAS scores talked less to the robot and avoided touching it. Schermerhorn et al. (2008) also reported gender effects on people's ratings of social presence toward robots, with males perceiving a robot as more human-like and females perceiving it as more machine-like and less socially desirable. These findings are in line with results obtained by MacDorman and Entezari (2015) in their investigation into whether individual differences can predict sensitivity to the uncanny valley. They found significant correlations between gender and android eerie ratings; females in this study perceived android robots as more eerie than males.

Individual differences have been explored as a way to better understand proxemic preferences between people and robots. Walters et al. (2005) investigated the effects of people's personality traits on their comfortable social distances while approaching a robot. Results showed that more proactive people felt more comfortable standing further away from the robot. In a follow-up study (Syrdal et al., 2007), researchers from the same group found that people with high extroversion and low conscientiousness scores let a robot get closer when they were in control of the robot, as opposed to when they believed the robot to be in control of itself. Takayama and Pantofaru (2009) confirmed the hypothesis that pet owners felt more comfortable with being closer to robots, a result that also held true for people with past experience with robots. Additionally, they found that proxemic comfort levels were related to the agreeableness personality trait, with more agreeable people experiencing higher levels of comfort closer to a robot than participants rated as less agreeable in the personality questionnaire. More recently, Mumm and Mutlu (2011) reported significant differences in the effects of gender on proxemics, with males distancing themselves significantly further than females. Another interesting factor that played an effect in this study was robot likability: people who reported disliking the robot positioned themselves further away in a condition where the robot tried to establish mutual eye gaze with the subjects.

4. RESEARCH QUESTIONS AND HYPOTHESES

The main goal of the research presented in this paper is to investigate whether the social context of the interaction, i.e., children interacting with robots alone or in a small group, has an impact on information recall and understanding of the learning content. Our research goals can be translated into two main research questions:

RQ1 *How does interaction context impact children's information recall?*

RQ1 *How does interaction context impact children's emotional understanding and vocabulary?*

As previously outlined, socially assistive robotic applications are typically one-on-one, but educational research suggests that learning gains may increase in a group setting (Hill, 1982; Pai et al., 2015). To further understand these questions, we explored individual factors that may impact how children perform in learning environments and/or how users perceive the robots. Considering previous findings on individual differences presented in sections 2 and 3, as well as our particular application domain, we took into account gender, grade level (first vs. second), perception of the robot (peer vs. adult), and empathy level (low vs. high). To further explore the human individual factors that influence recall and understanding in single versus group interactions, we outlined the following hypotheses:

H1 *Second graders will achieve higher performance than first graders in narrative recall and emotional understanding.*
In comparison to first graders, second graders are more developmentally advanced both cognitively and emotionally, so we hypothesize that second-grade students will perform better since narrative recall abilities and emotional understanding tend to develop with age (Morrow, 1985; Crais and Lorch, 1994; Bliss et al., 1998; John et al., 2003). Although a number of previous studies report age instead of grade, we determined that grade is more fitting for our study as the social and emotional learning curriculum (see section 5.2) employed in the school where our study was conducted is grade-dependent. For this reason, we predict that grade level could be a better explanatory factor than age.

H2 *Females will achieve higher performance than males in narrative recall and emotional understanding.*
Previous research has shown that females tend to tell more accurate (Pohl et al., 2005) and detailed (Ross and Holmberg, 1992) narratives, while accounting for the emotions of the narrative characters more often (Buckner and Fivush, 1998; John et al., 2003). Additionally, several authors found that females scored higher in emotional intelligence tests (Collis, 1996; Petrides and Furnham, 2000). For these reasons, we hypothesize that females will perform better than their male counterparts.

H3 *Higher empathy students will achieve higher scores in emotional understanding.*
Because we anticipate a positive correlation between high empathy and high emotional intelligence (Salovey and Mayer, 1990), we hypothesize that individuals with higher empathy will be better at emotional understanding.

H4 *Children's perceived role of the robot will affect their narrative recall and emotional understanding abilities.*
Considering the extensive HRI literature showing that perception of robots changes how individuals perform and interact with them (Nomura et al., 2008; Mumm and Mutlu, 2011; MacDorman and Entezari, 2015), we expect perception of robots to affect our main measures. As most robots used in educational domains are viewed by students as either peers or adult tutors (Mubin et al., 2013), we will gage perception within these two opposite roles.

5. AFFECTIVE NARRATIVES WITH ROBOTIC CHARACTERS

We developed an interactive narrative system such that any number of robotic characters can act out stories defined in a script. This system prompts children to control the actions of one of the robots at specific moments, allowing the child to see the impact of their decision on the course of the story. By exploring all the different options in these interactive scenarios, children have the opportunity to see how the effects of their decisions play out before them, without the cost of first having to make these decisions in the real world. This section describes the architecture of this system and introduces RULER, a validated framework for promoting emotional literacy that inspired the interactive stories developed for this system.

5.1. System Architecture

The central component of the narrative system is the story manager, which interprets the story scripts and communicates with the robot controller modules and the tablet (see diagram in **Figure 1**). The scripts contain, in a representation that can be interpreted by the story manager, every possible scene episode. A scene contains the dialog lines of each robot and a list of the next scene options that can be selected by the user. Each dialog line contains an identifier of the robot playing that line, the path to a sound file, and a descriptor of a non-verbal behavior for the robot to display while "saying" that line (e.g., happy, bouncing). When the robots finish playing out a scene, the next story options are presented on the tablet as text with an accompanying illustration. When the user selects a new story option on the tablet, the story manager loads that scene and begins sending commands to the robots based on the scene dialog lines.

The system was implemented on Robot Operating System (ROS) (Quigley et al., 2009). The story manager is a ROS node that publishes messages subscribed by the active robot controller nodes. Each robot controller node is instantiated with a robotID parameter, so that each node can ignore the messages directed to the other characters in the scene. The tablet communicates with the story manager module using a TCP socket connection over Wi-Fi.

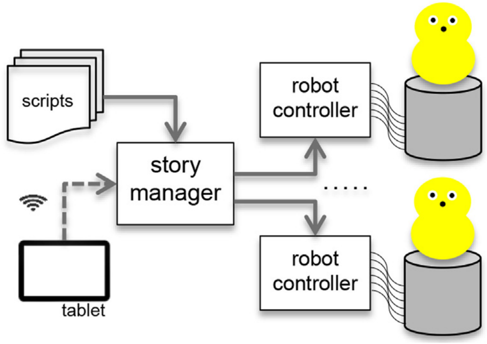


FIGURE 1 | Diagram of the different modules and hardware comprising our system.

The robot platforms used in this implementation were two MyKeepon robots (see **Figure 2**) with programmable servos controlled by an Arduino board (Admoni et al., 2015). MyKeepon is a 32 cm tall, snowman-like robot with three dots representing eyes and a nose. Despite their minimal appearance, these robots have been shown to elicit social responses from children (Kozima et al., 2009). Each robot has four degrees of freedom: it can pan to the sides, roll to the sides, tilt forward and backward, and bop up and down. To complement the prerecorded utterances, we developed several non-verbal behaviors such as idling, talking, and bouncing. All the story authoring was done in the script files, except the robot animations and tablet artwork. In addition to increased modularity, this design choice allows non-expert users (e.g., teachers) to develop new content for the system.

5.2. The RULER Framework

RULER is a validated framework rooted in emotional intelligence theory (Salovey and Mayer, 1990) and research on emotional development (Denham, 1998) that is designed to promote and teach emotional intelligence skills. Through a comprehensive approach that is integrated into existing academic curriculum, RULER focuses on skill-building lessons and activities around Recognizing, Understanding, Labeling, Expressing, and Regulating emotions in socially appropriate ways (Rivers et al., 2013). Understanding the significance of emotional states guides attention, decision-making, and behavioral responses, and is necessary in order to navigate the social world (Lopes et al., 2005; Brackett et al., 2011).

This study employs components of RULER, including the Mood Meter, a tool that students and educators use as a way to identify and label their emotional state, and the Feeling Words Curriculum, a tool that centers on fostering an extensive feelings vocabulary that can be applied in students' everyday lives. The story scripts are grounded in the Feeling Words Curriculum and are intended to encourage participants to choose the most appropriate story choice after considering the impact of each option. Our target age group was first to second graders (6–8 years old). Prior to beginning the study, we gathered feedback from elementary school teachers to ensure that the vocabulary and difficulty levels of story comprehension were age-appropriate. A summary of the scenes forming the scripts of each session is displayed in **Table 1**. All three stories followed the same structure: introduction scene, followed by three options. Each option impacted the story and the characters' emotional state in different ways.

6. EXPERIMENTAL METHOD

In order to investigate the research questions and hypotheses outlined earlier, we conducted a user study using the system described in the previous section.

6.1. Participants

The participants in the study were first- and second-grade students from an elementary school where RULER had been implemented. A total of 46 participants were recruited in the school where the study was conducted, but six participants were excluded for various reasons (i.e., technical problems in collecting data or participants missing school). For this analysis, we

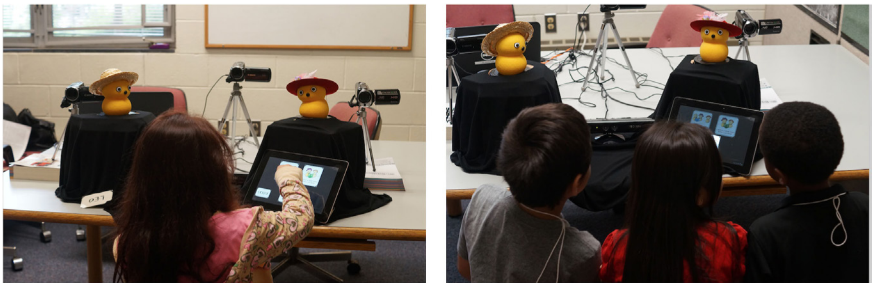


FIGURE 2 | Children interacting with the robots in the single (left) and group (right) conditions.

TABLE 1 | Summary of the story scenes in each session.

	Session 1	Session 2	Session 3
Feeling word	Included	Frustration	Cooperation
Difficulty level	Easy	Hard	Medium
Intro scene	Leo is new at school and does not know anyone. Another student in class, Marlow, called Leo's hat stupid. What should Berry do to help Leo feel included?	Berry tells Leo that he just started a new book as part of an assignment, but some of the words are too hard for him to read. What should Berry do to get through his frustration?	Berry has just mastered a big, hard book on his own. Leo asks Berry to be his reading buddy. Leo wants to read an easier book that's on his reading level, while Berry wants to try reading the hardest books. What should Berry do to be cooperative?
Optional scenes	A. Talk bad about Marlow	A. Ask Leo to read the book	A. Find another reading buddy
	B. Tell Leo how cool Marlow is	B. Wait for the teacher	B. Choose a book both can read
	C. Ask Leo to play	C. Try again	C. Choose a hard book anyway

considered a total of 40 children (22 females, 18 males) between the age of 6 and 8 years (mean (M) = 7.53, standard deviation (SD) = 0.51). Out of these 40 students, 21 were first graders and 19 were second graders.

Ethnicity, as reported by guardians, was as follows: 17.5% African American, 17.5% Caucasian, 25% Hispanic, and 27.5% reported more than one ethnicity (12.5% missing data). The annual income reported by guardians was as follows: 30% in \$0–\$20,000, 42.5% in \$20,000–\$50,000, and 10% in the \$50,000–\$100,000 range (17.5% missing data).

6.2. Design

We used a between-subjects design with participants randomly sorted into one of two conditions: *single* (one participant interacted alone with the robots) or *group* (three participants interacted with the robots at the same time). We studied groups of three children as three members is the smallest number of members considered to be a group (Moreland, 2010). Our main dependent metrics focused on participants' recall abilities and emotional interpretation of the narrative choices.

Each participant or group of participants interacted with the robots three times, once per week. Participants in the group condition always interacted with the robots in the same groups. The design choice to use repeated interactions was not to measure learning gains over time, but to ensure that the results were not affected by a novelty effect that robots often evoke in children (Leite et al., 2013).

6.3. Procedure

The study was approved by an Institutional Review Board. Parental consent forms were distributed in classrooms that had agreed to participate in the study. Participants were randomly assigned to either the single condition (19 participants) or group condition (21 participants). Each session lasted approximately 30 min with each participant. The participant first interacted with the robots either alone or in a small group (approximately 15 min), and then was interviewed individually by an experimenter (approximately 15 min).

Participants were escorted from class by a guide who explained that they were going to interact with robots and then would be asked questions about the interaction. In addition to parental consent, the child was introduced to the experimenter and asked for verbal assent. The experimenter began by introducing the participants to Leo and Berry, the two main characters (MyKeepon robots) in the study. The first half of each session involved the participants interacting with the robots as the robots autonomously role-played a scenario centered around a RULER feeling word. After observing the scenario introduction, participants were presented with three different options. Participants were instructed to first select the option they thought was the best choice and were told they would then have the opportunity to choose the other two options. In the group condition, participants were asked to make a joint decision. The experimenter was present in the room at all times, but was outside participants' line of sight.

After interacting with the robots, participants were interviewed by additional experimenters. The interviews had the same format for both conditions, which means that even participants in the group condition were interviewed individually. Interviews were conducted in nearby rooms. Experimenters followed a protocol that asked the same series of questions (one open-ended question, followed by two direct questions) for each of the four scenes (i.e., Introduction, Option A, Option B, and Option C) that comprised one session. The same three repeated questions were asked in the following order:

1. What happened after you chose <option>?
2. After you chose <option>, what color of the Mood Meter do you think <character> was in?
3. What word would you use to describe how <character> was feeling?

These questions were repeated for a total of 36 times (3 questions × 4 scenes per session × 3 sessions) over the course of the study. If a participant remained silent for more than 10 s after being asked a question, the experimenter asked, "Would you like me to repeat the question or would you like to move on?" The interviewer used small cards with artwork representing the different scene choices similar to the ones that appeared on the tablet near the robots. Interviews were audio-recorded and transcribed verbatim for coding.

All three sessions followed the same format (i.e., robot interaction followed by the series of interview questions). Additionally, in the second session we employed an adapted version of Bryant's Empathy Scale (Bryant, 1982) to measure children's empathy index, and in the third session we measured perception of the robots (peer vs. adult) using a scale specifically developed for this study. For the empathy assessment, the interviewer asked participants to sort each one of the scale items, printed on small cards, between two boxes, "me" or "not me." A similar box task procedure was followed in the third session for collecting perception of the robots, but this time children were asked to sort cards with activities they would like to do with Leo and Berry.

6.4. Interview Coding

6.4.1. Word Count

The number of words uttered by each participant during the interview was counted using an automated script. Placeholders such as "umm" or "uhh" did not contribute toward word count. This metric was mainly used as a manipulation check for the other measures.

6.4.2. Story Recall

Responses to the open-ended question "What happened after you chose <option>?" were used to measure story recall through the Narrative Structure Score (NSS). Similar recall metrics have been previously used in HRI studies with adults (Szafir and Mutlu, 2012).

We followed the coding scheme used in previous research by McGuigan and Salmon (2006) and McCartney and Nelson (1981), in which participants' verbal responses to open-ended questions were coded for the presence or absence of core characters (e.g., Leo, Berry) and core ideas (e.g., Leo does not know anyone, everyone is staring at Leo's clothes). This score provides a snapshot of the participants' "ability to logically recount the fundamental plot elements of the story." For session *S* and participant *i*, NSS was computed using the following formula:

$$NSS_{S,I} = \frac{Mentioned(CoreCharacters + CoreIdeas)}{All(CoreCharacters + CoreIdeas)}$$

A perfect NSS of 1.0 would indicate that the participant mentioned all the core characters and main ideas in all four open-ended questions of that interview. The first mention of core characters and core ideas was given a point each, with additional mentions not counted. The sum of core characters and core ideas for each interview session were combined to generate the Narrative Structure Score. The average number of characters in each story was three (Leo, Berry, and Marlow or the teacher), while the number of core ideas varied depending on the difficulty of the story, ranging from an average of four in the easiest story to six in the hardest.

6.4.3. Emotional Understanding

The Emotional Understanding Score (EUS) represents participants' ability to correctly recognize and label character's emotional states, a fundamental skill of emotional intelligence (Brackett et al., 2011; Castillo et al., 2013). Responses to the two direct questions "After you chose <option>, what color of the Mood Meter do you think <character> was in?" and "What word would you use to describe how <character> was feeling?" were coded based on RULER concepts and combined to comprise EUS.

Appropriate responses for the first question were based on the Mood Meter colors and included Yellow (pleasant, high energy), Green (pleasant, low energy), Blue (unpleasant, low energy), or Red (unpleasant, high energy), depending on the emotional state of the robots at specific points in the role-play. Responses to the second direct question were based on the RULER Feeling Words Curriculum with potential appropriate responses being words such as excited (pleasant, high energy), calm (pleasant, low energy), upset (unpleasant, low energy), or angry (unpleasant, high energy), depending on which color quadrant the participant "plotted" the character. Since participants were recruited from schools implementing RULER, they use the Mood Meter daily and are accustomed to these types of questions. Most participants answered with one or two words when asked to describe the character's feelings.

For the ColorScore, participants received +1 if the correct Mood Meter color was provided, and −1 if an incorrect color was given. In the FeelingWordScore, participants received +1 or −1 depending on whether the feeling word provided was appropriate or not. If participants provided additional appropriate or inappropriate feeling words, they were given +0.5 or −0.5 points for each, respectively. The total EUS was calculated using the following formula:

$$EUS_{S,I} = ColorScore + FeelingWordScore$$

Higher EUS means that participants were able to more accurately identify the Mood Meter color and corresponding feeling word associated with the character's emotional state. For each interview session, EUS scores for each scene were summed to calculate an aggregate EUS score.

6.5. Reliability between Coders

Two researchers independently coded the interview transcriptions from the three sessions according to the coding scheme described in the previous section. Both coders first coded the interviews from the excluded participants to become familiar with the coding scheme. Once agreement between coders was reached, coding began on the remaining data. Coding was completed for the 120 collected interviews (40 participants × 3 sessions), overlapping 25% (30 interviews) as a reliability check.

Reliability analysis between the two coders was performed using the Intraclass Correlation Coefficient test for absolute agreement using a two-way random model. All the coded variables for each interview session had high reliabilities. The lowest agreement was found in the number of correct feeling words ($ICC(2, 1) = 0.85$, $p < 0.001$), while the highest agreement was related to the total number of core characters mentioned by each child during one interview session ($ICC(2, 1) = 0.94$, $p < 0.001$). Given the high agreement between the two coders in the overlapping 30 interviews, data from one coder were randomly selected to be used for analyses.

6.6. Data Analysis Plan

We first calculated the story recall and emotional understanding metrics according to the formulas described above. Narrative Structure Score (NSS) and Emotional Understanding Score (EUS) were computed for each participant in every session (1, 2, and 3) and averaged across the three sessions. The empathy and perception of the robots indices were also calculated and a median split was used to categorize participants in two empathy levels (low vs. high) and perception of the robots (peer vs. adult). With regard to the empathy scale, 19 participants were classified in the low empathy category and 21 were classified in the high empathy category. Regarding perception of the robots, 19 children perceived the robots more as adults and 21 perceived the robots more as peers.

Our data analysis consisted of two main steps. First, we explored our main research questions (RQ1 and RQ2) about how interaction context (single vs. group) affects story recall and emotional understanding. We started by testing the difference between the two study conditions collapsed across the three sessions using between-subjects univariate analyses of variance (ANOVA). Next, ANOVA models were conducted with interaction context (single vs. group) as the between-subjects factor and session (1, 2, and 3) as the within-subjects factor. For all the dependent measures, we planned to test the single versus group differences in each session.

We then tested our formulated hypotheses to identify which individual factors (*grade*: first or second; *gender*: female or male; *empathy*: low or high; and *perception of the robots*: peer or adult) play a major role in our measures of interest. There are not enough children in our study for an analysis including all the individual factors in the same model. As a compromise, we explored the impact of interaction type (single vs. group) and one individual difference variable at a time on participants' story recall and emotional understanding abilities. Separate planned comparisons from ANOVA models are reported below.

7. RESULTS

The results concerning our research questions (RQ1 and RQ2) on the effects of interaction context are presented in the first subsection, and the results on the hypotheses about individual differences (H1 to H4) are reported in the second subsection.

Before analyzing our two main measures of interest, story recall and emotional understanding, we examined whether there were any differences between single and group conditions in the number of words spoken by the participants during the interview sessions. An ANOVA model was run with the number of words spoken as the dependent measure. No significant difference was found, which indicates that overall, there was no significant difference in word count between the two groups. The average number of words per interview was 124.82 (standard error (SE) = 16.01). This result is important because it serves as a manipulation check for other reported findings.

7.1. Effects of Interaction Context

7.1.1. Story Recall

We investigated the impact of interaction context (single vs. group) on participants' story recall abilities (RQ1), measured by the Narrative Structure Score (NSS). An ANOVA model was run with NSS as the dependent measure. We found a significant effect of interaction context (collapsed across sessions), with students interacting alone with the robots achieving higher scores on narrative structure ($M = 0.49$, $SE = 0.03$) than the group condition ($M = 0.38$, $SE = 0.02$), $F(1, 28) = 7.71$, $p = 0.01$, and $\eta^2 = 0.22$ (see **Figure 3**).

Planned comparisons were conducted to test the role of interaction context in each particular session. No significant differences were found for session 1. For session 2, students in the single condition ($M = 0.49$, $SE = 0.05$) had a higher NSS than the students in the group condition ($M = 0.36$, $SE = 0.03$), $F(1, 36) = 7.35$, $p = 0.01$, and $\eta^2 = 0.17$. Similarly, for session 3, students in the single condition ($M = 0.50$, $SE = 0.04$) had a higher score than in the group condition ($M = 0.35$, $SE = 0.03$), $F(1, 38) = 6.59$, $p = 0.01$, and $\eta^2 = 0.15$.

These findings suggest that overall, the narrative story-related recall rate was higher in the single versus the group interaction with the robots. In the easiest session (session 1), there was no effect on interaction context, but during the more difficult sessions (sessions 2 and 3), students were found to perform better in individual than group level interactions.

7.1.2. Emotional Understanding

To investigate our second research question (RQ2), we tested whether students' emotional understanding differed in the single versus group condition. The ANOVA model with EUS as the dependent measure suggested that there was no effect of interaction context. The effect of session was significant $F(2, 62) = 7.39$, $p = 0.001$, and $\eta^2 = 0.19$, which aligns with our expectation given that the three sessions had different levels of difficulty (see **Figure 4**). Planned comparisons also yielded no significant differences between single versus group in any of the three sessions. In sum, the degree of emotional understanding did not seem to be affected by the type of interaction in this setting, but varied across sessions with different levels of difficulty.

7.2. Effects of Individual Differences

7.2.1. Grade

We tested how grade level (first vs. second) and interaction context influenced NSS (see **Figure 5**). Planned comparisons suggested that first graders scored higher in narrative structure when interacting alone with the robots than in the group condition, $F(1, 36) = 4.44$, $p = 0.04$, and $\eta^2 = 0.11$. However, for second graders, this effect was non-significant.

A similar trend was found with EUS as an outcome, as depicted in **Figure 6**. Planned comparisons suggested that for the first

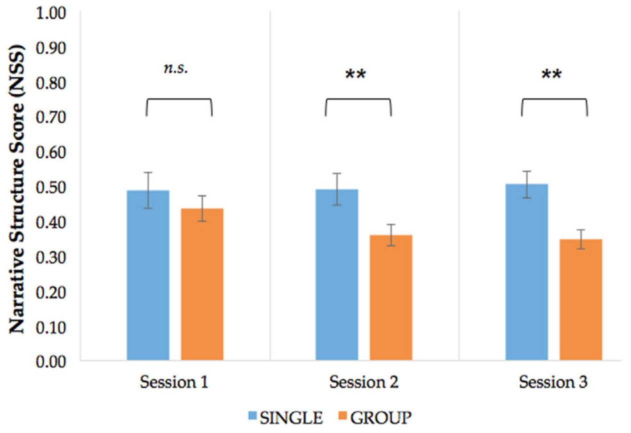


FIGURE 3 | Average Narrative Structure Scores (NSS) for participants in each condition on every interaction session. **$p < 0.01$ and *n.s.* non-significant differences.

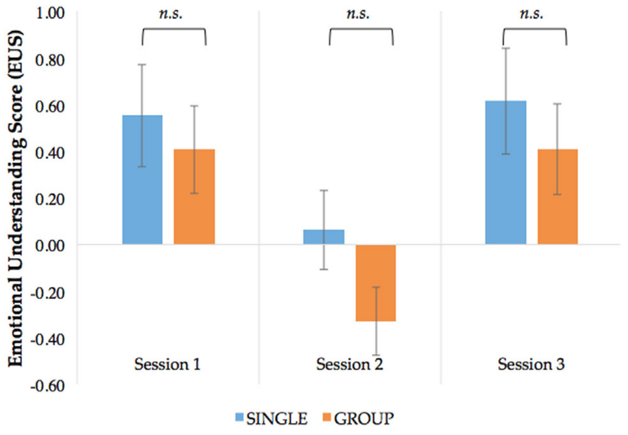


FIGURE 4 | Average Emotional Understanding Scores (EUS) for participants in each condition for sessions 1 (easy), 2 (advanced), and 3 (medium). No significant differences (*n.s.*) were found between conditions.

graders, emotional understanding is higher in the single than in the group condition, $F(1, 36) = 4.45$, $p = 0.04$, and $\eta^2 = 0.11$, but no significant differences were found for second graders. These results support Hypothesis 1, in which we predicted that second graders would have higher performance than first graders.

7.2.2. Gender

Hypothesis 2, which predicted that females would achieve higher performance scores, was not supported. We started by testing how gender differences and interaction context influenced NSS. Planned comparisons revealed only a marginal significance, with females in the single condition recalling more story events than in the group condition, $F(1, 36) = 4.09$, $p = 0.05$, and $\eta^2 = 0.10$. No significant differences were found for male students between the single versus group conditions for this variable.

Similarly, no overall significant gender differences were found in emotional understanding (EUS) or with single vs. group interaction contexts.

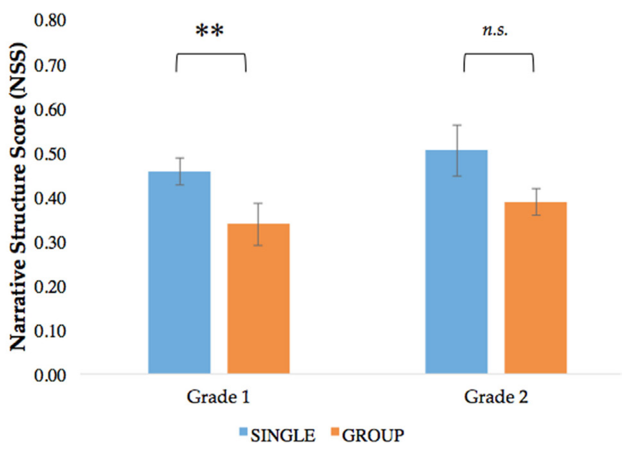


FIGURE 5 | Average Narrative Structure Scores (NSS) for participants in each condition and grade. **$p < 0.01$.

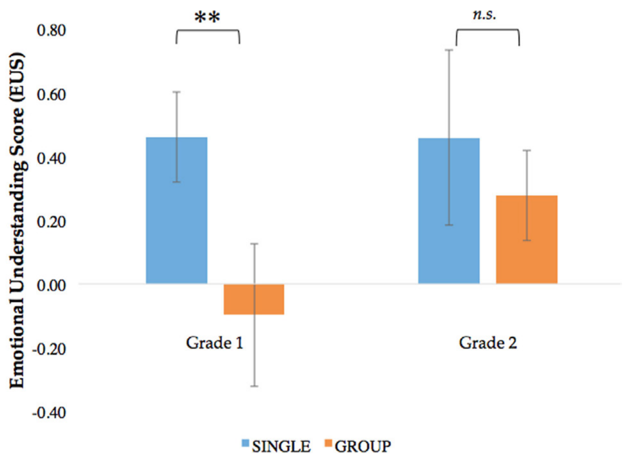


FIGURE 6 | Average Emotional Understanding Scores (EUS) for participants in each condition and grade. **$p < 0.01$.

7.2.3. Empathy

Hypothesis 3 predicted that higher empathy students would achieve higher scores in emotional understanding. Overall, high empathy students had significantly higher EUS than low empathy students, $F(1, 36) = 4.58$, $p = 0.04$, and $\eta^2 = 0.11$, confirming our hypothesis (see **Figure 7**). Furthermore, in the single condition, high empathy students performed higher on emotional understanding than those with low empathy, $F(1, 36) = 4.14$, $p = 0.049$, and $\eta^2 = 0.10$.

Planned comparisons suggested that among low empathy individuals, those in the single condition had a higher NSS than those in the group condition, $F(1, 36) = 5.98$, $p = 0.02$, and $\eta^2 = 0.14$ (see **Figure 8**).

7.2.4. Perception of the Robots

Finally, we investigated Hypothesis 4, in which we expected the perceived role of the robot to affect children's recall and

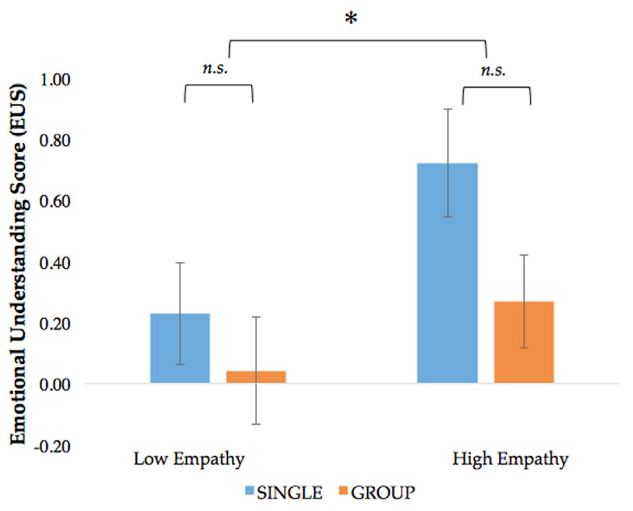


FIGURE 7 | Average Emotional Understanding Scores (EUS) for participants in each condition and empathy level. *$p < 0.05$.

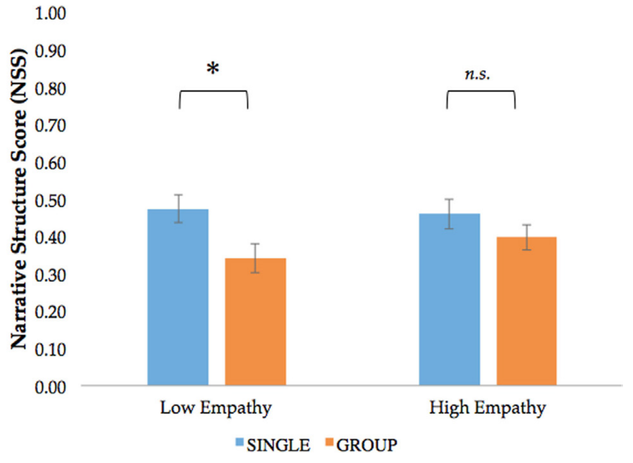


FIGURE 8 | Average Narrative Structure Scores (NSS) for participants in each condition and empathy level. *$p < 0.05$.

understanding abilities. Planned contrasts suggested that those who perceived the robots as adults recalled more story events when alone than in a group, $F(1, 36) = 11.54$, $p = 0.002$, and $\eta^2 = 0.24$ (see **Figure 9**). However, for participants who perceived robots as peers, no significant differences were found between interaction context. Among the participants in the group condition, those who perceived robots as peers (rather than adults) had higher NSS, $F(1, 36) = 4.26$, $p = 0.046$, and $\eta^2 = 0.11$.

Perception of the robots in single versus group interactions did not seem to predict the emotional understanding of the students (EUS), as none of the planned comparisons were significant. Therefore, Hypothesis 4 was only partially supported.

8. DISCUSSION

We separate this discussion into the two main steps of our analysis: exploratory analysis of interaction context and effects of individual differences in children's story recall and emotional understanding.

8.1. Effects of Interaction Context (RQ1)

Our study yielded interesting findings about the effects of interaction context on children's recall and understanding abilities. Participants interacting with the robots alone were able to recall the narrative structure (i.e., core ideas and characters) significantly better than participants in the group condition.

We offer three possible interpretations from these results. First, while the child was solely responsible for all choices when interacting alone, decisions were shared when in the group, thereby affecting how the interaction was experienced. A second interpretation is that in individual interactions, children may be more attentive since social standing in relation to their peers is not a factor. Third, the peers might simply be more distracting.

At first glance, our results may seem to contradict previous findings highlighting the benefits of learning in small groups (Hill, 1982; Pai et al., 2015). However, recalling story details is different than increasing learning gains. In fact, no significant differences were found between groups in our main learning metric, Emotional Understanding Score (participants' ability to interpret the stories using the concepts of the RULER framework), despite average individual condition scores being slightly higher for every session. Other than session 2, which had the most difficult story content, all participants performed quite well despite the type of interaction in which they participated. One possible explanation, in line with the findings from Shahid et al. (2014), is that participants in the individual condition might have benefited from some of the effects of a group setting since they were interacting with multiple autonomous agents (the two robots), but further research is needed to verify this. Moreover, several authors argue that group interaction and subsequent learning gains do not necessarily occur just because learners are in a group (Kreijns et al., 2003). An analysis of the participants' behavior while in the group during the interaction could clarify these alternative explanations.

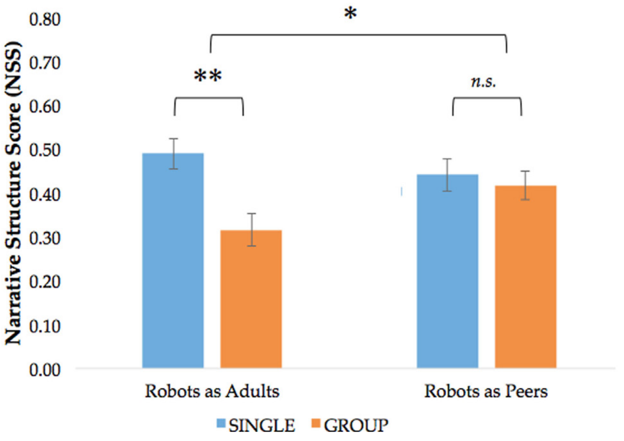


FIGURE 9 | Average Narrative Structure Scores (NSS) for participants in each condition and perception of the robots (peer vs. adult). **$p < 0.01$ and *$p < 0.05$.

8.2. Effects of Individual Differences (RQ2)

Our hypotheses about individual differences proved to be useful to further understand the effects of children interacting with robots alone or in small groups. The results suggest that interaction context is not the only relevant predictor for children's success in story recall and emotional understanding.

Grade level, for example, seems a good predictor of children's recall and understanding in these two contexts (H1). First graders interacting alone with the robots scored higher on our two main metrics (NSS and EUS) than first graders in the group condition, but no significant differences were found in second graders. While a more comprehensive analysis is necessary to validate this result, our trend suggests that first graders might not have developed the necessary skills to learn in small groups, but second graders (and potentially higher grade levels) are ready to do so.

Contrasting previous research, no significant gender differences were found in our data and, therefore, were unable to validate H2. In the existing HRI studies where gender differences were found, participants were adults and most of the effects were related to preferences rather than performance. While other reasons like a different robot or type of task might explain this result, one possibility is that children at this age might not have developed gender bias. The previous literature suggesting gender differences in narrative accuracy and emotional understanding in children might not apply as much to the present generation, as gender neutrality is currently promoted more widely in classrooms. In fact, one of the most recent meta-reviews in this area concludes that there is little evidence for gender differences in episodic memory (Andreano and Cahill, 2009).

Recall abilities seem to be affected by empathy levels in specific interaction contexts, with lower empathy individuals scoring higher on story recall in the single condition compared to the group condition (H3). A possible explanation is that lower empathy students need to be in a less distracting environment to achieve similar recall as high empathy students. Not surprisingly, our hypothesis confirmed the relation between high empathy and higher emotional understanding (Salovey and Mayer, 1990).

Like in other HRI experiments, the way participants perceived the robots had an impact on the collected measures (H4). In this domain, higher story recall is more likely to occur when participants perceive the robots as adults while interacting alone with the robot and when they perceive the robots as peers while interacting in small groups. More importantly, these findings suggest that researchers should design robot behaviors tailored to a specific interaction context and make sure that the robot's behavior is coherent with the role they are trying to portray (e.g., teacher or peer).

9. IMPLICATIONS FOR FUTURE RESEARCH

There are potential implications for the future design of socially assistive robotic scenarios based on the results obtained in this study. First, considering how well children reacted to the robots and reflected on the different choices in the postinterviews in both study conditions, affective interactive narratives using multiple robots seem to be a promising approach in socially assistive robotics.

Regarding the optimal type of interaction for these interventions, while single interactions seem to be slightly more effective in the short-term, group interventions might be more suitable in the long-term. Previous research has shown that children have more fun interacting with robots in groups rather than alone (Shahid et al., 2014). Since levels of engagement are positively correlated with students' motivation for pursuing learning goals (Ryan and Deci, 2000), influence concentration, and foster group discussions (Walberg, 1990), future research in this area should study the effects of different interaction contexts in long-term exposure to robots.

Our results have also shown that specific interaction contexts might be more suitable for particular children based on their individual differences such as grade, empathy level, and the way they perceive the robots. Therefore, in order to maximize recall and understanding gains, it might be necessary to implement more sophisticated perception and adaptation mechanisms in the robots. For example, the robots should be able to detect disengagement and employ recovery mechanisms to keep children focused in the interaction, particularly in group settings (Leite et al., 2015b). Similarly, as group interactions seem more effective when participants perceive the robots as peers, the robots could portray different roles depending on whether they were interacting with one single child or a small group.

10. CONCLUSION

The effective acquisition of social and emotional skills requires constant practice in diverse hypothetical situations. In this paper, we proposed a novel approach where multiple socially assistive robots are used in interactive role-playing activities with children. The robots acted as interactive puppets; children could control the actions of one of the robots and see the impact of the selected actions on the course of the story. Using this scenario, we investigated the effects of interaction context (single child versus small groups) and individual factors (grade, gender, empathy level, and perception of the robots) on children's story recall and emotional interpretation of three interactive stories.

Results from this repeated interaction study showed that although participants who interacted alone with the robot remembered the stories better than participants in the group condition, no significant differences were found in children's emotional interpretation of the narratives. This latter metric was fairly high for all participants, except in the session with the hardest story content. To further understand these results, we investigated the effects of participants' individual differences in the two interaction contexts for these metrics. We found that single settings seem more appropriate to first graders than groups, empathy is a very important predictor for emotional understanding of the narratives, and children's performance varies depending on the way they perceive the robots (peer vs. adult) in the two interaction contexts. In addition to the promising results of this study, further research is required to more thoroughly understand how context of interaction affects children's learning gains in longer-term interactions with socially assistive robots, as well as how participants' individual differences interplay with each other.

AUTHOR CONTRIBUTIONS

All the authors contributed to the system and experimental design. IL, MM, ML, DU, and NS ran the experiment and conducted data analysis.

ACKNOWLEDGMENTS

We thank Bethel Assefa for assistance in data collection and coding, Rachel Protacio and Jennifer Allen for recording the voices for the robots, Emily Lennon for artwork creation, and the students and staff from the school where the study was conducted.

FUNDING

This work was supported by the NSF Expedition in Computing Grant #1139078 and SRA International (US Air Force) Grant #13-004807.

REFERENCES

Admoni, H., Nawroj, A., Leite, I., Ye, Z., Hayes, B., Rozga, A., et al. (2015). *Mykeepon Project*. Available at: http://hennyadmoni.com/keepon

Al Moubayed, S., Beskow, J., Skantze, G., and Granstrom, B. (2012). "Furhat: a back-projected human-like robot head for multiparty human-machine interaction," in *Cognitive Behavioural Systems, Volume 7403 of Lecture Notes in Computer Science*, eds A. Esposito, A. Esposito, A. Vinciarelli, R. Hoffmann, and V. Muller (Berlin, Heidelberg: Springer), 114–130.

Andreano, J. M., and Cahill, L. (2009). Sex influences on the neurobiology of learning and memory. *Learn. Mem.* 16, 248–266. doi:10.1101/lm.918309

Andrist, S., Mutlu, B., and Tapus, A. (2015). "Look like me: matching robot personality via gaze to increase motivation," in *Proceedings of the 33rd Annual ACM*

Conference on Human Factors in Computing Systems, CHI '15 (New York, NY: ACM), 3603–3612.

Bainbridge, W. A., Hart, J. W., Kim, E. S., and Scassellati, B. (2011). The benefits of interactions with physically present robots over video-displayed agents. *Int. J. Soc. Robot*. 3, 41–52. doi:10.1007/s12369-010-0082-7

Baxter, P., de Greeff, J., and Belpaeme, T. (2013). "Do children behave differently with a social robot if with peers?" in *International Conf. on Social Robotics (ICSR 2013)* (Bristol: Springer).

Bliss, L. S., McCabe, A., and Miranda, A. E. (1998). Narrative assessment profile: discourse analysis for school-age children. *J. Commun. Disord.* 31, 347–363. doi:10.1016/S0021-9924(98)00009-4

Bohus, D., Saw, C. W., and Horvitz, E. (2014). "Directions robot: in-the-wild experiences and lessons learned," in *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems, AAMAS'14* (Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems), 637–644.

Bolla-Wilson, K., and Bleecker, M. L. (1986). Influence of verbal intelligence, sex, age, and education on the rey auditory verbal learning test. *Dev. Neuropsychol.* 2, 203–211. doi:10.1080/87565648609540342

Brackett, M. A., Rivers, S. E., and Salovey, P. (2011). Emotional intelligence: implications for personal, social, academic, and workplace success. *Soc. Personal. Psychol. Compass* 5, 88–103. doi:10.1111/j.1751-9004.2010.00334.x

Breazeal, C., Brooks, A., Gray, J., Hancher, M., McBean, J., Stiehl, D., et al. (2003). Interactive robot theatre. *Commun. ACM* 46, 76–85. doi:10.1145/792704.792733

Briton, N. J., and Hall, J. A. (1995). Beliefs about female and male nonverbal communication. *Sex Roles* 32, 79–90. doi:10.1007/BF01544758

Bruce, A., Knight, J., Listopad, S., Magerko, B., and Nourbakhsh, I. (2000). "Robot improv: using drama to create believable agents," in *Proc. of the Int. Conf. on Robotics and Automation, ICRA'00* (San Francisco, CA: IEEE), 4002–4008.

Bryant, B. K. (1982). An index of empathy for children and adolescents. *Child Dev.* 53, 413–425. doi:10.2307/1128984

Buckner, J. P., and Fivush, R. (1998). Gender and self in children's autobiographical narratives. *Appl. Cogn. Psychol.* 12, 407–429. doi:10.1002/(SICI)1099-0720(199808)12:4<407::AID-ACP575>3.0.CO;2-7

Camras, L. A., and Allison, K. (1985). Children's understanding of emotional facial expressions and verbal labels. *J. Nonverbal Behav.* 9, 84–94. doi:10.1007/BF00987140

Capitani, E., Laiacona, M., and Basso, A. (1998). Phonetically cued word-fluency, gender differences and aging: a reappraisal. *Cortex* 34, 779–783. doi:10.1016/S0010-9452(08)70781-0

Capps, L., Losh, M., and Thurber, C. (2000). The frog ate the bug and made his mouth sad: narrative competence in children with autism. *J. Abnorm. Child Psychol.* 28, 193–204. doi:10.1023/A:1005126915631

Cassell, J. (2000). *Embodied Conversational Agents*. MIT Press.

Castillo, R., Fernández-Berrocal, P., and Brackett, M. A. (2013). Enhancing teacher effectiveness in Spain: a pilot study of the RULER approach to social and emotional learning. *J. Educ. Train. Stud.* 1, 263–272. doi:10.11114/jets.v1i2.203

Collis, B. (1996). The internet as an educational innovation: lessons from experience with computer implementation. *Educ. Technol.* 36, 21–30.

Crais, E. R., and Lorch, N. (1994). Oral narratives in school-age children. *Top. Lang. Disord.* 14, 13–28. doi:10.1097/00011363-199405000-00004

Crow, T., Crow, L., Done, D., and Leask, S. (1998). Relative hand skill predicts academic ability: global deficits at the point of hemispheric indecision. *Neuropsychologia* 36, 1275–1282. doi:10.1016/S0028-3932(98)00039-6

Denham, S. A. (1986). Social cognition, prosocial behavior, and emotion in preschoolers: contextual validation. *Child Dev.* 57, 194–201. doi:10.2307/1130651

Denham, S. A. (1998). *Emotional Development in Young Children*. New York, NY: Guilford Press.

Dillenbourg, P. (1999). What do you mean by collaborative learning? *Collab. Learn. Cogn. Comput. Approaches* 1–19.

Dunn, J., Bretherton, I., and Munn, P. (1987). Conversations about feeling states between mothers and their young children. *Dev. Psychol.* 23, 132. doi:10.1037/0012-1649.23.1.132

Eisenberg, N., and Lennon, R. (1983). Sex differences in empathy and related capacities. *Psychol. Bull.* 94, 100. doi:10.1037/0033-2909.94.1.100

Foster, M. E., Gaschler, A., Giuliani, M., Isard, A., Pateraki, M., and Petrick, R. P. (2012). "Two people walk into a bar: dynamic multi-party social interaction with a robot agent," in *Proceedings of the 14th ACM International Conference on Multimodal Interaction, ICMI '12* (New York, NY: ACM), 3–10.

Gomez, R., Kawahara, T., Nakamura, K., and Nakadai, K. (2012). "Multi-party human-robot interaction with distant-talking speech recognition," in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction, HRI '12* (New York, NY: ACM), 439–446.

Greif, E., Alvarez, M., and Ulman, K. (1981). Recognizing emotions in other people: sex differences in socialization. *Poster Presented at the Biennial Meeting of the Society for Research in Child Development*, Boston, MA.

Griffith, P. L., Ripich, D. N., and Dastoli, S. L. (1986). Story structure, cohesion, and propositions in story recalls by learning-disabled and nondisabled children. *J. Psycholinguist. Res.* 15, 539–555. doi:10.1007/BF01067635

Harris, P. L. (1983). Children's understanding of the link between situation and emotion. *J. Exp. Child Psychol.* 36, 490–509. doi:10.1016/0022-0965(83)90048-6

Harris, P. L. (1989). *Children and Emotion: The Development of Psychological Understanding*. Basil Blackwell.

Harris, P. L. (2008). Children's understanding of emotion. *Handb. Emotions* 3, 320–331.

Hill, G. W. (1982). Group versus individual performance: are n+1 heads better than one? *Psychol. Bull.* 91, 517. doi:10.1037/0033-2909.91.3.517

Hoffman, G., Kubat, R., and Breazeal, C. (2008). "A hybrid control system for puppeteering a live robotic stage actor," in *Proc. of RO-MAN 2008* (IEEE), 354–359.

Hoffman, M. L. (1977). Sex differences in empathy and related behaviors. *Psychol. Bull.* 84, 712. doi:10.1037/0033-2909.84.4.712

Iaria, G., Petrides, M., Dagher, A., Pike, B., and Bohbot, V. D. (2003). Cognitive strategies dependent on the hippocampus and caudate nucleus in human navigation: variability and change with practice. *J. Neurosci.* 23, 5945–5952.

Johansson, M., Skantze, G., and Gustafson, J. (2013). "Head pose patterns in multiparty human-robot team-building interactions," in *Social Robotics, Volume 8239 of Lecture Notes in Computer Science*, eds G. Herrmann, M. Pearson, A. Lenz, P. Bremner, A. Spiers, and U. Leonards (New York, NY: Springer), 351–360.

John, S. F., Lui, M., and Tannock, R. (2003). Children's story retelling and comprehension using a new narrative resource. *Can. J. Sch. Psychol.* 18, 91–113. doi:10.1177/082957350301800105

Kanda, T., Sato, R., Saiwaki, N., and Ishiguro, H. (2007). A two-month field trial in an elementary school for long-term human-robot interaction. *IEEE Trans. Robot.* 23, 962–971. doi:10.1109/TRO.2007.904904

Kelleher, C., Pausch, R., and Kiesler, S. (2007). "Storytelling Alice motivates middle school girls to learn computer programming," in *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems, CHI '07* (New York, NY: ACM), 1455–1464.

Kim, Y., and Baylor, A. L. (2006). A social-cognitive framework for pedagogical agents as learning companions. *Educ. Technol. Res. Dev.* 54, 569–596. doi:10.1007/s11423-006-0637-3

Kozima, H., Michalowski, M., and Nakagawa, C. (2009). Keepon: a playful robot for research, therapy, and entertainment. *Int. J. Soc. Robot.* 1, 3–18. doi:10.1007/s12369-008-0009-8

Kramer, J. H., Delis, D. C., Kaplan, E., O'Donnell, L., and Prifitera, A. (1997). Developmental sex differences in verbal learning. *Neuropsychology* 11, 577. doi:10.1037/0894-4105.11.4.577

Kreijns, K., Kirschner, P. A., and Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: a review of the research. *Comput. Human Behav.* 19, 335–353. doi:10.1016/S0747-5632(02)00057-2

Leite, I., Martinho, C., and Paiva, A. (2013). Social robots for long-term interaction: a survey. *Int. J. Soc. Robot.* 5, 291–308. doi:10.1007/s12369-013-0178-y

Leite, I., McCoy, M., Lohani, M., Ullman, D., Salomons, N., Stokes, C., et al. (2015a). "Emotional storytelling in the classroom: individual versus group interaction between children and robots," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction, HRI '15* (New York, NY: ACM), 75–82.

Leite, I., McCoy, M., Ullman, D., Salomons, N., and Scassellati, B. (2015b). "Comparing models of disengagement in individual and group interactions," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction, HRI '15* (New York, NY: ACM), 99–105.

Leyzberg, D., Spaulding, S., Toneva, M., and Scassellati, B. (2012). "The physical presence of a robot tutor increases cognitive learning gains," in *Proc. of the 34th Annual Conf. of the Cognitive Science Society* (Austin, TX: Cognitive Science Society).

Linn, M. C., and Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: a meta-analysis. *Child Dev.* 56, 1479–1498. doi:10.2307/1130467

Lopes, P. N., Salovey, P., Coté, S., and Beers, M. (2005). Emotion regulation abilities and the quality of social interaction. *Emotion* 5, 113–118. doi:10.1037/1528-3542.5.1.113

Lou, Y., Abrami, P. C., and d'ÄôApollonia, S. (2001). Small group and individual learning with technology: a meta-analysis. *Rev. Educ. Res.* 71, 449–521. doi:10.3102/00346543071003449

Lu, D., and Smart, W. (2011). "Human-robot interactions as theatre," in *RO-MAN, 2011 IEEE*, 473–478.

Maccoby, E. E., and Jacklin, C. N. (1974). *The Psychology of Sex Differences*, Vol. 1. Stanford University Press.

MacDorman, K. F., and Entezari, S. O. (2015). Individual differences predict sensitivity to the uncanny valley. *Interact. Stud.* 16, 141–172. doi:10.1075/is.16.2.01mac

McCabe, A., Bliss, L., Barra, G., and Bennett, M. (2008). Comparison of personal versus fictional narratives of children with language impairment. *Am. J. Speech Lang. Pathol.* 17, 194–206. doi:10.1044/1058-0360(2008/019)

McCartney, K. A., and Nelson, K. (1981). Children's use of scripts in story recall. *Discourse Process.* 4, 59–70. doi:10.1080/01638538109544506

McGuigan, F., and Salmon, K. (2006). The influence of talking on showing and telling: adult-child talk and children's verbal and nonverbal event recall. *Appll. Cogn. Psychol.* 20, 365–81. doi:10.1002/acp.1183

Mercer, N. (1996). The quality of talk in children's collaborative activity in the classroom. *Learn. Instr.* 6, 359–377. doi:10.1016/S0959-4752(96)00021-7

Moreland, R. L. (2010). Are dyads really groups? *Small Group Res.* 41, 251–267. doi:10.1177/1046496409358618

Morrow, D. G. (1985). Prominent characters and events organize narrative understanding. *J. Mem. Lang.* 24, 304–319. doi:10.1016/0749-596X(85)90030-0

Mubin, O., Stevens, C. J., Shahid, S., Al Mahmud, A., and Dong, J.-J. (2013). A review of the applicability of robots in education. *J. Technol. Educ. Learn.* 1, 209–215. doi:10.2316/Journal.209.2013.1.209-0015

Mumm, J., and Mutlu, B. (2011). "Human-robot proxemics: physical and psychological distancing in human-robot interaction," in *Proceedings of the 6th International Conference on Human-Robot Interaction, HRI '11* (New York, NY: ACM), 331–338.

Mutlu, B., Forlizzi, J., and Hodgins, J. (2006). "A storytelling robot: modeling and evaluation of human-like gaze behavior," in *Humanoid Robots, 2006 6th IEEE-RAS International Conference on* (Genova, Italy: IEEE), 518–523.

Nomura, T., Kanda, T., Suzuki, T., and Kato, K. (2008). Prediction of human behavior in human–robot interaction using psychological scales for anxiety and negative attitudes toward robots. *IEEE Trans. Robot.* 24, 442–451. doi:10.1109/TRO.2007.914004

Pai, H.-H., Sears, D., and Maeda, Y. (2015). Effects of small-group learning on transfer: a meta-analysis. *Educ. Psychol. Rev.* 27, 79–102. doi:10.1007/s10648-014-9260-8

Pereira, A., Prada, R., and Paiva, A. (2014). "Improving social presence in human-agent interaction," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14* (New York, NY: ACM), 1449–1458.

Petrides, K., and Furnham, A. (2000). On the dimensional structure of emotional intelligence. *Pers. Individ. Dif.* 29, 313–320. doi:10.1016/S0191-8869(99)00195-6

Pohl, R. F., Bender, M., and Lachmann, G. (2005). Autobiographical memory and social skills of men and women. *Appl. Cogn. Psychol.* 19, 745–760. doi:10.1002/acp.1104

Pons, F., Harris, P. L., and de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: developmental periods and hierarchical organization. *Eur. J. Dev. Psychol.* 1, 127–152. doi:10.1080/17405620344000022

Quigley, M., Conley, K., Gerkey, B. P., Faust, J., Foote, T., Leibs, J., et al. (2009). "Ros: an open-source robot operating system," in *ICRA Workshop on Open Source Software* (Kobe, Japan).

Renz, K., Lorch, E. P., Milich, R., Lemberger, C., Bodner, A., and Welsh, R. (2003). On-line story representation in boys with attention deficit hyperactivity disorder. *J. Abnorm. Child Psychol.* 31, 93–104. doi:10.1023/A:1021777417160

Rivers, S. E., Brackett, M. A., Reyes, M. R., Elbertson, N. A., and Salovey, P. (2013). Improving the social and emotional climate of classrooms: a clustered randomized controlled trial testing the RULER approach. *Prev. Sci.* 14, 77–87. doi:10.1007/s11121-012-0305-2

Ross, M., and Holmberg, D. (1992). Are wives' memories for events in relationships more vivid than their husbands' memories? *J. Soc. Pers. Relat.* 9, 585–604. doi:10.1177/0265407592094007

Ryan, R. M., and Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am. Psychol.* 55, 68. doi:10.1037/0003-066X.55.1.68

Saerbeck, M., Schut, T., Bartneck, C., and Janse, M. D. (2010). "Expressive robots in education: varying the degree of social supportive behavior of a robotic tutor," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10* (New York, NY: ACM), 1613–1622.

Salovey, P., and Mayer, J. D. (1990). Emotional intelligence. *Imagination Cogn. Pers.* 9, 185–211. doi:10.2190/DUGG-P24E-52WK-6CDG

Schermerhorn, P., Scheutz, M., and Crowell, C. R. (2008). "Robot social presence and gender: do females view robots differently than males?" in *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction* (Amsterdam, The Netherlands: ACM), 263–270.

Schoenau-Fog, H. (2011). "Hooked! – evaluating engagement as continuation desire in interactive narratives," in *Proceedings of the 4th International Conference on Interactive Digital Storytelling* (Berlin, Heidelberg: Springer-Verlag), 219–230.

Schultze, T., Mojzisch, A., and Schulz-Hardt, S. (2012). Why groups perform better than individuals at quantitative judgment tasks: group-to-individual transfer as an alternative to differential weighting. *Organ. Behav. Hum. Decis. Process* 118, 24–36. doi:10.1016/j.obhdp.2011.12.006

Shahid, S., Krahmer, E., and Swerts, M. (2014). Child-robot interaction across cultures: how does playing a game with a social robot compare to playing a game alone or with a friend? *Comput. Human Behav.* 40, 86–100. doi:10.1016/j.chb.2014.07.043

Smedler, A.-C., and Törestad, B. (1996). Verbal intelligence: a key to basic skills? *Educ. Stud.* 22, 343–356. doi:10.1080/0305569960220304

Syrdal, D. S., Koay, K. L., Walters, M. L., and Dautenhahn, K. (2007). "A personalized robot companion? The role of individual differences on spatial preferences in hri scenarios," in *Robot and Human Interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on* (IEEE), 1143–1148.

Szafir, D., and Mutlu, B. (2012). "Pay attention! Designing adaptive agents that monitor and improve user engagement," in *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems, CHI '12* (New York, NY: ACM), 11–20.

Takayama, L., and Pantofaru, C. (2009). "Influences on proxemic behaviors in human-robot interaction," in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on* (St. Louis, MO: IEEE), 5495–5502.

Tapus, A., Matarić, M., and Scassellatti, B. (2007). The grand challenges in socially assistive robotics. *IEEE Robot. Auto. Mag.* 14, 35–42. doi:10.1109/MRA.2007.339605

Vannini, N., Enz, S., Sapouna, M., Wolke, D., Watson, S., Woods, S., et al. (2011). Fearnot! Computer-based anti-bullying programme designed to foster peer intervention. *Eur. J. Psychol. Educ.* 26, 21–44. doi:10.1007/s10212-010-0035-4

Walberg, H. J. (1990). Productive teaching and instruction: assessing the knowledge base. *Phi Delta Kappan*. 71, 470–478.

Walters, M. L., Dautenhahn, K., Te Boekhorst, R., Koay, K. L., Kaouri, C., Woods, S., et al. (2005). "The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment," in *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on* (IEEE), 347–352.

Yuill, N. (1984). Young children's coordination of motive and outcome in judgements of satisfaction and morality. *Br. J. Dev. Psychol.* 2, 73–81. doi:10.1111/j.2044-835X.1984.tb00536.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Leite, McCoy, Lohani, Ullman, Salomons, Stokes, Rivers and Scassellati. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 22 June 2017
doi: 10.3389/frobt.2017.00027

Spatiotemporal Aspects of Engagement during Dialogic Storytelling Child–Robot Interaction

Scott Heath†, Gautier Durantin†, Marie Boden, Kristyn Hensby, Jonathon Taufatofua, Ola Olsson, Jason Weigel, Paul Pounds and Janet Wiles*

School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, QLD, Australia

The success of robotic agents in close proximity of humans depends on their capacity to engage in social interactions and maintain these interactions over periods of time that are suitable for learning. A critical requirement is the ability to modify the behavior of the robot contingently to the attentional and social cues signaled by the human. A benchmark challenge for an engaging social robot is that of storytelling. In this paper, we present an exploratory study to investigate dialogic storytelling—storytelling with contingent responses—using a child-friendly robot. The aim of the study was to develop an engaging storytelling robot and to develop metrics for evaluating engagement. Ten children listened to an illustrated story told by a social robot during a science fair. The responses of the robot were adapted during the interaction based on the children's engagement and touches of the pictures displayed by the robot on a tablet embedded in its torso. During the interaction the robot responded contingently to the child, but only when the robot invited the child to interact. We describe the robot architecture used to implement dialogic storytelling and evaluate the quality of human–robot interaction based on temporal (patterns of touch, touch duration) and spatial (motions in the space surrounding the robot) metrics. We introduce a novel visualization that emphasizes the temporal dynamics of the interaction and analyze the motions of the children in the space surrounding the robot. The study demonstrates that the interaction through invited contingent responses succeeded in engaging children, although the robot missed some opportunities for contingent interaction and the children had to adapt to the task. We conclude that (i) the consideration of both temporal and spatial attributes is fundamental for establishing metrics to estimate levels of engagement in real-time, (ii) metrics for engagement are sensitive to both the group and individual, and (iii) a robot's sequential mode of interaction can facilitate engagement, despite some social events being ignored by the robot.

Keywords: engagement, human–robot interaction, storytelling, social robotics, immediacy

OPEN ACCESS

Edited by:
Bilge Mutlu,
University of Wisconsin-Madison,
United States

Reviewed by:
Fulvio Mastrogiovanni,
Università di Genova, Italy
Erol Sahin,
Middle East Technical University,
Turkey

****Correspondence:***
Scott Heath
scott.heath@uq.edu.au

†These authors have contributed equally to this work.

Specialty section:
This article was submitted to Humanoid Robotics, a section of the journal Frontiers in Robotics and AI

Received: *13 January 2017*
Accepted: *06 June 2017*
Published: *22 June 2017*

Citation:
Heath S, Durantin G, Boden M, Hensby K, Taufatofua J, Olsson O, Weigel J, Pounds P and Wiles J (2017) Spatiotemporal Aspects of Engagement during Dialogic Storytelling Child–Robot Interaction. Front. Robot. AI 4:27. doi: 10.3389/frobt.2017.00027

1. INTRODUCTION

As robots become more prevalent in our lives, it is important to design robots that can share spaces with humans and to evaluate how robots can engage in social interactions. A robot's social abilities will affect whether the robot is allowed into spaces occupied by humans, as well as the types of tasks that the robot will be trusted to perform. A key ability for social robots is that of establishing and maintaining human engagement.

Engagement during a social interaction is defined as a combination of attention and understanding of this interaction (Tomasello et al., 2005). Building social robots that can maintain user engagement has been recognized as one of the main challenges of human–robot interaction (HRI) (Sidner et al., 2005), in particular when interacting with children in a learning context (Walters et al., 2008 and Ioannou et al., 2015). A major component of HRI affecting engagement is the level of *immediacy* of the interaction (Kennedy et al., 2015). *Immediacy behaviors* are defined as "… those which increase the sensory stimulation between two interaction partners" (Mehrabian, 1968), where high immediacy is implemented as a greater number of socially contingent responses from the robot.

The level of immediacy of a social robot can be varied by using different forms of responsiveness to the user's actions (Yanco and Drury, 2004). *Open-loop* modes of interaction correspond to the lowest level of immediacy, where the robot executes scripted actions without processing any of the user's actions. In contrast, higher levels of immediacy are achieved when the robot implements *closed-loop* control by processing and adapting to user inputs. User inputs can either be invited by the robot at certain times (termed a *synchronous* mode of interaction) or provided whenever the user wants (an *asynchronous* mode of interaction).

Robot storytelling presents a benchmark challenge for creating engaging interactions, where different levels of immediacy can be used. As with other interactions, the level of immediacy during storytelling can be controlled through open-loop scripted responses, or closed-loop responses with synchronous or asynchronous modes of interaction. Closed-loop storytelling is also called "dialogic" storytelling (Whitehurst et al., 1988) and requires the storyteller to contingently respond and change how the story is delivered based on children's reactions. Rather than considering the child as a passive listener, the aim of this approach is to give an active role to the child by initiating richer interactions with them. Dialogic storytelling by human storytellers has been shown to increase the level of engagement in children during storytelling (Whitehurst et al., 1988 and Mol et al., 2008). We hypothesize that implementing dialogic storytelling using robots has the potential to increase the level of engagement in a similar way.

In their study of the interaction between a robot and preschoolers Ioannou et al. (2015) used dancing, moving, and storytelling activities. They noted that the engagement of children remained high during dancing and moving activities that exhibited higher levels of interactivity and were enriched with emotions and gestures. In contrast, the storytelling activity was performed in an open-loop mode of interaction and used few gestures and emotions, which resulted in disengagement of the children. This study demonstrated the effect of different levels of immediacy on engagement in different types of human–robot interaction. Similarly, the majority of previous attempts at building robots capable of telling stories have used open-loop modes of interaction (e.g., Mutlu et al. (2006); Gelin et al. (2010); and Fridin (2014)), where the motions, utterances, and emotions displayed by the robot were not dependent on user inputs. Few studies have implemented closed-loop (dialogic) storytelling with synchronous interaction by allowing a human to program or trigger robot actions through a control interface (Ryokai et al., 2009 and Kory, 2014) and by measuring the location of the user in the space surrounding the robot at specific points in time (Pitsch et al., 2009). Both methods resulted in good levels of engagement. Finally, to the authors' knowledge, only one storytelling study can be considered both closed-loop and asynchronous. The approach used measurements of engagement estimated in real-time from brain signals to modify the behavior of the storytelling robot (Szafir and Mutlu, 2012), which also resulted in high levels of immediacy and comprehension of the story.

In each implementation of closed-loop storytelling robots, enabling interactivity required preliminary programming of the robot by the user (Ryokai et al., 2009), control by an experimenter (Kory, 2014), or invasive measures of engagement (electroencephalography, see Szafir and Mutlu (2012)) to manipulate the responses of the robot. None of these solutions are suitable for a robot capable of interacting with children in a public space with a high level of autonomy.

In the OPAL project, we aim to build a child-sized robot (Opie) that is capable of socially interacting in public spaces through a variety of activities including storytelling. Opie is inspired by the RUBI project at UCSD (Malmir et al., 2013) and is designed to be a safe social robot for children that encourages the use of haptic modalities such as touching, leaning on, and interacting in close proximity. In the current study, we implement and evaluate dialogic storytelling using Opie in a public setting (a science fair). The specific aims of our study are to explore reaction times (through the modality of touch), touch patterns, and location of the child in space around the robot during the course of a dialogic storytelling interaction. We aim to explore the following research questions:

1. *What level and duration of engagement can Opie facilitate?*
2. *How individual or stereotypical are the spatial and temporal reactions across different participants?*
3. *How do the patterns of spatial and temporal reactions relate to Opie's synchronous behavior?*

These research questions will be explored by studying the patterns that are present within the different spatial and temporal reactions and what these patterns show. The location and task create a challenging context for gaining, maintaining, and estimating engagement. In order to create a responsive robot for a public location, we require methods to evaluate engagement that are non-invasive, provide high temporal resolution, and have the potential to be automated. As both immediacy and proxemics have been argued to play a role in user engagement (Mehrabian, 1972), we aim to develop methods for evaluating engagement that are based on temporal and spatial features.

In this paper, we present an implementation of dialogic storytelling using synchronous inputs of touch, which were designed to maintain the engagement of the children during the course of the interaction. The impact of dialogic storytelling on the engagement of children was evaluated while they took turns interacting with the robot during a science fair. The children's behavior was monitored by a video camera and by recording their actions. Analysis focused on spatial and temporal features (such as response times, patterns of touches and of motion in the space of the robot) extracted from these non-invasive recordings to provide estimates of children's engagement during the interaction.

2. MATERIALS AND METHODS

2.1. Robot

The robot platform used in this study is the child-friendly robot Opie, designed as a social robot for social interaction with children across different modalities. Opie is the result of a multidisciplinary, iterative design process (Wiles et al., 2016) and was previously used for interaction with children to investigate language (Heath et al., 2016) and elements of spatial proximity such as touch patterns (Hensby et al., 2016 and Rogers et al., 2016). Opie is intended to explore how robots can be used to facilitate social tasks, such as educating, conversing, or playing. Opie's torso, head, arms, and single neck actuator are intended to enable social functions.

Opie's torso is manufactured from soft materials to facilitate interaction through touch and increase safety. Opie's torso and child-sized stature were designed to make the robot appear friendly. The shape of Opie and the surrounding cushions and mat contribute to the creation of a safe area for children to occupy in front of the robot (see **Figure 1A**). Opie incorporates two tablets, one mounted on the head and one mounted on the torso. The 8-inch head tablet displays animated eyes that are capable of moving and expressing emotions. Opie's head is tilted slightly forward to help create an inviting space in front of the robot. The 12-inch torso tablet displays media and runs a speech synthesizer, while allowing children to interact through touch. The inclusion of two tablets allows the face tablet to be dedicated to social interactions. This version of Opie contains a single actuator in the neck, which allows Opie's head to yaw left and right around the neck. Opie also has arms, which rotate around the shoulder, but are not actuated. The robot's behavior can be controlled both by games that run on the torso tablet or by a Wizard of Oz (*WoZ*) through a phone interface. In this study, robot behaviors were autonomous during storytelling.

Opie's other electronics include a router that enables wireless information transfer between the robot parts and a Raspberry Pi running core server software. All the components are integrated using the robot operating system (ROS) middleware (Quigley et al., 2009), which allows the tablets, neck motors, and the WoZ phone to communicate with each other. The software running on the head and torso tablets is written using the Unity game engine and uses the Android native text-to-speech API to tell the story.

Opie was installed at a science fair in Brisbane (Australia) in a 1 m × 2 m space delimited by rugs. The robot had two separator panels of approximately 2 m height on the back and left sides and separating the robot's area from other activities at the science fair. The space around the robot was monitored by two video cameras: one facing the front of the robot at a distance of 1.5 m and the other attached to the separator panel above the robot's head and looking down at the mat from behind the robot (see **Figure 1B**).

2.2. Storytelling Game

A storytelling game was designed for Opie's torso tablet. The aim of the game developed for this study was to present an interactive narrative combined with a simple object finding task for children to perform. The storytelling game was built to facilitate (i) presentation of narrative content to children, (ii) the robot responding to touches on the tablet (allowing dialogic storytelling), (iii) a temporal measure of engagement based on touches, and (iv) expression of emotions from the robot that accompany narration. The storytelling game consisted of the presentation of a

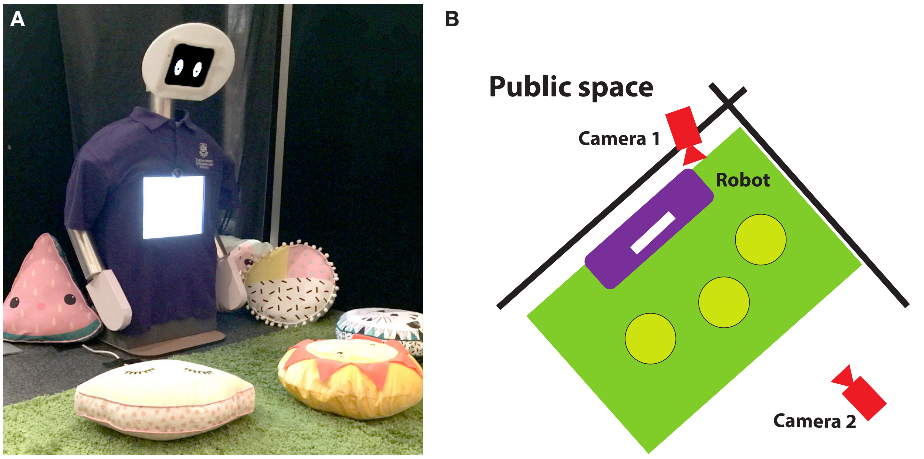


FIGURE 1 | The experimental setup. **(A)** Opie in the experimental setup—Opie is set up to create a safe space in front of the robot with a mat and cushions. The angle of the robot's head helps to shape the space in front of Opie. Opie is constructed from foam and pool noodles, which provides a soft exterior and allows children to interact with Opie through touch. The height of the robot is short to allow interactions with children that are sitting on the cushions. **(B)** Opie's position in the space with the location of the cameras and cushions (where the children sit) shown.

scene and an accompanying narrative. Each of the three scenes within the game consisted of a background, a target animal, and several non-target elements (called distractors). Each of the scenes was designed to be as similar as possible in terms of difficulty. A target animal was presented on the torso tablet first, while Opie named and described the animal, and explained why the target animal was visiting that scene. The first level would then start.

2.2.1. Levels

Each scene within the game consisted of six levels of increasing difficulty. Each time a level started, Opie said a sentence about the target animal running away and asked the child to find that animal. When the child selected the target animal, the level ended and the next level would begin. The task (finding the animal) was then repeated with increased difficulty as the target became smaller or partially occluded in each successive level.

2.2.2. Overlay

Every time Opie started a level or a child pressed an object in a scene, Opie used an "overlay" to decrease the saliency of the scene or increase the saliency of an object. The *overlay* is a semi-transparent black rectangle that is used to increase the saliency of objects in the scene relative to the rest of the scene by darkening all other objects and the background. The event of adding or removing the overlay to the scene was significant, as in addition to changes in saliency, it designated when Opie started or stopped reacting to children's touches. Opie did not respond to touches when the overlay was displayed.

2.2.3. Attentional Countermeasures

If 10 s elapsed since the last detected touch or utterance from Opie an "attentional countermeasure" was presented. An *attentional countermeasure* consisted of Opie telling the child that help was needed and reiterating to the child to find the target animal.

2.3. Ethics

Testing of the robot at the science fair was approved by a local ethics committee. Parents provided consent for their child's participation in the study, and experimenters engaged parents prior to the children entering Opie's space. The consent form was completed on an iPad and also included an optional media release consent. Parents were able to stay with their children during the study, either watching from behind the child or sitting with their child in front of the robot.

2.4. Procedure

The procedure of the study consisted of three phases—an introductory phase (which required the intervention of a human facilitator and WoZ), a storytelling phase (which was completely autonomous), and then a quiz phase (conducted by the human experimenters). During the entire procedure, the role of the human facilitator was to familiarize the children with the robot and supervise the interaction without taking part in it. As all robot behaviors were autonomous during the storytelling game, the only role of the WoZ was to trigger the start of the story.

2.4.1. Introductory Phase

After obtaining consent, a human facilitator took up to three children and their parents over to Opie and introduced them to the robot. Any additional children had to wait until the end of the current interaction (out of sight). The children were encouraged to sit down on a cushion each. The facilitator started a pregame consisting of colored shapes displayed on Opie's torso tablet. This pregame was designed to familiarize the children with the robot and prime them to touch Opie's tablet during the storytelling game. The facilitator encouraged the children to touch the shapes. Upon touching a shape, that shape would become salient for a short period by darkening the rest of the screen, and then the shapes would return to their initial colors. After each child touched the screen once, the WoZ would press a button to begin Opie's storytelling.

2.4.2. Storytelling Phase

During the storytelling phase, Opie would run the storytelling game. The facilitator remained next to the robot for the storytelling game and would select a child to take a turn if more than one child was present. During each story the facilitator did not interrupt the story or the interaction. The children remained sitting in front of Opie for the storytelling phase. The storytelling game proceeded as follows (see **Figure 2**):

1. Opie began with a narrative while the children sat and listened. The torso tablet initially was blank.
    (a) Opie introduced itself and verbally greeted the child/children.
    (b) Opie presented an image of the target animal with a white background onto the torso tablet and introduced the story.
    (c) Opie then showed the first scene on the torso tablet while continuing to narrate.
2. The children's active participation (dialogic storytelling) began when Opie verbally asked them to find the target animal by name.
3. The child would choose an object by touching the object on the torso tablet. The touched object would become salient by darkening the rest of the scene.
    (a) If the object was a distractor, then Opie would tell the child that the object they had selected was not the target, briefly describe the distractor, and then ask the child to try again. The darkening would then be removed.
    (b) If the object was the target animal, then the robot would congratulate the child and the game would move to the next level.
    (c) If the child did not choose an object during a given time limit (10 s), Opie would attempt to regain attention using an attentional countermeasure by telling the child that help was needed and asking them to find the target again.
4. When changing level, the screen would be briefly darkened again, and Opie would tell the child that the animal had "… run away again." The target animal was repositioned in the scene to increase the difficulty and the process was repeated from step 2. The process was repeated an additional five times (six levels per scene).

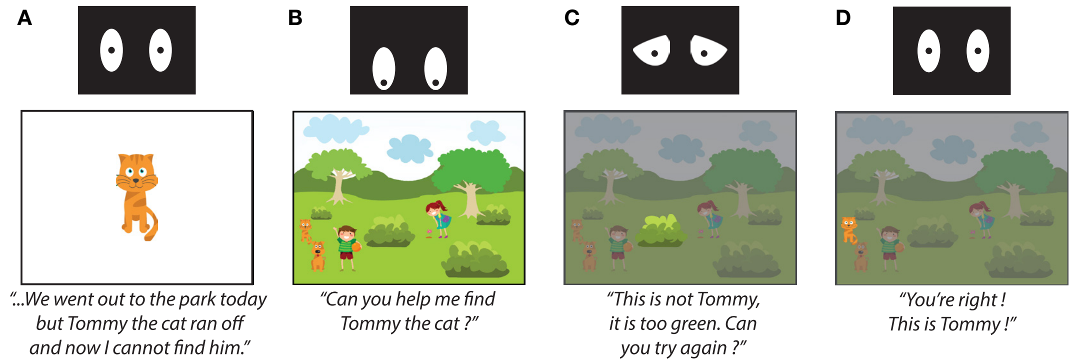

A
"...We went out to the park today but Tommy the cat ran off and now I cannot find him."
B
"Can you help me find Tommy the cat ?"
C
"This is not Tommy, it is too green. Can you try again ?"
D
"You're right ! This is Tommy !"

FIGURE 2 | The storytelling game. **(A)** The introduction to a scene, **(B)** searching for the target animal, **(C)** selecting a distractor, and **(D)** selecting the target. Opie changes expression in response to different events in the game. When the child is invited to touch the torso tablet, Opie looks down at that tablet. When the child makes a mistake, Opie adopts sad looking eyes. **(C)** and **(D)** show the effect of the overlay on the screen.

5. For each child present (up to three children), Opie would change to the next scene and repeat the process from step 1c. The start of a new scene from the beginning was the only action controlled by the WoZ.

2.4.3. Quiz Phase

After the experiment, each child's comprehension of the story was estimated by asking a series of questions. One experimenter asked the set of questions to each child present. The questions were presented on additional tablets (one per child) and consisted of recognizing visual and audio content. Children were asked to visually identify the main character and the background scene of the story from sets of five pictures and identify the name of the main character from a set of five names read out by the experimenter. After the questions were answered, the study was complete. The order of the options for each question was shuffled across participants.

2.5. Data Analysis

The behavioral data (touches, performance, and spatial movements around the robot) measuring the interaction with the first story (cat story) of ten participants (five males and five females; mean age = 54.4 months; SD = 13.7 months) were collected and analyzed.

2.5.1. Touch Patterns

Screen touch data were collected using ROS logging functionality (rosbag) and processed using Matlab. The location of each touch on the screen was recorded as well as the touch duration. Touches were automatically classified into four types:

- **target touches**, when the child touched the target object (i.e., the cat);
- **distractor touches**, when the child touched another object (distractor object) in the scene;
- **background touches**, when the child touched the background of the picture (which did not trigger any reaction from the robot); and
- **overlay touches**, when the child touched the overlay.

Each of these different touch types was expected to give different information about the interaction. Target touches suggest that the child is understanding and completing the task given by the story. Distractor touches suggest that the child understands part of the task, but is not able to find the correct object. Background touches suggest that the child does not understand the task at all. Overlay touches suggest that the child does not understand the synchronous interaction mode of the robot and that it is not possible to interrupt the robot during this time. Touch patterns were analyzed by comparing the percentage of touches that were classified as each of these four types.

2.5.2. Spatial Movement of Children

Spatial position data were extracted from the camera looking down at the scene from behind Opie's head, in order to characterize the motions of the children in the space surrounding the robot. The relative position of the child with respect to the robot was extracted from the video every 2 s, using the center of the child's forehead. Due to the noise contained in the video data, this extraction was performed manually using Manual Video Analysis (MVA) software. The data were then used to create spatial heatmaps for each participant to look at the area the participant occupied during the study. The distance between the child and the robot over the course of the interaction was computed from the spatial position data (in pixels). We applied a linear regression model with robust fitting options in Matlab (*fitlm* function) to estimate the direction of evolution of child proximity during the interaction. The significance of the linear fit was estimated by applying an analysis of variance to the model (Matlab *anova* function).

2.5.3. Quiz Data

Quiz data were aggregated for each participant to give a score out of three. Data were also aggregated for each question to give the number of participants that answered correctly so that the questions could be compared against each other to better understand what elements of the story the children recalled best.

3. RESULTS

Out of the ten participants, nine approached the robot individually, while the remaining participant was part of a group of three (P1—see spatial results in Section 3.2 and **Figure 9** for differences in position). The storytelling interaction lasted on average 373 s (6 min 13 s; SD = 96 s), from the start of Opie narrating a scene (Opie's first utterance) until the end of the scene (Opie changing scenes) (see **Figure 4** for a typical interaction). During the interaction children frequently looked at the face of the robot, which indicated that they were attending to the social functions of the robot. The robot only used attentional countermeasures (when the children did not touch the screen for more than 10 s) with two participants (respectively, one and six countermeasures used). For the quiz at the end of the story children got 2.1 questions correct on average (SD = 0.99); with nine correctly identifying who the main character of the story was; four correctly indentifying the name of the character; and eight correctly indicating the place where the story was located. A χ^2 test on these data showed that participants recalled the name of the character significantly less than its appearance ($\chi^2 = 5.5$; $p < 0.05$ after correction for comparisons across the three quiz questions).

All participant response times during the story (except one outlier) tended to converge to a stable value of level duration, after decreasing from a maximum value for the first level (see **Figure 3**).

3.1. Touch Patterns

Touches of the children mainly focused on the targets of the story. Touch data show the salience of zones containing a target at a moment of the story compared to other areas of the picture (see **Figure 5**). On average, 91.1% of the time spent touching the screen was in areas containing targets. In addition, touches in areas containing targets lasted on average 456.2 ms (SD = 276.4 ms) and were significantly longer than touches outside of these areas (average = 43.3 ms; SD = 16.6 ms; Mann–Whitney U $p < 0.01$; adjusted Z = 2.87). The significant amount of extra time that child spent touching targets indicates that children engaged with the task of finding the target.

Despite the salience of the target zones, target touches represented only 63.1% of the touches observed during the experiment. All participants completed the story and, therefore, did exactly six target touches during the story. On average, participants also did 2.90 overlay touches (SD = 1.41). Among the 29 overlay touches, 24 were measured while Opie was speaking (with 23 of them being in the first 5 s of Opie's utterance). Only three participants did background touches (five touches in total), and one participant did one distractor touch.

A variety of touch behaviors were observed during the study (see **Figure 6**). There is a concentration of the overlay and background touches at the beginning of the story (scenes one and two) (82.3% of the total number of overlay and background touches, see **Figure 7**), representing interaction opportunities missed by the robot at these instants. There is also a contrast between the first two levels of the story (with a large concentration of touches that would not trigger a robot reaction) and the rest of the interaction (with less touches not triggering reactions and longer target touches).

The temporal dynamics of the interaction between the child and the robot shows changes across the levels: the interquartile interval of the level durations appears to increase in the last

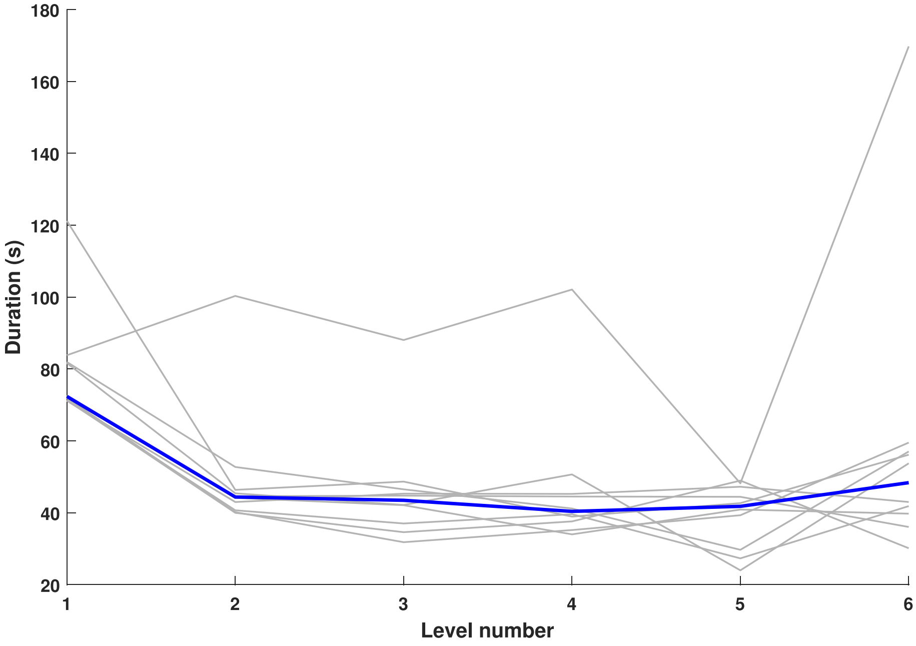


FIGURE 3 | Level duration (in seconds) per participant. The gray curves represent individual data. The blue curve shows the median across participants.

three levels, together with a stabilization of the median value of duration (see **Figure 7**). Finally, the large area covered by the interval between the third and fourth quartiles of response times for levels one, two, three, four, and six emphasizes the presence of an outlier (cf. **Figure 3**) for response times.

Among the small number of overlay touches when the robot was not talking, three were located in the fastest 25% of trials (first quartile) and four were below the median time taken (i.e., happening on shorter trials) (see **Figure 7**). Similarly, half of the background and distractor touches were located in the longest 25% of trials (fourth quartile) and four out of six were above median time taken (i.e., happening on longer trials).

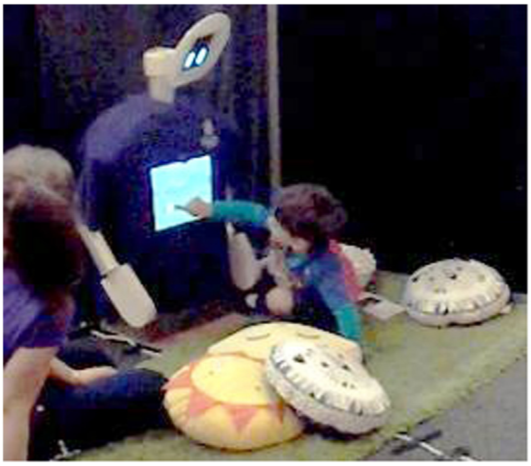


FIGURE 4 | A typical interaction between Opie and a child. The child presses Opie's torso tablet within the interactive story.

3.2. Spatial Motion

Spatial position heatmaps (see Section 2.5.2) were extracted from the camera view to show preferred locations of children over time (see **Figures 8** and **9**). All the locations corresponded to a distance of less than 1 m away from the robot (the participants stayed on the rugs), and all children that approached the robot alone (from P2 to P10) stayed on the left-hand side of the robot. The participant that was part of a group of three—P1—had no outlying temporal behavior but was an outlier in spatial motion due to the constraints caused by the other children in the area. In particular, P1 was constrained to the right side of the robot, while all the other participants approached on the left.

For seven out of ten participants, there was a statistically significant decreasing linear trend for the distance, suggesting that the children got closer to the robot over time (see **Figure 10**). All participants' distances exhibited a large variance over time, due to the back and forth motions between active participation and listening to the story.

4. DISCUSSION

In this article, we describe an implementation of dialogic storytelling on a child-friendly robot (Opie), based on interaction with the robot through touch during a story. The interaction was implemented in a closed-loop synchronous way, as the robot invited the children to touch the screen at some moments and did not process touch inputs the rest of the time. We explored the interaction of children with the robot in a science fair environment and measured the time and space aspects of children's engagement with the robot, based on their response time, touch patterns, and motions in the peripersonal space of the robot.

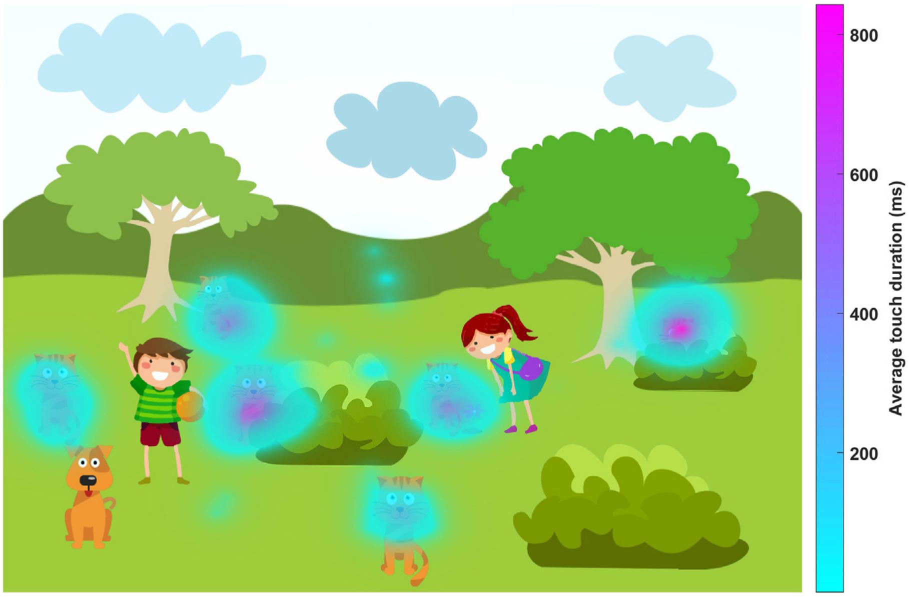


FIGURE 5 | Average time spent touching the different areas of the picture, showing the focus of children's attention on the target of the scene (cat character). Note that the targets from all six levels are shown here, while a child only sees one target in each level.

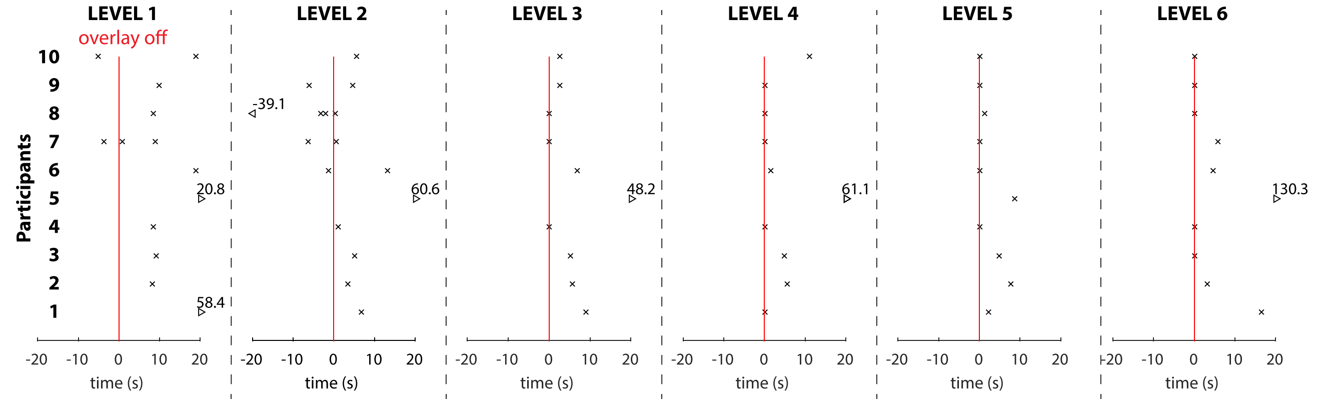


FIGURE 6 | Event-related touch raster showing touches (as x's) arranged around the removal of the overlay (the red line) on each level. The arrows indicate values that lie outside of the range of the plot. Each participant occupies one row on the Y-axis, and time is represented on the X-axis. The removal of the overlay is an important event within the interactive story as it indicates when the robot starts reacting to touches presented by the child. Children appear to adapt to the event of the overlay being removed—in earlier levels they touch the screen prior to removal and in later levels they do not.

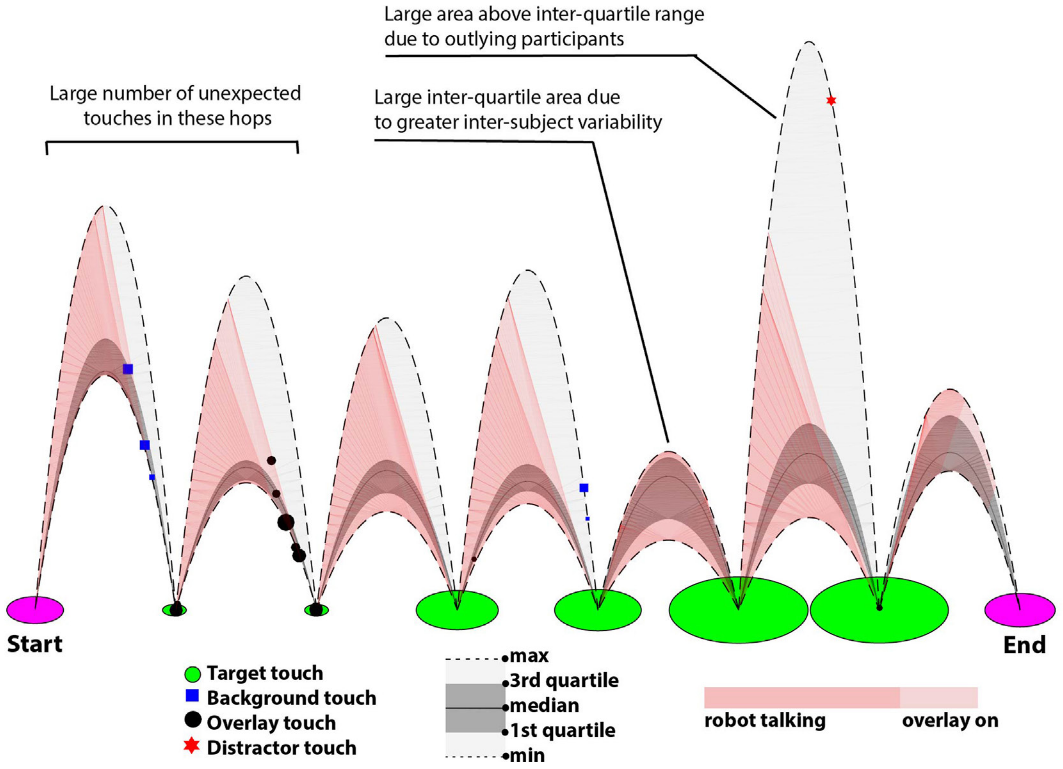


FIGURE 7 | Frog-hop plot showing the timeline of the interaction of all participants with the robot. The frog-hop plot shows the study as a series of "hops"—regions bounded by parabolas that represent a single level in the storytelling game. Each green circle in between hops represents a target touch (and, therefore, the end of a level; the size of the circle is proportional to the duration of the touch). Each hop shows the accumulated touches of all the participants, where the length of the curves representing a hop is proportional to the participant's time spent on that level (e.g., the parabola at the top of the hop shows the time taken by the slowest participant and the parabola at the bottom of the hop shows the fastest participant). The gray shaded areas represent the interquartile range of the data. The pink shaded areas represent moments when the children could not interrupt the robot, i.e., when the robot was talking or when the overlay was on. The markers in each hop represent the unexpected events: respectively, background touches (blue square), overlay touches (black circle), and non-target touches (red star). The size of the markers is proportional to the duration of the touch. The plot exhibits the variability of the behaviors observed among the participants, with a high number of unexpected events (black circles and blue squares) that would not trigger a response from the robot, particularly in the first two levels.

In a survey of human participants, de Graaf et al. (2015) listed the top three important abilities for a robot to appear social as (i) participating in two-way interactions with users—both synchronous and asynchronous, (ii) displaying thoughts and feelings, and (iii) exhibiting social awareness. In our experiment, we combined those requirements in a dialogic storytelling context. The robot not only told the story by combining speech, head motions, and displayed emotions but also engaged in a social interaction with the children by responding to their touches on the screen and using attentional countermeasures to maintain their engagement.

4.1. What Level and Duration of Engagement Can Opie Facilitate?

Within this study, we investigated performance at the interaction task (reaction times, completion of the task, and comprehension) and position and motion within close proximity of the robot. All participants completed the storytelling activity successfully, receiving directions from the social robot only. During the interaction, children remained in close proximity of the robot (<1 m) and touched the target (main character) preferentially and for a significantly longer time (cf. **Figure 5**). In addition, seven out of ten participants got closer to the robot during the interaction (the other three did not exhibit significant linear trends), suggesting greater engagement.

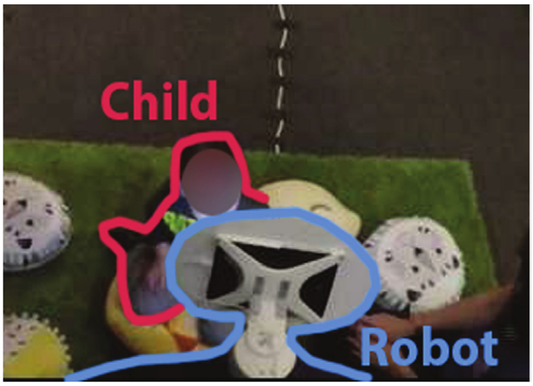


FIGURE 8 | Extraction of the children's location from the camera view. The location of the children was determined by identifying the point at the middle of their forehead.

Temporal and spatial data together indicate that the robot succeeded in creating and maintaining engagement with children during the experiment. Spatial proximity data show the existence of preferred locations for interaction with the robot (see **Figure 9**), and that the children remained in the "personal" space (Hall, 1966) of the robot, which is an optimal distance for social interaction. Previous studies on human–robot interaction have supported the hypothesis that presence in the space less than 1 m away from the robot and greater closeness can be associated with engagement (Vázquez et al., 2014).

Touch data revealed a greater attentional focus directed toward the target, which shows that the robot succeeded at sharing the goal of the interaction with the children during the interaction. As shared intentionality has been argued to be a major correlate of engagement (Tomasello et al., 2005), the result suggests that children successfully engaged with the robot. In most cases, this maintenance of engagement did not require attentional countermeasures (only two participants out of ten received a countermeasure), which also supports the idea that the engagement was the result of a shared intentionality rather than forced by the use of countermeasures. Furthermore, the greater touch duration on the target compared to distractors or background areas also reinforces this conclusion, as duration of touch has been associated with greater engagement levels (Baek et al., 2014 and Silvera-Tawil et al., 2014).

Similarly, the performance in the quiz showed good recall of the elements of the story as a result of engagement. The poor performance at recalling the name of the main character could be due to the difficulty in recalling auditory compared to visual information (Jensen, 1971 and Cohen et al., 2009). The synthetic

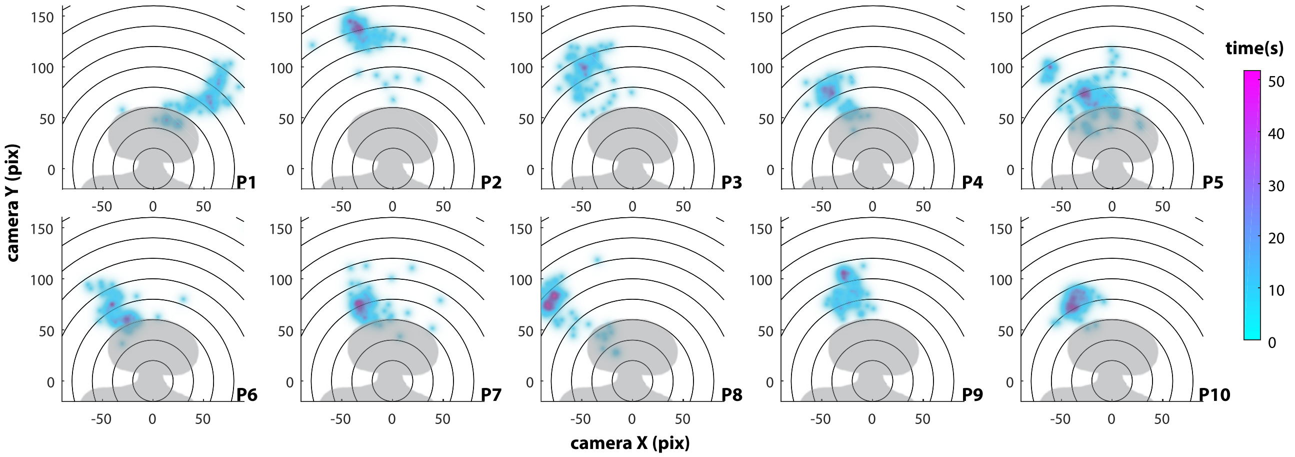


FIGURE 9 | Representation of all children's motion in Opie's peripersonal space (in pixels, measured on the image extracted from the camera overlooking the scene). Colored areas represent the time spent in each location (in seconds).

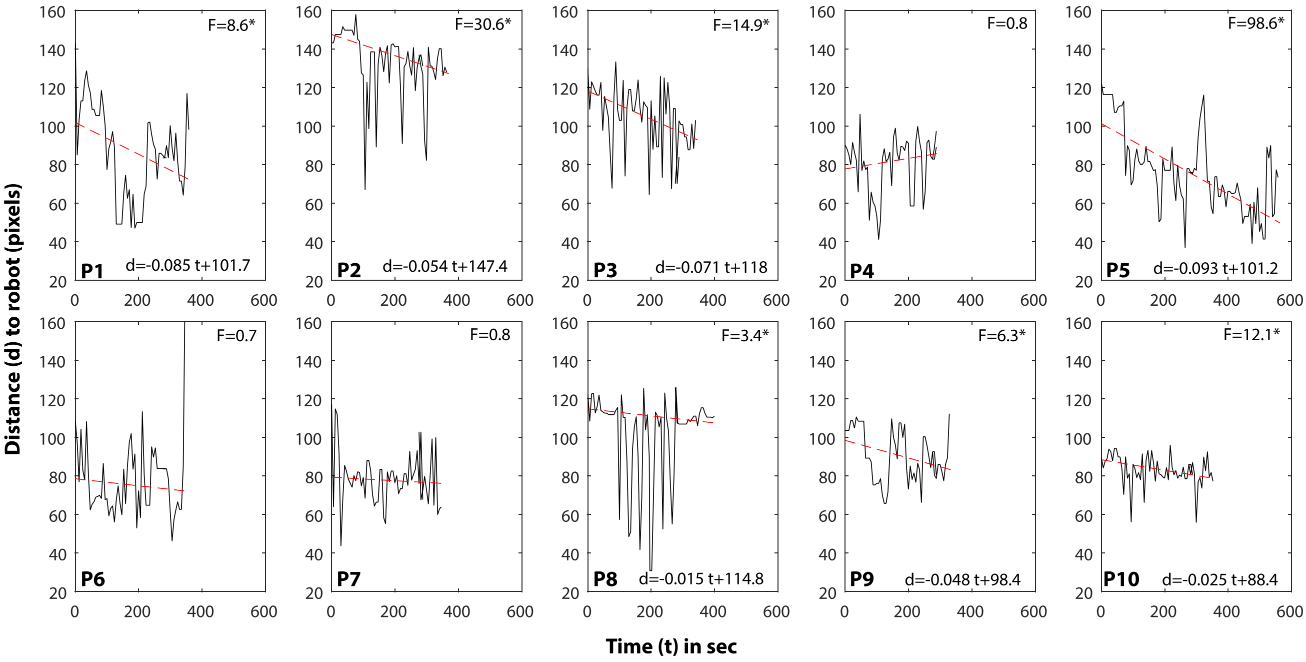


FIGURE 10 | Children's distance from the robot over time. The distance of each participant to the robot is shown in one plot. The Y-axis is distance (in pixels from the facing camera) while the X-axis is the interaction time. The red dashed line indicates the linear trend that is derived from *robust* linear regression. The ANOVA results to estimate significance of the fit are given in terms of F score for each participant. * indicates significance, and the linear equation is given (at the bottom of the graph) for significant trends.

speech used by the robot or difference to the other quiz questions (the experimenter read this question to the child) could also explain this result.

The robot maintained engagement with the children for an average of 6 min and 13 s, which is a long duration compared to other interactive settings in public spaces (between 3 and 4 min, see Hornecker and Stifter (2006)). However, this length of time is still short for educational purposes, where difficulties in engaging children can appear after a longer period of interaction (Ioannou et al., 2015). In particular, the implementation of dialogic storytelling proposed in this paper involves directional feedback and consideration of the turn taking rhythms, which are two of the three elements identified by Robins et al. (2005) to effectively maintain engagement. The current state of the robotic platform used in the study did not allow us to consider the third element: interaction kinesics (which would require moving limbs). Further investigation is required to study the impact interaction kinesics would have on sustained engagement.

4.2. How Individual or Stereotypical Are the Spatial and Temporal Reactions across Different Participants?

A possible explanation of the high level of engagement seen in this study is the adaptability of the robot's behavior during dialogic storytelling, as the robot was able to produce socially contingent responses to some of the children's actions by using verbal and emotional responses. Bartneck (2008) argued that one of the major bottlenecks of social robotics is that practical implementation often requires producing a system that has generalizable features, but having an impact on society requires the capability to adapt to each user independently of group behaviors.

In our study, we introduced a novel visualization (the frog-hop plot, see **Figure 7**) which is intended to show an overview of different touch events and how they relate across participants and storytelling state. The frog-hop plot exhibits the unexpected individual behaviors of the children and reveals the temporal dynamics of the interaction across the different levels of a scene.

From a spatial perspective, the location in the space surrounding the robot also showed that patterns of engagement were different across children. In particular, although our data suggest that the children got closer to the robot over time, we also observed a large variance of the location of the children around their linear trends (see **Figure 10**). The variance observed was likely a result of the turn-taking dynamics of the interaction, which required the child to alternatively touch the robot or listen to it. This is similar to Michalowski et al. (2006)—despite the existence of optimal areas for social interaction (the personal space), individual patterns of motion in the space should be considered to fully understand the dynamics of engagement.

Implementing engaging social behaviors in social robots can benefit from an awareness of the spatial and temporal features at the individual level. This recommendation is akin to previous studies that exhibited physical, social, and cultural aspects of engagement with interactive technologies (Dalsgaard et al., 2011). We suggest that a multimodal approach will help account for all these aspects when designing for engagement.

4.3. How Do the Patterns of Spatial and Temporal Reactions Relate to Opie's Synchronous Behavior?

A large number of touches did not lead to a socially contingent response (background or overlay touches) during the beginning of the interaction (see **Figures 6** and **7**). Overlay touches were associated with smaller response times. They could be indicative of a high level of engagement: as the objective of the task remained the same during the story, some children could have had such a high level of understanding and performance that they responded before being prompted to. The concentration of background and overlay touches on levels one and two suggests that the children who touched these regions then changed their behavior as the story advanced and adapted to the limitations of the robot (as the robot would not respond to these touches). Each of these touches are missed opportunities of interaction for the robot, to which the children had to adapt. Interestingly, these missed opportunities suggest that higher levels of immediacy could be obtained by implementing storytelling in a closed-loop asynchronous manner and this modification would likely result in even higher levels of engagement. This issue is left for further investigation.

5. CONCLUSION

We proposed an implementation of dialogic storytelling using a closed-loop synchronous mode of interaction in a child-friendly robot. Based on spatial and temporal features of the interaction, we conclude that our robot succeeded in engaging children in a dialogic storytelling interaction. However, one outlying child disengaged during the story, and some touches of the children did not produce a response from the robot.

Consideration of spatial and temporal attributes of the interaction is important for evaluating the engagement of participants. Our study results show that touch timing data and spatial position data demonstrate different trends over the course of the study and provide insight into the child's engagement toward the task and robot. While the spatial, temporal, and quiz data collected generally suggest engagement, each of these measures is sensitive to both the group and individual level. Temporal touch responses reveal group trends such as a concentration of overlay and background touches during the first stages of the story, while also showing unique unexpected touches and response times for individuals. Spatial data have properties that reflect not only engagement at the group level (decreasing distance with the robot over time) but also show the existence of preferred areas for each child during the interaction.

In addition to providing insight into engagement of children at the group and individual level, spatial and temporal measures also reflect the synchronous nature of the robot. This study demonstrates that closed-loop synchronous robots can facilitate engaging interactions; however, there is a distinct limitation created by the number of events that are not processed by the robot. It is not always feasible to have a reaction for every input that a robot receives, but each input could still be used to modify the robot's current state and future reactions. A future goal of the project is to implement asynchronous, closed-loop immediacy to enrich child–robot interactions with Opie, by processing all identified metrics automatically and online.

ETHICS STATEMENT

This study was carried out in accordance with the NHMRC statement on ethical conduct in human research (created in 2007—updated May 2015) and the protocol was approved by the UQ School of Psychology Ethics Committee (clearance no. 15-PSYCH-PHD-42-JH). Written informed consent was given by the parents of all subjects in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

All the authors contributed to prototype design, results discussion, and paper redaction. The experiment was designed by SH, GD, MB, KH, JT, OO, and JWi. Data collection was conducted by SH, GD, MB, KH, JT, OO, JWe, and JWi. Analysis of the data and initial drafting of the article were performed by SH, GD, and JWi.

ACKNOWLEDGMENTS

The authors would like to acknowledge Khoi Dhang (University of Queensland) for his help during the development of the storytelling game used in the study; Wilson Kong, Rollsy Ponmattam-Madassery, Sharon Koipuram, Jia Lim, Tamara Van Der Zant, Kelly Kirkland, Abigail Kelso, and Phoebe Mayne (University of Queensland) for their help during data collection.

FUNDING

The OPAL project was funded by the Australian Research Council (ARC) Centre of Excellence for the Dynamics of Language (CoEDL) (grant no. CE140100041).

REFERENCES

Baek, C., Choi, J. J., and Kwak, S. S. (2014). "Can you touch me? The impact of physical contact on emotional engagement with a robot," in *Proceedings of the Second International Conference on Human-Agent Interaction* (Tsukuba: ACM), 149–152.

Bartneck, C. (2008). "What is good? A comparison between the quality criteria used in design and science," in *CHI'08 Extended Abstracts on Human Factors in Computing Systems* (Florence: ACM), 2485–2492.

Cohen, M. A., Horowitz, T. S., and Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6008–6010. doi:10.1073/pnas.0811884106

Dalsgaard, P., Dindler, C., and Halskov, K. (2011). "Understanding the dynamics of engaging interaction in public spaces," in *IFIP Conference on Human-Computer Interaction* (Lisbon: Springer), 212–229.

de Graaf, M., Allouch, S. B., and van Dijk, J. (2015). "What makes robots social? A users perspective on characteristics for social human-robot interaction," in *International Conference on Social Robotics* (Paris: Springer), 184–193.

Fridin, M. (2014). Storytelling by a kindergarten social assistive robot: a tool for constructive learning in preschool education. *Comput. Educ.* 70, 5364. doi:10.1016/j.compedu.2013.07.043

Gelin, R., d'Alessandro, C., Le, Q. A., Deroo, O., Doukhan, D., Martin, J.-C., et al. (2010). "Towards a storytelling humanoid robot," in *AAAI Fall Symposium: Dialog with Robots*, Arlington.

Hall, E. (1966). *The Hidden Dimension. A Doubleday Anchor Book*. Anchor Books.

Heath, S., Hensby, K., Boden, M., Taufatofua, J., Weigel, J., and Wiles, J. (2016). "Lingodroids: investigating grounded color relations using a social robot for children," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction* (Christchurch: IEEE Press), 435–436.

Hensby, K., Wiles, J., Boden, M., Heath, S., Nielsen, M., Pounds, P., et al. (2016). "Hand in hand: tools and techniques for understanding children's touch with a social robot," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction, HRI '16* (Piscataway, NJ: IEEE Press), 437–438.

Hornecker, E., and Stifter, M. (2006). "Learning from interactive museum installations about interaction design for public settings," in *Proceedings of the 18th Australia conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments* (Sydney: ACM), 135–142.

Ioannou, A., Andreou, E., and Christofi, M. (2015). Pre-schoolers interest and caring behaviour around a humanoid robot. *TechTrends* 59, 2326. doi:10.1007/s11528-015-0835-0

Jensen, A. R. (1971). Individual differences in visual and auditory memory. *J. Educ. Psychol.* 62, 123. doi:10.1037/h0030655

Kennedy, J., Baxter, P., Senft, E., and Belpaeme, T. (2015). "Higher nonverbal immediacy leads to greater learning gains in child-robot tutoring interactions," in *International Conference on Social Robotics* (Paris: Springer), 327–336.

Kory, J. M. (2014). *Storytelling with Robots: Effects of Robot Language Level on Children's Language Learning*. Master's thesis, Massachusetts Institute of Technology.

Malmir, M., Forster, D., Youngstrom, K., Morrison, L., and Movellan, J. (2013). "Home alone: social robots for digital ethnography of toddler behavior," in *Proceedings of the IEEE International Conference on Computer Vision Workshops* (Sydney: IEEE), 762–768.

Mehrabian, A. (1968). Some referents and measures of nonverbal behavior. *Behav.Res. Methods Instrum.* 1, 203–207. doi:10.3758/BF03208096

Mehrabian, A. (1972). *Nonverbal Communication*. Transaction Publishers.

Michalowski, M. P., Sabanovic, S., and Simmons, R. (2006). "A spatial model of engagement for a social robot," in *9th IEEE International Workshop on Advanced Motion Control, 2006* (Istanbul: IEEE), 762–767.

Mol, S. E., Bus, A. G., de Jong, M. T., and Smeets, D. J. (2008). Added value of dialogic parent–child book readings: a meta-analysis. *Early Educ. Dev.* 19, 7–26. doi:10.1080/10409280701838603

Mutlu, B., Forlizzi, J., and Hodgins, J. (2006). "A storytelling robot: modeling and evaluation of human-like gaze behavior," in *6th IEEE-RAS International Conference on Humanoid Robots* (Genoa: Institute of Electrical and Electronics Engineers (IEEE)).

Pitsch, K., Kuzuoka, H., Suzuki, Y., Sussenbach, L., Luff, P., and Heath, C. (2009). "'The first five seconds': contingent stepwise entry into an interaction as a means to secure sustained engagement in HRI," in *RO-MAN 2009 – The 18th IEEE International Symposium on Robot and Human Interactive Communication*, Toyama.

Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., et al. (2009). "ROS: an open-source robot operating system," in *ICRA Workshop on Open Source Software*, Vol. 3 (Kobe, Japan), 5.

Robins, B., Dautenhahn, K., Nehaniv, C. L., Mirza, N. A., François, D., and Olsson, L. (2005). "Sustaining interaction dynamics and engagement in dyadic child-robot interaction kinesics: lessons learnt from an exploratory study," in *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005* (Nashville: IEEE), 716–722.

Rogers, K., Wiles, J., Heath, S., Hensby, K., and Taufatofua, J. (2016). "Discovering patterns of touch: a case study for visualization-driven analysis in human-robot interaction," in *ACM/IEEE Int Conf on Human-Robot Interaction*, Christchurch.

Ryokai, K., Lee, M. J., and Breitbart, J. M. (2009). "Childrenś storytelling and programming with robotic characters," in *Proceeding of the Seventh ACM Conference on Creativity and Cognition – C&C 09* (Berkeley: Association for Computing Machinery (ACM)).

Sidner, C. L., Lee, C., Kidd, C. D., Lesh, N., and Rich, C. (2005). Explorations in engagement for humans and robots. *Artif. Intell.* 166, 140–164. doi:10.1016/j.artint.2005.03.005

Silvera-Tawil, D., Rye, D., and Velonaki, M. (2014). Interpretation of social touch on an artificial arm covered with an eit-based sensitive skin. *Int. J. Soc. Robot.* 6, 489–505. doi:10.1007/s12369-013-0223-x

Szafir, D., and Mutlu, B. (2012). "Pay attention!," in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems – CHI 12*, Austin.

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28, 675–691. doi:10.1017/S0140525X05000129

Vázquez, M., Steinfeld, A., Hudson, S. E., and Forlizzi, J. (2014). "Spatial and other social engagement cues in a child-robot interaction: effects of a sidekick," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (Bielefeld: ACM), 391–398.

Walters, M. L., Syrdal, D. S., Dautenhahn, K., Te Boekhorst, R., and Koay, K. L. (2008). Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. *Auton. Robots* 24, 159–178. doi:10.1007/s10514-007-9058-3

Whitehurst, G. J., Falco, F. L., Lonigan, C. J., Fischel, J. E., DeBaryshe, B. D., Valdez-Menchaca, M. C., et al. (1988). Accelerating language development through picture book reading. *Dev. Psychol.* 24, 552. doi:10.1037/0012-1649.24.4.552

Wiles, J., Worthy, P., Hensby, K., Boden, M., Heath, S., Pounds, P., et al. (2016). "Social cardboard: pretotyping a social ethnodroid in the wild," in *ACM/IEEE Int Conf on Human-Robot Interaction*, Christchurch.

Yanco, H. A., and Drury, J. (2004). "Classifying human-robot interaction: an updated taxonomy," in *SMC* (The Hague: IEEE), 2841–2846.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Heath, Durantin, Boden, Hensby, Taufatofua, Olsson, Weigel, Pounds and Wiles. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 30 June 2017
doi: 10.3389/fpsyg.2017.00950

Affective and Engagement Issues in the Conception and Assessment of a Robot-Assisted Psychomotor Therapy for Persons with Dementia

*Natacha Rouaix[1], Laure Retru-Chavastel[2], Anne-Sophie Rigaud[2, 3, 4, 5], Clotilde Monnet[2, 3, 4], Hermine Lenoir[2, 3, 4] and Maribel Pino[2, 3, 4]**

[1] Sciences and Technology, Université Pierre et Marie Curie, Paris, France, [2] Arts et Métiers ParisTech, Paris, France, [3] Broca Hospital, Assistance Publique-Hôpitaux de Paris, Paris, France, [4] LUSAGE Living Lab, Research Unit EA4468, Faculty of Medicine, Paris Descartes University, Paris, France, [5] CEN STIMCO, Paris, France

The interest in robot-assisted therapies (RAT) for dementia care has grown steadily in recent years. However, RAT using humanoid robots is still a novel practice for which the adhesion mechanisms, indications and benefits remain unclear. Also, little is known about how the robot's behavioral and affective style might promote engagement of persons with dementia (PwD) in RAT. The present study sought to investigate the use of a humanoid robot in a psychomotor therapy for PwD. We examined the robot's potential to engage participants in the intervention and its effect on their emotional state. A brief psychomotor therapy program involving the robot as the therapist's assistant was created. For this purpose, a corpus of social and physical behaviors for the robot and a "control software" for customizing the program and operating the robot were also designed. Particular attention was given to components of the RAT that could promote participant's engagement (robot's interaction style, personalization of contents). In the pilot assessment of the intervention nine PwD (7 women and 2 men, *M* age = 86 y/o) hospitalized in a geriatrics unit participated in four individual therapy sessions: one classic therapy (CT) session (patient- therapist) and three RAT sessions (patient-therapist-robot). Outcome criteria for the evaluation of the intervention included: participant's engagement, emotional state and well-being; satisfaction of the intervention, appreciation of the robot, and empathy-related behaviors in human-robot interaction (HRI). Results showed a high constructive engagement in both CT and RAT sessions. More positive emotional responses in participants were observed in RAT compared to CT. RAT sessions were better appreciated than CT sessions. The use of a social robot as a mediating tool appeared to promote the involvement of PwD in the therapeutic intervention increasing their immediate wellbeing and satisfaction.

Keywords: dementia, social robots, engagement, geriatrics, psychomotor therapy, control software

OPEN ACCESS

Edited by:
Hatice Gunes,
University of Cambridge,
United Kingdom

Reviewed by:
Hatice Kose,
Istanbul Technical University, Turkey
Francisco Martín Rico,
Universidad Rey Juan Carlos, Spain

****Correspondence:***
Maribel Pino
maribel.pino@aphp.fr

Specialty section:
This article was submitted to
Human-Media Interaction,
a section of the journal
Frontiers in Psychology

Received: *13 January 2017*
Accepted: *23 May 2017*
Published: *30 June 2017*

Citation:
Rouaix N, Retru-Chavastel L, Rigaud A-S, Monnet C, Lenoir H and Pino M (2017) Affective and Engagement Issues in the Conception and Assessment of a Robot-Assisted Psychomotor Therapy for Persons with Dementia. Front. Psychol. 8:950. doi: 10.3389/fpsyg.2017.00950

INTRODUCTION

Psychosocial interventions, such as cognitive stimulation, physical activities and art-mediated therapies, play a key role in dementia care. Several studies show a positive impact of these interventions on the well-being, cognition, social life and daily functioning of persons with dementia (PwD) (Hulme et al., 2010; Vernooij-Dassen et al., 2010; Dickson et al., 2012;

Oyebode and Parveen, 2016). In recent years a growing number of studies have focused on the use of social robots in interventions for PwD. Social robots offer the possibility of engaging and stimulating the user through social interaction (speech, gestures, behavior). A wide range of robots interpreted as communicative and socially aware fall under this category (Ess et al., 2014), including humanoid, animal-like and some machine-like robots (**Figure 1**). Most social robots offer a great flexibility of programming allowing the creation of diverse behaviors and customization. For this reason, they have a great potential to support care interventions taking into account inter-individual differences, a well-known success factor in dementia care.

A good number of robot-assisted therapies (RAT) for PwD have used the seal robot PARO (AIST, Japan). Several studies have reported beneficial effects of PARO in PwD, such as an improvement on general well-being and social interaction (Wada and Shibata, 2007), a reduction of stress (Broekens et al., 2009; Mordoch et al., 2013), and diminished use of psychoactive and pain medications (Petersen et al., 2017). Fewer studies have explored the effects of RAT using humanoid robots with elderly persons with cognitive impairment.

López Recio et al. (2013) evaluated the feasibility of using the NAO robot (Softbank robotics, Japan) as an assistant in an individual physiotherapy program with 13 older adults in an assisted living facility. Three conditions were compared: (a) "classic therapy" in which the physiotherapist worked alone, (b) "ViNAO therapy" in which the therapist used a virtual NAO, displayed on a screen, to show the movements the inpatients should mimic and to provide them with feedback; and (c) "PhyNAO therapy" in which the therapist used a real NAO robot for the same purpose. Based on the requirements of the therapist some software modules and a user interface were developed to program NAO's movements and operate it during sessions. A good acceptance by participants was observed. Participants tried to synchronize their movements with those of the robot indicating a good compliance with RAT. One of the advantages of using the robot as an external model was that it allowed the therapist to be more available to mobilize directly the patient. Therefore, the robot contributed to reduce the therapist's workload and improve his interactions with the patients. All participants agreed that the robot's movements were natural and preferred unanimously the real robot to the virtual one. However, it was noted that technical limitations of the robot's hardware affected sometimes the way it performed the exercises (e.g., movements with less amplitude), an inaccuracy that was also mimicked by participants.

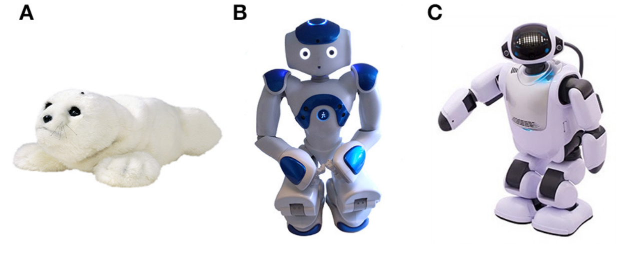


FIGURE 1 | Examples of social robots. **(A)** PARO (AIST, Japan); **(B)** NAO (Softbank robotics, Japan); and **(C)** PALRO (Fujisoft, Japan).

Martín et al. (2013) and Valentí Soler et al. (2015) evaluated the use of the NAO robot in cognitive and occupational therapy with 50 elderly PwD in two settings, a day care center and an assisted living facility. NAO was used in individual and group therapy sessions to assist the therapist by playing audio contents and carrying small objects used for the activities. Specific robot's scripts developed for the activity included speech, music and movement. A mobile device was used as remote control by the therapist to operate the robot. Main results from this 3-month experience were a good acceptance of the robot and the improvement of neuropsychiatric symptoms of dementia, such as apathy and irritability, in the group who benefited from the RAT with the NAO.

Results from previously cited studies show that humanoid robots have the potential to provide assistance for psychosocial interventions in dementia care, particularly, when the robot's role and behavior has been defined according to the needs of care professionals and PwD. However, further work is needed to identify the elements of RAT using humanoid robots that are likely to result in clinical improvements in PwD. Moreover, published studies have not dealt in detail with the quality of human-robot interaction (HRI) between PwD and humanoid robots.

In this respect, the assessment of participant's *engagement* in RAT could prove useful. Indeed, one of the factors contributing to the effectiveness of dementia care interventions is their ability to engage participants and ensure their adherence. Engagement in this context has been defined by the act of being occupied or involved with an external stimulus (Cohen-Mansfield et al., 2009). Factors such as the person's characteristics and his/her personal history, the type of stimulus and the environmental conditions in which the activity takes place, all have been found to influence the engagement that a specific individual may have with an activity (Cohen-Mansfield et al., 2010). In recent years, some models for studying engagement of PwD when participating in an activity have been developed and applied to different interventions, for instance the *Observational Assessment of Engagement* (OME) (Cohen-Mansfield et al., 2009) and the *Menorah Park Scale* (Judge et al., 2000). More recently, Jones et al. (2015) developed the *Video Coding Protocol- Incorporating Observed Emotion* (VC-IOE), a specific approach, particularly useful for RAT, to assess engagement in PwD using video coding.

Another aspect that has been little discussed is how to program a humanoid robot to provide PwD with a natural and positive interaction, and consequently, to improve the acceptance of the robot. The work by Hamada et al. (2016) provides some elements in this respect. In their research, they used the social robot PALRO (Fujisoft, Japan) as an assistant in a physical activity therapy for PwD. The robot was used to provide the instructions on how to perform the exercises and to model the movements for the person to follow. The assessment of clinical effects of the intervention was not an objective of this study. Nevertheless, better engagement and satisfaction of participants were reported when the robot's dialogues were accompanied by gestures, when it repeated instructions to enable user's comprehension and when it verbally encouraged and complimented participants. The robot exhibiting a kindly and compassionate attitude proved advantageous in this context.

In their analysis of main challenges of socially assistive robotics, Tapus et al. (2007) explained how giving an empathetic attitude to an assistive robot would benefit HRI. Considering that empathy, the capacity of understanding other's emotions and perspectives, is as a key factor for successful therapeutic relationships, it has been recommended that RAT integrates this aspect. Tisseron et al. (2015) have also suggested that the acceptance of social robots depends on their empathic qualities. These authors proposed a model of empathy extended to four dimensions (i.e., auto-empathy, direct empathy, reciprocal empathy, and intersubjective empathy) and to four components (action, emotion, thought, and assistance) aiming at better understanding HRI.

The main objective of the present study is to investigate the feasibility of using a humanoid robot as an assistant in psychomotor therapy for PwD. The robot's potential to incite the engagement of PwD in the activity and its effect on their emotional state will also be studied. In order to increase RAT acceptance, particular attention will be given to the definition of some components of the RAT: defining a highly acceptable and empathic interaction style for the robot, tailoring the program contents to the preferences and capacities of participants, and creating a framework for RAT based on the triad composed by the therapist, the patient and the robot.

This paper is structured as follows; first we describe the design process of the robot-mediated psychomotor therapy program, including general technical aspects of contents creation and robot programming. Then, we present the experimental pilot study conducted to assess feasibility and immediate effects of the intervention. The last section of the paper provides a general discussion of results and some suggestions for future studies in RAT for dementia care.

CONCEPTION AND DEVELOPMENT OF THE RAT

The Psychomotor Therapy Program

A psychomotor therapist conceived a short therapeutic program for PwD structured in four individual sessions: a classic therapy (CT) session, in which the patient was alone with the therapist, and three RAT sessions, in which the therapist was assisted by the robot NAO. Each session comprised five sections described as follows:

(1) *Introduction:* Time for greetings and introduction of the robot (RAT).
(2) *Motor section:* The section begins with a warm-up exercise by which the person is brought to rediscover and move different parts of his/her body (e.g., head, hands, arms, legs). This exercise should contribute to raise patient's alertness and allows him/her to be physically and mentally available for the session. Then, a sequence of gestural movements is modeled by the therapist (CT), or the robot (RAT), to be repeated step by step and learnt. By stimulating the patient's motor capacities, the therapist also seeks to improve his/her awareness of preserved functional and interaction abilities.
(3) *Cognitive stimulation section:* The section begins with some personalized questions tailored to the patient's life history and interests being formulated by the therapist (CT) or the robot (RAT). The second part of this session is devoted to ask the patient some questions about his own body. The purpose of this activity is to elicit verbal exchanges in fields that were familiar to and enjoyed by the patient and to help him/her increase his/her body awareness.
(4) *Body expression section*: The patient is invited to imitate a choreography in three steps, associating a sequence of movements to a series of brief meaningless sounds such as "BA, DA, KA." The sequence is presented and modeled by the therapist (CT) or the robot (RAT). The aim of this section is to stimulate body expression through movement, voice and emotion.
(5) *Conclusion:* The session ends with a series of breathing exercises allowing the participant to relax. The exercises are presented and modeled by the therapist (CT) or the robot (RAT). A time of verbal exchange is proposed to the patient at the end of the session.

Different scenarios were created in order to anticipate possible interaction sequences involving the patient, the therapist and the robot. Verbal and non-verbal robot behaviors required for each sequence were carefully defined taking into account the technical possibilities of the robot (**Figure 2**). During this process were also identified the "personalization parameters" needed to adapt the program contents to the specific requirements of each participant.

Once the therapeutic program was defined it was submitted for validation by a multi-disciplinary team (two geriatricians,

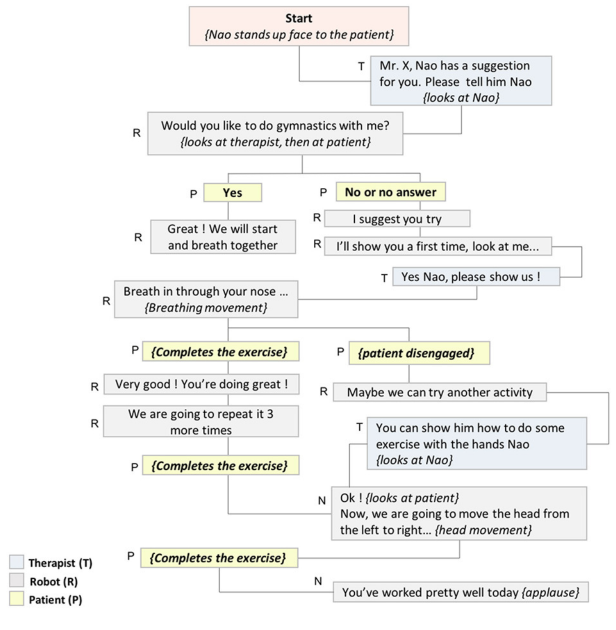

Start
{Nao stands up face to the patient}
T
Mr. X, Nao has a suggestion for you. Please tell him Nao {looks at Nao}
R
Would you like to do gymnastics with me? {looks at therapist, then at patient}
P
Yes
P
No or no answer
R
I suggest you try
R
Great ! We will start and breath together
R
I'll show you a first time, look at me...
T
Yes Nao, please show us !
R
Breath in through your nose ... {Breathing movement}
P
{Completes the exercise}
P
{patient disengaged}
R
Very good ! You're doing great !
R
Maybe we can try another activity
R
We are going to repeat it 3 more times
T
You can show him how to do some exercise with the hands Nao {looks at Nao}
P
{Completes the exercise}
N
Ok ! {looks at patient} Now, we are going to move the head from the left to right... {head movement}
Therapist (T)
Robot (R)
Patient (P)
P
{Completes the exercise}
N
You've worked pretty well today {applause}

FIGURE 2 | Example of RAT interactive scenario.

a neuropsychologist and a cognitive psychologist). Then, a computer engineer proceeded to program the robot including its behaviors and personality features. A "control software" allowing the personalization of the therapy sessions and the operation of the robot was also created. During this conception and development phase of the program, the psychomotor therapist and the engineer worked together enabling continuous feedback on the quality of the robot's movements and interactions.

Robot Programming

Presentation of the Control Software

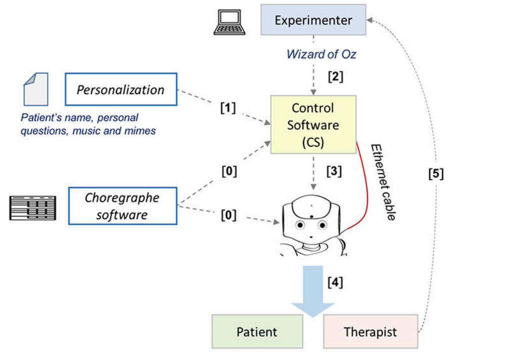
The design of the control software for operating the robot took into account two main criteria: *customization* and *intuitiveness.* Regarding *customization,* the software was designed to adapt the contents of the therapy program and some robot's features to each participant's capabilities and preferences. Customization is a key aspect in dementia care interventions to foster engagement and positive emotional responses. With this purpose, the following customization parameters were implemented: (a) adding the person's name so the robot could use it to address each person in an individualized manner, (b) selecting individual and familiar contents for the therapy activities (music, cognitive stimulation themes, adapted physical exercises,...); (c) adjusting some general robot parameters (e.g., rhythm, voice pitch, volume,...) according to each person's preferences and needs to provide the best possible user experience. *Intuitiveness* of the control interface was highly desired to ensure an easy navigation during therapy sessions, and so to allow the operator to smoothly initiate and stop robot's behaviors. The control software was created using Python language and the user interface was created with the program Qt Designer.

The control software encompassed two kind of files: structure and design files. The *structure files* which contained the raw code to run the software were: (a) the core module, and (b) the associated modules, used to define the functionalities related to movements, audio contents, properties, and software buttons. *Design files* contained the code to set and view the user interfaces. The connection to a virtual NAO robot (Choregraphe software) was set up in order to facilitate the implementation and testing of the robot's behaviors without having to connect the robot in real-time. **Figure 3** shows a schematic diagram of the system and the principles of its operation within the context of the present study.

Main control interface

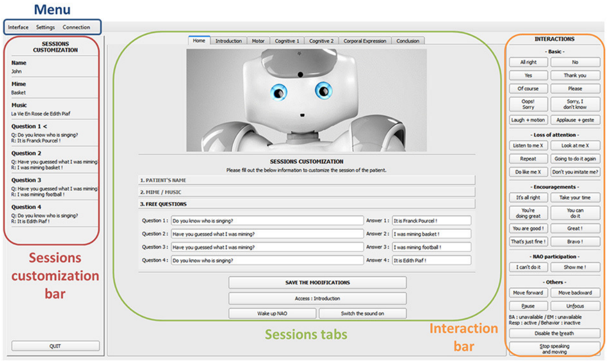
The main control interface's *central menu* (green box in **Figure 4**) included seven tabs controls: one tab to personalize the session and six tabs to manage each session section. Functionalities handled by each control tab are described in **Table 1**.

The control interface, at the top of the screen, included a menu (blue box in **Figure 4**) with three options: (a) "Interface," (b) "Settings," for customizing the robot's parameters and the session contents, and (c) "Connection," for connecting the robot. On the left (red box in **Figure 4**) a "Session customization bar" contained pre-programmed information recorded for a particular session for each individual participant. On the right (orange box in **Figure 4**) an "Interaction bar" allowing the operator to make the robot quickly react to various user's requests or responses. Options from this interaction bar allowed to make HRI smooth, for instance giving continuity to the conversations between the robot and the therapist or the participant, using basic transition words and accompanying gestures (e.g., *"All right," "Sorry, I didn't know," "Laugh + motion," "Applause + gesture")*.



[0] Preliminary creation of NAO behaviors in the Choregraphe software and installation in the NAO and the CS
[1] Choice and input of individual patient data into CS
[2] Experimenter presses a button on the CS to trigger a NAO action (Wizard of Oz technique)
[3] CS sends the command to NAO to perform the action
[4] NAO performs the requested action
[5] Experimenter will adapt to the reactions of the patient and the therapist and choose the next action to trigger

FIGURE 3 | Diagram of the system and principles of operation.



Name
John
Mime
Basket
Music
La Vie En Rose de Edith Piaf
Question 1 <
Q: Do you know who is singing?
R: It is Franck Pourcel !
Question 2
Q: Have you guessed what I was miming
R: I was miming basket !
Question 3
Q: Have you guessed what I was miming
R: I was miming football !
Question 4
Q: Do you know who is singing?
R: It is Edith Piaf !

SESSIONS CUSTOMIZATION
Please fill out the below information to customize the session of the patient.
1. PATIENT'S NAME
2. MIME / MUSIC
3. FREE QUESTIONS

Question 1 : Do you know who is singing? Answer 1 : It is Franck Pourcel !
Question 2 : Have you guessed what I was miming? Answer 2 : I was miming basket !
Question 3 : Have you guessed what I was miming? Answer 3 : I was miming football !
Question 4 : Do you know who is singing? Answer 4 : It is Edith Piaf !

BA : unavailable / EM : unavailable
Resp : active / Behavior : inactive

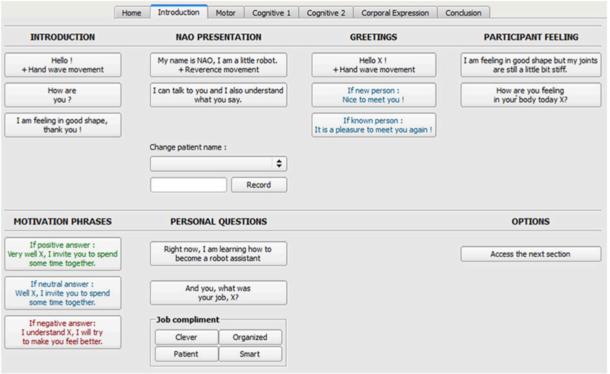
FIGURE 4 | Screen capture of the main control interface.

Additional options were proposed to deal with the loss of attention of the user (e.g., *"Don't you imitate me?" "Listen to me X (name of the person)," "Look at me X (name of the person)"*), to regularly encourage and praise the user (e.g., *"You're doing great," "Take your time," "You can do it"*), to react when the user requested the robot to make something the robot wasn't programmed for (e.g., *"I can't do it," "Show me"*), and finally to operate other robot's behaviors (e.g., walking toward and backward, making a pause, stop speaking and moving).

Each tab of the main interface (**Figure 5**) corresponding to different parts of the session included a set of buttons sorted by categories allowing a flexible leading of the session according to the participant's responses. Robot's actions were summarized on each button of the interface following a logic "dialogues to say" and "movements to achieve." For

TABLE 1 | Description of control tabs from the main interface.

Control tab	Description
Home	Set of parameters allowing personalization of the session : *participant's name, personalized content for music themes, themes for cognitive stimulation (questions/answers), and mimes for the session.*
Introduction	Set of parameters allowing the robot to greet participants, introduce itself and make a first "well-mannered" contact with the user: *asking the participant how he feels, or what he did for a living; the robot can laugh if the user touches its head.*
Motor	Set of parameters used by the robot to introduce and model the physical exercises: *the robot explains and performs breathing exercises (inhale and exhale), warm-up exercises, and various sequences of movements.*
Cognitive stimulation 1	Set of parameters used by the robot to introduce and formulate cognitive exercises: *playing music themes, performing a mime, asking questions, giving the answer to a question when the participant is not able to answer.*
Cognitive stimulation 2	Set of parameters used by the robot to ask user questions about his/her body knowledge according to his/her level of cognitive impairment (three levels of difficulty) and to provide guidance in case of error: *"touch my head," "touch my right shoulder with your right index finger"; "I think this is my left shoulder," or robot showing the answer using its body.*
Body expression	Set of parameters used by the robot to explain and perform a sequence of movements associated with sounds.
Conclusion	Set of parameters used by the robot to thank the user for participating in the activity, say *"goodbye"* with a yawn, bending and switching off.



INTRODUCTION

Hello ! + Hand wave movement

How are you ?

I am feeling in good shape, thank you !

NAO PRESENTATION

My name is NAO, I am a little robot. + Reverence movement

I can talk to you and I also understand what you say.

Change patient name :

Record

GREETINGS

Hello X ! + Hand wave movement

If new person : Nice to meet you !

If known person : It is a pleasure to meet you again !

PARTICIPANT FEELING

I am feeling in good shape but my joints are still a little bit stiff.

How are you feeling in your body today X?

MOTIVATION PHRASES

If positive answer : Very well X, I invite you to spend some time together.

If neutral answer : Well X, I invite you to spend some time together.

If negative answer: I understand X, I will try to make you feel better.

PERSONAL QUESTIONS

Right now, I am learning how to become a robot assistant

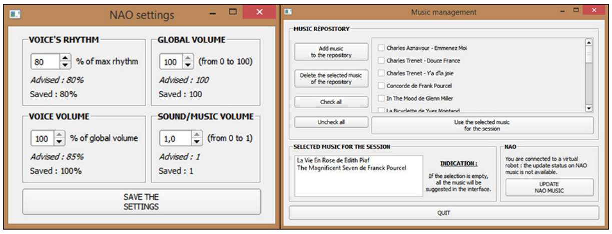
And you, what was your job, X?

Job compliment: Clever, Organized, Patient, Smart

OPTIONS

Access the next section

FIGURE 5 | Detail of the control tab for the "Introduction" section.



NAO settings

VOICE'S RHYTHM: 80 % of max rhythm
Advised : 80%
Saved : 80%

GLOBAL VOLUME: 100 (from 0 to 100)
Advised : 100
Saved : 100

VOICE VOLUME: 100 % of global volume
Advised : 85%
Saved : 100%

SOUND/MUSIC VOLUME: 1,0 (from 0 to 1)
Advised : 1
Saved : 1

SAVE THE SETTINGS

Music management

MUSIC REPOSITORY

Add music to the repository

Delete the selected music of the repository

Check all

Uncheck all

□ Charles Aznavour - Emmenez Moi
□ Charles Trenet - Douce France
□ Charles Trenet - Y'a d'la joie
□ Concorde de Frank Pourcel
□ In The Mood de Glenn Miller
□ La Bicyclette de Yves Montand

Use the selected music for the session

SELECTED MUSIC FOR THE SESSION

La Vie En Rose de Edith Piaf
The Magnificent Seven de Franck Pourcel

INDICATION : If the selection is empty, all the music will be suggested in the interface.

NAO

You are connected to a virtual robot : the update status on NAO music is not available.

UPDATE NAO MUSIC

QUIT

FIGURE 6 | Music settings manager and Audio settings manager.

example, *"Hello! X + Hand wave"* means that the robot says *"Hello! X"* and waves its hand to say hello (where X is the name of the patient). See Supplementary Material for the presentation of control interfaces for each subsection of the program.

Secondary interfaces

Three managers were accessible from the "Settings" menu on the main control interface to handle mime exercises, music and audio settings in an easy way (**Figure 6**). For example, the music settings manager allowed adding and deleting music themes to the music folder of the software and selecting the musical themes for the session according to each participant's preferences, without using the Choregraphe software. The mime exercises manager worked in a similar way but it required having created an associated behavior via Choregraphe beforehand. The audio settings manager allowed the modification in real time of volume and voice parameters of the robot. Personalized parameters, once registered, were held in the software memory and displayed when reopening each individual session.

Personality Features of the Robot

Effort was put on giving the robot an empathic and a positive attitude (e.g., being warm, polite, supportive, tolerant, gracious...). Some empathy signs, such as (a) the ability to recognize other person's emotions; (b) to communicate with persons; (c) to display emotions; and (d) to take perspective (Tapus et al., 2007), were considered when defining the robot's behavior and personality. Three other principles proposed in the field of HRI were also used in this process: (a) *interactivity*, the robot coexists with an interactive person in the same time-space continuum; (b) *equifinality*, the robot is able to adapt to each person and the same objective may be reached in different ways; and (c) *multimodality*, the robot is able to interact with a human using different communication channels (e.g., verbal, tactile, kinesthetic, or emotional) (Libin and Libin, 2004). **Table 2** presents a summary of robot's behaviors and personality traits related to the aforementioned dimensions that were implemented in this work.

MATERIALS AND METHODS

Study Design

An exploratory study aiming to assess the feasibility and immediate effects of a psychomotor therapy program for PwD using the NAO robot as an assistant was conducted between February and May 2016 in the Broca Geriatric Hospital (Paris). The intervention program consisted in 4 individual sessions of psychomotor therapy including: one classical psychomotor therapy session (CT) (therapist-patient) and 3 RAT sessions (therapist-patient-robot).

TABLE 2 | Robot's behaviors related to different HRI dimensions.

Dimension	Behaviors, attitudes, personality traits
Empathy	• Displays an emotional state and is able to acknowledge the participant's emotions and feelings. • Programmed to exhibit empathic gestures such as giving confirmation signs by head movements. • Expresses its own opinions. • Gives positive feedback and frequently acknowledges the participant's performance, boosting his/her confidence and motivation.
Interactivity	• Robot's embodiment is exploited in order to inspire participants the attribution of intentions, goals, and a personality to the robot. • The robot, often compared to a child for its size and appearance, is designed to answer and behave like a "well-mannered" child using simple sentences and childlike gestures. • The robot is programmed to automatically move its upper limbs when speaking to support verbal communication through body language. • When the robot is not talking, it is programmed to slightly undulate, giving the impression of breathing and being alive. • Regarding *proxemics,* the robot is placed on the ground so that the user has a higher view on it and dominates it. The robot is placed at a distance of about 1.50 m from the person which represents the social distance of interactions with friends and colleagues (Hall, 1966). This distance can be adjusted during the interaction to fit the dynamic of the session. • Before walking, the robot warns the person communicating the adjustment of the interactive distance.
Multimodality	• The robot shows engagement to its interlocutor through gaze and speech (e.g., *"do as I do X"* or *"look at me X"*). If the participant interrupts it, the robot is programmed to stop talking or making a movement and return to its initial position. • Robot's speech and gaze are programmed to face directly its interlocutor using the Face Detection application. • When the user touches the robot, it is programmed to laugh. At the end of the session, it is programmed to stretch and yawn before switching off.
Equifinality	• Before each session, the robot's behavior and RAT contents were customized for each user. • A set of basic and transition answers like *"yes," "no," "thank you," "please," "I don't know,"* were implemented to ensure the robot provides appropriate responses to each participant's requests. • The communication style of the robot was tailored to the abilities of older adults with cognitive disorders (e.g., simple vocabulary, short sentences). When the robot's comments are not understood by the participant the robots is programmed to repeat the sentence.

Participants

Nine persons (7 women and 2 men, mean age 86 years) hospitalized in a geriatrics unit, took part in the study. Inclusion criteria were: having a clinical diagnosis of neurodegenerative dementia and having signed a consent form. Exclusion criteria were: severe dementia (MMSE < 10/30), sensory deficit (vision and hearing) and severe acute illness impeding the participation in RAT sessions.

Tools

- A NAO robot, Version V4 (Softbank robotics).
- The "Choregraphe" software (Softbank robotics), a multi-platform application allowing the creation of behaviors for the NAO robot, its monitoring and control (version 2.1).
- A "home-made" software developed to create robot's behaviors, customize sessions, and monitoring and control the robot. The software is described in Section The Psychomotor Therapy Program.
- "The Observer XT" software, version 11.5 (Noldus), for video-based behavioral analysis.

Psychosocial Assessment Tools

- The "Mini Mental State Examination" (MMSE) (Folstein et al., 1975), for general cognitive assessment. Scores range from 0 (major cognitive impairment) to 30 (normal cognitive functioning).
- The "Neuropsychiatric Inventory-Nursing team version" (NPI-ES) (Sisco et al., 2000) for the assessment of behavioral symptoms in PwD by the nursing staff. NPI comprises 10 dimensions: delusions, hallucinations, dysphoria, apathy, euphoria, disinhibition, aggressiveness and agitation, irritability, anxiety, aberrant motor activity. Scores range from 0 to 120. Highest scores correspond to major behavioral disturbances.
- The "Self-Identity Questionnaire" (SQI) (Judge et al., 2000), used to establish a profile of customized activities for PwD, taking into account their interests and preferences.
- The "International Positive and Negative Affect Schedule Short-Form" (I-PANAS-SF) (Karim et al., 2011), used to quantify a person's emotional state in the short term, with 10 items representing either positive or negative affects (two scores ranging from 0 to 25).
- The "Instant Assessment of Wellbeing Tool" (EVIBE), for assessing immediate wellbeing and quality of life of elderly people in nursing homes (Kuhnel et al., 2014). Scores range from 1 (sadness) to 5 (happiness).
- The "Menorah Park Engagement Scale" (MPES) (Judge et al., 2000), for measuring the amount and types of engagement by PwD in the course of an activity based on behavioral analyses. Two adaptations were made to the MPES for the present study: (a) a *"robot engagement"* category was created to specify participant's emotional and behavioral responses denoting an exclusive engagement toward the robot (i.e., unrelated to the target activity), (b) an *"at ease/relaxed"* category was added to the emotional engagement dimension in order to take into account the flat affect and limited facial emotion responses commonly observed in PwD. **Table 3** presents a summary of the MEPS engagement categories and examples of responses within the context of this study.

Additionally, two Visual Analogic Scale (VAS) were built for the purposes of this study. One to assess the satisfaction of participants regarding each therapy session (Question was: *Did you enjoy the session?*); and the other to evaluate the pleasure while using the robot in RAT sessions (Question was: *Did you enjoy the presence of the robot?)* Each VAS was scored between 1 and 5 (highest values translated most positive opinions).

Procedure

The protocol of the study was explained to the Geriatrics Unit nursing staff and the geriatrician (MD) responsible for the unit who helped to identify the patients who met the criteria to take part in the trial. Two researchers contacted each potential participant and his/her relatives and gave them details on the study and the intervention. If the patient had given verbal consent to participate, an appointment was scheduled in order to make the inclusion. This study was carried out in accordance with the recommendations of Paris Descartes ethical procedures and included written informed consent from all subjects according to the Declaration of Helsinki.

On the day of the inclusion, after written consent was obtained, a clinician collected socio-demographic data and conducted the baseline neuropsychological assessment for the definition of the participant's profile (see **Table 4**). The experimental protocol consisted of four individual non-consecutive sessions over a period of 5 weeks: one CT session and three RAT sessions. **Figure 7** illustrates the different moments of the RAT sessions. Outcome variables were measured throughout the experimentation according to the schedule shown in **Table 5**.

Therapy sessions were held in the patient's hospital room. The patient was seated on a chair facing the therapist, and the robot

TABLE 3 | Summary of the Menorah Park Engagement Scale (MEPS) dimensions and examples of coding.

Type of engagement	Definition	Example of response coded
BEHAVIORAL DIMENSION		
Constructive Engagement (CE)	The person participates in the target activity. This includes motor and verbal responses in response to the target activity (e.g., commenting or making a gesture/action)	Participant responds to the therapist questions or instructions either verbally or by executing the physical movement required
Passive Engagement (PE)	The person listens to or looks at the target activity without making the actions required by the activity (repeating a movement/gesture or answering a question)	Participant watches the physical movement exercise presented by the therapist but does not reproduce the movement at his/her turn
Other Engagement (OE)	The person pays attention to something other than the target activity or does something not related to the target activity (speaking, gesturing, watching or listening to)	Participant looks out the window and talks about what he/she sees
Engagement with the robot not related to the target activity (RE)	The person is disengaged from the target activity and focuses his/her attention on the robot (touches the robot, speaks to the robot...)	Participant disengages from the therapy to interact verbally or physically with the robot in a way not related to the target activity: *"NAO, do you have a girlfriend?*
Non-engagement (NE)	The person does not participate in the target activity in any way	Participant sleeps, closes his/her eyes or stares into space
Emotion	**Definition**	**Example of coding**
EMOTIONAL DIMENSION		
Pleasure	The person clearly laughs, smiles or verbalizes a positive response/emotion during the activity	Participant distinctly shows and/or verbalizes a positive emotion: "*I'm happy,*" *"It makes me feel good"*
Anxiety/sadness	The person cries, looks sad, looks down, shows a tight facial expression, or verbalizes a negative response/emotion during the activity	Participant shows and/or verbalizes a negative emotion *"I feel useless," "it makes me feel sad"*
At ease/relaxed	The person is calmed, peaceful, comfortable at the activity	Person appears serene, shows a neutral expression

TABLE 4 | Demographic and clinical characteristics of the sample.

N°	Gender	Age	Education level	Diagnostic	MMSE (0–30)	NPI –ES dominant profile	NPI-ES (0–120)
1	Female	68	6	Alzheimer's disease	15	Agitation	5
2	Female	88	6	Parkinson's disease	22	Anxiety	15
3	Female	90	4	Mixed dementia	16	Agitation	12
4	Female	95	3	Mixed dementia	12	Dysphoria/depression	7
5	Female	92	7	Alzheimer's disease	16	Apathy	15
6	Male	92	7	Lewy body dementia	12	Agitation	12
7	Male	84	7	Mixed dementia	13	Apathy	14
8	Female	89	4	Neurodegenerative disease	19	Anxiety	7
9	Female	76	4	Neurodegenerative disease	19	Apathy	3

EL, Education level, ranging from 1 (validation of primary school) to 7 (higher education degree); MMSE,Mini Mental State Examination; NPI-ES, Neuropsychiatric Inventory-Nursing team version.

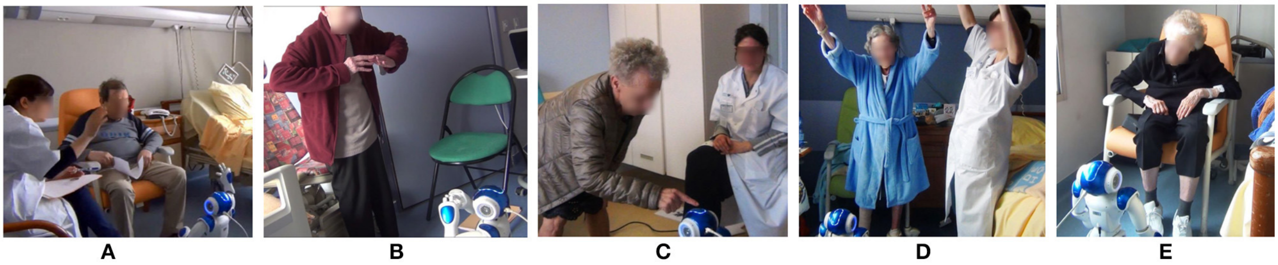


FIGURE 7 | Robot-assisted psychomotor therapy sessions. **(A)** Introduction, **(B)** motor section **(C)** cognitive Stimulation, **(D)** body expression section and **(E)** conclusion.

TABLE 5 | Evaluation criteria and schedule of assessments throughout the experimentation.

Assessment criteria	Tool	Baseline	CT	RAT1	RAT2	RAT3	Post
Cognitive functioning	MMSE	✓	-	-	-	-	-
Neuropsychiatric symptoms	NPI	✓	-	-	-	-	-
Life history and preferences	SQI	✓	-	-	-	-	-
Emotional state	PANAS	✓					✓
Immediate wellbeing	EVIBE	-	*Pre Post*	*Pre Post*	*Pre Post*	*Pre Post*	-
Engagement	MPES	-	✓	✓	✓	✓	-
Satisfaction with intervention	VAS	-	✓	✓	✓	✓	-
Appreciation of robot	VAS	-	-	✓	✓	✓	-
Verbal and nonverbal empathy related behaviors	Video analysis	-	-	✓	✓	✓	-

CT, Classic Therapy; RAT, Robot-Assisted Therapy (1,2,3 for sessions 1,2,3 respectively); Post, assessment after intervention; MMSE, Mini Mental State Examination; NPI–ES, Neuropsychiatric Inventory-Nursing team version; SQI, Self-Identity Questionnaire; PANAS, International Positive and Negative Affect Schedule; EVIBE, Instant Assessment of Wellbeing Tool; MPES, Menorah Park Engagement Scale; VAS,Visual Analogic Scales; Pre Post, assessment before and at the end of each therapy session, fields marked with a ✓ indicate that the variable was assessed at that time point; fields marked with a — indicate that the variable was not assessed at that time point.

in RAT sessions. The experimenter (engineer) who operated the robot was sitting back in the room with the computer which remained visible to the participant. The experimenter used the Wizard of Oz (WOZ) technique to remotely control the robot's movements, speech, and gestures (Kelley, 1984).

Data Analysis

The encoding and analysis of the video recordings was carried out by two researchers using the adapted form of the MPES (Judge et al., 2000). The order of video analysis was randomized. Analysis of the engagement was performed using time percentage with respect of the total time of each session's section (motor, cognitive stimulation, and body expression). Statistical analyses of neuropsychological measures were performed using the Wilcoxon test to compare means. For these analyzes, the significance level used was 95% (alpha = 0.05).

RESULTS

General Results

A total of 35 therapy sessions were conducted: 8 CT sessions and 27 RAT sessions. The sessions had a mean duration of 22.15 min, for a total of 770.19 min altogether that were video-analyzed. **Table 6** presents mean duration of the sessions detailing each subsection. All the participants underwent the four experimental sessions as stated in the protocol, except one participant who refused to take part in the CT session. **Table 7** presents a summary of a RAT session.

TABLE 6 | Mean duration of the sessions (total and each section's).

Session	Introduction	Motor	Cognitive stimulation	Body expression	Conclusion	Total
MEAN DURATION (MIN)						
CT	0.57	8.28	7.50	1.55	0.56	18.48
RAT 1	2.70	8.75	9.70	2.50	1.27	25.54
RAT 2	1.34	8.46	9.78	1.93	1.80	23.68
RAT 3	0.19	7.88	8.38	1.67	1.08	20.90
Total mean	1.2	8.34	8.84	1.91	1.18	22.15
SD	1.11	0.36	1.10	0.42	0.51	3.10

Engagement in the Psychomotor Intervention

Results indicated a high constructive engagement of participants in both CT and RAT sessions. **Table 8** shows the comparison of percentages in time of the different types of engagement for CT and RAT sessions, first for the entire session (all sections included) then for each subsection. To compare the engagement percentages in both conditions (CT and RAT), the values for the three RAT sessions were averaged.

TABLE 7 | Summary of a RAT session.

Dialogues	Behaviors
INTRODUCTION	
Therapist: Hello Mr. X, hello NAO.	*[looks at the patient, then at NAO]*
NAO: Hello Mr. X I think that we have already met. I am happy to see you again. *Personalized content: this is the second time Mr. X meets NAO*	*[looks at the patient, waves hand to say hello]*
Therapist: How do you feel in your body today NAO?	*[looks at NAO]*
NAO: I feel great in my body but my joints are not still well awaken. What about you Mr. X?	*[looks at the patient, head movement]*
Patient: I do not feel very well today.	
NAO: Ok Mr. X. then I will try to make you feel better with our therapist.	*[arms and head movement]*
MOTOR SECTION	
Therapist: We will begin by a short awakening, moving the different parts of our body. Which part of your body would you like to move first Mr. X?	*[looks at the patient]*
Patient: My hands.	
Therapist: NAO, do you have an idea for exercising our hands?	*[looks at NAO]*
NAO: Yes of course! We are going to open and close our hands, like this.	*[opens and closes its hands, looks at the patient]*
NAO: Now, let's do this together!	*[opens and closes its hands, looks at the patient]*
Patient:	*[opens and closes his hands like NAO]*
NAO: Very well done Mr. X.	*[affirmative head movement and applause]*
COGNITIVE STIMULATION SECTION	
Therapist: NAO, now that we have moved pretty well, I suggest that we take some time for speaking together and activating our brain.	*[looks at NAO]*
Therapist: Would you like Mr. X, if NAO asks us some riddles?	*[looks at the patient]*
Patient: Yes.	
Therapist: NAO, could you ask us a riddle about cooking?	*[Looks at NAO]*
NAO: Yes, of course! Which ingredients do we need to cook pancakes? *Personalized content: Mr. X. likes cooking*	*[looks at the patient, head and arms movement]*
Patient: eggs, flour, sugar, milk and salt!	
NAO: Well done! I would love to know as many things as you do once!	*[affirmative head and arms movement]*
BODY EXPRESSION SECTION	
Therapist: I suggest that we end the session with a shout of joy!	*[looks at the patient and then at NAO]*
Therapist: NAO, could you show us a choreography with movements and sounds to set up our shout of joy, please?	*[looks at NAO]*
NAO: with pleasure! I am going to show you how to do it for the first time: *"BA DA KA"*	*[NAO speaks loudly and shows the choreography to patient and therapist]*
NAO: Now, let's do it together Mr. X.	*[looks at the patient, inviting head and arms movement]*
Patient: yes.	*Together patient, therapist and NAO do the choreography and shout "BA DA KA"*
Therapist: Now, I suggest to do it again and shout louder!	*[looks at the patient and then at NAO]*
NAO: Yes, of course	*Together patient, therapist and NAO do the choreography and shout "BA DA KA" louder than the first time*
CONCLUSION	
Therapist: Now we have to say goodbye to NAO because it has to rest a little while.	*[looks at the patient, then at NAO]*
NAO: I had a very nice time with you Mr X. Goodbye Mr. X.	*[waves hand to say hello]*
Patient: Goodbye little boy.	*[looks at NAO]*
Therapist: Goodbye NAO.	*[looks at NAO]*
NAO:	*[NAO stretches and folds down]*

No significant difference between CT and RAT sessions was observed in any dimension of engagement, except for a significant increase in passive engagement in the Cognitive Stimulation section of RAT sessions. Robot engagement (i.e., participant disengaged from the target activity and focused on the robot) was observed in RAT but its duration was very short to consider the robot as a source of distraction.

We analyzed the relationship between Constructive Engagement, cognitive status (MMSE) and neuropsychiatric symptoms (NPI). The levels of Constructive Engagement in RAT sessions and the severity of neuropsychiatric symptoms were positively correlated ($r = 0.68$, $P < 0.05$, Spearman's rank correlation), showing that patients presenting behavioral symptoms such as apathy or agitation responded well to RAT. The correlation between Constructive Engagement and neuropsychiatric symptoms was not observed for the CT session. Furthermore, no association was observed between cognitive status (MMSE) and Constructive Engagement (independently of the condition).

TABLE 8 | Mean time percentage for the different types of engagement in CT and RAT sessions.

Type of engagement MPES	Entire session		Motor section		Cognitive section		Body Expression section	
	CT	RAT	CT	RAT	CT	RAT	CT	RAT
Constructive engagement	88%	81%	85%	79%	91%	83%	97%	84%
p-value	0.069		0.108		0.091		0.176	
Passive engagement	6%	12%	8%	15%	4%	12%	3%	10%
p-value	0.069		0.063		0.028*		0.138	
Robot engagement	/	5%	/	4%	/	4%	/	5%
Other engagement	5%	2%	7%	2%	4%	1%	0%	1%
p-value	0.344		0.075		0.593		0.18	
No engagement	1%	0%	0%	0%	1%	0%	0%	0%

*CT, Classic therapy; RAT, Robot-assisted therapy (mean of the 3 sessions); *Statistically significant values.*

Emotional Impact of the Intervention

The emotional impact of the intervention was assessed using the three kinds of responses from the "Emotional engagement" dimension of the MPES: *anxiety/sadness* (tearfulness, depressed affect), *relaxed/at ease* (neutral expression, calmed), and *pleasure-related* (evident manifestations of happiness, cheerfulness). In both conditions participants appeared to be most of the time relaxed and at ease (91% of the time in CT and 87% in RAT). Negative emotional responses were practically non-existent. Obvious pleasure-related responses were noticed during short periods of time, compared to the prevalent neutral/relaxed facial expression of participants during the therapy sessions. Nevertheless, results showed a significant statistical difference ($p = 0.018$) between CT and RAT sessions regarding the duration of pleasure-related responses (9 and 13% respectively) (**Figure 8**).

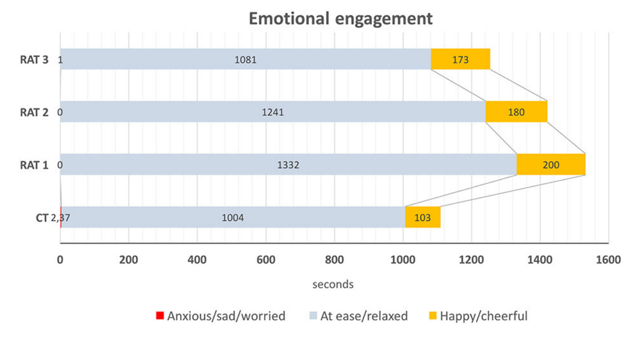


FIGURE 8 | Emotional engagement in CT and RAT sessions.

Immediate wellbeing (i.e., participant reporting feeling better after the end of the therapeutic session than before) was assessed using the difference in the EVIBE score after and before each therapy session. Highest scores indicate a highest improvement in immediate wellbeing. EVIBE scores showed a greater improvement in wellbeing in RAT sessions than in CT sessions (0.56 vs. 0.22 respectively), but this difference was not statistically significant.

The person's emotional state in the short term was analyzed by comparing the PANAS score at the baseline (baseline) and at the end of the intervention program. Results showed a significant improvement of *positive affects* (e.g., interested, excited, strong, enthusiastic, inspired, proud, alert, determined, attentive, active) (9.78 vs. 13.67, $p = 0.01$) and a decrease of negative affects (distressed, upset, guilty, ashamed, hostile, irritable, nervous, jittery, scared, afraid) (9.56 vs. 7.89, $p = 1.125$) that was not statistically significant.

Satisfaction of the Intervention and Appreciation of the Robot

Globally, all participants were satisfied with the intervention program. However, PwD preferred the RAT sessions rather than the CT one (RAT 4.31/5 vs. CT = 3.63/5). This difference regarding the modality of the therapy was statistically significant ($p = 0.027$). The robot was very well accepted by all participants as shown by a satisfaction score of 4.7/5.

Empathy Related Behavior in RAT Sessions

During the RAT sessions, various verbal and non-verbal empathy-related behaviors were observed in participants while interacting with the robot. **Table 9** provides an overview of the empathy-related behaviors exhibited by the participants. It also includes the number of participants who displayed these behaviors.

Qualitative analysis of video recordings showed that, when talking directly to the robot, three out of nine participants mostly used short sentences (e.g., *"yes"* or *"no"*) and initiated little or no dialogue with it. Among those three PwD, one participant rarely responded to the robot with a nod of his head and mostly answered the question looking at the therapist. The other six participants responded to the robot questions with complex sentences and spontaneously initiated conversations with it. As shown in **Table 9**, all the adjectives used by the participants to describe the robot were positive.

DISCUSSION

Technical Aspects

The main advantage derived from the control software created to operate the robot and customize therapy sessions was to conduct the therapeutic sessions in a smooth, fluid and natural way. The WOZ technique, used to tele-operate the robot during the experimentation, enabled the creation of natural, coherent,

TABLE 9 | Empathy-related behaviors observed in participants during RAT sessions.

Type of behavior	Examples	Number of participants (total $N = 9$)
Calling the robot by its name	*"Hello NAO"*	7
Giving an affective name (nickname) to the robot or expressing an affective feeling	*"My big one"; "My little one"; "My little chicken"; "I begin to love this little guy"*	4
Speaking directly to the robot without the intervention of the therapist	*"Yes"; "No"; "Thank You"*	8
Using an informal way of addressing the robot	*"You are cool"; "What's up?"*	7
Complementing the robot	*"You are nice"; "You are funny"; "You are cute"; "I like you very much"*	8
Contagious laughter	Smiles and laughs when the robot laughs; *"You make me laugh"*	8
Being receptive to robot's compliments	Smiles or laughs ; *"Thank you NAO"; "I am proud of your compliments"*	6
Attributing an emotional state to the robot	Asking the therapists what was the proper way to address the robot using *"Vous"* (formal) or *"Tu"* (familiar); *"Are you tired?"; "Are you happy?"; "Do you like this?"; "Are you laughing at me?"*	8
Attributing an environment or a life history to the robot	Asking whether NAO was a boy or a girl; "*Will you grow up"; "Do you have a girlfriend?"; "Your mother educated you very well"; "What do you eat?"*	4
Attributing the robot the ability to understand one's emotional state	*"I hope that I have not disappointed you"*	2
Positive behavioral manifestations	*Kissing, hugging, touching the robot*	8

and timely robot's verbal and non-verbal responses and thus to increase its capacities. However, this choice implied that the robot was not able to perform any automatic behavior. Operating the robot using the WOZ technique required thus a special sensitivity and sustained attention for achieving a high-quality HRI. Besides, the experimenter had to know well how to navigate the control interface and the location and contents of action buttons. In our case it was the developer of the software who played the role of "wizard," circumstance that simplified the task. However, the use of the control interface by an external user, despite its intuitiveness, would surely require extensive training.

In order to improve the operation of the robot in future work some possibilities can be considered:

(a) *Automatizing some of the robot's behaviors*, for instance by linking automatically the behaviors of the robot, one after the other, after triggering an action. By implementing this procedure, the number of buttons to handle in the control interface could be reduced and also the number of interventions required from the operator. Still, the risk of "over-automatizing" NAO's behavior is to greatly reduce the naturalness of the interaction.

(b) *Simplifying the control interface:* this option would require to group by categories different actions of the robot. Following this option, it could be possible to have an initial list of activity sections (e.g., introduction, motor, cognitive, etc.). The operator would then select the category wanted and a menu would display a page grouping again various subcategories of actions according to the choice. Adding a random option for some behaviors, such as the "Encouragements," that would be operated by using a single button instead of using a specific button for each phrase also goes in this direction;

(c) *Defining a decisional tree of actions* allowing to link automatically one action with the previous one, as proposed in the study of Sehili et al. (2014). However, although possible, this method would require an important work of reflection and planning to retain the flexibility of the control interface proposed in this study.

Finally, the technical setting used for this study resulted somehow complicated (e.g., transporting and installing the computer, connecting the robot by a cable, needing to accommodate the robot operator in the experimental setting). It would be interesting to adapt the control software to allow its use on a tablet, a smartphone, or any other mobile tool. After simplifying the software, the therapist could be able to operate the robot by himself. This solution has already been put into practice in other studies (Martín et al., 2013).

Factors of RAT Acceptance

Results from this experimental study showed a high level of constructive engagement among PwD throughout the intervention (indistinctly from the condition), increased manifestations of pleasure in RAT sessions, compared to CT sessions, a better appreciation of RAT sessions over CT sessions, and the exhibition of a wide range of empathy-related behaviors of PwD during RAT. All these findings represent good indicators of the advantage of using a humanoid robot for this kind of therapeutic intervention.

The choice of the humanoid robot NAO, the personalization of sessions, the "internal harmony" of the character created, empathy-related responses from the robot, and the characteristics of the therapeutic framework proposed, appeared to have contributed to create a well-accepted RAT intervention. In this section we discuss briefly these aspects:

(a) The *choice of a humanoid robot*: The humanoid aspect of NAO is a factor that facilitates its acceptance. Previous studies in this area had already confirmed the acceptance of this humanoid robot among elderly users (Wu et al., 2012; López Recio et al., 2013; Martín et al., 2013; Pino et al., 2015; Valentí Soler et al., 2015). Libin and Libin (2004) had also discussed that a key challenge of socially assistive robotics is to create robots that are able to imitate human behavior on the cognitive, motor and emotional level.

(b) *Personalization:* The flexibility of the NAO programming platform was an asset for the construction of personalized therapeutic sessions. Several studies have shown that dementia care interventions that have the greatest impact on behavioral disorders are those that are adapted to the person's cognitive, motor, and sensory abilities (Cohen-Mansfield et al., 2007) and tailored to the preferences of the person (Gerdner, 2000). The neuropsychological assessment and the use of the self-identity questionnaire (SQI) at the baseline of the experimental study, allowed us to accurately define participants' cognitive profile, and to identify their preferences and interests. This piece of information, used to program the content of the sessions, appeared to support RAT acceptance.

(c) The *"internal harmony" of the robot:* Another factor that could have contributed to RAT acceptance was the interaction style given to the NAO in our study. Regarding verbal and non-verbal communication, the robot was programmed to use simple sentences for facilitating understanding by elderly persons with cognitive impairment. Some of its behaviors were modeled also to be childlike and non-judgmental, in order to make the robot more likeable. This interactional style used to program robot's behavior was coherent with "childish" aspect of NAO. The concept of "internal coherence," suggested by Tisseron (2015), could explain the effects of our design choices on robot's acceptance. For this author, the acceptance of a social robot would strongly depend, not on its aspect but on its "internal harmony." This means, the coherence between its appearance and of its reactions.

(d) *Empathy-related responses:* For this study, NAO was designed to adapt to the cognitive level of PwD, for instance by adjusting the difficulty of exercises to each person's capacities, and by being supportive when the participant experienced some difficulties. For some participants NAO laughter facilitated the interaction with it. Fasola and Matarić (2010) have suggested that the motivation to interact with a social robot grows stronger if the interaction is adapted to the user's cognitive capacities. Being empathic, reassuring, and providing the participant with positive feedback (Vallerand, 1983) was in this perspective, another factor that could have added to the acceptance of the robot.

Several studies have highlighted as well the capacity humans have of empathic responses with artificial companions. Suzuki et al. (2015) demonstrated that humans can sympathize with the pain of a robot from a physiological point of view: in a painful situation for a robot, a neuronal response involved in empathic behavior was observed in a group of persons using an EEG (electroencephalogram) measure. Rosenthal-Von Der Pütten et al. (2014) showed an activation of the same emotional neuronal circuits when participants watched some videos showing either a human hurting another human or a human hurting a dinosaur-like robot. Activation was nevertheless more important in situations where humans were harming another human.

In order to better understand the quality of the interactions of PwD with NAO in our study, we used the model of empathy applied to HRI, proposed by Tisseron et al. (2015). This model is structured into four dimensions: (a) the *self-empathy*, empathic relationship with oneself; (b) the *direct empathy*, allowing the attribution of emotions and views to others; (c) the *reciprocal empathy,* thinking that another is able to feel our own emotions; and (d) the *intersubjective empathy*, thinking that others can bring us knowledge about ourselves and our emotional states. In our study eight participants showed direct empathy with the robot, that is, they attributed the robot emotional states and its own perspectives. Two persons showed reciprocal empathy, imagining that the robot was able to guess their emotions, or that the robot had emotions in their regard. One participant, showed intersubjective empathy by telling NAO that his compliments made him proud.

We observed conversely that when empathy-related behaviors toward the robot were absent, or uncommon, the adherence to the RAT appeared to be lower. In our study, the only participant who did not address the robot directly, did not attribute emotions to it, neither used qualifying adjectives when talking to/about the robot, appeared disengaged from the therapeutic activity. In sum, empathy toward the robot seems to be associated to engagement in RAT, but more research is needed to better measure and understand this association.

(e) *The therapeutic framework:* In our study, the therapist was a vehicle for constructive engagement in the CT sessions. The NAO robot, by its social characteristics, its humanoid aspect, and its social and affective behavior, also had the effect of engaging actively PwD. However, it is not possible to conclude that engagement observed in RAT is entirely due to the NAO itself. We observed that the therapist had an essential role in facilitating HRI as well. Indeed, at several times the therapist showed the participant how to talk to the robot or to touch it. The therapist in our framework created a true collaborative relationship with the robot as her assistant, contributing probably to help the participant accept and collaborate with the robot in a similar way. Further studies should explore this finding by comparing engagement of PwD in the three conditions: the therapist alone, the robot alone, and the therapist and the robot working together.

Studying Engagement in RAT

Overall results of this pilot study showed elevated levels of constructive engagement in both conditions (CT and RAT) comparatively higher in the first one. Conversely, passive engagement was more pronounced in RAT sessions. Though these results did not reach statistical significance, they are consistent with Cohen-Mansfield et al. (2010) study in which engagement toward 23 different stimuli, representing different levels of social attributes, was examined in 193 PwD. Results from their study showed higher levels of engagement and more positive attitude toward social, realistic and animated stimuli. Human and live stimuli appeared to be more engaging than non-human and non-alive stimuli. In our study the therapist was a vehicle for constructive engagement in the session. The robot NAO, encompassing most of the previously cited stimuli features that usually engage PwD, incited high levels of constructive engagement as well, even if it was a lower level than a real human (therapist).

From a methodological perspective, we found video-analysis to be a suitable method to examine and measure behavioral and emotional engagement in PwD during the course of an activity. However, the categories of engagement originally used in the MPES (Judge et al., 2000) resulted somehow too general in the context of RAT because they do not allow the distinction between the specific effect of the robot from the effect of the therapist or that from the environment. Also, in the MPES protocol it is not possible to differentiate the specific kind of behavior supporting engagement (e.g., visual, verbal, physical or emotional). This level of detail seems important in order to appreciate the analyze the contribution of robotic mediation. The Video Coding—Incorporating Observed Emotion (VC-IOE) tool developed by Jones et al. (2015) might provide a more coherent and comprehensive method for the assessment of engagement and merits to be tested in future studies.

Limitations of the Study

The present study presents some methodological limitations that should be taken into consideration when interpreting the above presented findings.

First, because of its exploratory nature it included a very limited number of participants and of therapy sessions. Further studies in this area should involve a larger number of subjects and a greater number of sessions in order to investigate RAT effects in the medium and long-term. Also, the sample group in this study was very heterogeneous regarding their clinical profile, aspect that limited the possibility of identifying profiles of respondents. This aspect would be an interesting dimension to examine in future work.

A third limitation refers to the absence of a valid control group. In our pilot study each patient participated only in one CT session but in three RAT sessions. This study design was chosen because of time constraints, with the idea of giving the priority to the observation of RAT sessions while keeping at the same time a baseline evaluation using a conventional therapeutic setting (patient-therapist). Since the assessment of clinical effects of the intervention was not the objective of the research, we accepted to keep the disparity between the two conditions; however, this choice impacted the quality of the results and limited the possibilities of analysis. Further studies should include a control condition truly comparable with the experimental one in terms of contents and frequency.

Finally, the results of this research should also be interpreted taking into consideration the technical possibilities of social robots today. In our experiment the robot NAO was completely controlled by an external operator who used the WOZ technique. Consequently, the observed interactions between NAO and the patients who took part in the study do not reflect to the current capabilities of such a robot. Indeed, we observed very positive HRI during RAT sessions. However, most of these interactions took place between humans: the patient, the therapist and the "wizard" who operated the robot. The fact that the robot behaved very "humanly" could explain why levels of engagement were very similar in the CT condition and in RAT sessions. We believe that this kind of "controlled" experiments are necessary to progress in the definition of the framework of RAT. Nevertheless, it seems important that future studies integrate progressively robot automation in order to examine the real possibilities of HRI with persons with cognitive impairment.

CONCLUSION

The results of this exploratory study confirmed the feasibility of robot-assisted psychomotor therapy for PwD. We were able to identify some encouraging indicators in favor of using the NAO robot in such kind of therapeutic program: a very good appreciation of the robot within this context, high positive emotional responses in RAT sessions, a better appreciation of RAT sessions, and a positive correlation between engagement of PwD in RAT sessions and the level of neuropsychiatric symptoms. Indeed, the robot NAO can be considered as a mediating tool favoring patients' engagement in psychomotor therapy when the therapist finds it difficult to motivate and involve the person in the intervention.

After improvement and simplification of the control software a larger trial would help to examine the clinical benefits of this kind of intervention, and to better understand the emotional impact of social robots in PwD. Future studies should also focus on the conception and assessment of other kinds of RAT for dementia care, such as physiotherapy or speech therapy.

AUTHOR CONTRIBUTIONS

NR and LR have equally contributed to the development of the study, data acquisition, analysis and interpretation. MP conceived and supervised the study. MP, NR, and LR drafted the article, AR, CM, and HL participated in revising it critically. All authors read and gave final approval of the version submitted.

FUNDING

This study was supported by the Assistance-Publique Hôpitaux de Paris (APHP), Directorat of Patients, Users and Associations, Department of Hospital Quality of Life (Direction des Patients, Usagers et Associations, Département Qualité de Vie, Assistance Publique-Hôpitaux de Paris).

ACKNOWLEDGMENTS

The present work benefited from the input of Grégory Legouverneur and Manon Demange who provided valuable assistance for the inclusion and neuropsychological assessment of subjects included in this study. Photo credits: PARO robot, courtesy of Cédric Maizières (INNO3MED, France); PALRO robot (Fujisoft, PALRO division).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2017.00950/full#supplementary-material

REFERENCES

Broekens, J., Heerink, M., and Rosendal, H. (2009). Assistive social robots in elderly care: a review. *Gerontechnology* 8, 94–103. doi: 10.4017/gt.2009.08.02.002.00

Cohen-Mansfield, J., Dakheel-Ali, M., and Marx, M. S. (2009). Engagement in persons with dementia: the concept and its measurement. *Am. J. Geriatr. Psychiatry* 17, 299–307. doi: 10.1097/JGP.0b013e31818f3a52

Cohen-Mansfield, J., Libin, A., and Marx, M. S. (2007). No pharmacological treatment of agitation: a controlled trial of systematic individualized intervention. *J. Gerontol. A Biol. Sci. Med. Sci.* 62:908. doi: 10.1093/gerona/62.8.908

Cohen-Mansfield, J., Thein, K., Dakheel-Ali, M., Regier, N. G., and Marx, M. S. (2010). The value of social attributes of stimuli for promoting engagement in persons with dementia. *J. Nerv. Ment. Dis.* 198, 586–592. doi: 10.1097/NMD.0b013e3181e9dc76

Dickson, K., Lafortune, L., Kavanagh, J., Thomas, J., Mays, N., and Erens, B. (2012). *Non-drug Treatments for Symptoms in Dementia: An Overview of Systematic Reviews of Non-pharmacological Interventions in the Management of Neuropsychiatric Symptoms and Challenging Behaviours in Patients with Dementia*. London: Policy Innovation Research Unit, London School of Hygiene and Tropical Medicine.

Ess, C., Sugiyama, S., Sandry, E., and Pfadenhauer, M. (2014). "Communication-theoretical issues in social robotics," in *Sociable Robots and the Future of Social Relations*, eds J. Seibt, R. Hakli, and M. Nørskov (Amsterdam: IOS Press), 153–156.

Fasola, J., and Matarić, M. J. (2010). "Robot motivator: increasing user enjoyment and performance on a physical/cognitive task," in *Proceedings: 2010 9th IEEE International Conference on Development and Learning (ICDL)* (Ann Arbor, MI), 274–279.

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatric Res.* 12, 189–198. doi: 10.1016/0022-3956(75)90026-6

Gerdner, L. A. (2000). Effects of individualized versus classical "relaxation" music on the frequency of agitation in elderly persons with Alzheimer's disease and related disorders. *Int. Psychogeriatr.* 12, 49–65. doi: 10.1017/S1041610200006190

Hall, E. (1966). *The Hidden Dimension*. Garden City, NY: Doubleday.

Hamada, T., Kawakami, H., Inden, A., Onose, K., Naganuma, M., Kagawa, Y., et al. (2016). "Physical activity rehabilitation trials with humanoid robot," in *Proceedings: 2016 IEEE International Conference on Industrial Technology* (Taipei).

Hulme, C., Wright, J., Crocker, T., Oluboyede, Y., and House, A. (2010). Non-pharmacological approaches for dementia that informal carers might try or access: a systematic review. *Int. J. Geriatr. Psychiatry* 25, 756–763. doi: 10.1002/gps.2429

Jones, C., Sung, B., and Moyle, W. (2015). Assessing engagement in people with dementia: a new approach to assessment using video analysis. *Arch. Psychiatr. Nurs.* 29, 377-382. doi: 10.1016/j.apnu.2015.06.019

Judge, K. S., Camp, C. J., and Orsulic-Jeras, S. (2000). Use of Montessori-based activities for clients with dementia in adult day care: effects on engagement. *Am. J. Alzheimers. Dis. Other. Demen.* 15, 42–46. doi: 10.1177/153331750001500105

Karim, J., Weisz, R., and Rehman, S. U. (2011). International positive and negative affect schedule short-form (I-PANAS-SF): testing for factorial invariance across cultures. *Procedia Soc. Behav. Sci.* 15, 2016–2022. doi: 10.1016/j.sbspro.2011.04.046

Kelley, J. F. (1984). "An iterative design methodology for user-friendly natural language office information applications," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '83)* (Boston, NA), 193–196.

Kuhnel, M. L., Sanchez, S., Hugault, F. B., de Normandie, P., and Dramé, M. (2014). Un outil simple pour mesurer le bien-être immédiat des personnes âgées en Ehpad. *Soins Gérontol.* 19, 9–13. doi: 10.1016/j.sger.2014.04.005

Libin, A. V., and Libin, E. V. (2004). "Person-robot interactions from the robopsychologists' point of view: the robotic psychology and robotherapy approach," in *Proceedings of the IEEE* (Piscataway, NJ), 1789–1803.

López Recio, D., Márquez Segura, E., Márquez Segura, L., and Waern, A. (2013). "The NAO models for the elderly," in *Proceedings: 8th ACM/IEE of the International Conference on Human-Robot Interaction (HRI)* (Tokyo), 187–188.

Martín, F., Agüero, C. E., Cañas, J. M., Valenti, M., and Martínez-Martín, P. (2013). Robotherapy with dementia patients. *Int. J. Adv. Robotic Syst.* 10, 1–7. doi: 10.5772/54765

Mordoch, E., Osterreicher, A., Guse, L., Roger, K., and Thompson, G. (2013). Use of social commitment robots in the care of elderly people with dementia: a literature review. *Maturitas* 74, 14–20. doi: 10.1016/j.maturitas.2012.10.015

Oyebode, J. R., and Parveen, S. (2016). Psychosocial interventions for people with dementia: an overview and commentary on recent developments. *Dementia* doi: 10.1177/1471301216656096. [Epub ahead of print].

Petersen, S., Houston, S., Qin, H., Tague, C., and Studley, J. (2017). The utilization of robotic pets in dementia care. *J. Alzheimers Dis.* 55, 569-574. doi:10.3233/JAD-160703

Pino, M., Boulay, M., Jouen, F., and Rigaud, A.-S. (2015). "Are we ready for robots that care for us?" Attitudes and opinions of older adults toward socially assistive robots. *Front. Aging Neurosci.* 7:141. doi: 10.3389/fnagi.2015.00141

Rosenthal-Von Der Pütten, A. M., Schulte, F. P., Eimler, S. C., Sobieraj, S., Hoffmann, L., Maderwald, S., et al. (2014). Investigations on empathy towards humans and robots using fMRI. *Comput. Hum. Behav.* 33, 201–212. doi: 10.1016/j.chb.2014.01.004

Sehili, M., Yang, F., Leynaert, V., and Devillers, L. (2014). "A corpus of social interaction between NAO and elderly people," in *Proceedings: International Workshop on Emotion, Social Signal, Sentiments, and Linked Open Data. Satellite of the Language Resources and Evaluation Conference (LREC)* (Reykjavik).

Sisco, F., Taurel, M., Lafont, V., Bertogliati, C., Baudu, C., Giordina, J., et al. (2000). Les troubles du comportement chez le sujet dément en institution: évaluation à partir de l'inventaire neuropsychiatrique pour les équipes soignantes (NPI/ES). *Année Gérontol.* 14, 151–171.

Suzuki, Y., Galli, L., Ikeda, A., Itakura, S., and Kitazaki, M. (2015). Measuring empathy for human and robot hand pain using electroencephalography. *Sci. Rep.* 5:15924. doi: 10.1038/srep15924

Tapus, A., Mataric, M. J., and Scassellati, B. (2007). The grand challenges in socially assistive robotics. *IEEE Robot. Autom.* 14, 35–42. doi: 10.1109/MRA.2007.339605

Tisseron, S. (2015). *Le jour où mon robot m'aimera: Vers l'empathie artificielle*. Paris: Albin Michel.

Tisseron, S., Tordo, F., and Baddoura, R. (2015). Testing empathy with robots: a model in four dimensions and sixteen items. *Int. J. Soc. Robot.* 7, 97–102. doi:10.1007/s12369-014-0268-5

Valentí Soler, M., Agüera-Ortiz, L., Olazarán Rodríguez, J., Mendoza Rebolledo, C., Pérez Muñoz, A., Rodríguez Pérez, I., et al. (2015). Social robots in advanced dementia. *Front. Aging Neurosci.* 7:133. doi: 10.3389/fnagi.2015.00133

Vallerand, R. J. (1983). The effect of differential amounts of positive verbal feedback on the intrinsic motivation of male hockey players. *J. Sport Psychol.* 5, 100–107. doi: 10.1123/jsp.5.1.100

Vernooij-Dassen, M., Vasse, E., Zuidema, S., Cohen-Mansfield, J., and Moyle, W. (2010). Psychosocial interventions for dementia patients in long-term care. *Int. Psychogeriatr.* 22, 1121–1128. doi: 10.1017/S1041610210001365

Wada, K., and Shibata, T. (2007). Living with seal robots: its sociopsychological and physiological influences on the elderly at a care house. *IEEE Trans. Robot* 23,972–980. doi: 10.1109/TRO.2007.906261

Wu, Y. H., Fassert, C., and Rigaud, A.-S. (2012). Designing robots for the elderly: appearance issue and beyond. *Arch. Gerontol. Geriatr.* 54, 121–126. doi: 10.1016/j.archger.2011.02.003

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Rouaix, Retru-Chavastel, Rigaud, Monnet, Lenoir and Pino. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers
in Robotics and AI

ORIGINAL RESEARCH
published: 27 July 2017
doi: 10.3389/frobt.2017.00036

Measuring Engagement in Robot-Assisted Autism Therapy: A Cross-Cultural Study

Ognjen (Oggi) Rudovic[1], Jaeryoung Lee[2], Lea Mascarell-Maricic[3], Björn W. Schuller[4,5] and Rosalind W. Picard[1]*

[1] *MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA, United States,* [2] *Department of Robotic Science and Technology, Chubu University, Kasugai, Japan,* [3] *Laboratory X, Je Klinik fur Psychiatrie and Psychotherapy, Charite Universitatsmedizin, Berlin, Germany,* [4] *Chair of Complex and Intelligent Systems, Universität Passau, Passau, Germany,* [5] *Department of Computing, Imperial College London, London, United Kingdom*

OPEN ACCESS

Edited by:
Hatice Gunes,
University of Cambridge,
United Kingdom

Reviewed by:
Hanan Salam,
University of Pittsburgh Medical Center, United States
Mohamed Chetouani,
Université Pierre et Marie Curie, France
Tony Belpaeme,
Plymouth University,
United Kingdom

****Correspondence:***
Ognjen (Oggi) Rudovic
orudovic@mit.edu

Specialty section:
This article was submitted to Humanoid Robotics, a section of the journal Frontiers in Robotics and AI

Received: *01 February 2017*
Accepted: *03 July 2017*
Published: *27 July 2017*

Citation:
Rudovic O, Lee J, Mascarell-Maricic L, Schuller BW and Picard RW (2017) Measuring Engagement in Robot-Assisted Autism Therapy: A Cross-Cultural Study. Front. Robot. AI 4:36. doi: 10.3389/frobt.2017.00036

During occupational therapy for children with autism, it is often necessary to elicit and maintain engagement for the children to benefit from the session. Recently, social robots have been used for this; however, existing robots lack the ability to autonomously recognize the children's level of engagement, which is necessary when choosing an optimal interaction strategy. Progress in automated engagement reading has been impeded in part due to a lack of studies on child-robot engagement in autism therapy. While it is well known that there are large individual differences in autism, little is known about how these vary across cultures. To this end, we analyzed the engagement of children (age 3–13) from two different cultural backgrounds: Asia (Japan, n = 17) and Eastern Europe (Serbia, n = 19). The children participated in a 25 min therapy session during which we studied the relationship between the children's behavioral engagement (task-driven) and different facets of affective engagement (valence and arousal). Although our results indicate that there are statistically significant differences in engagement displays in the two groups, it is difficult to make any causal claims about these differences due to the large variation in age and behavioral severity of the children in the study. However, our exploratory analysis reveals important associations between target engagement and perceived levels of valence and arousal, indicating that these can be used as a proxy for the children's engagement during the therapy. We provide suggestions on how this can be leveraged to optimize social robots for autism therapy, while taking into account cultural differences.

Keywords: autism, engagement, social robots, affective computing, human-robot interaction

1. INTRODUCTION

Autism spectrum conditions is a term for a group of complex neurodevelopmental conditions characterized by different challenges with social and reciprocal verbal and non-verbal communication, and repetitive and stereotyped behaviors (DSM-5, 2013). Social challenges are related to limitations in effective communication, social participation, social relationships, academic achievement, and/or occupational performance, individually or in combination. The onset of the symptoms occurs in the early developmental period, but deficits may not fully manifest until social communication demands exceed limited capacities. A recent meta-analysis based on 51 studies comparing children with autism and typically developing controls demonstrated a large effect size for motor issues in gait, postural control, motor coordination, upper limb control, and motor planning in autism (Fournier et al., 2010). Moreover, the overall motor performance is associated with the severity of diagnostic symptoms (Dziuk et al., 2007; Hilton et al., 2012), level of adaptive functioning (Kopp et al., 2010), and social withdrawal (Freitag et al., 2007). Children with autism also have difficulties in interpersonal synchrony (Marsh et al., 2013), which involves coordinating one's actions with those of social partners, requiring appropriate social attention, imitation, and turn taking skills (Isenhower et al., 2012; Vivanti et al., 2014).

Early engagement with the world provides opportunities for learning and practicing new skills, and acquiring knowledge critical to cognitive and social development (Keen, 2009; Kishida and Kemp, 2009). However, in children with autism, displays of engagement are usually perceived as of low intensity, particularly in their social world (Keen, 2009). This limits the learning opportunities that occur naturally in their typically developing peers. For instance, Ponitz et al. (2009) showed that higher levels of engagement in typically developing children are correlated with better learning outcomes: kindergartners who were identified as more engaged in classroom activities had higher literacy achievement scores at the end of the year than those with lower levels of engagement. Odom (2002) found that the engagement level of children with autism in inclusive settings was comparable to children with other disabilities, and slightly lower when compared to children who were developing typically. However, Kemp et al. (2013) found that children with autism were engaged during free play activities only half of the time compared to children with other disabilities, who were engaged in free play activities. Likewise, Wong and Kasari (2012) found that preschoolers with autism in self-contained classrooms were disengaged for a lower amount of time during classroom activities (e.g., free play, centers, circle, and self-care activities). Possible reasons for the variability of engagement across these research studies include various definitions of engagement used, different activities/tasks (e.g., free play, circle time, and routines), type of classroom (e.g., self-contained versus inclusive), and child to adult ratio. This is in part due to the lack of consensus about both definition and measurement of engagement in population with autism (see e.g., McWilliam et al. (1992) and Keen (2009)). Furthermore, all the studies mentioned above assess the engagement via external observers and/or questionnaires, which can be lengthy and tedious. All this poses limitations for educators primarily, but also computer scientists aiming to build the technology (i.e., social and affective robots) that can be used to assess and measure the children's engagement in a more effective and objective manner. To this end, we need first to find a suitable definition of engagement, investigate the related behavioral cues, and then build the tools and methods for its automatic measurement.

Russell et al. (2005) defines engagement as "the amount of time children spend interacting with the environment (with adults, children, or other materials) in a manner that is developmentally appropriate." Engagement is also defined as "energy in action"—the connection between a person and activity; an active, constructive, focused interaction with one's social and physical environment—consisting of three forms: behavioral, affective (emotional), and cognitive (Russell et al., 2005; Broughton et al., 2008). As described by Keen (2009), behavioral engagement refers to participation or involvement in learning activities and is related to on-task behavior, while affective engagement refers to the child's interest in the activities (also expressed by different emotions and moods). For example, in the autism therapy with robots, Kim et al. (2012) defined behavioral engagement on a 0–5 Likert scale, each level corresponding to a set of the predefined responses by the child to the tasks and prompts from the therapist. Likewise, in robot interaction with typically developing children, affective engagement is defined as "concentrating on the task at hand and willingness to remain focused" (Ge et al., 2016). Cognitive engagement can best be described as the child's eagerness or willingness to acquire and accomplish new skills and knowledge, and it relates to the goal directed behavior and self-regulated learning (Connell, 1990; Fredricks et al., 2004). As an example, Meece et al. (1988) measured the students' performance in various learning tasks, providing evidence for the better school performance as a consequence of being focused on mastering a task, persisting longer, and expressing positive affect toward the task, thus, a combination of behavioral, affective, and cognitive engagement. While these three-dimensional constructs of engagement have been widely accepted, little attention has been paid to their contextual dimension. The latter is particularly important, as in order to measure engagement, we need to know whether the child is actively participating in target activity in a *contextually* appropriate manner (McWilliam et al., 1992; Eldevik et al., 2012). For this, we need also to gather information about the background context (Appleton et al., 2006), which can be described by a number of variables, such as the child's demographics (age, gender, and cultural background), behavioral severity, individual vs. social interaction (Salam and Chetouani, 2015a), the use of tablets vs. robots, the type of therapy/tasks, and so on. To capture some aspects of this context taxonomy, Salam and Chetouani (2015b) proposed a model of human-robot engagement based on the context of the interaction (e.g., social, competitive, educative, etc.). A more recent work by Lemaignan et al. (2016) formalizes "with-me-ness," a concept borrowed from the field of computer-supported collaborative learning, to measure to what extent the human is engaged with the robot (on a Likert scale 0–5) over the course of an interactive task. While useful in measuring the attentional focus of the children interacting with a robot, "with-me-ness" does not quantify the behavioral engagement that we address in this work. Specifically, we adapt the engagement definition from Kim et al. (2012), focusing on the task-response time to define engagement levels on a 0–5 Likert scale. We study how levels of this behavioral engagement vary as a function of (i) context (the task, culture, and behavioral severity of the children) and (ii) different facets of the children affective engagement (the perceived valence and arousal levels and the face expressivity, as described in Sec. 2). Note that most of the works on engagement in human-robot interaction (HRI) report binary engagement (engaged vs. disengaged) mainly due to the difficulty in capturing subtle changes in engagement displays. However, when more complex interactions are considered, such coarse definition is insufficient to explain differences in the children's behavior. To the best of our knowledge, this is the first study that analyses the behavioral engagement (on a fine grained intensity scale) of children with autism, and in the context of assistive social robots deployed in different cultures.

2. ENGAGEMENT AND AFFECT

The most commonly used model of affect Russell and Pratt (1980) suggests that all affective states arise from two fundamental neurophysiological systems, embedded in a circumplex with two orthogonal dimensions: valence (pleasure–displeasure continuum) and arousal (sleepiness–excitement continuum). Buckley et al. (2004) found that both positive valence and arousal are good indicators of emotional engagement in learning tasks. Conversely, musical engagement has been shown to be a good predictor of perceived valence and arousal (Olsen et al., 2014). A study on happiness found that the more the subjects were engaged and satisfied, the more they experienced positive valence and high arousal (Pietro et al., 2014). In the context of HRI, Habib (2014) provides analysis of relationships between the self-reported levels of valence and arousal, and the task difficulty, which was directly related to the user's "mental" engagement with the task (engaged vs. being bored). Studies on the design of engaging personal robots emphasize the importance of these two dimensions for optimizing HRI (Breazeal, 2003). For instance, Castellano et al. (2014b) showed that in the child-robot interaction, the children's valence, interest, and anticipatory behavior are strong predictors of the (social) engagement with the robot. Motivated by these findings and considering that expressions of affect are expected to differ significantly in children with autism (Volkmar et al., 2005); in this work, we investigate the relationship between (perceived) valence and arousal, as two components of affective engagement, and the target behavioral engagement in the context of autism therapy with social robots. Note that arousal (and other facets of affect such as the face expressivity) can be measured from outward behavioral cues, e.g., facial expressions (Gunes et al., 2011), as well as inward cues, e.g., physiological signals such as the skin conductance response of the autonomic nervous system (Picard, 2009; Hedman et al., 2012). In this study, we limit our consideration to the outward behavioral characteristics of valence and arousal.

3. ENGAGEMENT IN AUTISM THERAPY WITH SOCIAL ROBOTS

Engagement with social robots for children with autism is about drawing their attention and interests toward both robot and social tasks, and maintaining the prolonged therapy sessions (Scassellati et al., 2012). Furthermore, educational, therapeutic, and assistive aspects of HRI are highly motivating environments for children with autism due to the simple, predictable, and non-intimidating nature of robots compared to humans (Robins et al., 2005; Scassellati, 2007). Also, the interaction mechanism in the field of assistive robots for children with autism is more focused on the social aspects of interaction than the physical interaction (Fong et al., 2003), such as joint attention, turn-taking, or imitation behavior, which are important target behavior for the children (Scassellati et al., 2012). In this context, the effectiveness of social interaction between robots and children increases when robotic systems have the capacity of generating coordinated and timely behaviors relevant to social surroundings (Breazeal, 2001). Such adaptive strategies for social interaction are expected to become the basis of a new class of interactive robots that act as "friends" and "mentors" to improve children's experience during, for instance, the hospital stay, and support their learning (Kanda et al., 2007; Belpaeme et al., 2013). It is, therefore, critical that the robots are able to engage the children in target activities. In social robotics, engagement is usually approached from the perspective of the design of the robot's appearance and its interaction capability. For instance, Tielman et al. (2014) showed that a robot that changed its voice, body pose, eye-color, and gestures in response to the emotions of children was perceived as more engaging than a robot that did not exhibit such adaptive behaviors. Similarly, Shen et al. (2015) showed that when the robot feedback based on the perceived user's sentiment is provided, as part of an emotion mimicry interaction, the users' were more engaged than when only a plain mimicking of the users was performed by the robot. In what follows, we review recent work providing evidence of engagement of children during interaction with robots and in the context of autism. We refer interested readers to Breazeal (2003, 2004) and Scassellati et al. (2012) for more detailed reviews of recent advances in social robotics.

The role of social robots in autism therapy is primarily (i) to act as a mediator between the therapist/caregiver and children with autism (Robins et al., 2010; Thill et al., 2012)—as in our study, (ii) to provide an interactive object to draw and maintain the children's attention (Robins et al., 2006), and (iii) to be a playful device facilitating the children's entertainment during the therapy (Scassellati et al., 2012). The advantages of using robots are, therefore, to help the children with autism to perceive and respond to the outside world through the least invasive exercises. This is mainly because the robots can modulate their behavioral responses according to the children's internal dynamics and are capable of repetitive behavior, in contrast to humans (Wainer et al., 2014). The application of interactive robots for development of communication skills in children with autism has been shown in many studies to be effective (Robins et al., 2008). Scassellati et al. (2012) and Diehl et al. (2012) observed repeatedly that children who suffer from difficulties in communication with other people surprisingly started to interact with them more easily when the communication was assisted with the robots. In a comparison of the responses to a robot vs. virtual agent environments, Dautenhahn and Werry (2004) showed that the children with autism were more engaged in playing a chasing game with the robot.

Imitative behaviors such as "reach-to-grasp" tasks performed by a human and by a robot were found more engaging and motivating when a robot was used (Pierno et al., 2008; Suzuki et al., 2017). In a pilot study of child-robot interactions (age 2–4) with a toy robot, capable of showing signs of attention by changing its gaze direction, and of articulating the emotional displays of pleasure and excitement, Kozima et al. (2007) showed that these positively engaged and influenced the children's emotional responses. Similarly, Stanton et al. (2008) reported that children with autism preferred to play with an interactive robotic dog (AIBO) rather than a toy having similar appearance but no interaction features. De Silva et al. (2009) proposed a therapeutic robot for children with autism, showing that children enjoyed interaction with the robot and that this approach enhanced their attention, based on an analysis of their eye-gaze. François et al. (2009) used a robot-assisted play, with the game designed in conformity with individual needs and abilities of each child. The authors validated their approach with a group of children with autism that were engaged in a non-directed play with a pet robot (Aibo ERS-7), which can assess the children's progress across three dimensions: play, reasoning, and affect, showing that each child exhibited highly individual patterns of play. Wainer et al. (2010) assessed collaborative behaviors in a group of children with autism, showing that the interaction with robots was more engaging, fostering collaboration among the groups through a more active interaction with their peers during the robot sessions. Focusing on behavioral cues of children with autism and those of their typical peers during interaction with NAO, Anzalone et al. (2015) showed that the children with autism exhibited significantly lower yaw movements and less stable gaze, while the posture variance was significantly lower in typical children, during a joint attention task. To summarize, all these studies evidence the benefits of using robots for facilitating the learning and interaction of children with autism. One limitation is the difficulty in generalizing these findings and comparing them across studies as different settings and performance measures were used. For this reason, in our analysis of engagement across the two cultures, we focus on two identical situations set in similar contexts (as described in Sec. 6).

4. CULTURAL DIFFERENCES

The importance of cultural diversity when studying different populations has been emphasized in a number of psychology studies (Russell, 1994; Scherer and Wallbott, 1994; Elfenbein and Ambady, 2002). For instance, the work by Scherer and Wallbott (1994) provides evidence for cultural variation in emotion elicitation, regulation, symbolic representation, and social sharing among populations from 37 countries. Likewise, Ekman (2005) found that whereas 95% of U.S. participants associated a smile with "happiness," only 69% of Sumatran participants did. Similarly, 86% of U.S. participants associated wrinkling of the nose with "disgust," but only 60% of Japanese did (Krause, 1987). Thus, subjective interpretation of specific emotions (i.e., primarily the cognitive component of emotion) differs across cultures (Uchida et al., 2004). These are seen as "cultural differences in perception, or rules about what emotions are appropriate to show in a given situation." Culture also influences expressiveness of emotions (Immordino-Yang et al., 2016). There are rather consistent patterns across Eastern and Western cultures, although differences also exist across cultures, and sometimes even within cultures (An et al., 2017; McDuff et al., 2017). Recently, Lim (2016) explored cultural differences in emotional arousal level between the East and West, focusing on the observation that high arousal emotions are valued and promoted more than low arousal emotions in the West. On the other hand, in the East cultures, low arousal emotions are valued more than high arousal emotions, with people preferring to experience low rather than high arousal emotions. Nevertheless, apart from a handful of works, virtually all studies on cultural differences focus on the typically developing population. Below, we focus our studies on cultural differences in autism.

Since autism also involves social challenges, its treatment and interventions need to be tailored to target cultures (Dyches et al., 2004; Kitzhaber, 2012; Cascio, 2015). Several cross-cultural studies highlight that the culture-based treatments are crucial for individuals with autism (Tincani et al., 2009; Conti et al., 2015). For example, Daley (2002) argues that the transcultural[1] supports are needed for the pervasive developmental conditions, including autism. So far, only a few studies have been conducted in this direction. Perepa (2014) conducted a study that investigated the cultural context in interventions for children with autism and with a diverse cultural background—British, Somali, West African, and south Asian. They found that the cultural background of the children's parents is highly relevant to their social behavior, emphasizing the importance of transcultural treatments for children with autism. However, one limitation of this study is that the target children all lived in the UK, and, thus, the role of the cultural context may have been reduced. Libin and Libin (2004) showed that the children's background, such as culture and/or psychological profile, can have a large impact on the robot therapy. Specifically, the authors conducted cross-cultural studies with Americans and Japanese in an interactive session using the robot cat called NeCoRo. Among other findings, they showed that, overall, Americans enjoyed more patting the robot than Japanese. Thus, accounting for cultural preferences is important when designing interactive games with robots. However, we are unaware of any published studies that looked into cultural differences in engagement, and, in particular, the social robots for autism therapy. A possible reason for the lack of such studies is that the heterogeneity in behavioral patterns of children with autism within cultures is already so pronounced (Happé et al., 2006). A famous adage says: "If you have met one person with autism, you have met one person with autism." Therefore, attempting analysis of these differences from a higher level (particularly, in terms of different cultures) is a far-fetched goal. Yet, it is necessary to look at these differences at multiple levels: within and between cultures, where the former would focuses on differences within and between the children with the same cultural background. This, in turn, would potentially allow the robot solutions to be adapted to each culture first by accounting for the differences that may exist among children, followed by individual adaptation to each child within a culture (e.g., by focusing on its age, gender, and psychological profile). In this work, we analyze multiple facets of engagement at each level mentioned above.

[1] In this paper, we use the term cross-cultural interchangeably with transcultural, as the latter was used in the cited works.

5. CONTRIBUTIONS AND PAPER OVERVIEW

We present a study aimed at analysis of behavioral engagement of children with autism in the context of occupational therapy assisted with a humanoid robot NAO. By focusing on two culturally diverse groups: Asia (Japan) and Eastern Europe (Serbia), we provide insights into cultural differences of engagement among these two groups in terms of (i) the task difficulty, its relationships to (ii) the affective dimensions (valence and arousal), and (iii) behavioral cues (facial expressivity). We chose these three because they are important for the design of child-robot interactions: (i) is important for the robot's ability to select a task respectful of the child's abilities, while (ii), (iii) are critical when building computer vision and machine learning algorithms that can automatically estimate the child's engagement, and, thus, enable robots to naturally engage the children in learning activities.

To the best of our knowledge, this is the first study of engagement in the context of social humanoid robots and therapy for children with autism across two cultures. Most previous work on engagement in autism focused on the discrete engagement (engaged vs. disengaged) (Hernandez et al., 2014) and within a culture. By contrast, we provide an analysis of engagement on a fine-grained scale (0–5) and in two cultural settings. Our exploratory analysis provides useful insights into the relationships between engagement dynamics as expressed within and between the two cultures, as well as its relationships to the perceived affect (valence and arousal). As one of the main findings, we provide evidence that outward displays of affect (valence and arousal) can be used as a proxy of target behavioral engagement. This confirms previous findings on the relationship between engagement and affect displays in typical individuals within a single culture (Buckley et al., 2004). Based on this, we provide suggestions for future research on automated measurement of children's engagement during robot-assisted autism therapy, which takes the cultural diversity of the children into account. We also provide an overview of the most recent efforts in the field of social and affective robots for autism.

The rest of the paper is organized as follows: we first describe the data and methods used to elicit and analyze engagement of children with autism. We then present and discuss our results. In the light of these results, we provide insights into the current challenges of engagement measurement (with the focus on automated methods) and provide suggestions for future research.

6. RESEARCH DESIGN, DATASET, AND CODING

We used the dataset produced by interactions between children with autism, specialized therapists, and NAO[2] robot. The interaction was recorded as part of occupational therapy for children with autism, following steps designed based on the Theory of Mind (ToM) (Cohen, 1993) teaching approach to emotion recognition and expression (Howlin et al., 1999). In the original version, the children are asked by the therapist to pair the images of people's expressive faces (see **Figure 1**) with four basic emotions (happiness, sadness, anger, and fear (Gross, 2004)) through storytelling. For the purpose of the study, the scenario was adapted to include NAO as an assistive tool in the tasks of emotion imitation and recognition.

6.1. Protocol

The interaction started with free play with NAO. Once the child felt comfortable, the following phases were attempted. (1) *Pairing* cards of static face images with the NAO's expressions: the therapist shows the card of an emotion and then activates NAO, via a remote control using the wizard-of-Oz approach (Scassellati et al., 2012), to display its (bodily) expression of that emotion. This was repeated for all four emotion categories. (2) *Recognition*: the therapist shows the NAO's expression of a target emotion and asks the child to select the correct emotion card. If the child selects the correct emotion card, the therapist moves to the next emotion, also providing a positive feedback; otherwise, the therapist moves to another emotion without the

[2] https://www.ald.softbankrobotics.com/en/cool-robots/nao.

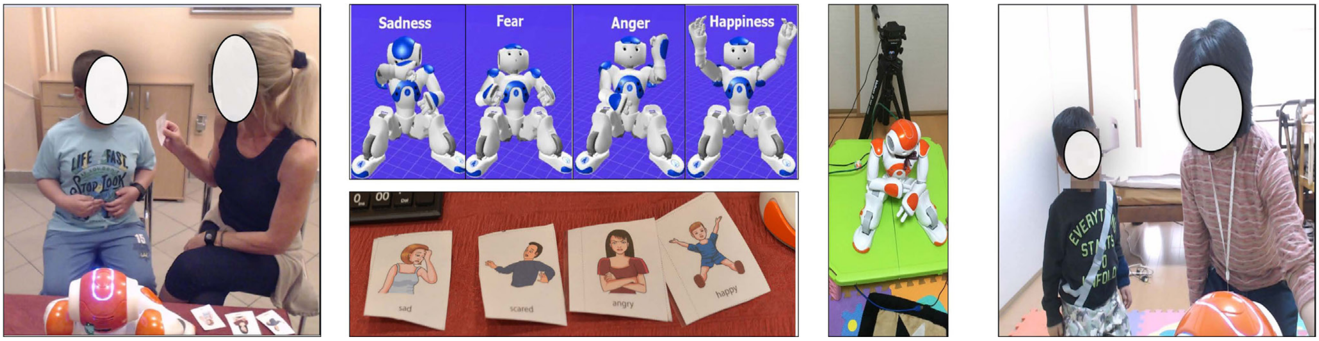


FIGURE 1 | The recording setting at the site in Japan (right) and Serbia (left). The NAO's expressions of four basic emotions (sadness, fear, anger, and happiness), paired with the expression cards, are depicted in the middle. Note that in Japan the children were seated on the floor, while children in Serbia were using a chair—a reflection of the cultural preferences.

feedback. This is iterated until either the child correctly paired all emotions, or the therapist decided that the child was unable to do so. (3) *Imitation*: the therapist asks the child to imitate the NAO's expressions. (4) *Story*: the therapist tells a short story involving NAO and asks the child to guess/show how NAO would feel in that particular situation. Note the increasing difficulty in executing each of the four phases: the Pairing phase requires minimal motor and cognitive performance. On the other hand, the Story phase is the most challenging, as it requires "social imagination," which has been shown to be limited in children with autism (Howlin et al., 1999).

The whole interaction lasted, on average, 25 min per child. In cases when a child was not interested in the play, the therapist would use occasional prompts, calling her/his name, and/or activate NAO, who then waved at the child, saying "hi, hello," and alike, to (re)engage the child. It is important to mention that the purpose of the designed scenario was not to improve existing therapies for autism, but rather induce a context in which the engagement can be measured more objectively (see **Table 2**). This protocol was reviewed and approved by relevant Institutional Review Boards (IRBs),[3] and informed consent was obtained in writing from the parents of the children participating in the study.

6.2. Participants

The children participating in the study included 17 from Asia (Japan) and 19 from Eastern Europe (Serbia), age 3–13 (see **Table 1**). They were all referred based on a previous diagnosis of autism. After the interaction with the robot, the child's behavioral severity was quantified using the Childhood Autism Rating Scale (CARS) (Baird, 2001), a diagnostic assessment method commonly used to differentiate children with autism from those with other developmental delays using scoring criteria (from non-autistic to severely autistic). The CARS is a practical and brief measure that encompasses both the social–communicative and the behavioral flexibility aspects of autism's diagnostic triad (Chen et al., 2012). The CARS were filled out by the Japanese and Serbian therapist who interacted with the children. They both have 20+ years of working with children with autism and see regularly the children who participated in this study. Scores 30–36.5 are considered mild-to-moderate autism and scores of 37–60 as moderate-to-severe autism (Chen et al., 2012). As can be observed from **Table 1**, Japanese participants were slightly younger than Serbian participants. In both samples, the boys outnumbered the girls, reflective of the gender ratio in autism. From average CARS scores, we note that, within the selected groups, the participants from Serbia exhibited more obvious autistic traits. Note also that some of the children were below the autism threshold of 30, despite their autism diagnosis. Again, CARS is an indicative measure of the behavioral severity, and we report it to show the group differences obtained using the same scoring test. Note that we did not include typically developing children as controls, as there are no claims being made here about autism vs. typical development.

TABLE 1 | The summary of participants.

	Serbia	Japan
Age	9.41 ± 2.46	7.59 ± 2.43
Age range	5–13	3–12
Gender (male:female)	15:4	15:2
CARS	40.3 ± 8.2*	31.8 ± 7.1

*The average CARS scores of the two groups are statistically different [t(34) = −3.35, *p = 0.002].*

6.3. Coding

The interactions were recorded using a high-resolution webcam with a microphone (see **Figure 1**). To measure the children's engagement during the interaction, each video was coded in terms of engagement levels defined based on the occurrence and relative timing of the children's behavior (including learning-related behaviors), and the therapist's requests and prompts. To this end, we adapted the coding approach proposed in Kim et al. (2012) and defined engagement on a 0–5 Likert scale, with 0 corresponding to the events when the child is fully non-compliant (evasive) and 5 when fully engaged (see **Table 2** for description). Note that in Kim et al. (2012), 10 sec long fragments of target videos were coded in terms of the engagement levels. By contrast, we find it more objective to code the whole engagement episode: starting with the target task, e.g., the therapist asking the child to select the card corresponding to the NAO's expression (the Recognition phase), until one of the conditions (**Table 2**) for scoring the engagement level has been met. We propose this task-driven coding of engagement[4] as it preserves the context

TABLE 2 | Engagement coding [adaptation of the engagement definition proposed in Kim et al. (2012)].

Level	Meaning	Example(s): wording given to the coders
5	High engagement	Child immediately responds to the question of therapist, following the interaction scenario and reacting with NAO spontaneously
4	Mid engagement	Child is to the first prompt/question to perform the task but needs a bit of boost from therapist (e.g., pointing with finger, calling by name, showing something to pay attention to, and so on)
3	Low engagement	Child complies with the instructions after 2–3 repetitions
2	Indifferent	Therapist repeats the question and/or attempts the task more than 3 times, until child complies with the instructions
1	Non-compliance	Child is not responding to questions and/or tasks by therapist (e.g., the child hung head and refused to participate in the interaction, was looking somewhere else, not paying attention to the interaction)
0	Evasive	Child is not responding to therapist and/or NAO's prompts at all and after the prompts, or immediately, walks away from NAO

[3] The approvals have been obtained from the IRBs of MIT, USA, Chubu University, Japan, and Institute of Mental Health, Serbia.

[4] By task, we refer to tasks in general, i.e., when the child is imitating the robot, as well as while paying attention to the therapist while matching the images with the robot's behaviour.

in which the engagement is measured. By coding short fixed-interval segments, as in Kim et al. (2012), the beginning and the end of target activity can easily be lost (since its duration varies across the tasks/children).

After the engagement episodes were coded in the videos, each episode was further coded in terms of the affective dimensions (valence and arousal) and the children's face-expressivity.[5] As explained in Sec. 2, valence and arousal are well-studied affective dimensions and they have been related to a number of emotional states and moods; however, this has not been investigated much in autism. To analyze the relationship of the perceived valence and arousal, and target engagement levels, the engagement episodes were coded for valence and arousal on a 5-point ordinal scale [−2,2]. For example, the episodes were coded with high negative valence (−2) in cases when the child showed clear signs of experiencing unpleasant feelings (being unhappy, angry, visibly upset, showing dissatisfaction, frightened), dissatisfaction and disappointment (e.g., when NAO showed an expression that the child did not anticipate). The very positive valence (+2) was coded when the child showed signs of intense happiness (e.g., clapping hands), joy (in most cases followed with episodes of laughter), and delight (e.g., when NAO performed). The observable cues of arousal are directly related to the child's level of excitement. The episodes in which the child seemed very bored or uninterested[6] (e.g., looking away, not showing interest in the interaction, sleepy, passively observing) were coded as a very low arousal (−2). Note that this outward expression of a very low arousal could also be a consequence of intense internal arousal that led to a shut-down state (Picard and Goodwin, 2008). However, in this work, we focus on outward expressions of target affective states. The levels in between (−1,+1) for both dimensions just varied in their intensity (thus, being of lower intensity than the aforementioned). The neutral state of valence/arousal (0,0) corresponded to cases where the child seemed alert and/or attentive, with no obvious signs of any emotion, and/or physical activity (head, hand, and/or bodily movements). Note that the coders were instructed to base their judgments solely on the children's outwards signs described above, and not their "intuition" about the children's internal states, in order to focus on most objectively visible data. It is important to mention again the key difference between these two dimensions (facets of affective engagement) and the directly measured engagement levels: while the former are purely based on the behavioral cues, as reflection of the children's level of joy (valence) and excitement (arousal), the latter is task driven (i.e., its score is based on a number of prompts and pre-defined activities, as defined in **Table 2**). Finally, the engagement episodes were also coded in terms of facial expressiveness of the children within the episodes. This was coded on a 0–5 Likert scale, from neutral (0) to very expressive (5) (regardless of the type of facial expression, such as positive or negative). Each episode was assigned a score based on the observed level of activation of facial muscles throughout the episode, thus, taking the total duration of the expressive video segments into account (and the parts where the face is mostly visible).

All video episodes of engagement were coded by two experienced occupational therapists (from Japan and Serbia, who did not participate in the recordings), with the percent agreement of 92.4%. This is expected as the coding scheme (**Table 2**) clearly defines the beginning/ending of the episodes. The disagreeing parts were caused by the language differences. For instance, in some cases, the coders needed the meaning of the vocalizations to make sure, e.g., that the therapist asked the child a question and not just made a statement. After the coding has been performed by each coder separately, the beginning/ending of each episode was adjusted by the coders together. The coding of the affective dimensions as well as face expressivity was done by the same coders (separately). Note that lower levels of agreements were obtained: valence (75.8%), arousal (67.4%), and face expressivity (69.8%). However, this is still widely accepted as a good level of agreement (Carmines and Zeller, 1979).

7. EXPERIMENTAL RESULTS AND ANALYSIS

For studying specifics of the participants' interactions with NAO, throughout this section we provide qualitative as well as quantitative (statistical) analysis of relationships between the engagement levels, the affect dimensions, and corresponding contextual variables (tasks and culture). To measure association between these variables, we report Pearson's correlation (r), as it is a commonly used dependence measure in HRI applications. Analysis of the group differences (within and between the two cultures) was performed using Welch's t-test (Welch, 1947) due to its robustness to the unequal variances. If not said otherwise, only outcomes with significance levels $p \leq 0.05$ were considered for interpretations.

We start by comparing the average engagement levels within each phase. In **Table 3**, we report the mean (M) and standard deviation (SD) for these values. Note that the average engagement in the Japanese did not vary as much as in the Serbs, in whom the highest (average) engagement levels occur in phases Pairing and Story. In the first phase, despite their behavioral severity, Serbs were able to perform the tasks fast because these were simple (see Sec. 6 (*Protocol*)). By contrast, children who reached the Story

TABLE 3 | Average engagement levels (with one SD) within each phase, and the phase duration, computed per culture.

	Engagement		Duration	
Phases	**Serbia**	**Japan**	**Serbia**	**Japan**
1—Pairing	3.85 ± 1.23*	4.68 ± 0.85	10.7 ± 7.43*	22.4 ± 13.0
2—Recognition	3.36 ± 1.54*	4.47 ± 1.09	31.7 ± 29.8*	23.9 ± 18.8
3—Imitation	3.23 ± 1.68*	4.07 ± 1.51	37.6 ± 33.6	26.6 ± 29.3
4—Story	4.54 ± 0.68	4.37 ± 1.37	53.7 ± 24.7	37.6 ± 21.0

*Note the change in the average duration (in sec) of each phase as the task difficulty increases: the average values increase much faster in Serbs compared to Japanese. This, again, may be related to Japanese being engaged for longer at the initial phases, while being able to finish faster the more difficult tasks in phases 3–4, as well as their CARS scores. Statistically significant differences are marked with *.*

[5] There was a time gap of two months between the two codings.

[6] Note that this relates to the child being uninterested in communication/interaction in general and not in performing the target task.

phase did not have much difficulty engaging, which explains the high average engagement (M = 4.54, SD = 0.68). By looking into the cross-cultural differences in the engagement levels per phase, we found that these were significantly different for all phases but the Story. Again, we suspect that this is because the majority of children, who reached the last phase, showed high levels of engagement. By looking into the mean duration (in seconds) of the engagement episodes per phase, we note that in phase Pairing, Serbs engaged much faster than their Japanese peers. Yet, we see in Serbs a much steeper increase in the time toward higher phases, with the duration of phases pairing/recognition being significantly different between the cultures. **Table 4** reveals that the higher levels of engagement, on average, last much shorter than lower levels (<3). The high duration of phase story in Serbs was also biased by a frequent presence of lower (longer) engagement levels (episodes) (see **Figure 2**). Note also the significant differences in duration of the engagement episodes across levels 3–5 of the two cultures. A reason for level 5 being longer in Japanese is possibly because they had lower CARS, and, therefore, took longer to engage in the target activity.

We next analyze the relationships between the engagement and perceived affect (valence and arousal, as well as the face expressivity) within target engagement episodes. **Figure 3** shows the (normalized) distributions of the four dimensions. We first observe that the affect distributions are very close in shape when compared across the cultures, with the valence being uniformly distributed at the intermediate levels (−1,0,1). This indicates that highly positive/negative expressions were not perceived during the interaction. Distributions of arousal, on the other hand, are more Gauss-like, signaling the majority of episodes with low arousal, with a few having very low/high (perceived) arousal. The distribution of the face expressivity levels is highly skewed to the right—thus, very low levels (or no facial activity) were observed. However, this in line with (DSM-5, 2013) emphasizing the presence of "deficits in nonverbal communicative behaviors used for social interaction, ranging, for example, from poorly integrated verbal and nonverbal communication […] to a total lack of facial expressions and nonverbal communication" in children with autism. While these distributions are similar across the cultures, notice the differences in distributions of the engagement levels.

To get better insights into the relationships between the engagement levels and the affect dimensions, in **Figure 4,** we depict the (normalized) co-occurrence matrices. First, there is a strong positive correlation between the affective dimensions (valence and arousal) observed in both cultures (r = 0.73 and r = 0.56, respectively) and is more pronounced in Serbs. Lo et al. (2016) showed that (neurotypical) participants were more capable of distinguishing valence than arousal changes in emotion expressions, thus, capturing these when occurring together may be easier. Also, Brewer et al. (2016) showed that neurotypical persons have, in general, a difficulty in recognizing emotional expressions of persons with autism. This could, in part, explain why we obtained high correlations between the two affective dimensions: when both increase/decrease, it might be more obvious to the coders to perceive the change in the level. Also, this can be attributed to the way the persons with autism express their valence and arousal, which looks more obvious if both are very high or low, e.g., the child is expressing happiness with smiles and laughter, and fast movements of arms (flapping). Again, note that the coders were instructed to judge these two dimensions based on behavioral signs commonly observed in a neurotypical population (see Sec. 6). The dependence between valence and engagement is more spread in Serbs than Japanese, which is in part due to the highly imbalanced levels of engagement in the latter. However, we observe that in Japan, the positive valence occurs frequently at higher levels (3–4) and the negative (−1) is more present at lower engagement levels. By contrast, in Serbs, this is not that pronounced, since we can see that the valence levels are more smoothly distributed, with the negative valence occurring even at higher levels of engagement (e.g., the child sitting calm, might look bored or sad, looking around in the room but still participating in the task). We draw similar conclusions from the arousal–engagement relationships. We also note that in both cultures, high engagement never occurs with very low arousal, but mainly with the neutral and/or low positive arousal, as in cases when the child is sitting calm, showing no significant movements, and is being focused on the tasks. This may also be due to the coding bias. Finally, we see from the facial activity, that regardless of the engagement levels, the average face-expressivity was very low (mainly 0) in both cultures, showing very low (and insignificant) correlation (r < 0.20) with the engagement levels. In addition to the lack or atypical facial expressiveness in autistic population, as mentioned above, this could also be, in part, the consequence of scoring the whole engagement episode level rather than the image frames. This, in turn, may result in the coders ignoring subtle changes in the children's facial expressions, the presence of which is obvious due to the perceived variation in the valence levels.

Table 5 shows average levels of target affective dimensions w.r.t. the engagement levels of the two cultures. We observe that within the higher levels of engagement (specifically, level 5), the average valence level is much higher than in the levels below and significantly different between the cultures. This indicates that, on average, Serbs showed more pronounced expressions of positive states (e.g., joy and interest), as can also be noted from the face expressivity levels. However, while their average arousal levels were similar at the peak of engagement, Japanese showed significantly lower arousal levels across all engagement levels, with much lower arousal when being evasive (level 0). This can

TABLE 4 | Average duration of engagement episodes per level.

Engagement	0	1	2	3*	4*	5*
Serbia	34.3 ± 30.8	75.4 ± 120	49.2 ± 38.9	22.9 ± 19.8	17.5 ± 27.1	11.2 ± 9.70
Japan	31.2 ± 15.1	59.6 ± 45.2	65.9 ± 34.8	40.2 ± 14.1	33.6 ± 18.1	20.3 ± 13.1

*Note that there is a significant difference in duration of higher engagement levels (3–5) between the two cultures. Statistically significant differences are marked with *.*

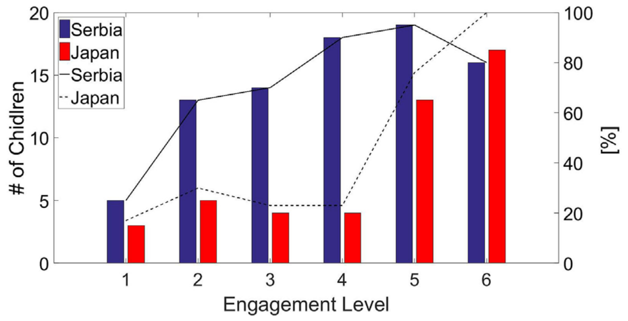


FIGURE 2 | The number (left)/percentage (right) of the participants from the two cultures that showed in at least one of their engagement episodes, the target level. Note that less than 30% of Japanese have levels 0–3, while the remaining 70% have at least one instance of the higher levels of engagement. By contrast, more than 70% of Serbs showed levels of very low engagement. We also observe that, in Japanese, the engagement distribution is largely skewed toward higher levels, in contrast to Serbs, who have more uniformly distributed levels.

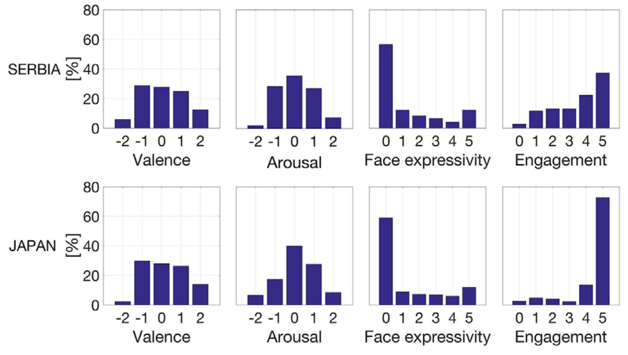


FIGURE 3 | The (normalized) distribution of the levels for each coded dimension: valence, arousal, face expressivity, and engagement. Note that in both cultures, the affective dimensions exhibit similar distributions, while engagement distributions are highly skewed to the highest level in Japanese. The face expressivity in both cultures is skewed toward the neutral, confirming the expected low face expressivity in children with autism (DSM-5, 2013).

be ascribed to the differences in behavioral dynamics in these two cultures as well as the individual expressions of resistance to specific interactions with NAO. For instance, some children preferred to imitate expressions of certain emotions only, and not all. This (inner) resistance to comply with the task can also be related to their (in)ability to recognize all of the four emotions. However, a detailed analysis of the target stimulus is out of the scope of this work.

So far, we focused mainly on the group (the between culture) differences of the children under the study. It is, however, important to assess some of these differences within the cultures and also by looking at the individual variation. **Figure 5** depicts the changes in the engagement levels, along with the affect dimensions, and the corresponding CARS for each child. Note the heterogeneity in the occurrences of the target levels per child. For example, we observe in Japanese (ID: 1, 16, and 12) and Serbs (ID: 7, 12, 15, and 18) that although valence and arousal are both (highly) negative, their average engagement was relatively high. This possibly is a consequence of idiosyncratic behavioral responses of the children: the same children had higher CARS, which means less functionality. This may be the reason for their expressions of valence and arousal being harder to perceive accurately (Brewer et al., 2016), although their engagement was high. We also observe from the children with ID:14 (Serb) and ID:3&15 (Japanese) that their high facial expressivity is typically followed with high engagement, valence, and arousal levels. This could indicate that these children are showing more obvious signs of positive emotions. We also report in **Figure 5** (above each plot) the results of the *t*-test for the cultural differences w.r.t. the four target variables. Note that no significant differences were found (with $p < 0.05$), in valence, arousal, and face expressivity. However, we found a significant difference in the distribution of the average engagement levels in Serbs and Japanese.[7]

[7] To test whether these differences also exist within the cultures, we split in half the children within culture, and performed 1,000 random permutations. In both cultures, there were no significant differences (with $p < 0.05$) between the children within two cultures.

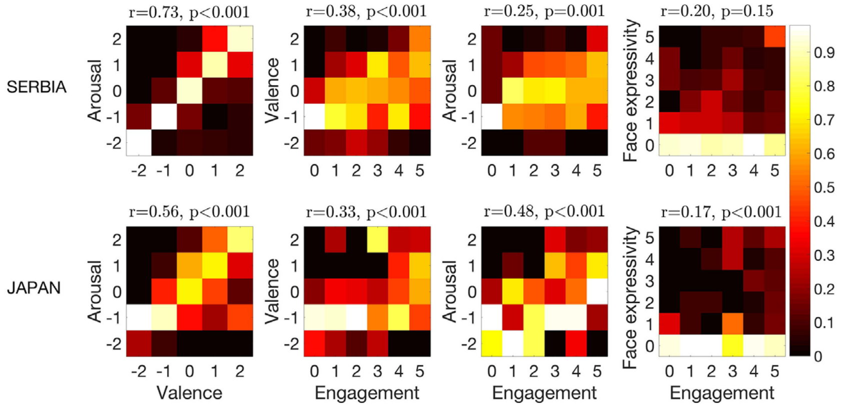


FIGURE 4 | Dependence structures of the coded dimensions, computed as co-occurrence matrices of the corresponding levels. These are normalized across columns to remove the bias toward the highest engagement level (as can be seen from **Figure 3**). This, in turn, allows us to analyze the distribution of each level of affect dimensions within the target engagement levels. The Pearson correlation (*r*) with significance levels is shown above each plot.

TABLE 5 | The average levels of valence, arousal, and face expressivity per engagement level.

Engagement	0	1	2	3	4	5
Valence						
Serbia	−0.90 ± 0.57	−0.39 ± 0.98	−0.48 ± 0.95	−0.02 ± 1.04	−0.01 ± 1.05	0.62 ± 1.11
Japan	−1.12 ± 0.64	−0.53 ± 1.18	−0.84 ± 0.55**	0.28 ± 1.70	0.06 ± 1.09	0.37 ± 1.00*
Arousal						
Serbia	−0.40 ± 1.07	−0.20 ± 0.67	−0.12 ± 0.93	0.06 ± 0.97	0.02 ± 0.83	0.43 ± 1.01
Japan	−1.25 ± 0.70*	−1.00 ± 1.07*	−1.15 ± 0.80*	0.14 ± 1.21	−0.35 ± 1.15	0.42 ± 0.81
Face expressivity						
Serbia	2.30 ± 0.48	2.28 ± 1.11	2.64 ± 1.00	2.60 ± 1.34	2.34 ± 1.21	2.49 ± 1.41
Japan	2.00 ± 0.76	1.27 ± 0.70*	1.61 ± 0.77*	1.71 ± 0.76*	1.97 ± 1.22	2.04 ± 1.08*

*The significant differences are marked with *(p = 0.05) and **(p = 0.08).*

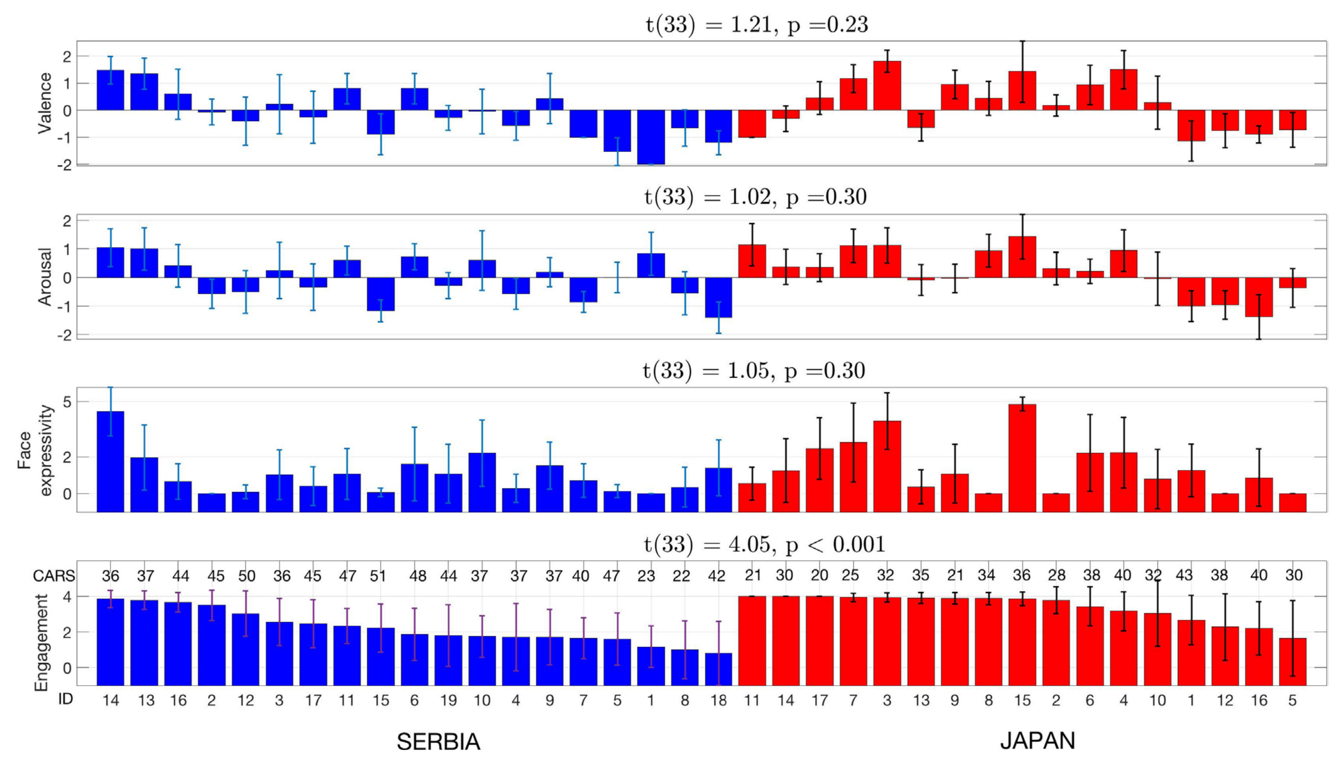


FIGURE 5 | Changes in the average frequency of the corresponding levels of target dimensions, per child. The children's ids were sorted according to the descending (average) engagement levels and the remaining dimensions aligned accordingly. The individual CARS are shown above the engagement bars.

We further investigate the between- and within-culture differences by looking into the Pearson (r) and Spearman scores (ρ) for dependencies between the average levels of valence, arousal, face expressivity, engagement, and CARS. We report both scores as the former assumes linear relationships and constant variance. As these can easily be violated due to the heterogeneity in the children, quantifying monotonic relationships (using the Spearman coefficient) may be a better indicator of target relationships (McDonald, 2009). For interpretations, we consider only the statistically significant scores (with $p < 0.05$). In **Figure 4**, the valence and arousal are highly (and linearly) correlated (as judged from similar r and ρ). There is also a high correlation between the perceived valence and face expressivity observed in both cultures ($r > 0.65$). This is expected, as both valence and arousal are directly related to the face modality (and facial expressions in particular), which provide the key cues when quantifying the (perceived) valence/arousal (Adolph and Georg, 2010). However, in contrast to the valence dimension, the face expressivity is not coded for the sign (positive/negative). Interestingly, despite the high (and significant) correlations between valence–face-expressivity ($r = 0.67$ in Serbs and $r = 0.74$ in Japanese) and valence-engagement ($r = 0.69$ in Serbs and $r = 0.49$ in Japanese), the correlations engagement–face-expressivity are relatively low and insignificant. This shows that the sign of valence (positive as when happy/negative as when sad) could be a good indicator of engagement. We also note that in Serbs, the engagement was highly correlated with the average valence levels per child.

In Japanese, we observe the opposite from **Table 6** (arousal is far more correlated with the engagement than valence). This may add evidence to the previous studies "showing variation as demonstrating cultural differences in 'display rules,' or rules about what emotions are appropriate o show in a given situation" (Ekman, 1972).

The findings described above can, in part, be attributed to the CARS of Japanese being lower, and, thus, them being more responsive (through timely body movements in response to target tasks). On the other hand, if we recall the engagement levels from **Figure 3**, in Serbs, the levels varied more than in Japanese, indicating that the former group showed more hesitation in doing the tasks. More importantly, note that there is a highly significant correlation between engagement and CARS ($r = -0.52$ and $\rho = -0.71$) in Japanese, while these are found to be insignificant in Serbs. To test whether these two groups are statistically different in engagement levels due to the cultural difference, and not due to the differences in the behavioral severity of the participants, we remove the CARS as a big causal factor, and one that is highly correlated with engagement on the Japanese side. After tossing out highest/lowest CARS, we match the Serbs and Japanese on CARS within the range [33–43], thus, with the mid behavioral severity. This range assured the best possible match between the CARS of the two cultures, which can also be seen from **Figure 6**, where we ranked the engagement from high to low CARS, resulting in 8 Serbs and 8 Japanese, with a very similar functionality levels. We ran the *t*-test on this sample and again obtained the statistically significant difference in the engagement levels between the two cultures [$t(12) = -2.1$, $p = 0.05$]. This shows that these differences are not only due to the variation in behavioral severity (CARS) solely but also due to other factors, the most likely being the culture. We also found highly strong relationships between CARS and engagement ($\rho = -0.86$, $p < 0.01$ for Serbs, and $\rho = -0.82$, $p = 0.01$ for Japanese—see **Figure 6**), thus, the Spearman scores are more consistent when the similar subgroups are matched based on CARS.

8. DISCUSSION AND FUTURE WORK

Before providing a further discussion of the study described, it is important to emphasize that this analysis was of the exploratory nature within a specific context: an occupational therapy for children with autism, using a social robot NAO, and recorded in two different cultures. Specifically, the participants in this study are 36 children from Japan and Serbia, who participated only once for a short duration. We note again that the aim of this study was not to propose a new therapy for autism, but induce a context in which the children's engagement can be measured in a structured way. Our analysis of the relationships between the behavioral engagement and affective components of engagement (the perceived valence and arousal, as well as face expressivity) showed significant differences in a number of parameters considered. However, we restrain from making any conclusive statements about the causal cultural differences in these two groups. This is for the following reasons. First, although the parametric tests used in our analysis did

TABLE 6 | The Pearson correlation (*r*) and Spearman rank correlation coefficient (ρ), with their significance levels (*p*), for the children's average levels of valence, arousal, face expressivity, and engagement, as well as their CARS.

	Serbia					Japan				
	Valence	**Arousal**	**Face express**	**Engagement**	**CARS**	**Valence**	**Arousal**	**Face express**	**Engagement**	**CARS**
Valence	–	$r = 0.57, p < 0.01$	$r = 0.67, p < 0.01$	$r = 0.69, p < 0.01$	$r = 0.16, p = 0.50$	–	$r = 0.69, p < 0.01$	$r = 0.74, p < 0.01$	$r = 0.49, p < 0.01$	$r = -0.16, p = 0.63$
	–	$\rho = 0.64, p < 0.01$	$\rho = 0.65, p < 0.01$	$\rho = 0.75, p < 0.01$	$\rho = -0.02, p = 0.94$	–	$\rho = 0.63, p < 0.01$	$\rho = 64, p < 0.01$	$\rho = 0.31, p < 0.01$	$\rho = -0.17, p = 0.50$
Arousal	$r = 0.57, p < 0.01$	–	$r = 0.53, p = 0.01$	$r = 0.38, p = 0.10$	$r = -0.26, p = 0.27$	$r = 0.69, p < 0.01$	–	$r = 0.55, p = 0.02$	$r = 0.74, p = 0.01$	$r = -0.40, p = 0.11$
	$\rho = 0.64, p < 0.01$	–	$\rho = 0.50, p = 0.03$	$\rho = 0.40, p = 0.08$	$\rho = -0.28, p = 0.23$	$\rho = 0.63, p < 0.01$	–	$\rho = 0.50, p = 0.04$	$\rho = 0.68, p = 0.01$	$\rho = -0.40, p = 0.10$
Face	$r = 0.67, p < 0.01$	$r = 0.53, p = 0.01$	–	$r = 0.31, p = 0.18$	$r = -0.12, p = 0.60$	$r = 0.73, p < 0.01$	$r = 0.55, p = 0.02$	–	$r = 0.38, p = 0.13$	$r = 0.02, p = 0.93$
Expressivity	$\rho = 0.65, p < 0.01$	$\rho = 0.49, p = 0.03$	–	$\rho = 0.17, p = 0.47$	$\rho = -0.25, p = 0.30$	$\rho = 0.64, p < 0.01$	$\rho = 0.50, p = 0.04$	–	$\rho = 0.36, p = 0.15$	$\rho = 0.02, p = 0.93$
Engagement	**$r = 0.69, p < 0.01$**	$r = 0.38, p = 0.10$	$r = 0.31, p = 0.18$	–	$r = 0.30, p = 0.19$	**$r = 0.49, p < 0.01$**	**$r = 0.74, p = 0.01$**	$r = 0.38, p = 0.13$	–	**$r = -0.52, p = 0.03$**
	$\rho = 0.75, p < 0.01$	$\rho = 0.40, p = 0.08$	$\rho = 0.17, p = 0.47$	–	$\rho = 0.22, p = 0.36$	**$\rho = 0.31, p < 0.01$**	**$\rho = 0.68, p = 0.01$**	$\rho = 0.36, p = 0.15$	–	**$\rho = -0.71, p = 0.01$**
CARS	$r = 0.16, p = 0.50$	$r = -0.26, p = 0.27$	$r = -0.13, p = 0.60$	$r = 0.30, p = 0.19$	–	$r = -0.13, p = 0.63$	$r = -0.40, p = 0.11$	$r = 0.02, p = 0.93$	$r = -0.52, p = 0.03$	–
	$\rho = -0.02, p = 0.94$	$\rho = -0.29, p = 0.23$	$\rho = -0.25, p = 0.30$	$\rho = 0.22, p = 0.36$	–	$\rho = -0.17, p = 0.50$	$\rho = -0.44, p = 0.10$	$\rho = 0.02, p = 0.93$	$\rho = -0.71, p = 0.01$	–

Compared to r, higher ρ indicates the prevalence of monotonic relationships (not linear), which is expected due to the heterogeneity in the children's expressions of engagement. The scores in bold denote the significant relationships ($p < 0.05$) between the engagement and the compared variables.

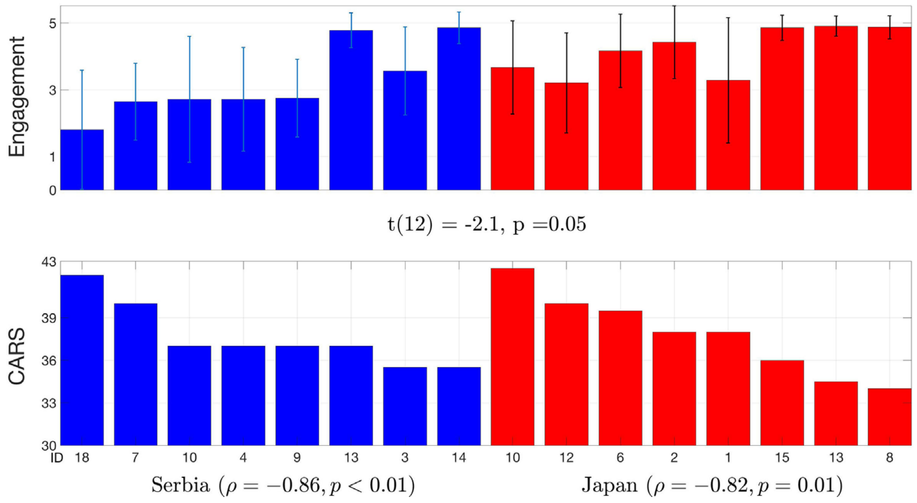


FIGURE 6 | Engagement levels of the subset of children that are matched based on their CARS, ranking the average engagement levels from high CARS to low CARS. Note that the differences in the engagement levels between the two cultural groups are still statistically different with 5% significance level. Note also the strong relationships between the CARS and average engagement levels.

find statistically significant relationships, a great variation was also found within the cultures and within the individuals (in terms of age, gender, and CARS). Therefore, larger-scale studies are needed to further confirm and extend our findings on the variations across cultures. Second, one of the goals of this study was to get better insights into the expression of engagement in naturalistic settings, where children with autism interact regularly with a therapist. This is in order to identify the key behavioral cues of engagement and its relationship to other parameters (in this case, task difficulty, affect, and CARS) that should be considered when designing social robots for autism therapy.

As noted by Keen (2009), "the ability to detect differences in engagement levels as a function of different program types, differences within programs, instructional methods, and other forms of 'intervention' is a first step toward establishing the validity of the measure for program evaluation purposes." Also, children with autism (and/or other neurodevelopmental differences) spend less time actively engaged with adults and their peers, and less time in mastery-level engagement with materials than do the typically developing children (McWilliam and Bailey, 1995). This is why it is very important to develop techniques where the children can master social skills through therapeutic settings, with the aim of being able to translate those in the play with their (typically developing) peers (Keen, 2009). Toward this end, the role of social and affective robots is twofold: (i) to provide more efficient and reliable (standardized) means of measuring children's engagement and (ii) to enable naturalistic interaction with the children by being able to automatically estimate their level of engagement and respond to it accordingly (e.g., timely giving a positive feedback and encouragement). Before this can be implemented, it is necessary to understand better the context: the children's behavioral and other parameters that relate to their engagement. In what follows, we briefly discuss our findings from this perspective and provide suggestions for future work in this direction.

The main findings relate to how engagement levels differed as a function of the cultures, tasks, affective dimensions, and CARS-based behavioral severity. Our results indicate that there are statistically significant differences in duration and average levels of engagement between the two cultures, when compared in terms of the task difficulty. Japanese were able to engage and complete easier tasks faster, while Serbs (those who reached the last phase in the interaction) were engaged for a shorter time in the interaction. This can be a consequence of the latter group being affected more severely by the condition, as also reflected in their CARS and the distribution of the engagement levels. By matching the two groups based on CARS, we were still able to find significant differences between the engagement levels in the two cultures. Another important aspect to be considered (especially when comparing the task execution time) and that is not explored in detail in our analysis is the children's age. While most of the typically developing children develop both motor and mental abilities by the age of four (Baron-Cohen et al., 1985), the lack of the same is indicated by the CARS scores in autism. By looking into the age of the children who performed the tasks faster than the others, we did not notice any significant dependencies on age.

These types of analyses are important because they provide a prior knowledge that can be used to adjust the dynamics of the robot interaction within each culture/age-range, with a possibility for further individual adjustment (Picard and Goodwin, 2008). For example, these could be used as priors for computer models, as part of the robot's perception. This has recently been attempted in terms of the personality adjustment, with the focus on typically developing adults (Salam et al., 2017). Adding to this, one of the important findings of our study is the relatively high correlation between the CARS and engagement levels (e.g., in Japanese, Spearman coefficient $\rho = -71\%$, $p = 0.01$). When the CARS range was matched across the cultures (**Figure 6**), these reached $\rho = -82-86\%$, $p = 0.01$, in Japan and Serbia, respectively. This could potentially be a good calibrating parameter for social robots, derived from an easy- and fast-to-obtain measure of the current behavioral condition. However, it should be investigated whether the other standardized tests (e.g., specific questions of ADI-R and/or ADOS (Le Couteur et al., 2008)) would provide more universal indicators for (expected) engagement levels, resulting in a more stable input for choosing the therapy mode, establishing therapy expectations and follow ups on the individual progress. Moreover, this could potentially allow the medical doctors and therapists to more easily assess the therapeutic value of the whole procedure, through the common protocols, which could be adjusted to each culture and with assistance of the social robots.

Our analysis of the relationships between the directly measured behavioral engagement and indirectly measured affective engagement (from the valence and arousal levels perceived from outward behavioral cues, including the face expressivity) revealed statistically significant ($p < 0.01$) dependencies between the valence and arousal levels (average per child) and the behavioral engagement, with Spearman coefficient $\rho = 68-75\%$. Therefore, these two affective dimensions can potentially be used as a proxy of the children's behavioral engagement. The benefit of this is that these two affective dimensions can automatically be measured in human-robot interaction by means of existing models for affective computing (Picard, 1997; Castellano et al., 2014a). Specifically, a number of works on automated estimation of valence and arousal have been proposed (Zeng et al., 2009; Gunes et al., 2011). For instance, Nicolaou et al. (2011) showed that automatically estimated levels of valence/arousal can achieve the agreement similar to that of human coders when multi-modal behavioral cues are used as input (facial expressions, shoulder gestures, and audio cues). As we showed in our experiments (**Table 6**), while the face expressivity was highly correlated with valence/arousal levels, it did not relate strongly to the behavioral engagement. Investigating other facial cues such as the eye-gaze (typically used in autism studies (Chen et al., 2012; Jones et al., 2016)) in the context of engagement is a promising way to go. Moreover, it is critical to take into account the multi-modal nature of human behavior when estimating engagement. In a recent work, Salam et al. (2017) measured individual and group engagement by an automated multi-modal system that exploits outward behavioral cues (face and body gestures) as well as contextual variables (the personality traits of a user). They were able to improve significantly the engagement estimation when the individual features (face and body gestures) were included. While in this work, we focused on outward measures of the target affective dimensions, note that significant advances have been made in measuring the same from inward expressions (biosignals such as the heart rate, electrodermal activity (EDA), body temperature, and so on) (Picard, 2009). For instance, Hernandez et al. (2013) showed that there are significant relationships between autonomic changes in arousal levels (measured using a wristband sensor) and behavioral challenges in children with autism in a school setting. In another work, Hernandez et al. (2014) showed that automatically estimating the ease of engagement on a scale (0–2) of typical children, participating in interactions with the educator and the parent, can be achieved with an accuracy of up to 68% (from EDA solely). Likewise, but in the context of children with autism and human-computer interaction, Liu et al. (2008) achieved automatic detection of liking, anxiety, and engagement, from physiological signals (EDA, heart rate and temperature) with an accuracy of 82.9%. Further research on the use of social robots in autism should also closely examine these modalities for automated analysis of engagement.

However, none of the methods that could potentially be used for automated estimation of engagement and/or valence and arousal, were evaluated before in the context of autism and child-robot interaction. Therefore, there are at least a few important questions for the future work toward building a system for automated estimation of engagement of children with autism, and in the context of therapies involving social robots. First, how can multiple modalities of children's behavior (including inward and outward expressions) be modeled efficiently using models for affective computing (Picard, 2009; Castellano et al., 2014a) to take the full advantage of their complimentary nature? This, in turn, would not only provide better insights into behavioral cues of engagement but also enable a more accurate and reliable perception of engagement by social robots. How to account for the contextual aspect of engagement is another important challenge in automating engagement estimation. As our results suggest, the culture (among other factors) may play an important role in modulating the time each child spends in a target activity, as well as the distributions and average levels of engagement. One approach would be to define affective computing models so that they embed this contextual information via priors on the model parameters, which can then be adjusted to each child as the therapy progresses. This brings us to our final and the most important challenge: how do we personalize the models to each child and obtain the child-specific estimation of target engagement levels? Although, in our study, we did not find statistically significant differences in engagement levels *within* either of the two cultures, we did, however, observe high levels of individual variation. Therefore, personalizing the models to each child with autism is, perhaps, the most challenging aspect of automated engagement estimation that the future work will face (Picard and Goodwin, 2008). Another important factor not addressed in this study is the influence of culture on the annotation process. While annotators in our study achieved a high level of agreement, as shown in several other studies (e.g., see Engelmann and Pogosyan (2013)), handling the annotators bias effectively is of paramount importance when designing social robots that can automatically estimate engagement. Likewise, CARS is the standard, and, thus, its scoring should not be affected by cultural background of the therapists (but it still may be affected). While the presence of cultural biases in autism screening is inevitable, Mandell and Novak (2005) showed that it is mostly due to the different perception of autistic traits by parents with different cultural backgrounds. Since the therapists did the CARS in our study, we do not expect these differences to be as affected by the culture.[8]

9. CONCLUSION

Taken together, the findings of this study and the questions raised clearly indicate that more research is needed in the field of social and affective robotics for autism. While our study focused on a single day recordings of the children, future research should focus on longitudinal studies of engagement, if more reliable conclusions are to be drawn and data for automating the engagement estimation collected. There is an overall lack of such studies and data (especially across cultures); yet, they are of critical importance for building more effective technology that could facilitate, augment, and scale, rather than replace, the efforts by medical doctors and therapists working directly with individuals with autism. We hope that this work will increase awareness for the need of such studies and also provide useful insights into computer scientists in the field of affective computing and social robotics, and also neuroscientists, psychologists, therapists, and educators working in the autism field. Finally, we must keep our sights on the main goal: building technology and insights that ultimately bring benefit to users on the autism spectrum, especially those who seek to sustain engagement more successfully in learning experiences.

[8] This is out of the scope of this study as a more detailed analysis of the codings/scorings, involving multiple coders/therapists from each culture, would need to be conducted in future.

ETHICS STATEMENT

This research has been approved by institutional review board (IRB) of Chubu University, Japan; Mental Health Institute, Serbia; and MIT COUHES, USA.

AUTHOR CONTRIBUTIONS

OR and JL are the lead authors of this work: they conducted the study and organized the data collection. LM-M contributed via constant discussions and literature review. BS provided useful insights about the study and feedback during the preparation of the manuscript. RP helped to define the structure of the manuscript, interpret the results, and improve the content of the manuscript.

ACKNOWLEDGMENTS

This work was supported by MEXT Grant-in-Aid for Young Scientists B grant no. 16763279 and Chubu University Grant I grant no. 27IS04I (Japan). The work of OR was funded by EU HORIZON 2020 under grant agreement no. 701236 (ENGAGEME)—Marie Skłodowska-Curie Individual Fellowship, and the work of BS under grant agreement no. 688835 (DE-ENIGMA). We thank the Serbian Autism Society (Ms. V. Petrovic and Ms. S. Petrovic), the occupational therapists: Ms. S. Babovic, Ms. N. Selic, Ms. Y. Takeuchi, Ms. K. Hara and Prof. M. Kojima, and Prof. M. Pejovic, MD (Mental Health Institute – MHI, Belgrade), for their help and discussions over the course of this study. We also thank Prof. G. Obinata, Prof. M. Pantic, and Dr. J. Shen for their support and feedback in the early stage of this study. Our special thanks go to all the children and their parents who participated in this study—without them, this research would not be possible.

REFERENCES

Adolph, D., and Georg, W. A. (2010). Valence and arousal: a comparison of two sets of emotional facial expressions. *Am. J. Psychol.* 123, 209–219. doi:10.5406/amerjpsyc.123.2.0209

An, S., Ji, L.-J., Marks, M., and Zhang, Z. (2017). Two sides of emotion: exploring positivity and negativity in six basic emotions across cultures. *Front. Psychol.* 8:610. doi:10.3389/fpsyg.2017.00610

Anzalone, S. M., Boucenna, S., Ivaldi, S., and Chetouani, M. (2015). Evaluating the engagement with social robots. *Int. J. Soc. Robot.* 7, 465–478. doi:10.1007/s12369-015-0298-7

Appleton, J. J., Christenson, S. L., Kim, D., and Reschly, A. L. (2006). Measuring cognitive and psychological engagement: validation of the student engagement instrument. *J. Sch. Psychol.* 44, 427–445. doi:10.1016/j.jsp.2006.04.002

Baird, G. (2001). Screening and surveillance for autism and pervasive developmental disorders. *Arch. Dis. Child.* 84, 468–475. doi:10.1136/adc.84.6.468

Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition* 21, 37–46. doi:10.1016/0010-0277(85)90022-8

Belpaeme, T., Baxter, P. E., Read, R., Wood, R., Cuayáhuitl, H., Kiefer, B., et al. (2013). Multimodal child-robot interaction: building social bonds. *J. Hum. Robot Interact.* 1, 33–53. doi:10.5898/JHRI.1.2.Belpaeme

Breazeal, C. (2001). "Affective interaction between humans and robots," in *Advances in Artificial Life*, Vol. 2159. (Berlin, Heidelberg: Springer), 582–591.

Breazeal, C. (2003). Emotion and sociable humanoid robots. *Int. J. Hum. Comput. Stud.* 59, 119–155. doi:10.1016/S1071-5819(03)00018-1

Breazeal, C. L. (2004). *Designing Sociable Robots*. Cambridge: MIT Press.

Brewer, R., Biotti, F., Catmur, C., Press, C., Happé, F., Cook, R., et al. (2016). Can neurotypical individuals read autistic facial expressions? atypical production of emotional facial expressions in autism spectrum disorders. *Autism Res.* 9, 262–271. doi:10.1002/aur.1508

Broughton, M., Stevens, C., and Schubert, E. (2008). "Continuous self-report of engagement to live solo marimba performance," in *Int'l Conference on Music Perception and Cognition*, Sapporo, 366–371.

Buckley, S., Hasen, G., and Ainley, M. (2004). *Affective Engagement: A Person-Centered Approach to Understanding the Structure of Subjective Learning Experiences*. Melbourne, Australia: Australian Association for Research in Education, 1–20.

Carmines, E. G., and Zeller, R. A. (1979). *Reliability and Validity Assessment*. London: SAGE, 17.

Cascio, M. A. (2015). Cross-cultural autism studies, neurodiversity, and conceptualizations of autism. *Cult. Med. Psychiatry* 39, 207–212. doi:10.1007/s11013-015-9450-y

Castellano, G., Gunes, H., Peters, C., and Schuller, B. (2014a). "Multimodal affect recognition for naturalistic human-computer and human-robot interactions," in *The Oxford Handbook of Affective Computing*, eds R. Calvo, S. D'Mello, J. Gratch, and A. Kappas (USA: Oxford University Press), 246.

Castellano, G., Leite, I., Pereira, A., Martinho, C., Paiva, A., and Mcowan, P. W. (2014b). Context-sensitive affect recognition for a robotic game companion. *ACM Trans. Interact. Intell. Syst.* 4, 1–25. doi:10.1145/2622615

Chen, G. M., Yoder, K. J., Ganzel, B. L., Goodwin, M. S., and Belmonte, M. K. (2012). Harnessing repetitive behaviours to engage attention and learning in a novel therapy for autism: an exploratory analysis. *Front. Psychol.* 3:12. doi:10.3389/fpsyg.2012.00012

Cohen, S. B. (1993). "From attention-goal psychology to belief-desire psychology: The development of a theory of mind and its dysfunction," in *Understanding Other Minds: Perspectives from Autism* (Oxford: Oxford University Press), 59–82.

Connell, J. P. (1990). Context, self, and action: a motivational analysis of self-system processes across the life span. *Self Trans. Infancy Child.* 8, 61–97.

Conti, D., Cattani, A., Di Nuovo, S., and Di Nuovo, A. (2015). "A cross-cultural study of acceptance and use of robotics by future psychology practitioners," in *IEEE Int'l Symposium on Robot and Human Interactive Communication (RO-MAN)*, Kobe, 555–560.

Daley, T. C. (2002). The need for cross-cultural research on the pervasive developmental disorders. *Transcult. Psychiatry* 39, 531–550. doi:10.1177/136346150203900409

Dautenhahn, K., and Werry, I. (2004). Towards interactive robots in autism therapy: background, motivation and challenges. *Pragmat. Cogn.* 12, 1–35. doi:10.1075/pc.12.1.03dau

De Silva, R. S., Tadano, K., Higashi, M., Saito, A., and Lambacher, S. G. (2009). "Therapeutic-assisted robot for children with autism," in *IEEE Int'l Conf. on Intelligent Robots and Systems (IROS)*, St. Louis, MO, 3561–3567.

Diehl, J. J., Schmitt, L. M., Villano, M., and Crowell, C. R. (2012). The clinical use of robots for individuals with autism spectrum disorders: a critical review. *Res. Autism Spectr. Disord.* 6, 249–262. doi:10.1016/j.rasd.2011.05.006

DSM-5. (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. American Psychiatric Association.

Dyches, T. T., Wilder, L. K., Sudweeks, R. R., Obiakor, F. E., and Algozzine, B. (2004). Multicultural issues in autism. *J. Autism Dev. Disord.* 34, 211–222. doi:10.1023/B:JADD.0000022611.80478.73

Dziuk, M. A., Larson, J. C. G., Apostu, A., Mahone, E. M., Denckla, M. B., and Mostofsky, S. H. (2007). Dyspraxia in autism: association with motor, social, and communicative deficits. *Dev. Med. Child Neurol.* 49, 734–739. doi:10.1111/j.1469-8749.2007.00734.x

Ekman, P. (1972). "Universals and cultural differences in facial expressions of emotion," in *Nebraska Symposium on Motivation*. ed. J. Cole (Lincoln, NE: University of Nebraska Press) 207–282.

Ekman, P. (2005). *Facial Expressions*. John Wiley & Sons, Ltd.

Eldevik, S., Hastings, R. P., Jahr, E., and Hughes, J. C. (2012). Outcomes of behavioral intervention for children with autism in mainstream pre-school settings. *J. Autism Dev. Disord.* 42, 210–220. doi:10.1007/s10803-011-1234-9

Elfenbein, H. A., and Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychol. Bull.* 128, 203–235. doi:10.1037/0033-2909.128.2.203

Engelmann, J. B., and Pogosyan, M. (2013). Emotion perception across cultures: the role of cognitive mechanisms. *Front. Psychol.* 4:118. doi:10.3389/fpsyg.2013.00118

Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2003). A survey of socially interactive robots. *Rob. Auton. Syst.* 42, 143–166. doi:10.1016/S0921-8890(02)00372-X

Fournier, K. A., Hass, C. J., Naik, S. K., Lodha, N., and Cauraugh, J. H. (2010). Motor coordination in autism spectrum disorders: a synthesis and meta-analysis. *J. Autism Dev. Disord.* 40, 1227–1240. doi:10.1007/s10803-010-0981-3

François, D., Powell, S., and Dautenhahn, K. (2009). A long-term study of children with autism playing with a robotic pet: taking inspirations from non-directive play therapy to encourage children's proactivity and initiative-taking. *Interact. Studies* 10, 324–373. doi:10.1075/is.10.3.04fra

Fredricks, J. A., Blumenfeld, P. C., and Paris, A. H. (2004). School engagement: potential of the concept, state of the evidence. *Rev. Educ. Res.* 74, 59–109. doi:10.3102/00346543074001059

Freitag, C. M., Kleser, C., Schneider, M., and von Gontard, A. (2007). Quantitative assessment of neuromotor function in adolescents with high functioning autism and Asperger syndrome. *J. Autism Dev. Disord.* 37, 948–959. doi:10.1007/s10803-006-0235-6

Ge, B., Park, H. W., and Howard, A. M. (2016). "Identifying engagement from joint kinematics data for robot therapy prompt interventions for children with autism spectrum disorder," in *Int'l Conference on Social Robotics*, 531–540.

Gross, T. F. (2004). The perception of four basic emotions in human and non-human faces by children with autism and other developmental disabilities. *J. Abnorm. Child Psychol.* 32, 469–480. doi:10.1023/B:JACP.0000037777.17698.01

Gunes, H., Schuller, B., Pantic, M., and Cowie, R. (2011). "Emotion representation, analysis and synthesis in continuous space: A survey," in *IEEE Int'l Conference on Automatic Face & Gesture Recognition (FG'W)*, 827–834.

Habib, M. K. (2014). *Handbook of Research on Advancements in Robotics and Mechatronics*. IGI Global.

Happé, F., Ronald, A., and Plomin, R. (2006). Time to give up on a single explanation for autism. *Nat. Neurosci.* 9, 1218–1220. doi:10.1038/nn1770

Hedman, E., Miller, L., Schoen, S., Nielsen, D., Goodwin, M., and Picard, R. (2012). "Measuring autonomic arousal during therapy," in *Proceedings of Design and Emotion*, 11–14.

Hernandez, J., Riobo, I., Rozga, A., Abowd, G. D., and Picard, R. W. (2014). "Using electrodermal activity to recognize ease of engagement in children during social interactions," in *ACM Int'l Joint Conference on Pervasive and Ubiquitous Computing*, 307–317.

Hernandez, J., Sano, A., Zisook, M., Deprey, J., Goodwin, M., and Picard, R. W. (2013). "Analysis and visualization of longitudinal physiological data of children with ASD," in *The Extended Abstract of IMFAR*, 2–4.

Hilton, C. L., Zhang, Y., Whilte, M. R., Klohr, C. L., and Constantino, J. (2012). Motor impairment in sibling pairs concordant and discordant for autism spectrum disorders. *Autism* 16, 430–441. doi:10.1177/1362361311423018

Howlin, P., Baron-Cohen, S., and Hadwin, J. (1999). *Teaching Children with Autism to Mind-Read: A Practical Guide for Teachers and Parents*. New York: Wiley.

Immordino-Yang, M. H., Yang, X.-F., and Damasio, H. (2016). Cultural modes of expressing emotions influence how emotions are experienced. *Emotion* 16, 1033–1039. doi:10.1037/emo0000201

Isenhower, R. W., Marsh, K. L., Richardson, M. J., Helt, M., Schmidt, R., and Fein, D. (2012). Rhythmic bimanual coordination is impaired in young children with autism spectrum disorder. *Res. Autism Spectr. Disord.* 6, 25–31. doi:10.1016/j.rasd.2011.08.005

Jones, R. M., Southerland, A., Hamo, A., Carberry, C., Bridges, C., Nay, S., et al. (2016). Increased eye contact during conversation compared to play in children with autism. *J. Autism Dev. Disord.* 47, 607–614. doi:10.1007/s10803-016-2981-4

Kanda, T., Sato, R., Saiwaki, N., and Ishiguro, H. (2007). A two-month field trial in an elementary school for long-term human-robot interaction. *IEEE Trans. Robot.* 23, 962–971. doi:10.1109/TRO.2007.904904

Keen, D. (2009). Engagement of children with autism in learning. *Aust. J. Spec. Educ.* 33, 130–140. doi:10.1375/ajse.33.2.130

Kemp, C., Kishida, Y., Carter, M., and Sweller, N. (2013). The effect of activity type on the engagement and interaction of young children with disabilities in inclusive childcare settings. *Early Child. Res. Q.* 28, 134–143. doi:10.1016/j.ecresq.2012.03.003

Kim, E., Paul, R., Shic, F., and Scassellati, B. (2012). Bridging the research gap: making HRI useful to individuals with autism. *J. Hum. Robot Interact.* 1, 26–54. doi:10.5898/JHRI.1.1.Kim

Kishida, Y., and Kemp, C. (2009). The engagement and interaction of children with autism spectrum disorder in segregated and inclusive early childhood center-based settings. *Topics Early Child. Spec. Educ.* 29, 105–118. doi:10.1177/0271121408329172

Kitzhaber, S. (2012). Interventions for multicultural children with autism. *Master Soc. Work Clin. Res. Papers* 118, 1–67.

Kopp, S., Beckung, E., and Gillberg, C. (2010). Developmental coordination disorder and other motor control problems in girls with autism spectrum disorder and/or attention-deficit/hyperactivity disorder. *Res. Dev. Disabil.* 31, 350–361. doi:10.1016/j.ridd.2009.09.017

Kozima, H., Nakagawa, C., and Yasuda, Y. (2007). Children–robot interaction: a pilot study in autism therapy. *Prog. Brain Res.* 164, 385–400. doi:10.1016/S0079-6123(07)64021-7

Krause, R. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *J. Pers. Soc. Psychol.* 5, 4–712.

Le Couteur, A., Haden, G., Hammal, D., and McConachie, H. (2008). Diagnosing autism spectrum disorders in pre-school children using two standardised assessment instruments: the ADI-R and the ADOS. *J. Autism Dev. Disord.* 38, 362–372. doi:10.1007/s10803-007-0403-3

Lemaignan, S., Garcia, F., Jacq, A., and Dillenbourg, P. (2016). "From real-time attention assessment to with-me-ness in human-robot interaction," in *The Int'l Conference on Human Robot Interaction (HRI)* (IEEE Press), 157–164.

Libin, A., and Libin, E. (2004). Person-robot interactions from the robopsychologists' point of view: the robotic psychology and robotherapy approach. *Proc. IEEE* 92, 1789–1803. doi:10.1109/JPROC.2004.835366

Lim, N. (2016). Cultural differences in emotion: differences in emotional arousal level between the east and the west. *Integr. Med. Res.* 5, 105–109. doi:10.1016/j.imr.2016.03.004

Liu, C., Conn, K., Sarkar, N., and Stone, W. (2008). Physiology-based affect recognition for computer-assisted intervention of children with autism spectrum disorder. *Int. J. Hum. Comput. Stud.* 66, 662–677. doi:10.1016/j.ijhsc.2008.04.003

Lo, L., Hung, N., and Lin, M. (2016). Angry versus furious: a comparison between valence and arousal in dimensional models of emotions. *J. Psychol.* 150, 949–960. doi:10.1080/00223980.2016.1225658

Mandell, D. S., and Novak, M. (2005). The role of culture in families' treatment decisions for children with autism spectrum disorders. *Ment. Retard. Dev. Disabil. Res. Rev.* 11, 110–115. doi:10.1002/mrdd.20061

Marsh, K. L., Isenhower, R. W., Richardson, M. J., Helt, M., Verbalis, A. D., Schmidt, R. C., et al. (2013). Autism and social disconnection in interpersonal rocking. *Front. Integr. Neurosci.* 7:4. doi:10.3389/fnint.2013.00004

McDonald, J. H. (2009). *Handbook of Biological Statistics*. Baltimore, MD: Sparky House Publishing, 2.

McDuff, D., Girard, J. M., and El Kaliouby, R. (2017). Large-scale observational evidence of cross-cultural differences in facial behavior. *J. Nonverbal Behav.* 41, 1–19. doi:10.1007/s10919-016-0244-x

McWilliam, R., Bailey, D., Bailey, D., and Wolery, M. (1992). Promoting engagement and mastery. *Teach. Infants Preschool. Disabil.* 2, 230–255.

McWilliam, R., and Bailey, D. B. Jr. (1995). Effects of classroom social structure and disability on engagement. *Topics Early Child. Spec. Educ.* 15, 123–147. doi:10.1177/027112149501500201

Meece, J. L., Blumenfeld, P. C., and Hoyle, R. H. (1988). Students' goal orientations and cognitive engagement in classroom activities. *J. Educ. Psychol.* 80, 514–523. doi:10.1037/0022-0663.80.4.514

Nicolaou, M. A., Gunes, H., and Pantic, M. (2011). Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space. *IEEE Trans. Affect. Comput.* 2, 92–105. doi:10.1109/T-AFFC.2011.9

Odom, S. L. (2002). *Widening the Circle: Including Children with Disabilities in Preschool Programs. Early Childhood Education Series*. Teachers College Press.

Olsen, K. N., Dean, R. T., and Stevens, C. J. (2014). A continuous measure of musical engagement contributes to prediction of perceived arousal and valence. *Psychomusicology* 24, 147–156. doi:10.1037/pmu0000044

Perepa, P. (2014). Cultural basis of social 'deficits' in autism spectrum disorders. *Eur. J. Spec. Needs Educ.* 29, 313–326. doi:10.1080/08856257.2014.908024

Picard, R. W. (1997). *Affective Computing*. Cambridge: MIT Press, 252.

Picard, R. W. (2009). Future affective technology for autism and emotion communication. *Philos. Trans. R. Soc. B* 364, 3575–3584. doi:10.1098/rstb.2009.0143

Picard, R. W., and Goodwin, M. S. (2008). Innovative technology: the future of personalized autism research and treatment. *Autism* 50, 32–39.

Pierno, A. C., Mari, M., Lusher, D., and Castiello, U. (2008). Robotic movement elicits visuomotor priming in children with autism. *Neuropsychologia* 46, 448–454. doi:10.1016/j.neuropsychologia.2007.08.020

Pietro, C., Silvia, S., and Giuseppe, R. (2014). The pursuit of happiness measurement: a psychometric model based on psychophysiological correlates. *Sci. World J.* 2014, 1–15. doi:10.1155/2014/139128

Ponitz, C., Rimm-Kaufman, S., Grimm, K., and Curby, T. (2009). Kindergarten classroom quality, behavioral engagement, and reading achievement. *School Psych. Rev.* 38, 102–120.

Robins, B., Dautenhahn, K., Boekhorst, R. T., and Billard, A. (2005). Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills? *Univ. Access Inform. Soc.* 4, 105–120. doi:10.1007/s10209-005-0116-3

Robins, B., Dautenhahn, K., and Dubowski, J. (2006). Does appearance matter in the interaction of children with autism with a humanoid robot? *Interact. Stud.* 7, 509–542. doi:10.1075/is.7.3.16rob

Robins, B., Ferrari, E., and Dautenhahn, K. (2008). "Developing scenarios for robot assisted play," in *IEEE Int'l Symposium on Robot and Human Interactive Communication (RO-MAN)*, 180–186.

Robins, B., Ferrari, E., Dautenhahn, K., Kronreif, G., Prazak-Aram, B., Gelderblom, G.-J., et al. (2010). Human-centred design methods: developing scenarios for robot assisted play informed by user panels and field trials. *Int. J. Hum. Comput. Stud.* 68, 873–898. doi:10.1016/j.ijhcs.2010.08.001

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychol. Bull.* 115, 102. doi:10.1037/0033-2909.115.1.102

Russell, J. A., and Pratt, G. (1980). A description of the affective quality attributed to environments. *J. Pers. Soc. Psychol.* 38, 311–322. doi:10.1037/0022-3514.38.2.311

Russell, V., Ainley, M., and Frydenberg, E. (2005). Student motivation and engagement. *School. Issues Digest* 2, 1–11.

Salam, H., Celiktutan, O., Hupont, I., Gunes, H., and Chetouani, M. (2017). Fully automatic analysis of engagement and its relationship to personality in human-robot interactions. *IEEE Access* 5, 705–721. doi:10.1109/ACCESS.2016.2614525

Salam, H., and Chetouani, M. (2015a). "Engagement detection based on mutli-party cues for human robot interaction," in *Int'l Conference on Affective Computing and Intelligent Interaction (ACII)*, 341–347.

Salam, H., and Chetouani, M. (2015b). "A multi-level context-based modeling of engagement in human-robot interaction," in *IEEE Int'l Conference on Automatic Face and Gesture Recognition (FG'W)*, Vol. 3, 1–6.

Scassellati, B. (2007). "How social robots will help us to diagnose, treat, and understand autism," in *In Robotics Research* (Springer), 552–563.

Scassellati, B., Admoni, H., and Matarić, M. (2012). Robots for use in autism research. *Annu. Rev. Biomed. Eng.* 14, 275–294. doi:10.1146/annurev-bioeng-071811-150036

Scherer, K. R., and Wallbott, H. G. (1994). Evidence for universality and cultural variation of differential emotion response patterning. *J. Pers. Soc. Psychol.* 66, 310–328. doi:10.1037/0022-3514.66.2.310

Shen, J., Rudovic, O., Cheng, S., and Pantic, M. (2015). "Sentiment apprehension in human-robot interaction with NAO," in *Int'l Conference on Affective Computing and Intelligent Interaction (ACII)*, 867–872.

Stanton, C. M., Kahn, P. H., Severson, R. L., Ruckert, J. H., and Gill, B. T. (2008). "Robotic animals might aid in the social development of children with autism," in *Int'l Conference on Human-Robot Interaction (HRI)*, 271–278.

Suzuki, R., Lee, J., and Rudovic, O. (2017). "Nao-dance therapy for children with ASD," in *Int'l Conference on Human-Robot Interaction (HRI)*, 295–296.

Thill, S., Pop, C. A., Belpaeme, T., Ziemke, T., and Vanderborght, B. (2012). Robot-assisted therapy for autism spectrum disorders with (partially) autonomous control: challenges and outlook. *Paladyn* 3, 209–217.

Tielman, M., Neerincx, M., Meyer, J.-J., and Looije, R. (2014). "Adaptive emotional expression in robot-child interaction," in *Int'l Conference on Human-Robot Interaction (HRI)*, 407–414.

Tincani, M., Travers, J., and Boutot, A. (2009). Race, culture, and autism spectrum disorder: understanding the role of diversity in successful educational interventions. *Res. Pract. Pers. Sev. Disabil.* 34, 81–90. doi:10.2511/rpsd.34.3-4.81

Uchida, Y., Norasakkunkit, V., and Kitayama, S. (2004). Cultural constructions of happiness: theory and empirical evidence. *J. Happiness Stud.* 5, 223–239. doi:10.1007/s10902-004-8785-9

Vivanti, G., Trembath, D., and Dissanayake, C. (2014). Mechanisms of imitation impairment in autism spectrum disorder. *J. Abnorm. Child Psychol.* 42, 1395–1405. doi:10.1007/s10802-014-9874-9

Volkmar, F. R., Paul, R., Klin, A., and Cohen, D. J. (2005). *Handbook of Autism and Pervasive Developmental Disorders, Diagnosis, Development, Neurobiology, and Behavior*, Vol. 1. John Wiley & Sons.

Wainer, J., Dautenhahn, K., Robins, B., and Amirabdollahian, F. (2014). A pilot study with a novel setup for collaborative play of the humanoid robot KASPAR with children with autism. *Int. J. Soc. Robot.* 6, 45–65. doi:10.1007/s12369-013-0195-x

Wainer, J., Ferrari, E., Dautenhahn, K., and Robins, B. (2010). The effectiveness of using a robotics class to foster collaboration among groups of children with autism in an exploratory study. *Pers. Ubiquit. Comput.* 14, 445–455. doi:10.1007/s00779-009-0266-z

Welch, B. L. (1947). The generalization of student's problem when several different population variances are involved. *Biometrika* 34, 28–35. doi:10.2307/2332510

Wong, C., and Kasari, C. (2012). Play and joint attention of children with autism in the preschool special education classroom. *J. Autism Dev. Disord.* 42, 2152–2161. doi:10.1007/s10803-012-1467-2

Zeng, Z., Pantic, M., Roisman, G. I., and Huang, T. S. (2009). A survey of affect recognition methods: audio, visual, and spontaneous expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* 31, 39–58. doi:10.1109/TPAMI.2008.52

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Rudovic, Lee, Mascarell-Maricic, Schuller and Picard. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

ORIGINAL RESEARCH
published: 28 June 2017
doi: 10.3389/fpsyg.2017.01029

Facilitating Social Play for Children with PDDs: Effects of Paired Robotic Devices

Soichiro Matsuda[1,2], Eleuda Nunez[1], Masakazu Hirokawa[1], Junichi Yamamoto[3] and Kenji Suzuki[1]*

[1] *Artificial Intelligence Laboratory, Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba, Japan,* [2] *Japan Society for the Promotion of Science, Tokyo, Japan,* [3] *Department of Psychology, Keio University, Tokyo, Japan*

Interacting with toys and other people is fundamental for developing social communication skills. However, children with autism spectrum disorder (ASD) are characterized by having a significant impairment in social interaction, which often leads to deficits in play skills. For this reason, methods of teaching play skills to young children with ASD have been well documented. Although previous studies have examined a variety of instructional strategies for teaching skills, few studies have evaluated the potential of using robotic devices. The purpose of the present study is to examine whether automatic feedback provided by colored lights and vibration via paired robotic devices, COLOLO, facilitates social play behaviors in children with ASD. We also explore how social play relates to social interaction. COLOLO is a system of paired spherical devices covered with soft fabric. All participants in this study were recruited as volunteers through the Department of Psychology at Keio University. The pilot study included three participants diagnosed with Pervasive Developmental Disorders (PDDs; 5- to 6-year-old boys), and compared experimental conditions with and without automatic feedback from the devices (colored lights and vibration). The results indicated that the participants in the condition that included feedback from the devices exhibited increased rates of ball contact and looking at the therapist's ball, but did not exhibit increased rates of eye contact or positive affect. In the experimental study, a systematic replication of the pilot study was performed with three other participants diagnosed with PDDs (3- to 6-year-old boys), using an A-B-A-B design. Again, the results demonstrated that, in the condition with colored lights and vibration, the children increased ball contact as well as looking at the therapist's ball. However, the results did not show the effect of automatic feedback consistently for three children. These findings are discussed in terms of the potential of paired robotic devices as a method to facilitate social play for children with ASD.

Keywords: autism spectrum disorder (ASD), social play, paired robotic devices, children, robot-mediated therapy, single subject design

OPEN ACCESS

Edited by:
Bilge Mutlu,
University of Wisconsin-Madison,
United States

Reviewed by:
Alda Troncone,
Università degli Studi della Campania "Luigi Vanvitelli" Caserta, Italy
Charlotte Magnusson,
Lund University, Sweden

****Correspondence:***
Soichiro Matsuda
matsuda@ai.iit.tsukuba.ac.jp

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Psychology

Received: *31 December 2016*
Accepted: *06 June 2017*
Published: *28 June 2017*

Citation:
Matsuda S, Nunez E, Hirokawa M, Yamamoto J and Suzuki K (2017) Facilitating Social Play for Children with PDDs: Effects of Paired Robotic Devices. Front. Psychol. 8:1029. doi: 10.3389/fpsyg.2017.01029

INTRODUCTION

Difficulties with play skills have been well documented in children with Autism Spectrum Disorder (ASD; Wuff, 1985; Baron-Cohen, 1987; Lewis and Boucher, 1988; Jarrold et al., 1993; Charman et al., 2000; Williams et al., 2001). These difficulties are seen in sensory motor play, manipulative play, physical play, pretend play, and social play (Boucher, 1999). Consistent with this view, many studies have focused on teaching play skills to children with ASD (Jung and Sainato, 2013).

Previous intervention studies have used video and live modeling (Jahr et al., 2000; MacDonald et al., 2005), pivotal response training (Stahmer, 1995; Thorp et al., 1995), activity schedules (Morrison et al., 2002; Machalicek et al., 2009), or social stories (Barry and Burlew, 2004). Researchers have also combined these strategies with contingent reinforcement (Jung and Sainato, 2013). These studies have found training increases engagement in appropriate play behavior and cooperative play in children with ASD. On the other hand, few studies have examined the effectiveness of robotic device use in teaching play skills to children with ASD, although robotic devices can automatically and immediately reinforce appropriate play behavior.

Robotic devices have been used to increase social-communication behaviors, such as joint attention (Warren et al., 2015; Simut et al., 2016) and imitation (Duquette et al., 2008), in children with ASD. These studies have focused on the use of both humanoid robots and non-humanoid toy-like robots, such as KASPER (Robins et al., 2009), Keapon (Costescu et al., 2015), NAO (Huskens et al., 2015; Warren et al., 2015), Probo (Simut et al., 2016), Robota (Billard et al., 2007), or Tito (Duquette et al., 2008). However, these robots mainly provide feedback as a result of the behavior of a child. Given that the facilitation of social play involves two people using toys, it may be necessary to consider directly providing feedback as a result of the behavior of both a child and the other individual.

Paired robotic devices might encourage cooperative behaviors, such as turn taking (Nunez et al., 2016). In this approach, remotely connected paired devices provided feedback separately as a result of child's own behavior as well as the other individual's behavior. Huskens et al. (2013) suggested robotic devices should be deployed as mediators to promote social interaction between a child with ASD and another individual. However, to our knowledge, no studies have examined how automatic feedback via paired robotic devices affects social play behaviors. In addition, as Diehl et al. (2012) have pointed out, most studies using robots for children with ASD have not used an experimental design, such as an experimental group design or single subject experimental design.

When considering play behaviors, we need to recognize two types. First are those related to social play, such as ball contact and looking at a therapist's ball (Bass and Mulick, 2007). Second are those related to social interaction, such as eye contact and positive affect. The purpose of the current study is to examine whether automatic feedback via paired robotic devices facilitates social play behaviors in children with ASD, and to explore how social play relates to social interaction.

If the paired robotic devices can immediately provide automatic feedback contingent on child's social play behaviors, it is possible that automatic feedback increases the social play behaviors in children with ASD. Therefore we hypothesized the following relationship between behavior contribution and feedback:

> Hypothesis 1: The child's ball contact and looking at the therapist's ball will increase with automatic feedback in the form of vibration and light.

It is possible that social play will also facilitate social interaction, and then we could expect that:

> Hypothesis 2a: The automatic feedback by vibration and light will increase behaviors associated with social interaction, such as eye contact and positive affect.

Alternatively, it is also possible that social play directs the child's attention away from the therapist toward the activity, and thus we could expect that:

> Hypothesis 2b: Automatic feedback in the form of vibration and light will decrease behaviors associated with social interaction, such as eye contact and positive affect.

To directly test these hypotheses a single AB design was used in a pilot study to make inferences about the effects of feedback made by colored lights and vibration via paired robotic devices on social play behaviors in three boys with PDDs. In this experiment, we used a rapidly changing reversal design with the same experimental condition as the pilot study. By using this experimental design, we further evaluated whether and what types of social play behaviors are facilitated by the feedback provided by remotely connected paired devices in children with PDDs.

GENERAL METHOD

Paired Robotic Devices: COLOLO

In the experiments, we used a system composed of paired devices, COLOLO. The devices have embedded sensors to detect when they are being manipulated, sending a message to the paired devices. This message is represented by visual cues made by colored lights and movements. Each device is made of a plastic spherical case covered by soft material. Inside there is a plate attached to the rotational axis of a motor by a microcontroller. A weight is attached to the motor and allows the sphere to wiggle by unbalancing the device. On the plate, there is a circuit board where a microcontroller, wireless communication module, tilt sensor, battery, and full color LEDs are installed. Each device is connected to a server via TCP/IP protocol. The server is a stationary computer that identifies the client device by a predefined ID. The roles of the server are to mediate communication among clients, pair/group clients, and log clients' communication history. The microcontroller changes the color of the LEDs and sends a message to the server when the tilt sensor detects the user's manipulation. Then, when the paired devices receive the message, the sphere starts wiggling and the color of the LEDs change according to the information in the message. In this way, users can perceive others' actions by visualizing color changes and wiggling motions. More details on the device can be found in our previous work (Nunez et al., 2016).

Experimental Condition

Both conditions (with and without automatic feedback) were implemented on the carpeted floor of a testing room at a university. In order to improve the visibility of light under the feedback condition, direct illumination was turned off and indirect lighting set at the two corners of the room (**Figure 1**). All sessions were videotaped.

There were two experimental conditions. The first condition was the *with automatic feedback condition* (Phase A). The sensors embedded in the devices detected contact (e.g., handling, bouncing, or tossing) and displayed feedback using colored lights and vibration according to the interaction rule (**Figure 2**). Under this rule it is necessary to use two devices that send and receive messages triggered by the users actions (paired configuration). When the sender device is manipulated, the visual/tactile feedback is transferred to the receiver device. By doing this, the roles of the devices are switched. If a receiver device is manipulated, it will not respond to the actions until it receives the turn from the sender device. The second condition was the *without automatic feedback condition* (Phase B), in which the devices were turned off. Therefore, the child and the therapist used them as regular balls.

The study examined the differences in child social play behaviors within the two experimental conditions: with and without automatic feedback. In both conditions, the interaction took place in the following format: (1) the therapist introduced balls to the child; (2) the therapist modeled how to play with the balls (e.g., rolling, shaking, and catching them); (3) the child manipulated the balls; and (4) the child's ball manipulating behavior was verbally/physically praised by the therapist (e.g., "You're great!" and tickling). In addition, the therapist verbally/physically praised whenever the child made eye contact, exhibited positive affect, or approach to the therapist, throughout the session.

Diagnosis Procedure

This study was approved by the affiliate university's Institutional Review Board and was, therefore, completed in accordance with the ethical standards established in the 1964 Declaration of Helsinki. All participants had a diagnosis of autistic disorder, PDD-NOS, or ASD by an outside medical doctor. Diagnosis of Pervasive Developmental Disorders (PDDs) was further confirmed using the *Pervasive Developmental Disorders Autism Society Japan Rating Scale* (PARS; Kamio et al., 2006; Ito et al., 2012a). PARS, developed in Japan, is an interview-based instrument for evaluating PDDs according to DSM-IV-TR (American Psychiatric and Association, 2000). The sub and total scores of PARS have correlations with the domain and total scores of the Autism Diagnostic Interview-Revised (ADI-R; Le Couteur et al., 1989; Lord et al., 1994). All participants with PDDs met the threshold for a diagnosis of PDDs on a total peak symptom scale score (>9).

Dependent Variables

Four dependent variables (eye contact, positive affect, ball contact, and looking at the therapist's ball) were scored using occurrence/non-occurrence data in 15-s intervals. For each session, 20 intervals were recorded. Videotape scoring was completed by a scorer who was naïve to the purpose of the study. *Eye contact*: Eye contact was defined as the child's looking at the therapist's facial region. *Positive affect*: Positive affect was defined as visible and/or audible indications of happiness and enjoyment, including smiling and laughing. *Ball contact*: Ball contact was defined as the child's contact with the ball, including handling, bouncing, and tossing the ball. *Looking at the therapist's ball*: Looking at the therapist's ball was scored when the child was looking at the ball that the therapist held.

Inter-observer Agreement

Inter-observer agreements (i.e., agreements divided by agreements plus disagreements and multiplied by 100) were calculated for both the pilot study and the experimental study. The second observer was the first author, who independently scored 33% (for pilot study) and 25% (for experiment) of the sessions for four dependent variables. Agreement was calculated as the average percentage of agreement across sessions.

Procedural Fidelity

To assess the degree to which all sessions were executed according to procedure, reliability indices for fidelity of implementation (i.e., agreements divided by agreements plus disagreements and multiplied by total number of sessions) were collected for both the pilot study and the experimental study. A research assistant and the second author completed procedural fidelity checklist on three different variables for all sessions.

PILOT STUDY

Participants

All participants were recruited as volunteers through the Department of Psychology at Keio University. Participants were three boys with PDD, "Taro," "Sabu," and "Jiro," between the ages of 5 and 6 years. Names of participants have been changed to protect the participants' identities. Informed consent was obtained from the parents before the children were included in the study.

Table 1 displays the participants' characteristics. The participants' initial profiles (i.e., language, communication, motor, perceptual, and adaptive behavior skills) were assessed using standardized assessment tools: the *Kyoto Scale for Psychological Development 2001* (KSPD; Ikuzawa et al., 2002), the *Vineland Adaptive Behavior Scales,* 2nd edition Japanese version (Vineland-II; Ito et al., 2012b), and the *MacArthur Communicative Development Inventories*, Japanese version (MCDIs; Ogura, 2007). The KSPD yields standard scores for physical-movement (P-M), language-sociability (L-S), and cognitive-adaptive (C-A) subscales and total developmental quotient (DQ). The KSPD was developed for use with typically developing infants and low-function children with ASD and other developmental disorders in Japan.

Design and Procedure

A single AB design was used in the pilot study. By contrasting the *with* automatic feedback condition (Phase A) and the *without* automatic feedback condition (Phase B), we could make

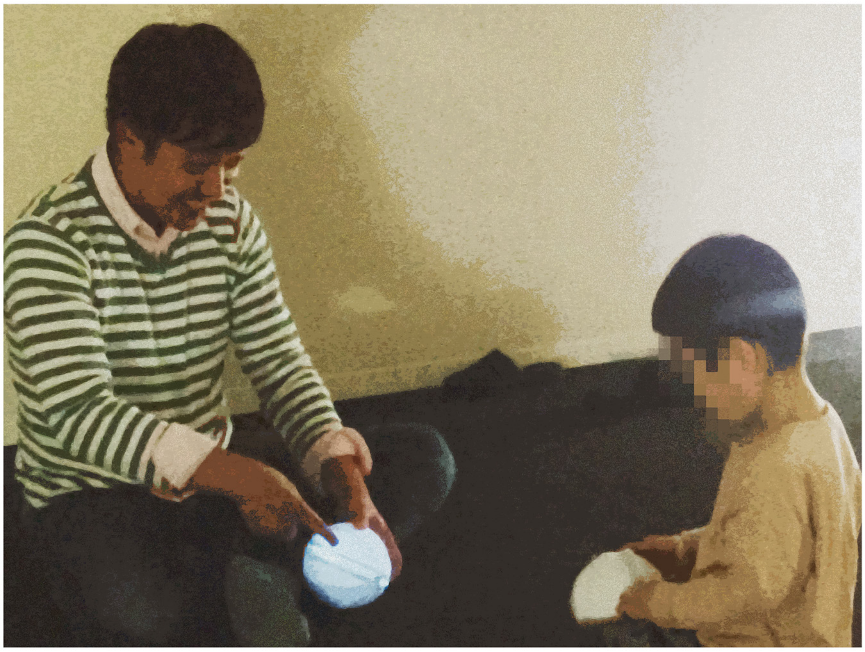


FIGURE 1 | Basic image of a session in the *with* automatic feedback condition. The therapist and the child both hold COLOLO.

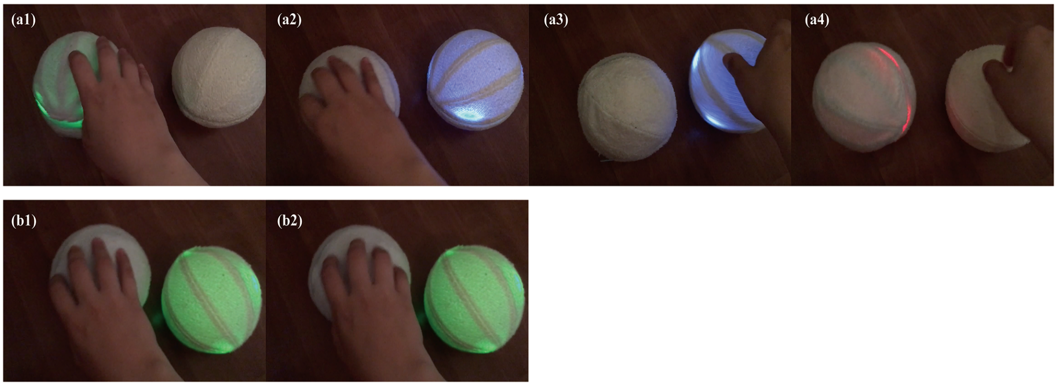


FIGURE 2 | Example of the transferring lights rule. **(a1)** When a user manipulates Device A, **(a2)** the paired device (Device B) will provide feedback of light and vibration. **(a3)** When the user manipulates Device B, **(a4)** the paired device (Device A) will provide feedback of light and vibration. **(b1)** When the user manipulates Device A during the feedback of Device B, **(b2)** Device A will not give any feedback, and the user will need to wait for a response.

inferences about differences of the dependent variables between the experimental conditions.

Each phase consisted of a 5-min session, and both phases were conducted in a same day for each participant. First, the *with* automatic feedback condition (Phase A) was presented, and then, after a short break, the *without* automatic feedback condition (Phase B) followed.

Results

For eye contact, the average observer agreement value was 97% (range 95–100%); for positive affect, 97% (range 95–100%); for ball contact, 90% (range 85–95%); and for looking at the therapist's ball, 82% (range 80–85%). Fidelity of implementation for socially/physically reinforcing the child's eye contact, positive affect, and approach to the therapist averaged 67%; fidelity of implementation for socially/physically reinforcing the child's ball contact averaged 100%; and fidelity of implementation for modeling and prompting ball play averaged 100%. Results are shown in **Figure 3**.

Eye Contact

The percentage of intervals with eye contact in the *with* automatic feedback condition was 0% for Taro, 0% for Jiro, and 15% for Sabu. On the other hand, in the *without* automatic feedback condition, these numbers increased to 10, 15, and 55%, respectively.

TABLE 1 | Participant profiles in the pilot study.

Child		Taro	Jiro	Sabu
Chronological age		6;9	5;6	5;6
PARS	Total peak symptom scale score	51	28	24
KSPD	Full DQ	77	33	38
	P-A DQ	56	56	55
	L-S DQ	76	29	34
	C-A DQ	79	45	41
VAB-II-J	Adoptive behavior composite	48	45	51
	Communication	63	34	58
	Daily living skills	47	54	61
	Socialization	38	36	45
	Motor	51	51	51
J-MCDIs	Words understood	418	74	376
	Words said	413	3	180
	Total gestures produced	37	22	35

PARS, pervasive developmental disorders autism society Japan rating scale; KSPD, Kyoto scales of psychological development 2001; DQ, developmental quotient; Full, total scale; P-A, physical-movement; L-S, language-sociability; C-A, cognitive-adaptive; VAB-II-J, Vineland adaptive behavior scales 2nd edition Japanese version; J-MCDIs, MacArthur Communicative Development Inventories, Japanese version.

Positive Affect

Taro and Sabu demonstrated almost the same levels of positive affect in both conditions. Jiro exhibited positive affect in 5% of the intervals in the *with* automatic feedback condition and 35% of the intervals in the *without* automatic feedback condition.

Ball Contact

All three children demonstrated increased levels of ball contact in the *with* automatic feedback condition. Specifically, the percentage of intervals with ball contact in the *with* automatic feedback condition was 65% for Taro, 95% for Jiro, and 60% for Sabu. In contrast, in the *without* automatic feedback condition, these figures decreased to 50, 10, and 45%, respectively.

Looking at the Therapist's Ball

Similarly, all three children exhibited increased levels of looking at the therapist's ball in the *with* automatic feedback condition. Specifically, the percentage of intervals with looking at the therapist's ball during the *with* automatic feedback condition was 40% for Taro, 60% for Jiro, and 15% for Sabu. In contrast, during the *without* automatic feedback condition, these numbers decreased to 5, 15, and 0%, respectively.

EXPERIMENTAL STUDY

Participants

All participants were recruited as volunteers through the Department of Psychology at Keio University. The participants were three boys with ASD, "Shiro," "Goro," and "Riku," between the ages of 3 and 6 years. Names of participants have been changed to protect the participants' identities. Informed consent was obtained from the parents before the children were included in the study. **Table 2** displays the participants' characteristics.

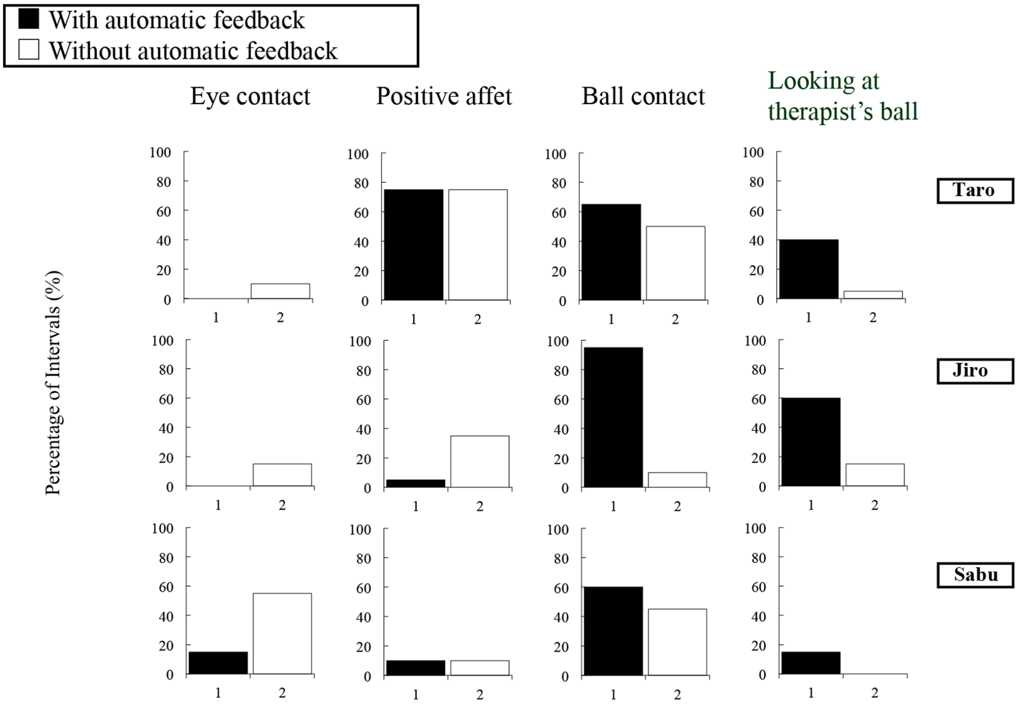


FIGURE 3 | Percentage of 15-s intervals with eye contact, positive affect, ball contact, and looking at the therapist's ball in the *with* and *without* automatic feedback conditions in the pilot study.

TABLE 2 | Participant profiles in the experiment.

Child		Shiro	Goro	Riku
Chronological age		5;8	6;8	3;8
PARS	Total peak symptom scale score	52	46	28
KSPD	Full DQ	45	74	44
	P-A DQ	54	57	65
	L-S DQ	41	68	25
	C-A DQ	49	79	45
VAB-II-J	Adoptive behavior composite	55	53	49
	Communication	51	68	43
	Daily living skills	60	45	56
	Socialization	69	45	41
	Motor	51	67	65
J-MCDIs	Words understood	199	418	47
	Words said	181	338	11
	Total gestures produced	52	51	25

PARS: pervasive developmental disorders autism society Japan rating scale; KSPD, Kyoto scales of psychological development 2001; DQ, developmental quotient; Full, total scale; P-A, physical-movement; L-S, language-sociability; C-A, cognitive-adaptive; VAB-II-J, Vineland adaptive behavior scales 2nd edition Japanese version; J-MCDIs, MacArthur Communicative Development Inventories, Japanese version.

Design and Procedure

The Council for Exceptional Children (CEC) Division of Research established a task force to develop guidelines for evidence-based practices (Odom et al., 2004). The task force identified four types of research methodologies: qualitative, correlational, experimental group, and single subject designs (Odom et al., 2004). Single subject designs have been used to compare the causal relationship between independent and dependent variables (Barlow et al., 2009). In this experiment, we used a single subject experimental design in a particular, rapidly changing reversal design (Cooper et al., 1990, 1993; Dunlap et al., 1991; Ishizuka and Yamamoto, 2016) over a total of two experimental days to compare the effects of lighting and vibration as automatic feedback. For all children, the experiment consisted of four 5-min sessions. Each participant had two 5-min sessions per day.

Results

For eye contact, the average observer agreement value was 80% (range 75–85%); for positive affect, 88% (range 85–90%); for ball contact, 95% (range 85–100%); and for looking at the therapist's ball, 88% (range 75–100%). Fidelity of implementation for socially/physically reinforcing the child's eye contact, positive affect, and approach to the therapist averaged 92%; fidelity of implementation for socially/physically reinforcing the child's ball contact averaged 92%; and fidelity of implementation for modeling and prompting ball play averaged 100%. Results of the reversal analyses for each of the dependent variables are presented in **Figure 4**.

Eye Contact

Shiro exhibited eye contact with a mean of 20% of the intervals in the *with* automatic feedback condition and a mean of 12.5% in the *without* automatic feedback condition. Goro showed no eye contact in either condition. In the *with* automatic feedback condition, Riku exhibited eye contact for a mean of 40% of the intervals. On the other hand, in the *without* automatic feedback condition, his eye contact decreased to a mean of 20% across sessions.

Positive Affect

Shiro and Goro demonstrated a similar response pattern for positive affect. Specifically, in the initial *with* automatic feedback probe, they exhibited low positive affect. With the introduction of the *without* automatic feedback condition, their levels of positive affect increased to 45% (for Shiro) and 60% (for Goro) of the intervals. The reintroduction of the *with* automatic feedback condition was accompanied by a drop in positive affect levels to 5% and 10% of the intervals, respectively. The final *without* automatic feedback condition phase resulted in positive affect for 25 and 50% of the intervals, respectively, for the two boys.

In the first *with* automatic feedback probe, Riku exhibited positive affect in 15% of the intervals. Following the introduction of the *without* automatic feedback condition, his positive affect decreased slightly to 10% of the intervals. During the reintroduction of the *with* automatic feedback condition, Riku exhibited positive affect in 60% of the intervals. In the final *without* automatic feedback condition phase, Riku did not exhibit any positive affect.

Ball Contact

All three children demonstrated similar response patterns for ball contact. The initial *with* automatic feedback phase resulted in ball contact in 100% (for Shiro), 85% (for Goro), and 95% (for Riku) of the intervals. With the introduction of the *without* automatic feedback condition, the levels of ball contact decreased to 75%, 40%, and 5%, respectively. The reintroduction of the *with* automatic feedback condition was accompanied by a rise in ball contact levels to 100, 85, and 95% of the intervals, respectively. The final *without* automatic feedback condition phase resulted in ball contact for 75, 70, and 30% of the intervals, respectively, for the three boys.

Looking at the Therapist's Ball

Shiro exhibited looking at the therapist's ball with a mean of 75% in the *with* automatic feedback condition and a mean of 47.5% in the *without* automatic feedback condition. For Goro, the means were 25% in the *with* automatic feedback condition and 10% in the *without* automatic feedback condition. In the *with* automatic feedback condition, Riku exhibited looking at the therapist's ball for a mean of 82.5% of the intervals. In contrast, during the *without* automatic feedback condition, his looking at the therapist's ball decreased to a mean of 15% across sessions.

GENERAL DISCUSSION

This study investigated the effects of automatic feedback in the form of colored lights and vibration produced via paired robotic devices, COLOLO, in social play and interaction in children with ASD. The frequency of ball contact and looking at

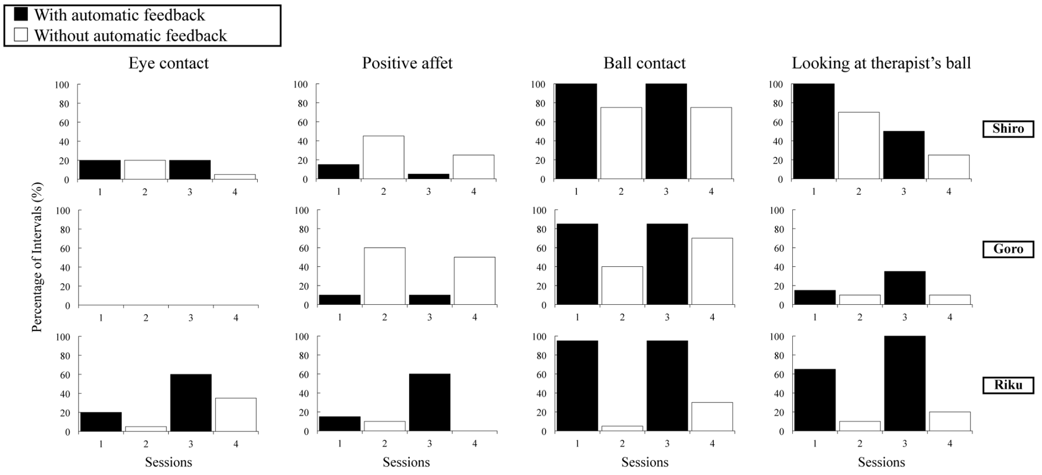


FIGURE 4 | Percentage of 15-s intervals with eye contact, positive affect, ball contact, and looking at the therapist's ball in the *with* and *without* automatic feedback conditions in the experimental study.

the therapist's ball were higher in the *with* automatic feedback condition than in the *without* automatic feedback condition, supporting Hypothesis 1. On the other hand, the frequencies of eye contact and positive affect for all children with ASD did not consistently increase or decrease in the *with* automatic feedback condition, thus the results indicated lack of support for both Hypothesis 2a and Hypothesis 2b. Therefore, when using the paired robotic devices, the children with ASD appear to have exhibited increases in social play behaviors using toys and but not increases in behaviors associated with social interaction.

Considering ball contact, Hypothesis 1 was positively supported. The findings are in lines with one of the pioneering works, which has demonstrated that a spherical mobile robot, Roball, may increase a child's interaction with a ball by providing automatic feedback consisting of motion, messages, sounds, and an illuminating interface (Michaud et al., 2005). This suggests that automatic feedback of vibration (tactile stimulus) might function as a reinforcer for ball contact behavior. However, we also used light feedback (visual stimulus). There is a possibility that light feedback also functions as a reinforcer for child's ball contact. Therefore, in a future study, we would evaluate which modality of feedback has a stronger effect on increasing ball contact.

Considering frequency of looking at the therapist's ball, the first Hypothesis was also positively supported. This indicates that attention to shared play materials might be increased by light feedback via paired robotic devices. Although we used vibration feedback, this feedback was not contingent upon the child looking at the therapist's ball, but contingent on the child looking at his own ball. Thus, light feedback provided via remotely connected paired devices may increase attention to the play materials of peers in children with ASD.

Concerning Hypotheses 2a and 2b of the current study, our results did not support either of these hypotheses. Neither the child's eye contact nor their positive affect consistently increased as a result of the feedback in the form of light and vibration. The result can be easily interpreted because the feedback was not contingent upon the child's responses. In addition, however, increases in eye contact and positive affect were observed in the *without* automatic feedback condition for two children. A potential explanation for this outcome could be the frequency of the reinforcement provided by the therapist. As far as ecologically validity is concerned, in the procedure of this experiment, the therapist provided verbal/physical praise for the child's eye contact and positive affect throughout the session. This may have led to increased opportunity for praise for the therapist in the *without* automatic feedback condition in which the frequencies of child's ball contact was lower. To improve this aspect of the intervention, we recommended that future studies include the combined use of other wearable devices, such as an eye tracker (Ye et al., 2012) or a face reader device we have developed for detecting smiles from facial EMG signals (Gruebler and Suzuki, 2014), in order to provide contingent feedback for child eye contact and/or positive affect.

There were several limitations to the current study. First, we used a single subject experimental design with three children with ASD in this study, and this limits the generalizability of the results to the larger population due to limitations inherent in single subject experimental designs, such as absence of statistical analysis and inference. Further studies are required, including use of a group experimental design with larger sample sizes. Second, although we used an ABAB design to minimize carryover effects, because the experiment sessions were administered across 2 days, we were not able to eliminate ordering and novelty effects. It is possible that the novelty of the interaction affected the increase in the dependent variables on the 1st day (first set of AB trials) and on the 2nd day (second set of AB trials), due to the time that has elapsed between the first and the second session. Further studies must seek to eliminate ordering and novelty effects through blocked and longitudinal study designs. Third, we need to be cautious about interpreting the observed increases in children's ball contact and looking at the therapist's ball as the result of automatic feedback functioning as a reinforcer. It was unclear whether automatic feedback functioned as an antecedent stimulus or a reinforcer for children's ball contact and looking at the therapist's ball. Further research is warranted to identify the function of automatic feedback via the implementation of a yoked condition. Fourth, we only used the feedback of light (visual stimulus) and vibration (tactile stimulus). Future studies will be required to use other modalities, such as sound (auditory stimulus).

Nevertheless, the current findings establish that feedback via paired robotic devices can facilitate some aspects of social play behaviors in children with ASD, whereas previous studies have focused on examining differences between a human and a robot as an interaction partner (e.g., Costescu et al., 2015; Srinivasan et al., 2015; Simut et al., 2016), or investigating the effects of teaching by the robot (e.g., Billard et al., 2007; Warren et al., 2015). As Huskens et al. (2013) have suggested, it would be interesting to see more studies on this topic; in other words, there is a wide range of necessities for further investigation. While we are hopeful that clinical applications of paired robotic devices may demonstrate significant enhancement of social play for children with ASD at an early developmental stage, it is important to note that future research should reveal both whether and how the paired robotic devices contribute to increasing various forms of social play behaviors in children with ASD.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of Keio University's Institutional Review Board with written informed consent from all parents of participants.

AUTHOR CONTRIBUTIONS

SM, EN, MH, JY, and KS designed the research. SM and EN performed the research. SM analyzed the data and wrote the article.

FUNDING

This work was supported by the CREST research project on Social Imaging, Japan Science and Technology Agency (JST) and a Grant-in-Aid for the Japan Society for the Promotion of Science (JSPS) Fellows.

ACKNOWLEDGMENT

The authors would like to thank Joseph P. McCleery for the English language review.

REFERENCES

American Psychiatric and Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC: American Psychiatric Association.

Barlow, D. H., Nock, M., and Hersen, M. (2009). *Single Case Experimental Designs: Strategies for Studying Behavior for Change*, 3rd Edn. Boston, MA: Pearson Education.

Baron-Cohen, S. (1987). Autism and symbolic play. *Br. J. Dev. Psychol.* 5, 139–148. doi: 10.1111/j.2044-835X.1987.tb01049.x

Barry, L. M., and Burlew, S. B. (2004). Using social stories to teach choice and play skills to children with autism. *Focus Autism Other Dev. Disabil.* 19, 45–51. doi: 10.1007/s10803-008-0628-9

Bass, J. D., and Mulick, J. A. (2007). Social play skill enhancement of children with autism using peers and siblings as therapists. *Psychol. Sch.* 44, 727–735. doi: 10.1002/pits.20261

Billard, A., Robins, B., Nadel, J., and Dautenhahn, K. (2007). Building robota, a mini-humanoid robot for the rehabilitation of children with autism. *Assist. Technol.* 19, 37–49. doi: 10.1080/10400435.2007.10131864

Boucher, J. (1999). Editorial: interventions with children with autism–methods based on play. *Child Lang. Teach. Ther.* 15, 1–5. doi: 10.1191/026565999676029298

Charman, T., Baron-Cohen, S., Swettenham, J., Baired, G., Cox, A., and Drew, A. (2000). Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. *Cogn. Dev.* 15, 481–498. doi: 10.1016/S0885-2014(01)00037-5

Cooper, L. J., Wacker, D. P., Millard, T., Derby, K. M., Cruikshank, B. M., and Rogers, L. (1993). Assessing environmental and medication variables in an outpatient setting: a proposed model and preliminary results with ADHD children. *J. Dev. Phys. Disabil.* 5, 71–85. doi: 10.1007/BF01046599

Cooper, L. J., Wacker, D. P., Sasso, G. M., Reimers, T. M., and Donn, L. K. (1990). Using parents as therapists to evaluate appropriate behavior of their children: application to a tertiary diagnostic clinic. *J. Appl. Behav. Anal.* 23, 285–296. doi: 10.1901/jaba.1990.23-285

Costescu, C. A., Vanderborght, B., and David, D. O. (2015). Reversal learning task in children with autism spectrum disorder: a robot-based approach. *J. Autism Dev. Disord.* 45, 3715–3725. doi: 10.1007/s10803-014-2319-z

Diehl, J. J., Schmitt, L. M., Villano, M., and Crowell, C. R. (2012). The clinical use of robots for individuals with autism spectrum disorders: a critical review. *Res. Autism Spectr. Disord.* 6, 249–262. doi: 10.1016/j.rasd.2011.05.006

Dunlap, G., Kern-Dunlap, L., Clarke, S., and Robbins, F. R. (1991). Functional assessment, curricular revision, and severe behavior problems. *J. Appl. Behav. Anal.* 24, 387–397. doi: 10.1901/jaba.1991.24-387

Duquette, A., Michaud, F., and Mercier, H. (2008). Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism. *Auton. Robots* 24, 147–157. doi: 10.1007/s10514-007-9056-5

Gruebler, A., and Suzuki, K. (2014). Design of a wearable device for reading positive expressions from facial EMG signals. *IEEE Trans. Affect. Comput.* 5, 227–237. doi: 10.1109/TAFFC.2014.2313557

Huskens, B., Palman, A., Van der Werff, M., Lourens, T., and Barakova, E. (2015). Improving collaborative play between children with autism spectrum disorders and their siblings: the effectiveness of a robot-mediated intervention based on Lego therapy. *J. Autism Dev. Disord.* 45, 3746–3755. doi: 10.1007/s10803-014-2326-0

Huskens, B., Verschuur, R., Gillesen, J., Didden, R., and Barakova, E. (2013). Promoting question-asking in school-aged children with autism spectrum disorders: effectiveness of a robot intervention compared to a human-trainer intervention. *Dev. Neurorehabil.* 16, 345–356. doi: 10.3109/17518423.2012.739212

Ikuzawa, M., Matsushita, Y., and Nakase, A. (eds). (2002). *Kyoto Scale for Psychological Development* 2001. Kyoto: Kyoto International Social Welfare Exchange Centre.

Ishizuka, Y., and Yamamoto, J. (2016). Contingent imitation increases verbal interaction in children with autism spectrum disorders. *Autism* 20, 1011–1020. doi: 10.1177/1362361315622856

Ito, H., Tani, I., Yukihiro, R., Adachi, J., Hara, K., Ogasawara, M., et al. (2012a). Validation of an interview-based rating scale developed in Japan for pervasive developmental disorders. *Res. Autism Spectr. Disord.* 6, 1265–1272. doi: 10.1016/j.rasd.2012.04.002

Ito, H., Tani, I., Yukihiro, R., Uchiyama, K., Ogasawara, M., and Tsujii, M. (2012b). Development of the Japanese version of the vineland adaptive behavior scales. Second edition: reliability and validity of the maladaptive behavior scales. *Seishin Igaku* 54, 537–548.

Jahr, E., Eldevik, S., and Eikeseth, S. (2000). Teaching children with autism to initiate and sustain cooperative play. *Res. Dev. Disabil.* 21, 151–169. doi: 10.1016/S0891-4222(00)00031-7

Jarrold, C., Boucher, J., and Smith, P. (1993). Symbolic play in autism: a review. *J. Autism Dev. Disord.* 23, 281–307. doi: 10.1007/BF01046221

Jung, S., and Sainato, D. M. (2013). Teaching play skills to young children with autism. *J. Intellect. Dev. Disabil.* 38, 74–90. doi: 10.3109/13668250.2012.732220

Kamio, Y., Yukihiro, R., Adachi, J., Ichikawa, H., Inoue, M., Uchiyama, T., et al. (2006). Reliability and validity of the pervasive developmental disorder (PDD) autism society Japan rating scale: a behavior checklist for adolescents and adults with PDDs. *Clin. Psychiatry* 48, 495–505.

Le Couteur, A., Rutter, M., Lord, C., Rios, P., Robertson, S., Holdgrafer, M., et al. (1989). Autism diagnostic interview: a standardized investigator-based instrument. *J. Autism Dev. Disord.* 19, 363–387. doi: 10.1007/BF02212936

Lewis, V., and Boucher, J. (1988). Spontaneous, instructed and elicited play in relatively able autistic children. *Br. J. Dev. Psychol.* 6, 325–339. doi: 10.1111/j.2044-835X.1988.tb01105.x

Lord, C., Rutter, M., and Le Couteur, A. (1994). Autism diagnostic interview–revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism Dev. Disord.* 24, 659–685. doi: 10.1007/BF02172145

MacDonald, R., Clark, M., Garrigan, E., and Vangala, M. (2005). Using video modeling to teach pretend play to children with autism. *Behav. Interv.* 20, 225–238. doi: 10.1002/bin.197

Machalicek, W., Shogren, K., Lang, R., Rispoli, M., O'Reilly, M. F., Franco, J. H., et al. (2009). Increasing play and decreasing the challenging behavior of children with autism during recess with activity schedules and task correspondence training. *Res. Autism Spectr. Disord.* 3, 547–555. doi: 10.1016/j.rasd.2008.11.003

Michaud, F., Laplante, J. F., Larouche, H., Duquette, A., Caron, S., Letourneau, D., et al. (2005). Autonomous spherical mobile robot for child development studies. *IEEE Trans. Syst. Man Cybern. A* 35, 471–480. doi: 10.1109/TSMCA.2005.850596

Morrison, R. S., Sainato, D. M., Benchaaban, D., and Endo, S. (2002). Increasing play skills of children with autism using activity schedules and correspondence training. *J. Early Interv.* 25, 58–72. doi: 10.1177/105381510202500106

Nunez, E., Matsuda, S., Hirokawa, M., Yamamoto, J., and Suzuki, K. (2016). "An approach to facilitate turn-taking behavior with paired devices for children with autism spectrum disorders," in *Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication*, Roman, 837–842. doi: 10.1109/roman.2016.7745216

Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. D., Thompson, B., and Harris, K. (2004). *Quality Indicators for Research in Special Education and Guidelines for Evidence-Based Practices: Executive Summary*. Arlington, VA: Council for Exceptional Children Division for Research.

Ogura, T. (2007). Early lexical development in Japanese children. *Gengo Kenkyu* 132, 29–53.

Robins, B., Dautenhahn, K., and Dickerson, P. (2009). "From isolation to communication: A case study evaluation of robot assisted play for children with autism with a minimally expressive humanoid robot," in *Proceedings of the Second International Conferences on Advances in Computer-Human Interactions, ACHI'09* (Cancun: IEEE Computer Society Press), 205–211. doi: 10.1109/achi.2009.32

Simut, R. E., Vanderfaeillie, J., Peca, A., Van de Perre, G., and Vanderborght, B. (2016). Children with autism spectrum disorders make a fruit salad with Probo, the social robot: an interaction study. *J. Autism Dev. Disord.* 46, 113–126. doi: 10.1007/s10803-015-2556-9

Srinivasan, S. M., Park, I. K., Neelly, L. B., and Bhat, A. N. (2015). A comparison of the effects of rhythm and robotic interventions on repetitive behaviors and affective states of children with autism spectrum disorder (ASD). *Res. Autism Spectr. Disord.* 18, 51–63. doi: 10.1016/j.rasd.2015.07.004

Stahmer, A. C. (1995). Teaching symbolic play skills to children with autism using pivotal response training. *J. Autism. Dev. Disord.* 25, 123–141. doi: 10.1007/BF02178500

Thorp, D. M., Stahmer, A. C., and Schreibman, L. (1995). Effects of sociodramatic play training on children with autism. *J. Autism Dev. Disord.* 25, 265–282. doi: 10.1007/BF02179288

Warren, Z. E., Zheng, Z., Swanson, A. R., Bekele, E., Zhang, L., Crittendon, J. A., et al. (2015). Can robotic interaction improve joint attention skills? *J. Autism Dev. Disord.* 45, 3726–3734. doi: 10.1007/s10803-013-1918-4

Williams, E., Reddy, V., and Costall, A. (2001). Taking a closer look at functional play in children with autism. *J. Autism Dev. Disord.* 31, 67–77. doi: 10.1023/A:1005665714197

Wuff, S. B. (1985). The symbolic and object play of children with autism: a review. *J. Autism Dev. Disord.* 15, 139–148. doi: 10.1007/BF01531600

Ye, Z., Li, Y., Fathi, A., Han, Y., Rozga, A., Abowd, G. D., et al. (2012). "Detecting eye contact using wearable eye-tracking glasses," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, (New York, NY: ACM), 699–704. doi: 10.1145/2370216.2370368

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Matsuda, Nunez, Hirokawa, Yamamoto and Suzuki. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

REVIEW
published: 09 June 2017
doi: 10.3389/fpsyg.2017.00958

Ethorobotics: A New Approach to Human-Robot Relationship

Ádám Miklósi[1,2], Péter Korondi[3], Vicente Matellán[4] and Márta Gácsi[2]*

[1] Department of Ethology, Eötvös Loránd University, Budapest, Hungary, [2] Magyar Tudományos Akadémia – Eötvös Loránd University Comparative Ethology Research Group, Budapest, Hungary, [3] Department of Mechatronics, Optics and Information Engineering, Budapest University of Technology and Economics, Budapest, Hungary, [4] Departamento Ingeniería Mecánica, Informática y Aeroespacial, Universidad de León, León, Spain

Here we aim to lay the theoretical foundations of human-robot relationship drawing upon insights from disciplines that govern relevant human behaviors: ecology and ethology. We show how the paradox of the so called "uncanny valley hypothesis" can be solved by applying the "niche" concept to social robots, and relying on the natural behavior of humans. Instead of striving to build human-like social robots, engineers should construct robots that are able to maximize their performance in their niche (being optimal for some specific functions), and if they are endowed with appropriate form of social competence then humans will eventually interact with them independent of their embodiment. This new discipline, which we call *ethorobotics*, could change social robotics, giving a boost to new technical approaches and applications.

Keywords: social robotics, ethology, human-robot interaction, niche, social competence, dog, uncanny valley

OPEN ACCESS

Edited by:
Ginevra Castellano,
Uppsala University, Sweden

Reviewed by:
Iolanda Leite,
Royal Institute of Technology, Sweden
Giovanna Varni,
Université Pierre et Marie Curie, France

****Correspondence:***
Márta Gácsi
marta.gacsi@gmail.com

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Psychology

Received: *27 December 2016*
Accepted: *24 May 2017*
Published: *09 June 2017*

Citation:
Miklósi Á, Korondi P, Matellán V and Gácsi M (2017) Ethorobotics: A New Approach to Human-Robot Relationship. Front. Psychol. 8:958. doi: 10.3389/fpsyg.2017.00958

THE MORE HUMAN-LIKE THE BETTER?

Motto: "You climb to reach the summit, but once there, discover that all roads lead down."

Stanislaw Lem, The Cyberiad

Social robotics is the science for developing and building robots that can be integrated into human groups, and are able to engage in complex social interactions with humans, including communication and collaboration (e.g., Fong et al., 2003; Dautenhahn, 2007).

The recent increased interest by the media to introduce and popularize such robots to the public (e.g., Saya) and general interest in science fiction (e.g., AI, Robocop) seems to make both lay persons and many scientists to believe that social robotics should produce robots (so called androids) that match perfectly humans both in their embodiment (e.g., DiSalvo et al., 2002) and in their communicative and problem solving skills (some improved version of C-3PO). Although the emergence of everyday social robots on the markets is still decades away, marketing pressure, grant agencies (in the United States, EU, and China), and the challenges of engineering also push applications toward building human-like robots.

Subjectively one may feel that humans like to be and interact with agents of closely similar kind and may avoid more machine-like creatures. However, the only serious hypothesis, which was put forward by Mori (1970), argues the opposite: the more similar robots are to humans the more humans avoid them.

THE RETHINKING OF THE 'UNCANNY VALLEY' HYPOTHESIS AND ITS PREDICTIONS

The 'uncanny valley' hypothesis articulated by Mori in 1970 was the first theoretical evaluation of the predicted relationship between humans and non-living agents, including robots. **Figure 1** presents a modified reproduction of Mori's (1970) original idea by showing the humans' reaction only to moving agents. It is assumed that social robots getting very similar to humans (measured by some complex variable) are being more and more rejected by people. Very similar robots are rejected much more than less similar ones. Social robots may never reach the 'Maximum peak' which represents humanness. Implicitly this figure also suggests that social robotics develops from left to right aiming specifically at designing human-like robots. Thus the X axis represents both "human likeness" and "time."

Mori's hypothesis suggests a complex relationship between the agent's (biological or artificial) similarity to a human and the human's affinity toward the agent. Accordingly, the dependent variable (in Japanese 'shinwakan'), called affinity (MacDorman and Minato, 2005) has two local maximum values. The first one on the left (**Figure 1**) is referred to as the "Medium Peak." It emerges at a point where similarity between the agent and a typical human is substantial but still relatively low (approx. 60–75%). The other one is at the right part of the figure when the agents reach (nearly) perfect similarity with humans. This is the "Maximum Peak." Most importantly, it is claimed that for a narrow range of very close similarity to humans, values of affinity will obtain very low or even negative values, labeled as the uncanny valley.

In the original paper Mori left open the question of causation, and subsequent scientific discussions focused on either (1) evolutionary explanations (e.g., avoidance of threat, or death; see MacDorman and Ishiguro, 2006; Moosa and Minhaz Ud-Dean, 2010), (2) developmental effects (e.g., babies show this effect only after 12 months of age; Lewkowicz and Ghazanfar, 2012), or (3) perceptual and mental mechanisms (e.g., activation of competing mental representations; Chen et al., 2009; Ferrey et al., 2015). While these explanations are not mutually exclusive they all assume that the phenomenon is specific to humans (or non-human primates) (MacDorman et al., 2009; Steckenfinger and Ghazanfar, 2009) and researchers investigate it only in relation to artificial creatures (cf. robots) (Mathur and Reichling, 2016).

One may consider that the phenomenon may have a more wide-spread biological (functional) basis, the recognition of which leads to a different perspective. Here we argue that the present trend in social robotics is misguided. We show that an ethological approach, considering functional aspects of behavior and human-robot interaction, can provide a more plausible theoretical background for social robotics. We aim to establish an interdisciplinary science of ethorobotics, which relies on evolutionary, ecological, and ethological concepts for developing social robots. We suggest that while the similarity of the agent's characteristics may enhance the efficiency of the interactions, the social identification/categorization of the agent also plays a crucial role in respect of affinity and expectations.

THE IMPORTANCE OF RECOGNITION OF OTHERS

We propose that in humans the avoidance of very closely similar others reflects a more widely distributed skill in animal species, which is aimed to precisely categorize and recognize other potentially significant biological agents. The specific function of this ability depends on the ecology of the species but this process is invaluable for survival (Mateo, 2004). In general biological agents should be able to discriminate others at three different levels: (1) conspecifics (same species) versus heterospecifics (other species, e.g., predators); (2) familiar conspecifics (e.g., group members) versus unfamiliar conspecifics (e.g., strangers/intruders); (3) familiar conspecifics versus individuals (e.g., mate, friends, and pups). The rapid and precise discrimination of others is important because it determines what kind of actions should be taken and what kind of responses could be expected. Animals may rely on different set of features (e.g., visual, auditory and olfactory) for this discrimination but generally it can be assumed that the computational need is the highest at the 3rd level.

Biological agents achieve this performance by being sensitive to some simple but specific pattern of cues (e.g., sign stimuli) early in their development, and this attraction provides the basis for further learning about the peculiarities of others. Such learning usually takes place during a specific sensitive phase when some neural structures acquire selective responsiveness to recognize and discriminate specific set of cues. Such perceptual learning is based on selective elimination of not-stimulated pre and post-synaptic connections. Although such learning can take place also later in development or adulthood, the stronger and less reversible effects probably happen when the neural system matures. The ability to discriminate others has been investigated in several species (Colgan, 1983), and also on humans.

Sensitive Period of Social Recognition in Humans

Recently, it has been hypothesized that early experience with human faces provides the basis of the uncanny valley effect in infants (Ferrey et al., 2015). The comparison of 6 to 12 month old infants showed that only the oldest group avoided unrealistic faces.

It has been long known that few hour old newborns show preference toward face-like patterns (Johnson et al., 1991). More recent results have indicated that 3-day-old newborns look longer at faces gazing at them directly, and they also prefer to look at faces presenting two eye-like patterns on the top rather than on the bottom (Farroni et al., 2005). It seems that there is a genetically canalized preference for some visual features (sign stimuli) that make the infant focus on the (human) face. This interest helps the infant to learn about other components of the face that is made possible by the parallel improvement of visual and neural processing (e.g., Gliga and Csibra, 2007; Pascalis and

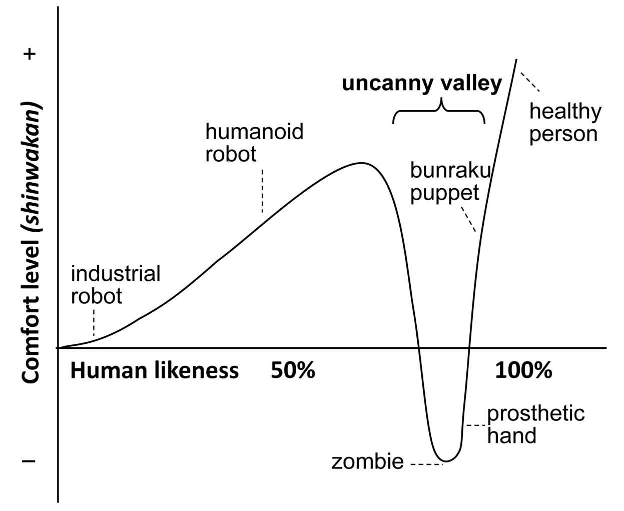


FIGURE 1 | A modified reproduction of Mori's (1970) original 'uncanny valley' hypothesis by showing only the reaction to moving agents.

Kelly, 2009). As a result infants become experts in discriminating and recognizing individuals from the same category (familiar faces in the group). Babies are much better in making such discriminations in the case of their own race than in other races ('other-race' effect; e.g., Kelly et al., 2009), although this effect is smaller if babies are exposed to members of different races early on (Sangrigoli and De Schonen, 2004).

This natural process of emerging social recognition in humans suggests that only by massively exposing babies to (future) social robots can we avoid that they 'fall in the uncanny valley.' Such forced exposure seems unrealistic and would be also unethical, moreover, it could also confuse the social recognition system of humans, and lead to misguided social and sexual preferences.

RE-INTERPRETATION OF THE 'UNCANNY VALLEY'

We argue that in Mori's landscape, the similarity measure (*X*-axis) relates to the interaction of heterogenic agents when one type of agent is used as point of reference. This is equivalent to a biological scenario with conspecifics and heterospecifics. Thus the Medium Peak refers to interactions with a specific group of heterospecifics that share many attributes with humans (e.g., domesticated animals) and the Maximum Peak refers to interaction among conspecifics (**Figure 1**). Note that heterospecific agents represent a much larger and diverse category than conspecific agents, and many heterospecific agents fall to the left from the Medium Peak. For example, from the humans' point of view dogs and Rhesus monkeys can be both placed on an arbitrary *similarity* scale on Mori's figure but it is questionable whether the same measure could be applied to *familiarity* with humans.

Importantly, the mental and behavioral mechanisms activated in the case of the Medium Peak and Maximum Peak are quite different, because biological agents possess a dedicated mechanism to detect individuals belonging to their own species but probably much less detailed discrimination is needed in the case of very different heterospecific species. Thus in the case of the Maximum Peak (distinguishing among conspecifics) the agent has to be more choosy and focused than when contacting heterospecific agents (Medium Peak). Biologically speaking this means that members of a species must avoid to get in close contact with non-conspecifics, e.g., hybrids, or closely related species because such mistakes can be fatal, especially with regard to reproduction (mating with hybrids reduces the fitness). This interpretation fits well with the depiction of the figure in which the Maximum Peak has a much narrower basis then the Medium Peak. Intuitively this suggests that conspecifics are evaluated more selectively then heterospecifics.

Strategies for Social Robotics

In the light of recent research on social recognition learning (e.g., Lewkowicz and Ghazanfar, 2009), Mori's hypothesis offers two options for developing optimal social robots. Social robots should achieve perfect humanness or humans (infants) should be exposed to social robots as soon as possible (before 1st year of age), which would probably decrease later uncanny feelings toward them. While the first option is quite unrealistic and counterintuitive (see also below), the second option may lead to serious problems because the exposure to such social robots during the sensitive period of infant development could lead to misguided learning about the human species, confusing species recognition and preferences at some later life (see debate initiated by Sharkey and Sharkey, 2010). Humans socialized as infants with robots (e.g., Tanaka et al., 2007) may prefer them later as social companions or sexual partners (Levy, 2010).

Androids and *Trans*-Humans

Let's assume for a moment that modern information technology continues to develop at least with the speed we have experienced in the last two decades. Then, there is little doubt that this technology will be able to surpass biologically evolved human traits in social robots, partly including new features not present in humans or any other naturally evolved agent. Just one example: gaze following is an automatic skill by which a bystander can perceive the focus of interest of the subject. Thus the head turn of one subject elicit head turn in others. A wide range of mammals and birds share this skill, which is based on visual perception and rapid processing of head orientation and movement. While such ability can be easily mimicked in an android robot, there is technically no restriction (even today) to equip a social robot with 360° vision capacities (just like in jumping spiders). This skill is certainly more advantageous for the robot but very likely it will change also how the robot behaves (no need to turn to follow the other's gaze) and also how it processes visual information. Thus it is not difficult to envisage that even very much human-like robots may at some point over-perform and transcend human performance.

Thus "perfect" human-like robots would represent only a relatively short and transient period in the technical development

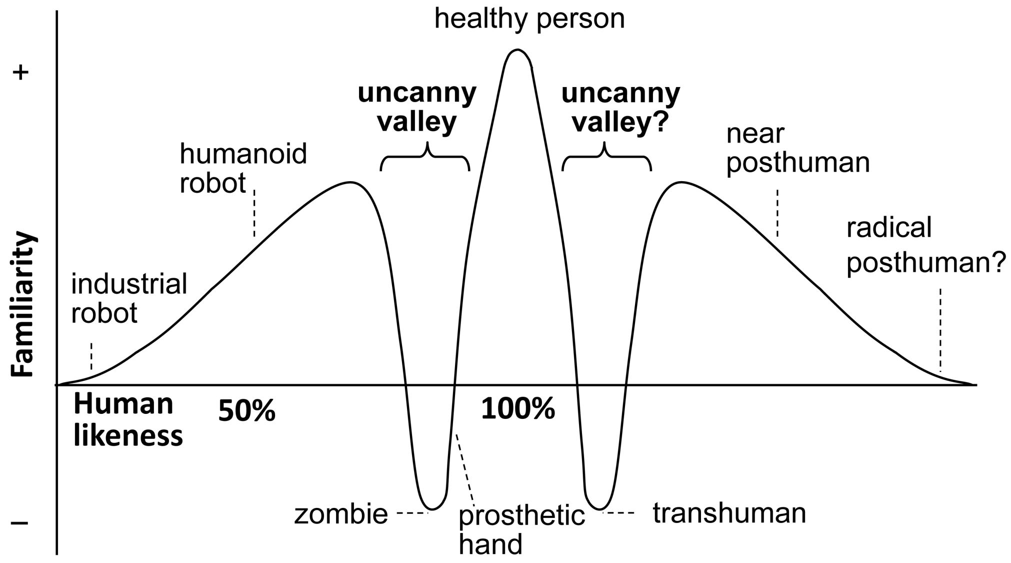


FIGURE 2 | An extended version of Mori's idea by Jamais Cascio (from http://www.openthefuture.com/2007/10/the_second_uncanny_valley.html). The second valley shows a similar effect related to robots evolved *from* perfect humanlike agents, as they become *less* similar to humans – following the path of *trans*-human and, eventually, *post*-human robots. The hill after the valley is when differentiation is strong enough to create a new category.

of social robots, which would be followed by robots to which some people may refer to as "*trans*-humans" during a transitional period and then moving away from human likeness, as "post-humans" (see Jamais Cascio unpublished source[1]). **Figure 2** shows this extended version of the original idea, indicating that technical development may not end at reaching maximum humanness and social robots may "fall" into a second uncanny valley. For today's social robotics this situation presents a real paradox.

In this sense, post-humans can be envisioned as "improved" humans but some of these agents may also fall into another uncanny valley to the right side of the "healthy person." For example, it has been shown that humans may have problems in predicting the behavior of robots that look like us but behave differently (Saygin et al., 2012).

Thus, Mori's hypothesis can be extended to a symmetrical landscape where there are two uncanny valleys on both sides of "perfect humanness" and humans may avoid both the lesser and the overly humanlike robots. Looking at this landscape it becomes clear that after the Maximum Peak has been reached there would be a narrow range of biological and artificial humans, in a largely extended world of heterospecific agents. Thus the notion of convergence in the direction of perfect humanness should be replaced by a more general view of divergence with regard to artificial systems, notwithstanding that such divergent processes may parallel a development of a specific class of agents which show very close resemblance to humans, and some of which may be able to evade the biological and cultural mechanisms of human social recognition system.

In summary, the paradox of the uncanny valley is that passing the valley successfully does not seem to solve the problem of social robotics because it is likely that robots will soon fall into another uncanny valley and/or in any case they will diverge from humanness. In addition, such *trans*-human robots that achieve or transcend human performance would very likely disrupt typical (natural) human social systems (Kubinyi et al., 2010).

ETHOLOGICAL APPROACH TO SOCIAL ROBOTICS

The ethological approach is centered on the function of behavior in relation to the specific environment in which the species evolved (Tinbergen, 1963). The application of this general concept to social robotics means that the robot should have a function, and in terms of embodiment, behavior, and problem solving (cognitive) abilities it should fit its specific environment. Instead of aiming to build more and more human-like robots and trying to "climb" the Maximum Peak, we may start robot construction by determining their function and their environment and design the must suited agent independently from its similarity to humans. Note that robot engineering can proceed by 'jumps' from one type of agent to a radically different one because it is not constrained by evolutionary continuity like biological agents. Moreover, humans may be not adequately 'designed' for a range of tasks thus uncritical copying of humans could turn out as wasted effort.

Solving the Paradox of the 'Uncanny Valley' Hypothesis

With regard to the uncanny valley metaphor this would mean that we go around the Maximum Peak and avoid the uncanny valley on the other side (**Figure 3**). Ethologically, such a robot is

[1] http://ieet.org/index.php/IEET/more/2083

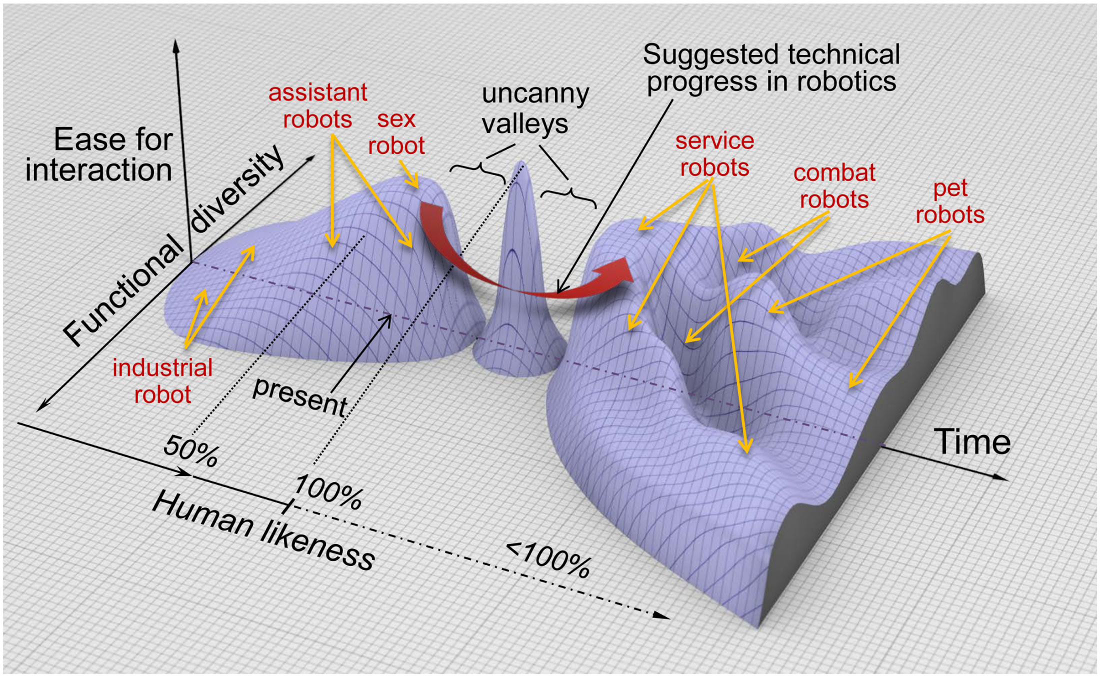


FIGURE 3 | An ethorobotic concept of emerging human-robot interaction. Based on Mori's idea, the present situation and the envisaged progress of social robotics are shown in a three-dimensional space to separate human-likeness, functionality and ease of interaction. After the peak and the second uncanny valley, robots are likely to evolve into a diversity of morphologies and behaviors that, depending on their functions, gradually move away from perfect human likeness. The wide curved arrow indicates the possible detour for social robotics by moving directly from the present state to less humanlike robots with diverse functionality retaining high-level capacity for social interaction with humans. The labels on the terrain are only for informative purposes and do not necessarily refer to actual existing robots.

occupying a different niche that is created by its specific function. This approach has several beneficial consequences: (1) robots can have their own evolution without interfering specifically with that of humans; (2) robots survive only if their niche exists and die out if they have not performed well to the expectation of humans; (3) no competition emerges between humans and robots.

This ethologically inspired functional perspective also shows that there is actually no need to 'climb' the uncanny valley.

Dogs Are Showing the Way

The viability of this approach is strongly supported by an analogous situation existing between humans and dogs for more then 18,800–32,100 years (e.g., Thalmann et al., 2013). The domestication of the dogs (from a wolf-like ancestor) resulted in several important morphological and behavior changes in dogs that enhanced the possibility of dog-human social interaction (Hare and Tomasello, 2005; Miklósi, 2014). Further steps in dog evolution led to dog breeds which occupy specific behavioral niches with regard to their specific function in collaborative interactions with humans (Miklósi, 2014). The large number of dogs sharing our life as companions, or working individuals (e.g., rescue dogs, dogs leading bind persons) shows the success of this evolutionary change. Thus with regard to the above points both dogs and humans retained their independent capacity to evolve, dogs have changed and can change if novel niches for interaction with humans emerge (e.g., Gácsi et al., 2013) and there is only limited competition between the two species.

Importantly, there are two critical features of the domestication process. Because of biological constrains (e.g., reproduction) dogs retained basic morphology and behavior of their ancestors but at the same time they acquired a level of social competence that allows them to be integrated into the human society (Miklósi and Topál, 2013). The history of dogs shows that humans are able to interact in very sophisticated ways with agents that are morphologically and behaviorally rather different, but show a specific human-like social competence. Dogs' social competence manifests in several cognitive domains including attachment, gestural and auditory inter-specific communication, inter-specific cooperation, ability to learn by observation (Topál et al., 2009b). Importantly, these components are supported by rather different mental mechanisms in dogs, and may show some important limitations when compared to analogous human skills (Lakatos et al., 2009; Topál et al., 2009a; Fugazza and Miklósi, 2014). Nevertheless, the connection and synergism that exists among these components lead to complex social competence in dogs, which allows them to perform efficiently in our societies.

Social Competence in Robots

Earlier we defined social competence as an individual's ability to generate social skills that conform to the expectations of others and the social rules of the group (Miklósi and Topál, 2013). Such complex level of interaction emerges if the individual wants to participate, has the means to participate, and is regarded by others as being able to participate in the life of the group (see also

Johnson, 2001). Thus the overarching goal for social robots is to gain some level of social competence that allows them to be integrated in the human group.

Several research teams in the field of social robotics have aimed to define the necessary and sufficient skills for such agents. Such approaches are problematic because they regard the components of human social competence as a starting point. For example, Fong et al. (2003) provide a long list of quite specific human skills that social robots should possess. Apart from the fact that at the moment there is no robust technical solution available for most of these social skills, the human model is less appropriate here because the biological foundations of the social interaction are obscured by the complexity of our social and cultural behaviors.

Bottom up Approach for Social Robotics

We suggest an alternative approach for the development of social robots using principles of dog-human interaction. First, the human-robot relationship should be represented as an inter-specific relationship rather than in an intra-specific one. As indicated above, such relationship is not unique among agents, and would most likely manifest some form of symbiosis in which humans experience positive fitness consequences (mutualism). Such functional approach to social robotics may also be helpful because it stresses that robots are constructed for a social process and not for a social state. Just like in the case of human-dog relationship, a social robot does not automatically become a social partner (e.g., companion) but it achieves this state of social affairs if it engages in the appropriate kind of social interactions with its partner (see Fujita, 2007; Miklósi and Gácsi, 2012). Any type of partnership is not an *a priori* attribute of the robot but actually an outcome of relevant social interactions between the agents. Accordingly, the social skills of the robot and the time devoted to the social interactions (by both parties) determine whether some type of partnership emerges or not.

We envisage that social robots should be able to show some basic social skills that are present in dogs. These may include, for example, attachment to humans (Topál et al., 2005), simple ways of communicative interaction (Miklósi et al., 2000; Gaunet and El Massioui, 2014), responsiveness to learning and training (Topál et al., 2009a) and being useful in some specific way (Naderi et al., 2001; Ostojic and Clayton, 2014). These commonalities between human and robot social competence are enough to form a basis for social interaction if there is time to gain experience mutually.

Importantly, there is no need to socialize humans to such social robots in any specific way or at any specific age and there is also no danger that humans develop unnatural preferences toward them.

PROMISES OF ETHOROBOTS

Social robotics aims to deliver various robots that serve human needs in modern societies but society may not accept many present day social robots because of their limited abilities which contradict their human-like appearance. We argue that ethorobotics offers a new approach by suggesting that social robots should be regarded as separate species that are highly adapted to their niche, and their similarity to humans both in terms of physical appearance and behavior in itself (without specific function) is irrelevant. This also includes that social robots can and should have human like features if this is required and optimal for their functions (e.g., simple verbal feedback, or human hand).

Simple insights from ethology can lead to a new generation of social robots. Ethorobots' basic social competence should ensure that humans eventually develop a social relation to them, which is sufficient for advantageous cooperation. We expect that these new ethorobots provide several advantages for the human society while avoiding possible dangers which may emerge if the present trend of technical development continues.

From the robots' perspective:

(1) Ethorobots are more efficient in their own niche because they are not constrained by expected similarity to humans.
(2) Considering the state of art in robotics, ethorobots are more acceptable social partners than imperfect androids.
(3) Ethorobots do not pose the problem of having a gender because they could be still regarded as part of the category of animals, where the actual gender is of secondary importance from the human point of view.

From the humans' perspective:

(1) Humans do not need to compete with ethorobots, instead, these robots would need to compete with each other (which of them is better at fulfilling a specific function).
(2) Humans can maintain control over ethorobots by controlling the nature of interaction, and whether they maintain or close down the actual niche for the robot.
(3) Humans have the necessary mental skills to learn to adjust their social behavior to robots with different embodiment and behavior if they show basic levels of social competence.

The validity and relevance of our claims and arguments can be tested by carrying out experiments that address the following questions. What is the minimally functioning social competence in ethorobots? Does it depend on embodiment and/or function? Would ethorobots be easier to accept by humans than humanoids, androids and any other type of human-like robots? What decides if embodiment and social behavior contradict or complement each other? Would humans develop different type of social relationships with ethorobots depending on their social competence? Under what condition would humans perceive an ethorobot as a living being? Experiments get started (e.g., Faragó et al., 2014; Lakatos et al., 2014; Takahashi et al., 2015; Gácsi et al., 2016; Paetzel et al., 2016; Tschöpe et al., 2017) but there is a long way to go.

CONCLUSION

Robotics has reached a stage when there is a demand for robots that can be considered as partners of humans. But without a clear theory built on biological (ecological and technological) knowledge, social robotics may fall in serious traps, will not be able to fulfill the societies' demand, and waste much money. We suggest robots that are developed on the basis of ethological concept: they (1) do not destroy natural human relationships, (2) do not get into a competitive situation with humans, (3) are able to develop a social partnership with humans, which matches the level of cooperation needed, and (4) are more acceptable for integration into our communities.

AUTHOR CONTRIBUTIONS

ÁM and MG: conception of the paper, drafting the work. PK and VM: conception of the paper, revising the draft. All: final approval of the version to be published, agreement to be accountable for all aspects of the work.

FUNDING

ÁM and MG were supported by the Hungarian Academy of Sciences (MTA 01 031), and PK was supported by Hungarian Research Fund (OTKA K100951).

ACKNOWLEDGMENT

We are grateful to the members of the LIREC consortium for enlightening discussions of the subject.

REFERENCES

Chen, H., Russell, R., Nakayama, K., and Livingstone, M. (2009). Crossing the "uncanny valley": adaptation to cartoon faces can influence perception human faces. *Perception* 39, 378–386. doi: 10.1068/p6492

Colgan, P. (1983). *Comparative Social Recognition*. New York, NY: Wiley.

Dautenhahn, K. (2007). Socially intelligent robots: dimensions of human–robot interaction. *Philos. Trans. R. Soc. B* 362, 679–704. doi: 10.1098/rstb.2006.2004

DiSalvo, C. F., Gemperle, F., Forlizzi, J., and Kiesler, S. (2002). "All robots are not created equal: the design and perception of humanoid robot heads," in *Proceedings of the 4th Conference On Designing Interactive Systems: Processes, Practices, Methods, And Techniques (Dis '02)* (New York, NY: ACM), 321–326. doi: 10.1145/778712.778756

Faragó, T., Miklósi, Á, Korcsok, B., Száraz, J., and Gácsi, M. (2014). Social behaviours in dog-owner interactions can serve as a model for designing social robots. *Interact. Stud.* 15, 143–172. doi: 10.1075/is.15.2.01far

Farroni, T., Johnson, M. H., Menon, E., Zulian, L., Faraguna, D., and Csibra, G. (2005). Newborns' preference for face-relevant stimuli: effect of contrast polarity. *Proc. Natl. Acad. Sci. U.S.A*. 102, 17245–17250. doi: 10.1073/pnas.0502205102

Ferrey, A. E., Burleigh, T. J., and Fenske, M. J. (2015). Stimulus-category competition, inhibition, and affective devaluation: a novel account of the uncanny valley. *Front. Psychol.* 6:249. doi: 10.3389/fpsyg.2015.00249

Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2003). A survey of socially interactive robots. *Robot. Autonom. Syst.* 42, 143–166. doi: 10.1016/S0921-8890(02)00372-X

Fugazza, C., and Miklósi, Á (2014). Deferred imitation and declarative memory in domestic dogs. *Anim. Cogn.* 17, 237–247. doi: 10.1007/s10071-013-0656-5

Fujita, M. (2007). How to make an autonomous robot as a partner with humans: design approach versus emergent approach Phil. *Trans. R. Soc. A* 2007, 21–47. doi: 10.1098/rsta.2006.1923

Gácsi, M., Kis, A., Faragó, T., Janiak, M., Muszyñski, R., and Miklósi, Á (2016). Humans attribute emotions to a robot that shows simple behavioural patterns borrowed from dog behaviour. *Comput. Hum. Behav.* 59, 411–419. doi: 10.1016/j.chb.2016.02.043

Gácsi, M., Szakadát, S., and Miklósi, Á (2013). Assistance dogs provide a useful behavioral model to enrich communicative skills of assistance robots. *Front. Psychol.* 4:971. doi: 10.3389/fpsyg.2013.00971

Gaunet, F., and El Massioui, F. (2014). Marked referential communicative behaviours, but no differentiation of the "knowledge state" of humans in untrained pet dogs versus 1-years old infants. *Anim. Cogn*. 17, 1137–1147. doi: 10.1007/s10071-014-0746-z

Gliga, T., and Csibra, G. (2007). Seeing the face through the eyes: a developmental perspective on face expertise. *Prog. Brain Res.* 164, 323–339. doi: 10.1016/S0079-6123(07)64018-7

Hare, B., and Tomasello, M. (2005). Human-like social skills in dogs? *Trends Cogn. Sci.* 9, 439–444.

Johnson, C. M. (2001). Distributed primate cognition: a review. *Anim. Cogn.* 4, 167–183. doi: 10.1007/s100710100077

Johnson, M. H., Dziurawiec, S., Ellis, H., and Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition* 40, 1–19. doi: 10.1016/0010-0277(91)90045-6

Kelly, D. J., Liu, S., Lee, K., Quinn, P. C., Pascalis, O., Slater, A. M., et al. (2009). Development of the other-race effect during infancy: Evidence toward universality? *J. Exp. Child Psychol.* 104, 105–114. doi: 10.1016/j.jecp.2009.01.006

Kubinyi, E., Pongrácz, P., and Miklósi, Á (2010). Can you kill a robot nanny? Ethological approach to the effect of robot caregivers on child development. *Interact. Stud.* 11, 214–219. doi: 10.1075/is.11.2.06kub

Lakatos, G., Gácsi, M., Konok, V., Brúder, I., Bereczky, B., Korondi, P., et al. (2014). Emotion attribution to a non-humanoid robot in different social situations. *PLoS ONE* 9:e114207. doi: 10.1371/journal.pone.0114207

Lakatos, G., Soproni, K., Dóka, A., and Miklósi, Á (2009). A comparative approach to dogs' (*Canis familiaris*) and human infants' comprehension of various forms of pointing gestures. *Anim. Cogn.* 12, 621–631. doi: 10.1007/s10071-009-0221-4

Levy, D. (2010). "Falling in love with a companion," in *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*, ed. Y. Wilks (Amsterdam: John Benjamins Publishing Company), 147–154.

Lewkowicz, D. J., and Ghazanfar, A. A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends Cogn. Sci.* 13, 470–478. doi: 10.1016/j.tics.2009.08.004

Lewkowicz, D. J., and Ghazanfar, A. A. (2012). The development of the uncanny valley in infants. *Dev. Psychobiol.* 54, 124–132. doi: 10.1002/dev.20583

MacDorman, K. F., Green, R. D., Ho, C. C., and Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Comput. Hum. Behav.* 25, 695–710. doi: 10.1016/j.chb.2008.12.026

MacDorman, K. F., and Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interact. Stud.* 7, 297–337. doi: 10.1075/is.7.3.03mac

MacDorman, K. F., and Minato, T. (2005). The uncanny valley. *Trans. Energy* 7, 33–35.

Mateo, J. (2004). Recognition systems and biological organisation: the perception component of socil recognition. *Ann. Zool. Fennici.* 41, 729–745.

Mathur, M. B., and Reichling, D. B. (2016). Navigating a social world with robot partners: a quantitative cartography of the uncanny valley. *Cognition* 146, 22–32. doi: 10.1016/j.cognition.2015.09.008

Miklósi, Á. (2014). *Dog Behaviour, Evolution and Cognition*, 2nd Edn. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199646661.001.0001

Miklósi, Á, and Gácsi, M. (2012). On the utilization of social animals as a model for social robotics. *Front. Psychol.* 3:75. doi: 10.3389/fpsyg.2012.00075

Miklósi, Á, Polgárdi, R., Topál, J., and Csányi, V. (2000). Intentional behaviour in dog-human communication: an experimental analysis of 'showing' behaviour in the dog. *Anim. Cogn.* 3, 159–166. doi: 10.1007/s100710000072

Miklósi, Á, and Topál, J. (2013). What does it take to become "best friends"? Evolutionary changes in canine social competence. *Trends Cogn. Sci.* 17, 287–294. doi: 10.1016/j.tics.2013.04.005

Moosa, M. M., and Minhaz Ud-Dean, S. M. (2010). Danger avoidance: an evolutionary explanation of uncanny valley. *Biol. Theory* 5, 12–14. doi: 10.1162/BIOT_a_00016

Mori, M. (1970). Bukimi no tani the uncanny valley. *Energy* 7, 33–35.

Naderi, S. Z., Csányi, V., Dóka, A., and Miklósi, Á (2001). Cooperative interactions between blind persons and their dog. *Appl. Anim. Behav. Sci.* 74, 59–80. doi: 10.1016/S0168-1591(01)00152-6

Ostojic, L., and Clayton, N. S. (2014). Behavioural coordination of dogs in a cooperative problem-solving task with a conspecific and a human partner. *Anim. Cogn.* 17, 445–459. doi: 10.1007/s10071-013-0676-1

Paetzel, M., Peters, C., Nyström, I., and Castellano, G. (2016). "Effects of multimodal cues on children's perception of uncanniness in a social robot," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI 2016)* (New York, NY: ACM), 297–301. doi: 10.1145/2993148.2993157

Pascalis, O., and Kelly, D. J. (2009). The origins of face processing in humans: phylogeny and ontogeny. *Perspect. Psychol. Sci.* 4, 200–209. doi: 10.1111/j.1745-6924.2009.01119.x

Sangrigoli, S., and De Schonen, S. (2004). Recognition of own-race and other-race faces by three-month-old infants. *J. Child Psychol. Psychiatry* 45, 1219–1227. doi: 10.1111/j.1469-7610.2004.00319.x

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., and Frith, C. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Soc. Cogn. Affect. Neurosci.* 7, 413–422. doi: 10.1093/scan/nsr025

Sharkey, N., and Sharkey, A. J. C. (2010). The crying shame of robot nannies: an ethical appraisal. *Interact. Stud.* 11, 161–190. doi: 10.1075/is.11.2.01sha

Steckenfinger, S. A., and Ghazanfar, A. A. (2009). Monkey visual behavior falls into the uncanny valley. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18362–18366. doi: 10.1073/pnas.0910063106

Takahashi, S., Gácsi, M., Korondi, P., Hashimoto, H., and Niitsuma M. (2015). "Leading a Person Using Ethologically Inspired Autonomous Robot Behavior," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts, Portland, Oregon, USA — March 02-05, 2015* (New York, NY: ACM), 87–88.

Tanaka, F., Cicourel, A., and Movellan, J. R. (2007). Socialization between toddlers and robots at an early childhood education center. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17954–17958. doi: 10.1073/pnas.0707769104

Thalmann, O., Shapiro, B., Cui, P., Schuenemann, V. J., Sawyer, S. K., Greenfield, D. L., et al. (2013). Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science* 342, 871–874. doi: 10.1126/science.1243650

Tinbergen, N. (1963). On aims and methods of ethology. *Z. Tierpsychol.* 20, 410–433. doi: 10.1111/j.1439-0310.1963.tb01161.x

Topál, J., Gácsi, M., Miklósi, Á, Virányi, Z., Kubinyi, E., and Csányi, V. (2005). Attachment to humans: a comparative study on hand-reared wolves and differently socialized dog puppies. *Anim. Behav.* 70, 1367–1375. doi: 10.1016/j.anbehav.2005.03.025

Topál, J., Gergely, G., Hegyi, Á, Csibra, G., and Miklósi, Á (2009a). Differential sensitivity to human communication in dogs, wolves, and human infants. *Science* 325, 1269–1271. doi: 10.1126/science.1176960

Topál, J., Miklósi, Á, Gácsi, M., Dóka, A., Pongrácz, P., Kubinyi, E., et al. (2009b). The dog as a model for understanding human social behavior. *Adv. Study Anim. Behav.* 39, 71–116. doi: 10.1016/S0065-3454(09)39003-8

Tschöpe, N., Reiser, J. E., and Oehl, M. (2017) "Exploring the uncanny valley effect in social robotics," in *Proceedings of the Companion of the 2017 ACM / IEEE International Conference on Human-Robot Interaction – HRI '17* (New York, NY: ACM), 307–308. doi: 10.1145/3029798.3038319

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Miklósi, Korondi, Matellán and Gácsi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

PERSPECTIVE
published: 13 June 2017
doi: 10.3389/frobt.2017.00024

Social Moments: A Perspective on Interaction for Social Robotics

Gautier Durantin†, Scott Heath† and Janet Wiles*

School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, QLD, Australia

During a social interaction, events that happen at different timescales can indicate social meanings. In order to socially engage with humans, robots will need to be able to comprehend and manipulate the social meanings that are associated with these events. We define social moments as events that occur within a social interaction and which can signify a pragmatic or semantic meaning. A challenge for social robots is recognizing social moments that occur on short timescales, which can be on the order of 10^2 ms. In this perspective, we propose that understanding the range and roles of social moments in a social interaction and implementing social micro-abilities—the abilities required to engage in a timely manner through social moments—is a key challenge for the field of human robot interaction (HRI) to enable effective social interactions and social robots. In particular, it is an open question how social moments can acquire their associated meanings. Practically, the implementation of these social micro-abilities presents engineering challenges for the fields of HRI and social robotics, including performing processing of sensors and using actuators to meet fast timescales. We present a key challenge of social moments as integration of social stimuli across multiple timescales and modalities. We present the neural basis for human comprehension of social moments and review current literature related to social moments and social micro-abilities. We discuss the requirements for social micro-abilities, how these abilities can enable more natural social robots, and how to address the engineering challenges associated with social moments.

Keywords: social moments, social robotics, timescales, responsiveness, social interaction, human–robot interaction

OPEN ACCESS

Edited by:
Bilge Mutlu,
University of Wisconsin-Madison,
United States

Reviewed by:
Tony Belpaeme,
Plymouth University,
United Kingdom
Fulvio Mastrogiovanni,
Università di Genova, Italy

****Correspondence:***
Gautier Durantin
g.durantin@uq.edu.au

†These authors have contributed equally to this work.

Specialty section:
This article was submitted to Humanoid Robotics, a section of the journal Frontiers in Robotics and AI

Received: *13 January 2017*
Accepted: *29 May 2017*
Published: *13 June 2017*

Citation:
Durantin G, Heath S and Wiles J (2017) Social Moments: A Perspective on Interaction for Social Robotics. Front. Robot. AI 4:24. doi: 10.3389/frobt.2017.00024

1. INTRODUCTION

For robots to develop social skills, they need to engage in interaction dynamics that convey social meanings. We term the events that occur within these interaction dynamics as *social moments*. Social interactions occur between social agents at multiple timescales. Conversations and other consciously considered social interactions typically span seconds, minutes, or longer. However, managing social exchanges also relies on the interpretation and manipulation of fast timescales (on the order of 10^2 ms) upon which the interaction is constructed.

For social robots, it is important to be able to understand the social significance of these fast interaction dynamics when participating in a social interaction. For a robot, social moments must be grounded both in the culture and personality of the interactant and also in the attributes of the interaction (environment), the roles of participants and robot, and the interaction task. The social skills required for a social robot include detecting, creating, and learning the meanings of social moments.

While the fields of human–robot interaction (HRI) and social robotics have investigated aspects of language (e.g., Kollar et al. (2010)), social interaction (e.g., Breazeal (2004)), and social motion (e.g., Hoffman and Ju (2014)), there is little or no investigation of social moments during these interactions and the short timescales associated. In this paper, we introduce and provide a definition for social moments; outline the related literature from psychology, neuroscience, and HRI; and present the practical challenges that need to be addressed to enable fast timescale social abilities.

2. DEFINITION OF SOCIAL MOMENTS

2.1. Definition

Social moments are brief events that occur during an interaction between two or more agents that have the potential to impact social dynamics.

Social moments have the potential to convey pragmatic and semantic information during an interaction that need not be deliberate or conscious actions. If a tutor gives a lecture to a group of students and briefly looks at one of the students, only to notice the student looking away out of boredom, the behavior of the tutor can be affected by this event, potentially for the rest of the lecture. The tutor might decide to focus more on this student or to make the lecture more interesting to capture attention. Alternatively, the tutor might instead decide to ignore the student and not to look that way again for the rest of the lecture. In practice, the set of events {*tutor looks at student; student looks away*}, although happening in a very short period of time, carries a social meaning that can potentially affect the rest of the interaction. We refer to such sets of events as *social moments*.

Social moments can evoke performative meanings (Condoravdi and Lauer, 2011) and can convey positive or negative valences. Communicating intentions is considered to be a prerequisite of the acquisition of language abilities in humans (Bates et al., 1975), with performative meanings conveyed in both verbal and non-verbal ways. For example, the speed of response in a social exchange can determine whether a speaker is being answered or a new comment generated (Newman et al., 2010; Maroni, 2011), and small delays can indicate the state of the responder as eager or reluctant to engage with an agent (Bögels et al., 2015) (see **Figure 1**). Larger delays in response can be considered to violate social norms and lead to the interpretation that an interactant is distracted, anti-social, or offensive. Additionally, movement patterns, particularly those in peripersonal space (within arms length), can convey social meanings (Ramenzoni et al., 2012). Different motions can indicate different intentions (Blythe et al., 1999), and changes in body posture can indicate different levels of engagement (Sanghvi et al., 2011).

In their foundational work in the field of social robotics, Dautenhahn et al. (2002) hypothesized that as robotic agents become socially embedded, they have to be able to observe, learn from, and adapt to their social environment, but they must also be able to influence it. Accordingly, the authors defined a set of micro-behaviors for hand-annotating the impact of robots on the humans surrounding them at the temporal resolution of 500 ms (Dautenhahn and Werry, 2002). In essence, Dautenhahn's hypothesis for social robots can be summarized as a need for them to have the ability to detect, interpret, and create social moments, but on a smaller timescale than that of micro-behaviors.

3. TIMESCALES FOR SOCIAL MOMENTS

Although a social moment may occur over any duration, social moments that occur over short timescales are difficult for a robot to detect, but can be just as significant. The importance of timing and short timescale responses is demonstrated in several social interaction studies. For tasks that require joint motor actions to achieve a goal, motor coordination becomes an inherent property of the interaction (Riley et al., 2011; Ramenzoni et al., 2012), requiring participants to be responsive to each other's actions. Synchronization can also occur between the motions of humans even when there is no joint motor coordination (Richardson et al., 2007), and deliberate synchronicity from an experimenter

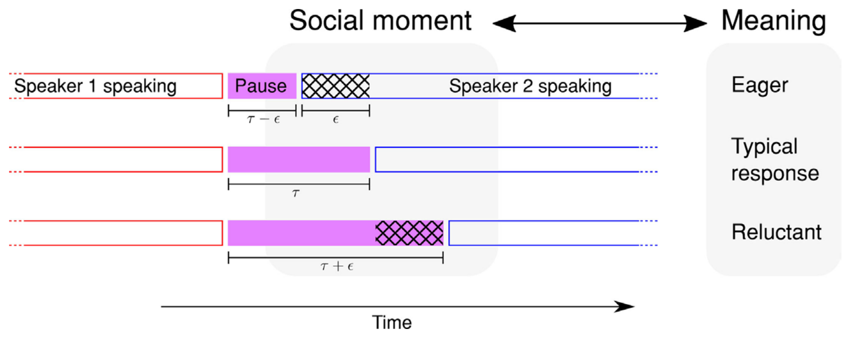


FIGURE 1 | A social moment that occurs between two conversing people from the changing length of a pause. There is a pause between one speaker's turn ending and the other speaker's turn starting. The pause length is controlled by speaker 2, and different lengths of the pause can indicate different pragmatic meanings to speaker 1. Given a typical pause length, τ, a pause that is shortened and has length $\tau - \epsilon$ might indicate eagerness to speak, while a pause that is lengthened has length $\tau + \epsilon$ might indicate reluctance to speak. Such an event can be considered a social moment, as the event can have an impact on the interaction dynamics. The social moment is caused by the divergence from the typical behavior by speaker 2. Speaker 1 can interpret the event and change his/her approach to the conversation. Alternatively, speaker 1 could miss or misinterpret the social moment and appear anti-social to speaker 2.

can increase affiliation from a participant (Hove and Risen, 2009). These studies suggest that synchronization is a key part of social normality.

Another modality of a social interaction that is greatly influenced by timing is that of conversation. Language processing occurs at many different timescales, and different events at different times can result in both changes to perception of the interaction and interlocutors. Consonant transitions can take less than 50 ms (O'Shaughnessy, 1974), and the difference in pronunciation can allow discrimination between a native and non-native speaker. For children, the inability to discriminate between two 43 ms tones is related to speech disorders (Tallal and Piercy, 1974). However, the timescales for each of these processes are below the level of even a single word, which is often the level that social robots work at (e.g., when using speech-to-text algorithms).

Additional challenges for social robots are found at the conversation turn-taking level. Turn-taking requires meeting timescales on the order of 10^2 ms, but this timescale varies depending on the culture (Stivers et al., 2009). For humans, the apparent time taken to process an utterance can be 500—700 ms, resulting in at least a 200 ms gap where an interlocutor would be expected to respond before having completely processed what was said previously (Stivers et al., 2009; Levinson and Torreira, 2015). Pauses between turns can take on further critical social meanings. Longer pauses have been demonstrated to be associated with non-preferred responses (Bögels et al., 2015), while pauses also reflect opportunities to enter conversation (Newman et al., 2010; Maroni, 2011) (see **Figure 1** for an example). For robots, the challenges for conversation turn-taking are to meet required social response times without complete knowledge of what was previously said and to understand the social meanings that are indicated through timing. Again, failing to meet these timing requirements can cause violations of social norms, which can carry a negative pragmatic meaning.

Altogether, a large set of events across multiple modalities can intertwine to create social meanings, given a specific context (Mondada, 2016). A good example is the McGurk effect (McGurk and MacDonald, 1976), where the visual and auditory channels are integrated during language perception. The timing of this multimodal integration is critical (Munhall et al., 1996), as delays of more than 180 ms across one modality can disrupt the perceived social moment. For a social robot, it is then critical to process social moments as spanning multiple timescales and modalities and as part of a broader context.

4. NEURAL BASIS FOR SOCIAL MOMENTS

Social moments can be constructed by societies through social norms, but they can also be grounded directly in human biology. The neural architecture required to detect social moments at multiple timescales and through different modalities is visible in the neural basis of human sensory processing. Processing of sensory (particularly visual) information in human cortical pathways shows differences linked to social information as early as 100 ms after presentation of the stimulus (Meeren et al., 2005; de Gelder, 2006). Other studies have shown that ultra-rapid categorization of visual stimuli is possible due to a likely parallel process in the visual cortex (Thorpe et al., 1996; van Rullen and Thorpe, 2001). As a consequence, the extraction of social information from visual stimuli can be done in less than 150 ms, thus being able to trigger a motor reaction to social stimuli in less than 300 ms (Thorpe et al., 1996).

Similarly, subcortical areas of the brain are believed to play a role in the processing of social stimuli (Morris et al., 1998; LeDoux, 2012), and their close link to motor structures suggests that they could play a role in the establishment of an automatic, reflex-like social behavior (de Gelder, 2006). A recent study on monkeys (Kuraoka et al., 2015) suggested that neurons were maximally informative of emotion and identity about 250 ms posterior to stimulation, which would be consistent with an extremely rapid reaction to strong social stimuli. In contrast, the maximum of information in cortical areas was observed after 300 to 1,000 ms, thus supporting a more elaborate but slower processing for emotion and an extremely rapid reaction to identity in the cortex.

Altogether, the organization of the processing of social stimuli inside human brains is consistent with a multi-scale approach to social moments. Accordingly, human behaviors are driven by the processing of social stimuli along two main scales: a very rapidly generated but very coarse representation of the social context, highly linked to motor structures and responsible for reflex-like social behavior, and a more elaborate but slower processing of social information. Although social robots do not have to implement social behavior in the same way, this organization emphasizes the different timescales and levels of processing that should be considered when designing robots.

5. PERSPECTIVES FOR SOCIAL ROBOTS

Awareness of social interactions is a critical component of social robots (de Graaf et al., 2015). In particular, the speed and timing of robot responses during social interactions have been identified as necessary prerequisites to engage users (Robins et al., 2005) and for the acceptance of the robot as a social interaction partner (Lee et al., 2006). Therefore, in a similar way to the mechanisms underlying the human interaction engine (Levinson, 2006), social robots need what we term *social micro-abilities*. Social micro-abilities are a set of abilities that augment social interactions and provide backchannels of communication (i.e., in parallel with symbolic communication). The following paragraphs describe the requirements of social micro-abilities.

5.1. Social Robots Require Sensitivity to Events at Very Short Timescales

Social moments can happen on the order of 10^2 ms. The latency and rates of robot sensors directly affect the detection of social moments, as a robot's response is constrained by hardware. For instance, robots that use standard web-cameras with latencies of 100–200 ms have restricted response times due to the time needed to capture an image during a control loop. A framerate of 30 Hz would lead to a maximum of 6 frames to capture an event of 200 ms length. Faster cameras exist, but their use comes at a cost of additional processing requirements. While there are currently constant advances in processing power, the increase comes at the cost of rethinking processing for each gain (Larus, 2009). Faster cameras have also been limited to physical applications and not considered for social interaction studies. For instance, motion capture cameras can achieve lower latencies and higher framerates and are often used for physical applications where timing is important (e.g., catching objects (Kim et al., 2014)). New event-based sensors such as the Dynamic Vision Sensor (DVS) (Lichtsteiner et al., 2008; Thorpe, 2012) can allow the detection of events at smaller timescales. There are also low-latency sensors for other modalities, such as audio, touch, and proprioception. The event-based silicon cochlea allows audio frequency data to be obtained with low latency and high frequency (Liu et al., 2010). The development of virtual whiskers (Schlegl et al., 2013) with high measurement frequency (1.25 kHz) allows rapid gathering of spatial information in the vicinity of the robot. The use of mechanical sensors such as torque sensors also allows detection of collision events in less than 15 ms (Haddadin et al., 2008). Despite the cost of the paradigm shift associated with using faster sensors, adapting such alternative approaches taken from industrial robotics or physical Human-Robot Interaction (pHRI) could augment the sensing capabilities of social robots.

5.2. Social Robots Require Rapid Response Capabilities

Achieving low latency responses to social moments requires processing events rapidly or maintaining a best guess representation of the social environment (Robins et al., 2005; Lee et al., 2006). When using control approaches that constantly update the knowledge of the environment and trigger motor actions from incomplete or uncertain knowledge, robots manage to catch flying objects whose time of flight does not exceed 700 ms (Kim et al., 2014) or react to attenuate collision forces in less than 100 ms (Haddadin et al., 2008), therefore matching or surpassing human capabilities. Such approaches have been restricted to the domain of physical robot dynamics and have not been considered in social robotics. Social robotics requires consideration of social dynamics, and models that take these dynamics into account have the potential to give social robots faster social responsiveness. In particular, a major challenge toward this goal is the difficulty of modeling the dynamics of social events compared to physical events. To achieve this goal, a better understanding of the dynamics of social interactions and social moments is required. Accordingly, using acquired knowledge of human reactions to their social environment can help predict the future occurrence of social events (e.g., Koppula and Saxena (2016)). Prediction of elements of the interaction in an anticipatory control system can also help reduce response times significantly (Huang and Mutlu, 2016) toward meeting fast response timescales. Notably, even if the amount of uncertainty contained in models of social interactions is greater than that in physical systems, it is important to note that interpretation of social interactions does not have to reach perfect accuracy. Rather, misinterpretation of social moments would contribute to give social robots human-like fallible traits that could improve their acceptance and long-term relationships with humans (Biswas and Murray, 2015). In addition to rapid processing of social events, responses at short timescales require fast actuators for robots. Social robots are often restricted to low-speed actuators to avoid harming humans—a trade-off that can restrict social ability.

5.3. Social Robots Require the Ability to Interpret Social Meanings and Maintain Social Awareness during Future Actions

The interaction with humans requires the robot to integrate social moments within their processing of the environment. This includes processing the social information and the context of the events together. As social moments are tied to social norms, detecting social moments will require the ability to predict typical behavior [e.g., motion from DVS, see Gibson et al. (2014)] and highlight deviations. Using neuromorphic processing of visual inputs, studies have achieved categorization of objects in less than 160 ms (Wang et al., 2017) or triggered robot responses in 4 cycles of a periodic event (Wiles et al., 2010). In addition, the interpretation of social moments requires the integration of information across multiple modalities, as multiple modalities contribute to the generation of social meaning (Mondada, 2016). Finally, as the social meanings interpreted from a social moment can affect long-term social interactions, robots need to be able to integrate social information obtained at short timescales into their cognitive architecture [see Lemaignan et al. (2017)] and memory to be able to represent the context against which future responses will occur.

6. CONCLUSION

In this paper, we have proposed the concept of a social moment and the social micro-abilities that are necessary for a robot to detect, interpret, predict, and respond to social moments. We believe that social micro-abilities are a fundamental requirement for social robots in order to gain acceptance in human societies. In particular, we believe that social robots need the ability to detect, predict, and respond to interaction dynamics across multiple modalities and at timescales as low as the order of 10^2 ms.

The implementation of social micro-abilities raises a set of compelling questions for the field of HRI and meets the definition of a new paradigm (Koschmann, 1996). We encourage social roboticists to consider social moments as part of their robot or architecture designs, and we anticipate new developments in robot hardware and cognitive architectures that will feature social micro-abilities. We intend to expand on the concepts of social moments and social micro-abilities and the required topics, tools, and methodologies in our future work.

From this perspective paper, we draw several current recommendations for social robotics:

(i) Although technology for implementing fast sensing and response already exists, use of such technology has been constrained to industrial robotics or pHRI. The interaction dynamics of social robotics needs to be considered just as temporally challenging as physical dynamics, with existing high speed sensors, actuators, and algorithms considered for social interactions.

(ii) Robots need to be able to respond quickly to maintain interaction dynamics even when there is missing or uncertain information about the social environment. For some interactions, there is a socially acceptable window in which a robot can respond, and no further incoming information or processing of information can compensate for responding too slowly.

(iii) Events on very short timescales and across multiple modalities can profoundly impact the current and future interactions, and therefore, it is essential for social robots to detect, predict, interpret, rapidly react to, and maintain long-term knowledge about social moments.

AUTHOR CONTRIBUTIONS

GD, SH, and JW all contributed to the content and writing of this paper.

ACKNOWLEDGMENTS

The ideas developed in this perspective have their roots in many conversations and collaborations across a spectrum of disciplines around the role of timing in learning and social dynamics. In particular, we thank colleagues from the Temporal Dynamics of Learning Center, the Kavli Institute for Brain and Mind at UCSD, and colleagues in the ARC Centre of Excellence for the Dynamics of Language.

FUNDING

This research is supported by an Asian Office of Aerospace Research and Development grant (FA2386-14-1-0017) and an Australian Research Council grant ARC Centres of Excellence (CE140100041).

REFERENCES

Bates, E., Camaioni, L., and Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill Palmer Q. Behav. Dev.* 21, 205–226.

Biswas, M., and Murray, J. (2015). “Towards an imperfect robot for long-term companionship: case studies using cognitive biases,” in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Hamburg: IEEE), 5978–5983.

Blythe, P. W., Todd, P. M., and Miller, G. F. (1999). “How motion reveals intention: categorizing social interactions,” in *Simple Heuristics That Make Us Smart* (New York, NY: Oxford University Press, Inc.), 257–285.

Bögels, S., Kendrick, K. H., and Levinson, S. C. (2015). Never say no … how the brain interprets the pregnant pause in conversation. *PLoS ONE* 10:e0145474. doi:10.1371/journal.pone.0145474

Breazeal, C. (2004). Social interactions in HRI: the robot view. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* 34, 181–186. doi:10.1109/tsmcc.2004.826268

Condoravdi, C., and Lauer, S. (2011). “Performative verbs and performative acts,” in *Proceedings of Sinn und Bedeutung*, Vol. 15 (Saarbrücken: Citeseer), 149–164.

Dautenhahn, K., Ogden, B., and Quick, T. (2002). From embodied to socially embedded agents – implications for interaction-aware robots. *Cogn. Syst. Res.* 3, 397–428. doi:10.1016/S1389-0417(02)00050-5

Dautenhahn, K., and Werry, I. (2002). “A quantitative technique for analysing robot-human interactions,” in *2002 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vol. 2 (Lausanne: IEEE), 1132–1138.

de Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nat. Rev. Neurosci.* 7, 242–249. doi:10.1038/nrn1872

de Graaf, M., Allouch, S. B., and van Dijk, J. (2015). “What makes robots social? A users perspective on characteristics for social human-robot interaction,” in *International Conference on Social Robotics* (Paris: Springer), 184–193.

Gibson, T. A., Henderson, J. A., and Wiles, J. (2014). “Predicting temporal sequences using an event-based spiking neural network incorporating learnable delays,” in *International Joint Conference on Neural Networks (IJCNN)* (Beijing: IEEE), 3213–3220.

Haddadin, S., Albu-Schaffer, A., De Luca, A., and Hirzinger, G. (2008). “Collision detection and reaction: a contribution to safe physical human-robot interaction,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (Nice: IEEE), 3356–3363.

Hoffman, G., and Ju, W. (2014). Designing robots with movement in mind. *J. Hum. Robot Interact.* 3, 89. doi:10.5898/jhri.3.1.hoffman

Hove, M. J., and Risen, J. L. (2009). Its all in the timing: interpersonal synchrony increases affiliation. *Soc. Cogn.* 27, 949–960. doi:10.1521/soco.2009.27.6.949

Huang, C.-M., and Mutlu, B. (2016). “Anticipatory robot control for efficient human-robot collaboration,” in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Christchurch: IEEE), 83–90.

Kim, S., Shukla, A., and Billard, A. (2014). Catching objects in flight. *IEEE Trans. Robot.* 30, 1049–1065. doi:10.1109/tro.2014.2316022

Kollar, T., Tellex, S., Roy, D., and Roy, N. (2010). “Toward understanding natural language directions,” in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, (Osaka: IEEE).

Koppula, H. S., and Saxena, A. (2016). Anticipating human activities using object affordances for reactive robotic response. *IEEE Trans. Pattern Anal. Mach. Intell.* 38, 14–29. doi:10.1109/TPAMI.2015.2430335

Koschmann, T. D. (1996). *CSCL, Theory and Practice of an Emerging Paradigm*. New York, NY: Routledge.

Kuraoka, K., Konoike, N., and Nakamura, K. (2015). Functional differences in face processing between the amygdala and ventrolateral prefrontal cortex in monkeys. *Neuroscience* 304, 71–80. doi:10.1016/j.neuroscience.2015.07.047

Larus, J. (2009). Spending Moore’s dividend. *Commun. ACM* 52, 62–69. doi:10.1145/1506409.1506425

LeDoux, J. (2012). Rethinking the emotional brain. *Neuron* 73, 653–676. doi:10.1016/j.neuron.2012.02.004

Lee, K. M., Peng, W., Jin, S.-A., and Yan, C. (2006). Can robots manifest personality? An empirical test of personality recognition, social responses, and social presence in human-robot interaction. *J. Commun.* 56, 754–772. doi:10.1111/j.1460-2466.2006.00318.x

Lemaignan, S., Warnier, M., Sisbot, E. A., Clodic, A., and Alami, R. (2017). Artificial cognition for social human–robot interaction: an implementation. *Artif. Intell.* 247, 45–69. doi:10.1016/j.artint.2016.07.002

Levinson, S. C. (2006). “On the human “interaction engine”,” in *Wenner-Gren Foundation for Anthropological Research, Symposium*, Vol. 134 (Duck, NC: Berg), 39–69.

Levinson, S. C., and Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Front. Psychol.* 6:731. doi:10.3389/fpsyg.2015.00731

Lichtsteiner, P., Posch, C., and Delbruck, T. (2008). A 128×128 120 dB 15 μs latency asynchronous temporal contrast vision sensor. *IEEE J. Solid-State Circuits* 43, 566–576. doi:10.1109/jssc.2007.914337

Liu, S.-C., van Schaik, A., Mincti, B. A., and Delbruck, T. (2010). “Event-based 64-channel binaural silicon cochlea with Q enhancement mechanisms,” in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems* (Paris: IEEE), 2027–2030.

Maroni, B. (2011). Pauses, gaps and wait time in classroom interaction in primary schools. *J. Pragmat.* 43, 2081–2093. doi:10.1016/j.pragma.2010.12.006

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi:10.1038/264746a0

Meeren, H. K., van Heijnsbergen, C. C., and de Gelder, B. (2005). Rapid perceptual integration of facial expression and emotional body language. *Proc. Natl. Acad. Sci. U.S.A* 102, 16518–16523. doi:10.1073/pnas.0507650102

Mondada, L. (2016). Challenges of multimodality: language and the body in social interaction. *J. Socioling.* 20, 336–366. doi:10.1111/josl.1_12177

Morris, J. S., Öhman, A., and Dolan, R. J. (1998). Conscious and unconscious emotional learning in the human amygdala. *Nature* 393, 467–470. doi:10.1038/30976

Munhall, K. G., Gribble, P., Sacco, L., and Ward, M. (1996). Temporal constraints on the McGurk effect. *Percept. Psychophys.* 58, 351–362. doi:10.3758/BF03206811

Newman, W., Button, G., and Cairns, P. (2010). Pauses in doctor–patient conversation during computer use: the design significance of their durations and accompanying topic changes. *Int. J. Hum. Comput. Stud.* 68, 398–409. doi:10.1016/j.ijhcs.2009.09.001

O'Shaughnessy, D. (1974). Consonant durations in clusters. *IEEE Trans. Acoust.* 22, 282–295. doi:10.1109/tassp.1974.1162588

Ramenzoni, V. C., Riley, M. A., Shockley, K., and Baker, A. A. (2012). Interpersonal and intrapersonal coordinative modes for joint and single task performance. *Hum. Mov. Sci.* 31, 1253–1267. doi:10.1016/j.humov.2011.12.004

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R., and Schmidt, R. (2007). Rocking together: dynamics of intentional and unintentional interpersonal coordination. *Hum. Mov. Sci.* 26, 867–891. doi:10.1016/j.humov.2007.07.002

Riley, M. A., Richardson, M. J., Shockley, K., and Ramenzoni, V. C. (2011). Interpersonal synergies. *Front. Psychol.* 2:38. doi:10.3389/fpsyg.2011.00038

Robins, B., Dautenhahn, K., Nehaniv, C. L., Mirza, N. A., François, D., and Olsson, L. (2005). "Sustaining interaction dynamics and engagement in dyadic child-robot interaction kinesics: lessons learnt from an exploratory study," in *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication*, Vol. 2005 (Nashville, TN: IEEE), 716–722.

Sanghvi, J., Castellano, G., Leite, I., Pereira, A., McOwan, P. W., and Paiva, A. (2011). "Automatic analysis of affective postures and body motion to detect engagement with a game companion," in *Proceedings of the 6th International Conference on Human-Robot Interaction – HRI'11* (Lausanne: Association for Computing Machinery (ACM)).

Schlegl, T., Kröger, T., Gaschler, A., Khatib, O., and Zangl, H. (2013). "Virtual whiskers – highly responsive robot collision avoidance," in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (Tokyo: IEEE), 5373–5379.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10587–10592. doi:10.1073/pnas.0903616106

Tallal, P., and Piercy, M. (1974). Developmental aphasia: rate of auditory processing and selective impairment of consonant perception. *Neuropsychologia* 12, 83–93. doi:10.1016/0028-3932(74)90030-X

Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature* 381, 520–522. doi:10.1038/381520a0

Thorpe, S. J. (2012). "Spike-based image processing: can we reproduce biological vision in hardware?" in *Computer Vision – ECCV 2012. Workshops and Demonstrations* (Florence: Springer Nature), 516–521.

van Rullen, R., and Thorpe, S. J. (2001). The time course of visual processing: from early perception to decision-making. *J. Cogn. Neurosci.* 13, 454–461. doi:10.1162/08989290152001880

Wang, R., Thakur, C. S., Hamilton, T. J., Tapson, J., and van Schaik, A. (2017). Neuromorphic hardware architecture using the neural engineering framework for pattern recognition. *IEEE Trans. Biomed. Circuits Syst.* 11, 574–584. doi:10.1109/TBCAS.2017.2666883

Wiles, J., Ball, D., Heath, S., Nolan, C., and Stratton, P. (2010). "Spike-time robotics: a rapid response circuit for a robot that seeks temporally varying stimuli," in *17th International Conference on Neural Information Processing (ICONIP)*, Sydney.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Durantin, Heath and Wiles. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: info@frontiersin.org | +41 21 510 17 00

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership