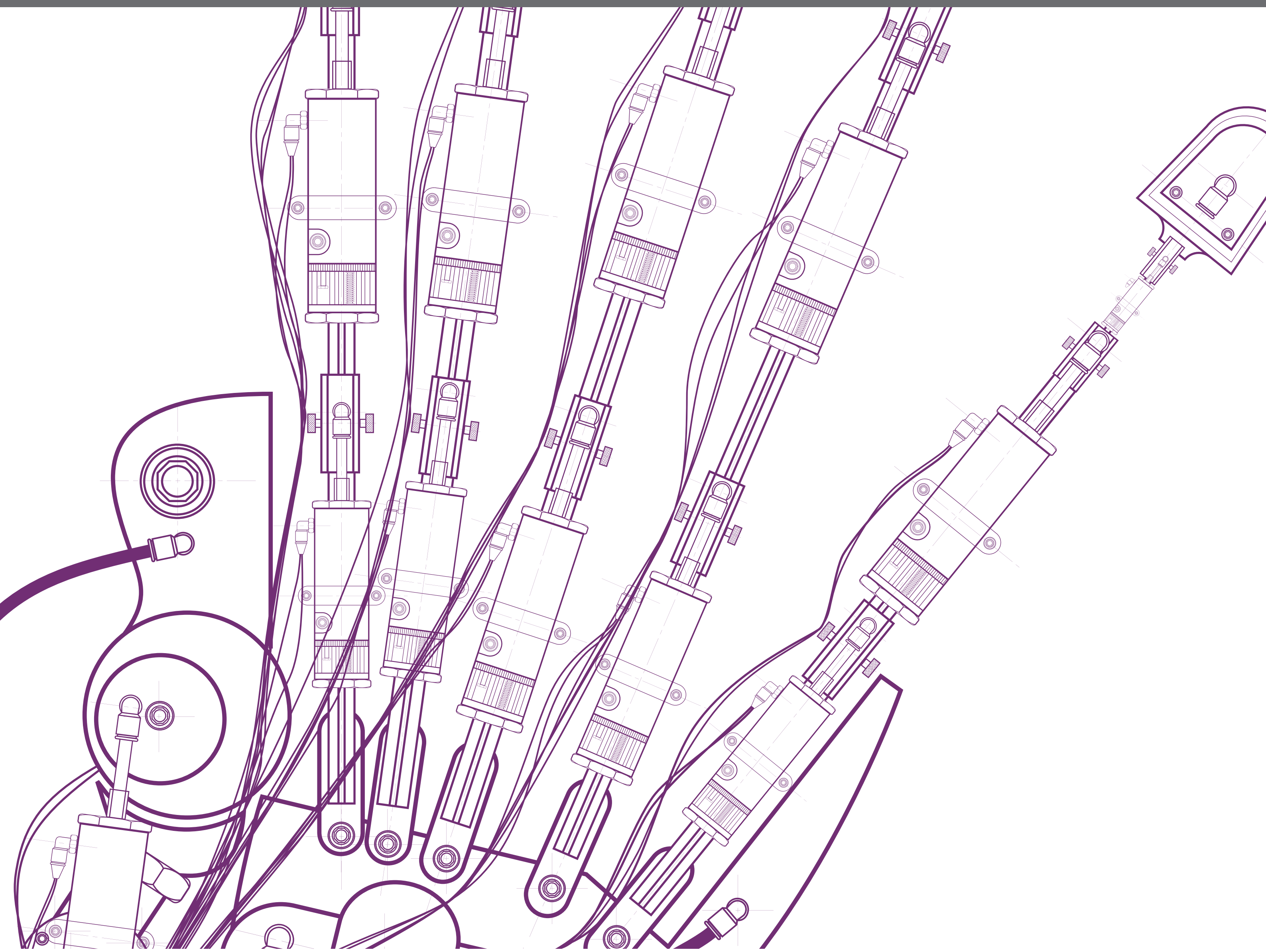
EDITORIAL: TOWARDS REAL WORLD IMPACTS: DESIGN, DEVELOPMENT, AND DEPLOYMENT OF SOCIAL ROBOTS IN THE WILD

EDITED BY: Chung Hyuk Park, Raquel Ros, Sonya S. Kwak, Chien-Ming Huang and Séverin Lemaignan
PUBLISHED IN: Frontiers in Robotics and AI



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88966-402-3
DOI 10.3389/978-2-88966-402-3

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

EDITORIAL: TOWARDS REAL WORLD IMPACTS: DESIGN, DEVELOPMENT, AND DEPLOYMENT OF SOCIAL ROBOTS IN THE WILD

Topic Editors:
Chung Hyuk Park, George Washington University, United States
Raquel Ros, Universitat Ramon Llull, Spain
Sonya S. Kwak, Korea Institute of Science and Technology (KIST), South Korea
Chien-Ming Huang, Johns Hopkins University, United States
Séverin Lemaignan, Bristol Robotics Laboratory, United Kingdom

Citation: Park, C. H., Ros, R., Kwak, S. S., Huang, C.-M., Lemaignan, S., eds. (2021). Editorial: Towards Real World Impacts: Design, Development, and Deployment of Social Robots in the Wild. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88966-402-3

Table of Contents

frontiers in Robotics and AI

EDITORIAL
published: 03 December 2020
doi: 10.3389/frobt.2020.600830

Editorial: Towards Real World Impacts: Design, Development, and Deployment of Social Robots in the Wild

Chung Hyuk Park[1], Raquel Ros[2], Sonya S. Kwak[3], Chien-Ming Huang[4] and Séverin Lemaignan[5]*

[1] Department of Biomedical Engineering, School of Engineering and Applied Science, The George Washington University, Washington, DC, United States, [2] Grup de Tecnologies Media, La Salle-Universitat Ramon Llull, Barcelona, Spain, [3] Center for Intelligent and Interactive Robotics, Korea Institute of Science and Technology (KIST), Seoul, South Korea, [4] Department of Computer Science, Johns Hopkins University, Baltimore, MD, United States, [5] Bristol Robotics Laboratory, Bristol, United Kingdom

Keywords: social robots in the wild, HRI theories, HRI computational modeling, longitudinal HRI, large-group HRI, HRI in healthcare and special education, emotions and ethics in HRI

Editorial on the Research Topic

Towards Real World Impacts: Design, Development, and Deployment of Social Robots in the Wild

OPEN ACCESS

Edited and reviewed by:
Alessandra Sciutti, Italian Institute of Technology (IIT), Italy

****Correspondence:***
Chung Hyuk Park
chpark@gwu.edu

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *31 August 2020*
Accepted: *11 November 2020*
Published: *03 December 2020*

Citation:
Park CH, Ros R, Kwak SS, Huang C-M and Lemaignan S (2020) Editorial: Towards Real World Impacts: Design, Development, and Deployment of Social Robots in the Wild. Front. Robot. AI 7:600830. doi: 10.3389/frobt.2020.600830

INTRODUCTION

Social robots have great potential to provide social, behavioral, emotional, and cognitive support to people with diverse characteristics and needs. Although still in its infancy, the field of social robotics has explored various aspects of human-robot interaction (HRI), such as multimodal communication and personalized interaction, and their applications in different domains including education and patient care. However, to evaluate the acceptance and efficacy of social robots and to understand their broader impacts in the real world, it is necessary to deploy these robots in the "wild" field for an extended period of time. Such deployment typically involves collaboration with different disciplines such as medicine, social psychology, clinical therapy, industrial design, public health, marketing, and education.

Thus, this Research Topic focuses on social robotics research with novel algorithms and computational modeling that have been or are being evaluated with intended users/consumers, patients, or individuals with special needs. A special focus has been given to results that arose from multidisciplinary studies in which the roles and impacts of social robots are evaluated in "real-world" settings, especially in collaboration between engineering, industrial design, clinical science, medicine, social psychology, marketing, and education.

RESEARCH TOPIC FORMATION

This Research Topic emerged from a discussion among young and active researchers in HRI—Dr. Park and Dr. Huang from the U.S.A., Dr. Ros from Spain, Dr. Kwak from South Korea, and Dr. Lemaignan from the United Kingdom—who have all realized the emergent and crucial needs to explore the topic further and get further input from the many researchers in HRI.

CONTENTS OF THE RESEARCH TOPIC

As anticipated, the submitted and accepted papers showed multi-disciplinary characteristics and combined themes from the proposed Research Topic outline. Out of the multiple mixed themes, however, we could identify the following four major important research themes and trends, where the social robots are taking active roles in the "wild" of the human-robot interaction frontiers.

HRI Theories and Computational Modeling

We have seen continuous efforts in applying theories from psychology with more sound experimental settings in the real world. One recent example is by Agrigoroaie et al. who have applied the regulatory focus theory in human-robot communication, with which a Tiago robot approached people with two communication methodologies based on either promotion type or prevention type, and evaluated its effectiveness in correlation with the regulatory focus types of the individuals ($N = 29$).

We have also found increased efforts in designing computational models for HRI in the wild, especially in the clinical domains. Clabaugh et al. propose a math tutoring system for children with Autism Spectrum Disorder (ASD) (aged 4–7), where Reinforcement Learning is used to personalize instruction and robot feedback. Javed et al. apply machine learning algorithms to derive personalized models for acquiring social engagement measures for children with ASD (aged 4–12).

Longitudinal and Large-Group HRI

For the past 10 years, longitudinal and larger-group studies in HRI have gained great interest from the community aiming at building systems that can engage and adapt to different users through time with the ambition of getting closer to a world where robots can truly be part of our daily lives. The educational setting is one of the typical areas where the application of social robots has been studied. A math tutoring system proposed by Clabaugh et al. has been evaluated with 17 children with ASD over month-long interventions at their homes. A peer-like social robot for language learning designed by Kory-Westlund and Breazeal has been developed, and the role of "rapport" has been investigated in their 2 months long study with 17 children. Besides education, social robots are used for elderly care; Van Maris et al. explored the longitudinal effects of older adults ($N = 17$) interacting with social robots.

HRI in public, especially interacting with a large group of people is another emerging Research Topic. Fraune et al. have studied the impact of group characteristics and norms in interacting with a robot in public settings and its influences on people's behavior changes.

HRI in Healthcare and Special Education

ASD is one of the areas in special education where socially assistive robot (SAR) systems have shown potential benefits, typically supporting social development, though not limited to. An example is a work proposed by Clabaugh et al., where they propose a tutoring system for space-themed mathematics problems addressed to young children with ASD while embedding social contexts in the learning environment. Another work by Javed et al. exhibits a robotic playmate with socio-emotional interventions with personalized engagement monitoring.

Management of diabetes is another healthcare-related domain where in-the-wild long-term interventions are key. Neerincx et al. present the results of a 4-years European project, where they deploy the Socio-Cognitive Engineering methodology to design and integrate a SAR, tested for several months with large groups of children.

Though evaluating SAR systems with end-users is vital, it is equally important to consider the views of other stakeholders to guide the design of such systems. As such, Alcorn et al. present an analysis of educators' views on the use of SAR systems in ASD suggesting guidelines to the HRI community, not only regarding the design of robotic systems but also proposing areas of research that should be further considered.

Emotions and Ethics

While socially assistive robots are uniquely characterized by their potential in participating in social-emotional interactions with people, these robots are currently not as emotionally capable as humans do. Indeed, emotive behaviors displayed by robots can be considered as emotional deception, possibly leading to broader ethical concerns. Van Maris et al. studied how deceptively emotive behaviors by a social robot might influence older adult's perceptions of the robot.

On the other hand, social robots' behaviors could possibly help mitigate human negative psychological states such as stress. Björling et al. explored how teens may interact with social robots in the school environment where teens might feel stressed. Little research on robotics and teens has been conducted to date: this article, relying on mixed-methods where the teens are in turn users, experimenters, and witnesses, offers a very novel glimpse into how socially assistive robots could support this population.

CONCLUSIONS

With the development of robotic technologies, it is imperative to develop social robots that support people in their daily lives. The papers published in this issue collectively show the recent and advanced application of theories from various academic fields such as healthcare, education, social psychology on social robots and HRI. More efforts are found to be focusing on child/adolescents and older adults who are in need of or can benefit from the company of robots. We believe that these collective endeavors will help to further extract knowledge regarding the nature of the interaction between humans and robots, which we hope to be utilized in building successful social robots "in the wild," expanding into wider areas of applications in diverse human domains (age, gender, culture, etc.). Specifically, topics presented by the papers including HRI Theories and Computational Modeling, Longitudinal and Larger-group HRI, HRI in Healthcare and Special Education, and Emotions and Ethics provide the readers with a vast array of applicable areas of social robots, focused on their effectiveness in various real-world settings. These studies will enlighten further research avenues regarding social robots that are closely connected to the users, which enrich their satisfaction with their lives by successfully fulfilling their needs. We believe that these collaborative efforts will contribute to the further development of the theory and knowledge regarding HRI, boosting the people's acceptance of social robots in their daily lives.

AUTHOR CONTRIBUTIONS

CP and C-MH conceived the idea for the Research Topic and recruited RR, SSK, and SL to form a team of guest editors for this Research Topic. Together, this team of guest editors have crafted the Research Topic abstract and the international call for papers. The editors have reviewed submitted papers within their expertise and accepted nine papers for this Research Topic. This editorial has been compiled through joint efforts of all five guest editors. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The guest associate editors of this Research Topic wish to express their gratitude to Dr. Bilge Mutlu, Specialty Chief Editor of frontiers in Robotics and AI: Human-Robot Interaction, for the opportunity to publish this Research Topic. We greatly appreciate the responsiveness of the Specialty Chief Editor, associate editors, and staff members on the timely support throughout the review and publication processes. We also express our sincere gratitude to the many reviewers for this Research Topic.

Conflict of Interest: 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.

frontiers
in Robotics and AI

ORIGINAL RESEARCH
published: 27 June 2019
doi: 10.3389/frobt.2019.00048

Check for updates

Human Group Presence, Group Characteristics, and Group Norms Affect Human-Robot Interaction in Naturalistic Settings

Marlena R. Fraune[1*], *Selma Šabanović*[2] *and Takayuki Kanda*[3]

[1] *Intergroup Human-Robot Interaction Laboratory, Department of Psychology, New Mexico State University, Las Cruces, NM, United States,* [2] *R-House, School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, United States,* [3] *HRI Laboratory, Graduate School of Informatics, Kyoto University, Kyoto, Japan*

As robots become more prevalent in public spaces, such as museums, malls, and schools, they are coming into increasing contact with groups of people, rather than just individuals. Groups, compared to individuals, can differ in robot acceptance based on the mere presence of a group, group characteristics such as entitativity (i.e., cohesiveness), and group social norms; however, group dynamics are seldom studied in relation to robots in naturalistic settings. To examine how these factors affect human-robot interaction, we observed 2,714 people in a Japanese mall receiving directions from the humanoid robot Robovie. Video and survey responses evaluating the interaction indicate that groups, especially entitative groups, interacted more often, for longer, and more positively with the robot than individuals. Participants also followed the social norms of the groups they were part of; participants who would not be expected to interact with the robot based on their individual characteristics were more likely to interact with it if other members of their group did. These results illustrate the importance of taking into account the presence of a group, group characteristics, and group norms when designing robots for successful interactions in naturalistic settings.

Keywords: human-robot interaction, social robotics, group dynamics, entitativity, group norms, gender

OPEN ACCESS

Edited by:
Chung Hyuk Park,
George Washington University,
United States

Reviewed by:
Wafa Johal,
École Polytechnique Fédérale de Lausanne, Switzerland
Sofia Serholt,
Chalmers University of Technology, Sweden

****Correspondence:***
Marlena R. Fraune
mfraune@nmsu.edu

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *04 April 2019*
Accepted: *05 June 2019*
Published: *27 June 2019*

Citation:
Fraune MR, Šabanović S and Kanda T (2019) Human Group Presence, Group Characteristics, and Group Norms Affect Human-Robot Interaction in Naturalistic Settings. Front. Robot. AI 6:48. doi: 10.3389/frobt.2019.00048

INTRODUCTION

Recent years have seen robots in wider use in public venues and organizational contexts, such as malls, airports, schools, and hospitals. Malls and stores around the world have deployed the humanoid robot Pepper to direct and guide people. In museums, the humanoid NAO guides guests through exhibits (Pitsch et al., 2013). The minimalistic robot Mugbot has been used in nursery schools to read to students and help implement classroom activities (Koike et al., 2009). Along with being used by individuals and families, the socially assistive robot Paro has also been placed in common areas of nursing institutions, where residents can interact with it when and how they like (Wada and Shibata, 2007; Chang et al., 2014).

When one person interacts with a public robot, they often draw other people to interact with it (Weiss et al., 2008; Fraune et al., 2015). Therefore, in public spaces such as those mentioned, robots interact with groups more often than with individual humans (Kanda et al., 2004; Sabanovic et al., 2006). However, such group interaction is seldom studied.

Group interaction introduces new factors that can affect and should be studied in HRI, such as characteristics of the human group (Sabanovic et al., 2006; Johansson and Skantze, 2015). Researchers have also begun to address solutions to group technical problems, such as tracking multiple people in group configurations (Holthaus et al., 2011; Taylor and Riek, 2016; Tseng et al., 2016) or switching attention between multiple people (Bennewitz et al., 2005). However, there are many open questions as to how varied group social dynamics in HRI should be addressed. For example, how should a robot respond when group members show it off to others (Sabanovic et al., 2006) or children debate over who the English-tutor robot liked more (Kanda et al., 2004)?—factors that do not arise in one-on-one interaction. For successful group interaction, it is critical to understand how social group dynamics change and affect perceptions of and behaviors toward robots.

Beyond the mere presence of groups, researchers have found that group behavior is affected both by relational characteristics of group members (e.g., family, coworker), and norms based on individual characteristics of group members (e.g., gender, age; Zanlungo et al., 2017). Although previous studies have also placed robots in public situations where they interact with multiple humans (Al Moubayed et al., 2012; Foster et al., 2012; Gomez et al., 2012; Johansson et al., 2013; Pereira et al., 2014), studies are only beginning to examine the group dynamics of the interaction (e.g., Admoni et al., 2013; Jung et al., 2015; Fraune et al., 2017; Alves-Oliveira et al., 2019).

In this research, we test how the presence of a group, group characteristics, and group norms relate to people's behavior toward a guide robot in a public mall (see **Figure 1**). We use behavioral and survey measures to answer our questions. Then, we discuss how these variables can be implemented in future robots to enhance interactions with humans.

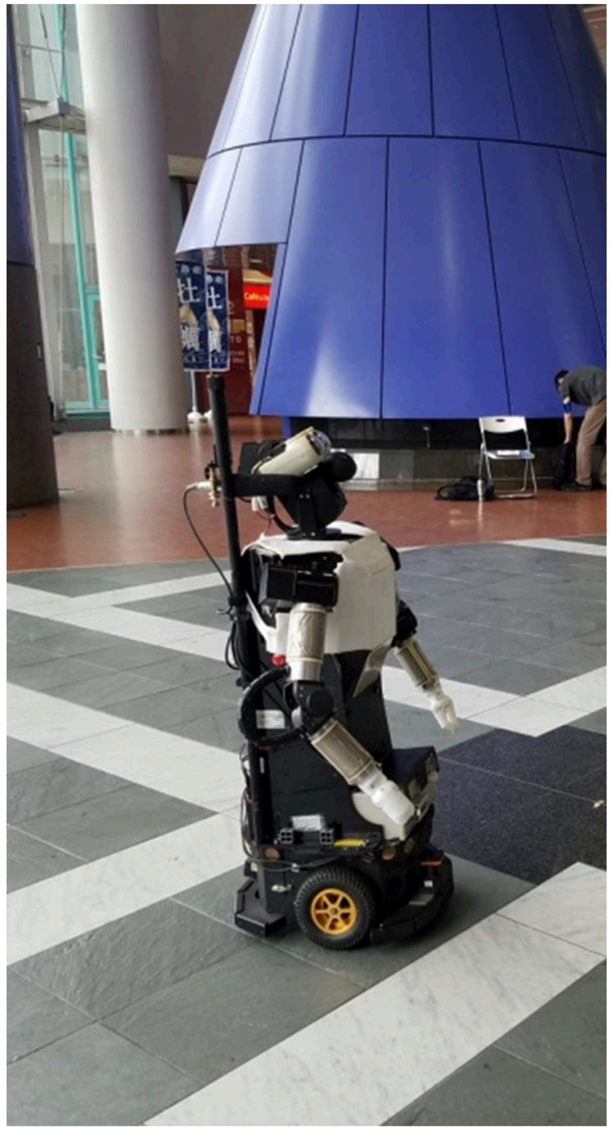


FIGURE 1 | The humanoid robot, Robovie, used in this study.

BACKGROUND

Groups Increase Following Group Goals

When people interact in groups, as opposed to individually, their motivation, and goals shift to be more similar to the group's goals (Reicher et al., 1995). For example, people in a group for a particular political ideology hold that group's ideology and goals more strongly when in that group or thinking of that group than when in a sewing or sports group. This even occurs when people are arbitrarily assigned to groups and had no interaction or common goals with them previously (i.e., minimal groups paradigm; Tajfel et al., 1971). In addition, the goals depend on the interaction context (Sherif, 1936; Gergen et al., 1973; Johnson and Downing, 1979; Fraune et al., 2019; e.g., competitive, collaborative).

When the group's goals are for competition, the discontinuity effect occurs—that is, groups are more aggressive and competitive than individuals (Sherif, 1936; Wildschut et al., 2002, 2003, 2007; Meier and Hinsz, 2004; Wildschut and Insko, 2007; Nawata and Yamaguchi, 2011; Insko et al., 2013). Conversely, when the group's goals are non-competitive or co-operative (e.g., groups must work together to accomplish a shared goal), separate groups co-operate with each other, potentially even combining groups or re-categorizing into one group to complete a shared goal (Sherif, 1936; Anastasio et al., 1997; Gaertner et al., 2000).

This pattern also occurs in HRI. In competitive situations, human groups competed more than individuals against robots (Chang et al., 2012; Fraune et al., 2019). In naturalistic environments, groups of unaccompanied children were more aggressive toward robots than individual children (Brscić et al., 2015). Conversely, interacting with robots in a learning context, groups of children were not more negative toward the robots (Leite et al., 2015). In naturalistic settings, groups of humans, rather than individuals, were more likely to stop to interact with robots (Weiss et al., 2008; Fraune et al., 2015).

In this study, a humanoid robot gave directions in a mall and actively sought to help participants. In this context, we expect that the typical participant goal while interacting with the robot will be to explore or seek guidance, and that people would be positive about the experience. The group should hold this goal more strongly than individuals. Therefore, we hypothesize:

H1. Presence of a group. Groups will be more positive and willing to interact with a robot than individuals, as measured in positive survey responses and duration of interaction with the robot.

Entitativity Increases Following Group Goals

Group entitativity magnifies certain characteristics of groups. Entitativity is defined as group cohesiveness, which includes group members sharing similar static traits (e.g., background, appearance, that are unlikely to change) and dynamic traits (e.g., goals, and outcomes, that may change frequently; Campbell, 1958). Entitativity increases group identification (Castano et al., 2003), and norms for behavior Lickel et al., 2000, which motivate members to achieve the goals of the group. The more entitative a group is, the more the group's behavior aligns with its goal (Gergen et al., 1973; Insko et al., 1988, 2013).

In competitive contexts, group entitativity magnifies the discontinuity effect, increasing competition, and aggression (Gaertner and Schopler, 1998; Insko et al., 2013) across cultures (Kumagai and Ohbuchi, 2009). In co-operative or positive contexts, group entitativity increases positivity (Gergen et al., 1973; Johnson and Downing, 1979). For example, when participants were inserted into cohesive groups, context cues of group harshness (e.g., KKK) influenced behavior to be more harsh than cues of group kindness (Gergen et al., 1973; Johnson and Downing, 1979; e.g., vs. nurse or hippie). In the case of a mall guidance robot, group entitativity should increase participants' exploratory and positive manner toward the robot. We hypothesize:

H2.Group characteristics. Entitative groups will be more positive and willing to interact with the robot than Diverse groups.

Although the effects of human group entitativity in naturally-occurring HRI have not yet been examined, robots can detect factors of entitativity. Robots have been capable of accurately predicting child friend groups based on proximity (Kanda et al., 2007) and adult groups based on how they interacted with each other (Giuliani et al., 2013). Therefore, if group membership and entitativity is useful in determining appropriate robot behavior, practitioners could develop algorithms to detect these patterns in naturalistic settings.

Group Type as a Natural Indicator of Group Entitativity

Group entitativity has been shown to vary across different types of groups. Entitativity is typically high in intimacy groups (e.g., family, friends), medium in task groups (e.g., coworkers), and low in loose associations (e.g., people standing in line; Lickel et al., 2000). Thus, in this study, we hypothesize that:

H2a. Family and friends will be more positive toward and interact more with the robot than colleagues because…

H2ai. Family and Friend groups will be more entitative than Other groups (e.g., coworkers) as measured in the survey.

A second likelihood is that intimacy groups will more commonly share a group goal of leisure and exploration in the mall, whereas task groups will share more group goals of getting a job done. We did not measure this, and future studies should examine goals specific to groups; however, in this study, we do have multiple measures of group entitativity.

Whereas, prior research investigated entitative groups that were created in the lab, this study examines naturally-occurring groups in a public space. This is critical because artificially-created lab groups may be loose associations or even task groups centered on a task, but are typically not intimacy groups. This is the first study to examine intimacy groups in intergroup HRI.

Gender as a Natural Indicator of Group Type in Japan

Literature in social psychology indicates that gender differences in behaviors and attitudes occur across cultures (Costa et al., 2001). Gender differences, while small on their own (e.g., explaining 5% of the variance aggression; Hyde, 1984), are increased (Hyde, 1984; Eagly and Wood, 1999) by differences in the social roles that people of each gender occupy (Rosario et al., 1988). In particular, in Japan (where we conduct this study), gender strongly correlates with occupation. That is, in Japan, males are more likely to be business people and managers, and females are more likely to be homemakers (Wright et al., 1995; Steinberg and Nakane, 2012) even in 2018, females made up only 43% of the labor force in Japan ("Labor Force., female (% of total labor force)," 2018). These gender differences are likely to account for differences in HRI in naturalistic settings. Therefore, although gender effects in HRI are mixed (Siino and Hinds, 2005; Schermerhorn et al., 2008; Siegel et al., 2009; Eyssel et al., 2012), we expect that in this situation, gender effects will be primarily driven by females being part of family and friend groups and therefore more entitative (argued above). Relatedly, past research indicates that females expected robots to be helpful in their personal lives (like family and friends), whereas males expected robots to be helpful in their work (Wang, 2014). This leads us to hypothesize:

H2b. Females will interact with the robot for longer and rate it more positively than males, especially coworkers. Because…

H2bi. Females will be in more family and friend groups (as measured by reported group type, video coded group type, and more children with them), and males in more coworker groups. Additionally, because…

H2bii. Females will rate their groups as more entitative (because they are more likely to be with family and friend groups).

Groups Influence People to Follow Group Norms

Group norms set expectations for typical behavior (Cialdini, 2007; Smith et al., 2007; Goldstein et al., 2008; Burger and Shelton, 2011). For example, how people respond to death and whether they take the stairs or elevator (Burger and Shelton, 2011) depends on ingroup norms (Goldstein et al., 2008) that are embedded in their culture or explicitly stated (Cialdini, 2007). In unfamiliar situations in which people do not otherwise know how to act, they are especially likely to follow norms they observe (Smith et al., 2007).

Within HRI, because robots are unfamiliar, people often look to norms when interacting with them (Lee et al., 2010; Chang et al., 2014). In the present study, seeing other people interact with the robot creates a norm of interacting and may induce more people to interact with it. Further, creating a norm that is more relevant to participants may increase following the norm (Goldstein et al., 2008). For example, participants with their family may be more likely to interact with the robot if they see their family member interact with it than if they see a group of friends interact. We hypothesize:

H3. Participants will be more likely to interact with the robot if they previously see someone from their group interact with the robot.

According to H2B, males will be least likely to interact with the robot. We hypothesize group norms will change this behavior as follows:

H3a. Males will interact more with the robot if others in their group interact than if others do not interact with it.

H3b. Males will interact more with the robot if there are more females in their group.

Overview

In the study, we examine how presence of groups, group characteristics, and group norms influence people's behavior toward robots. We do so by placing a humanoid robot in a public mall in Japan and using survey and video data to determine group and individual characteristics and the valence of participant responses toward the robot. The results will help practitioners account for and make adjustments to robot designs to enhance interaction depending on the context of interaction and characteristics of human groups involved in the interaction.

METHODS

This study was approved by the Institutional Review Board (IRB) at Indiana University (Protocol code 1606171019). Informed consent was not required for video recordings, as interactions occurred in a public setting, but a large sign was placed in the area indicating that it was being recorded. Verbal consent was obtained for survey participants.

Procedure

In a large open area in the Asian Trade Center (ATC) mall in Osaka, Japan, we placed the humanoid robot Robovie (**Figure 2**). Robovie remained stationary and waited for people to approach. When someone was ∼1 m in front of or to the side of Robovie, the robot detected and turned toward them.

The robot introduced itself and asked where visitors would like to go. It directed visitors to that location by turning, pointing, and describing the path they should take. Robovie allowed participants to ask for directions to multiple locations. As participants began to walk away, Robovie said, "Bye-bye." The entire time, Robovie switched between making eye contact with participants and looking toward where it was pointing to share joint attention in the direction participants should travel.

Between two and three researchers stood spread out on the outskirts of the open mall area and intercepted participants who had interacted with the robot long enough for Robovie to speak at least three sentences. If there were multiple people in a group, the researchers asked everyone to take a survey on a clipboard. The researchers requested that participants take the survey far from the robot so as to not interrupt anyone else's interactions with Robovie.

Video cameras recorded interactions with Robovie with a wide angle from above.

The experiment took place over the course of 21 days between October 2016 and February 2017. Each day, the robot was placed in the mall for ∼3 h at a time. For the purpose of this study, ∼20% of videos, randomly chosen, were coded for a total of approximately 12 h of analyzed video. Surveys were included in the analysis regardless of whether or not they overlapped with video that was coded.

The Humanoid Robot Robovie

Robovie (**Figure 2**) has two arms (each with four degrees of freedom [DOF]), a head (3 DOF), two eyes (each with 2 DOF), a mobile platform (two driving wheels and one free wheel), 10 tactile sensors, an omnidirectional vision sensor, two microphones to listen to human voices, and two laser rangefinders for detecting obstacles. The eyes have a pan-tilt mechanism with direct-drive motors, and they are used for stereo vision and gaze control.

Although Robovie can function fully autonomously, for the purposes of this study we controlled certain aspects of Robovie's behavior via a wireless local area network (IEEE 802.11a LAN), employing a Wizard of Oz (WoZ) technique. We employed this technique because in the loud mall environment, the robot has difficulty parsing human speech. In this study, Robovie's gaze and direction-giving behaviors were autonomous, but a Japanese researcher typed the locations participants wanted to go to into a computer so the robot could respond accordingly. From there, the robot autonomously directed them on how to get there.

Participants

Participants were people in the ATC Mall in Japan. Survey and video demographics are summarized in **Tables 1**, **2** below. In the video, participants were included if they were visible enough in the video frame that demographic information could be collected about them. Thus, a total of 2,714 participants were coded in the video (some of whom interacted with the robot) and 375 participants took the survey. Seventy-eight participants were both coded in the video and took the survey.

Measures

Survey

The survey took ∼2 min (see **Appendices A,B** for full survey in English and Japanese, respectively). Questions asked participants to report on:

FIGURE 2 | Robovie robot in the ATC Mall with Age-Diverse **(Left)** and Age-Similar **(Right)** groups.

TABLE 1 | Survey demographics.

	Demographics		Experience with…		Group type			
	Age (*M*)	Gender (%)	Computer experience (*M*)*	Robot experience (*M*)**	Alone (*N*)	Family (*N*)	Friend (*N*)	Coworker (*N*)
M/N/%	37.75	50.20 Female	2.15	3.00	12	219	60	43
Std. Deviation	14.22		1.53	5.17				

*Scale: from 1 (Novice) to 5 (Programmer). **Scale: 1 (None) to 5 (Build robots).

- Group Characteristics
  - Group size (number of members; free response)
  - Group type (family, friend, coworkers, acquaintances, alone)
  - Entitativity (i.e., cohesiveness with group, similarity of members in the group) was rated on a Likert scale from 1 (Strongly Disagree) to 7 (Strongly Agree) for human groups (Cronbach's $\alpha = 0.826$) and humans with the robot ($\alpha = 0.824$)
- Participant characteristics
  - Experience with computers (novice, comfortable for simple tasks, comfortable for moderately complex tasks, comfortable programming) and robots (seen none, in media/TV, interacted with, own one or more, work with/build)
  - Year born (free response)
  - Gender (free response)
- Ratings of Robot
  - Perceptions of the robot on a Likert scale was rated on a 9-point semantic differential Likert scale (from 0 to 8, i.e., negative-positive, scary-friendly, mean-kind, useless-useful, stupid-smart, non-social-social, machinelike-humanlike; Fraune et al., 2015)
  - Willingness to interact (enjoyment of the interaction, would interact with the robot in the future, would recommend for others to interact with the robot) was rated on a 9-point Likert scale from 0 (Definitely not) to 8 (Definitely yes).

Video

Video data were coded by four independent coders using ELAN. They were trained to code the data in the same way using 60 min of video from the study. They did this in 10 min segments, coding independently, and then meeting to discuss differences in codes. This process was continued until agreement was 80% or higher. Then video segments were assigned to the coders to work on independently based on the times they were available.

Approximately 20% of videos were coded by two of the four coders who did not know that someone else was coding the same video. Finally, we calculated percent agreement across all videos. We calculated percent agreement because other measures of interrater reliability were not feasible for the thousands of participants and coding method we used in this study. Due to the large number of participants, individual coders often gave different participant numbers to each participant, but generally coded them similarly (e.g., Coder 1 may have seen someone walk in from the left side of the camera first and labeled that person Participant 1, while Coder 2 saw a group walk in from the right and labeled them Participants 1–4, then labeled the participant from the left as Participant 5, but looking closely, Coder 1's Participant 1 and Coder 2's Participant 5 match up in terms of coded gender, time spent looking at the robot, etc.).

For deciding which codes to include in the data analysis for video sections that were coded by multiple coders, codes were taken from coders who had the highest interrater reliability across videos.

Videos were coded for the variables described in **Table 3**. Percent agreement in specific video segments ranged from 60 to 100%, but was averaged for overall percent agreement, included in **Table 3**.

RESULTS

Data were analyzed in SPSS 24. *P*-values of < 0.05 were considered statistically significant. When we conducted multiple

TABLE 2 | Video demographics.

	Age					Group type			
	Older adults	**Adult**	**Teenager**	**Child**	**Baby**	**Age-diverse**	**Age-similar**	**Business-dressed**	**Business-young**
Frequency	60	2,409	50	145	52	535	516	106	20
Percent	2.2	88.6	1.8	5.3	1.9	19.7	19	3.9	0.7

tests, we used Bonferroni corrections. To promote open science, deidentified video and survey data can be found at https://osf.io/ew8ta/?view_only=df9ef0f0919b48afb64a773ffa0251ac.

First, we examine if survey and video measures cohere. Next, we test corollary hypotheses (H2ai, H2bi, H2bii). Finally, we test the main hypotheses about main effects of group and gender (H1, H2a, H2b), linear regression of entitativity (H2), and effects of norms (H3, H3a, H3b).

Survey and Video Measures of Interaction and Gender Were Consistent. Measures of Group Type Differed

Interaction

Everyone who completed the survey had interacted with the robot as coded in the video. Of 2,714 video-coded participants who walked through the video, ∼15% interacted with the robot. Participants who interacted with the robot did so for an average of 47.8 s and a median of 37.3 s. The maximum time participants interacted with the robot was 289.3 s or 4 min and 49.3 s. A normality test revealed that skewness of interaction time was at an acceptable level (1.88).

Gender

Of 78 surveys (32 female, 46 male) that overlapped with coded video, one self-identified male was coded as female for a 98.7% accuracy rate for video coding. This false code was changed to match the survey data. In the video, there were 1,655 Males, 968 Females, and 91 Undefined either because they were too young to tell their gender or the coder did not have a good view of their face. In this study, Undefined were excluded from gender analyses because all but three Undefined were children.

Groups

Groups had an average (mean) of 2.69 members ($SD = 0.95$), with a minimum of two and maximum of six. In surveys, some participants indicated both Family and Friend as group types. For these, they were recoded as Family groups. Video coding and survey responses related to each other as described: Families and coworkers were typically coded as Age-Diverse groups (more than 50%) and sometimes as Age-Similar (about 25%). Coworkers were coded as Business-Dressed almost 50% of the time. On some occasions, loose acquaintances and strangers were coded as Age-Diverse or-Similar groups. Those who were alone were coded as Alone 100% of the time (**Table 4**).

Females Were in More Family and Friend Groups, and Males in More Coworker Groups (H2bi)

In this section, we included the variable 2 (Gender: Male, Female) for most of the tests.

Females Were in More Family and Friend, and Males in More Coworker Groups, According to Survey Responses

We ran a chi-squared test on 2 (Gender) × 4 (Group Type: Alone, Family, Friend, Coworker) reported in the survey. Loose Acquaintances, Stranger, and Other were excluded because they violated the expectation of having at least five counts per cell and they did not fit logically into the Alone, Family, Friend, or Coworker groups. Results indicated a significant relationship between Gender and Group Type ($X^2(3,\ N = 327) = 16.19$, $p = 0.001$) such that Males were more likely to be in groups with Coworkers (*Adjusted Standardized Residual*; $ASR = 3.5$) and slightly less likely to be in groups with Friends ($ASR = -1.9$) than Females (**Figure 3**).

Females Were Less Likely to be Alone or in Business-Dressed Groups Than Males, According to Video Responses

We ran a similar chi-squared test on video data: 2 (Gender) × 4 (Group Type: Alone, Age-Diverse, Age-Similar, Business-Dressed). We excluded Student Groups because it violated the expectation of having at least five counts per cell. Results indicated a significant interaction effect of Gender and Group Type ($X^2(3,\ N = 2{,}605) = 259.46$, $p < 0.001$) such that Males were much more likely to be Alone ($ASR = 11.9$) or in groups of Business-Dressed ($ASR = 6.5$) than in groups of Age-Diverse ($ASR = -10.7$) or Age-Similar ($ASR = -9.0$) compared to Females (**Figure 4**).

Females Were More Likely to Have Children in Their Group Than Males, Suggesting Family Ties

Gender proportion in groups, as coded on video, are reported in **Figure 5**. A 5 (Age: Older Adult, Adult, Teenager, Child, Baby) × 5 (Gender Proportion: All Male, Mostly Male, 50–50, Mostly Female, All Female) chi-squared test showed that Age interacted with Gender Proportion in Group ($X^2(16, N = 1{,}273) = 125.66$, $p < 0.001$). Primary-female groups contained more teenagers, children, and babies, but fewer adults than expected. Primary-male groups included more adults, teenagers, and fewer older

TABLE 3 | Description of video coding scheme and operational definitions of variables.

Variable	Percent agreement	Code	Operational definition
Interaction with robot	82.31%	No	
		Yes (duration)	Participants came close enough to Robovie that it would begin talking to them
Duration of interaction (seconds)	89.86%		Began when participants entered the interaction space with the robot and ending when participants left the space and had stopped looking at the robot.
Social gesture toward robots		No	
	76.94%	Yes	Participants made social gestures toward the robot (e.g., After interacting with the robot, participants turned back and waved).
Age	84.26%	Older adults	Looked to be approximately older than 55 years (e.g., moved more slowly).
		Adults	Looked to be between 18 and 55 years (e.g., tall, medium pace).
		Teenagers	Looked to be between 13 and 18 (e.g., short, often with adults or wearing school uniforms).
		Children	Looked younger than 13, but could walk on their own (e.g., were shorter, more likely to run, and often with adults or older adults).
		Babies	Could not walk on their own (e.g., in a stroller or carried the entire time on camera).
		*Three participants were excluded on analyses of age because their age is impossible to estimate given the camera angle.	
Gender	Matched survey 98.7% of the time	Male	Appeared to be male.
		Female	Appeared to be female.
		Undefined	Used when gender could not be determined, in particular with babies.
Gender Proportion	Calculated by computer based on the above codes.	All Male	
		Mostly Male	Between 50.1% male and 99.9% male
		50/50	
		Mostly Female	Between 50.1% female and 99.9% female
		All Female	
Group	75.75%	No	Participant was alone
		Yes (divided into categories below; **Figure 2**)	Participants walked in close formation with each other and spoke with each other during their time on the camera. The divisions below are mutually exclusive.
		Age-Diverse	Groups with diverse ages
		Age-Similar	Groups with similarly-aged participants
		Business-Dressed	Groups of adults dressed in suits or other business wear (Martin and Chaney, 2012). If participants fit the criteria for business-dressed, they were coded in this category rather than age-diverse or age-similar.
		Business-Young	Groups of children or young adults dressed in school uniform. If participants fit the criteria for business-young, they were coded in this category rather than age-diverse, age-similar, or business-dressed.
		*Data were excluded from Group Type analyses in 36 cases in which the group type could not be determined (e.g., because some group members were partially excluded from the camera frame resulting in being unable to tell what type of group it was).	
Group Size	76.36%	Participants who were in a group were coded to be in a group with a certain number of participants–one (alone) to six (the maximum group size coded).	
Seen Previous Interaction	Calculated by computer based on the above codes.	No	
		Yes	Participants were considered to have seen a previous interaction if another participant was coded as having interacted with the robot <10 s before the current participant. This included if the participant saw another person approach the robot and the participant approached the robot while the other person was still interacting with it.

(Continued)

TABLE 3 | Continued

Variable	Percent agreement	Code	Operational definition
Group of Previous Interaction	Calculated by computer based on the above codes.	No Previous	No interaction occurred 10 s before participants interacted
		Different Group	A group that was not the participant's group interacted within 10 s of the participant interacting
		Different Group and Individual	A group and individual, both not of the participant's group, interacted with the robot 10 s before the participant interacted
		Own Group	The participant's own group interaction with the robot 10 s or less before the participant appeared on screen

TABLE 4 | Comparing video coding with survey description of group type.

		Survey description of group					Total	
	Family	Friend	Coworker	Loose acquaintance	Strangers	Alone		
Video-coded group type	Age-diverse	31 (69%)	7 (58%)	2 (22%)	1 (50%)	1 (50%)	0	42 (55%)
	Age-similar	12 (27%)	3 (25%)	2 (22%)	1 (50%)	1 (50%)	0	19 (25%)
	Business-dressed	0	0	4 (44%)	0	0	0	4 (5%)
	Business-young	2 (4%)	2 (17%)	0	0	0	0	4 (5%)
	Alone	0	0	1 (11%)	0	0	7 (100%)	8 (10%)
Total		45	12	9	2	2	7	77

738 survey description of group.

adults and children than expected. 50/50 groups contained more elderly and adults and fewer teens and children (**Table 5**).

Families Were More Entitative Than Friends and Coworkers (H2ai), and for Friend Groups, Females Rated Groups as More Entitative Than Males (H2bii)

An ANOVA revealed a main effect of Group Type ($F(2, 298) = 3.45$, $p = 0.033$, $n_p^2 = 0.007$) such that Families rated themselves as more Cohesive than Friend groups did and as more similar ($F(2, 290) = 3.90$, $p = 0.021$, $n_p^2 = 0.026$) than Coworkers ($p = 0.017$). An interaction effect between Group Type and Gender ($F(2, 298) = 3.73$, $p = 0.025$, $n_p^2 = 0.024$) indicated that Female Friends rated their groups as more cohesive than Male Friends, but otherwise Males and Females rated their group similarly in relation to their cohesion.

Groups (H1), Especially Age-Diverse and Age-Similar (H2a), and Females (H2b) Were Typically More Positive Toward Interaction With the Robot

In this section, we included the variable 2 (Gender: Male, Female) for most of the tests. Categories were excluded in cases having 12 or fewer participants.

Females were more positive toward the robot than males according to *survey responses* (H2b), but there was no main effect of group (H1) or group type (H2a). We ran a series of 2 (Gender) × 3 (Family, Friend, Coworker) ANOVAs on survey responses. Main effects of Gender indicated that Females rated more enjoyment ($F(2, 306) = 5.59$, $p = 0.019$, $n_p^2 = 0.018$) and usefulness in the future ($F(2, 309) = 4.55$, $p = 0.030$, $n_p^2 = 0.015$) from the robot than males. No main effects of Group Type were found. An interaction effect revealed that Females rated the robot as less smart than Male when in Coworkers groups, but Females rated the robot as smarter than Males did in Family and Friend groups ($F(2, 306) = 4.55$, $p = 0.011$, $n_p^2 = 0.029$).

Groups (H1), especially Age-Diverse and Age-Similar (H2a) were *more likely to interact* with the robot than Alone participants. Females were more likely to interact than males (H2b). We ran a 2 (Gender) × 4 (Group Type: Alone, Age-Diverse, Age-Similar, Business-Dressed) × 2 (Interaction: Yes, No) Chi squared test on whether or not participants interacted with the robot. There were statistically significant differences ($X^2(3, N = 2{,}605) = 339.94$, $p < 0.001$; see **Table 6**). Groups of Age-Diverse or Age-Similar participants were more likely, and Alone participants were less likely, to interact than expected. When divided by Gender, the same was true of Females, but for Males, only Age-Diverse (not Age-Similar) were more likely, and those who were Alone were less likely, to interact than expected. Further, Females were more likely to interact than Males ($X^2(1, N = 2{,}605) = 36.68$, $p < 0.001$).

Groups (H1), especially Age-Diverse and Age-Similar (H2a) *interacted for longer* with the robot than Alone participants. Females were *not* more likely to interact for longer than males (H2b). Excluding participants who did not interact with the robot, a 2 (Gender) × 3 (Group Type: Alone, Age-Diverse, Age-Similar) ANOVA indicated a main effect of Group Type ($F(2, 343) = 4.85$, $p = 0.008$, $n_p^2 = 0.028$) such that participants who were Alone interacted

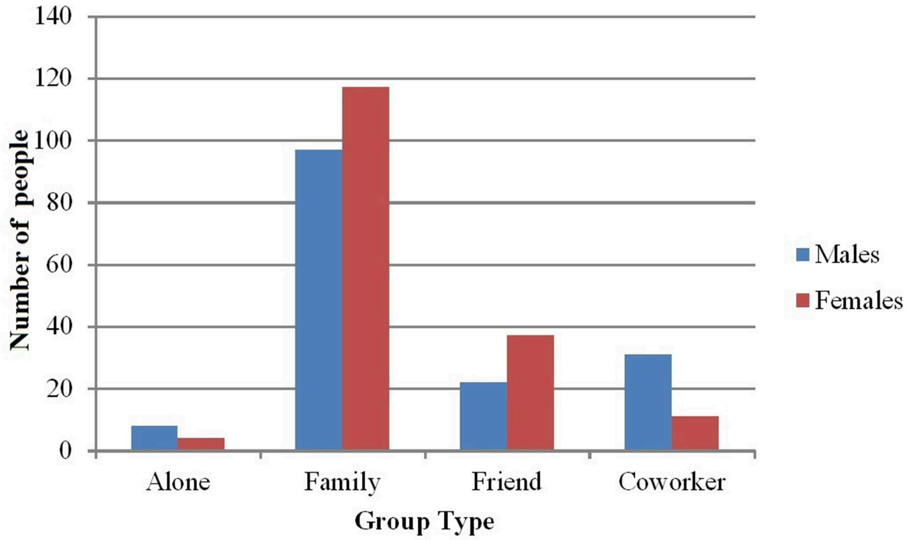


FIGURE 3 | Number of males and females in different group types according to survey ratings.

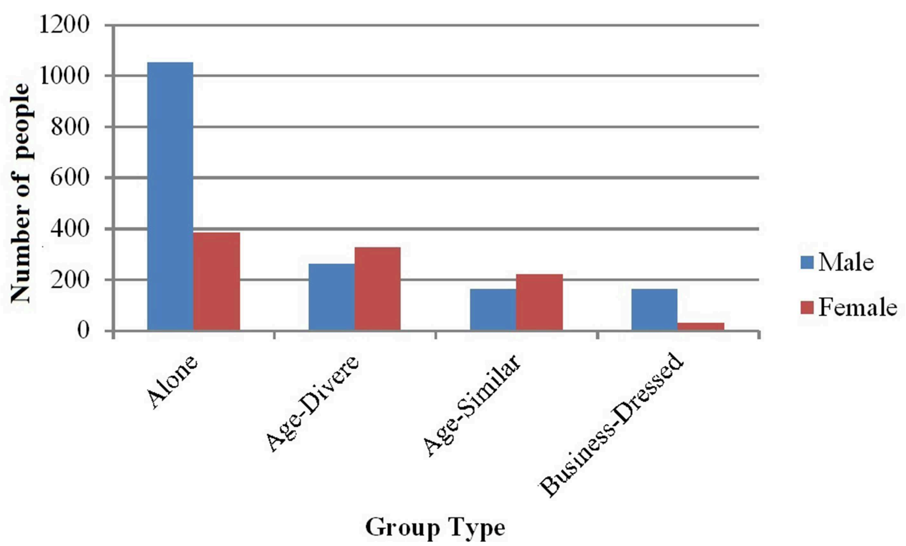


FIGURE 4 | Number of males and females in different group types as observed in the video.

for a shorter time with the robot than those in Age-Diverse ($p = 0.001$) and Age-Similar groups ($p = 0.013$). There were no differences among those who were in groups. No main effects of Gender or interaction effects were found (**Figure 6**).

Group (H1) and Group Type (H2a) did not affect *social gestures* toward the robot. Females were more likely to socially gesture than males (H2b). We ran a 2 (Gender) × 3 (Group Type: Alone, Age-Diverse, Age-Similar) chi-squared test on whether or not participants made social gestures toward the robot. There was a main effect of Gender ($X^2(1, N = 2{,}425) = 17.04$, $p < 0.001$) such that Females were more likely to make social gestures toward the robot than Males. No main effect of Group Type or interaction effects occurred (**Table 7**).

High-Entitative Groups Were Slightly More Positive Toward the Robot Than Low-Entitative Groups (H2)

Linear regression indicated no relation between perceived ingroup entitativity and ratings of the robot ($ps > 0.050$), except that participants who rated their group as highly entitative were more likely to recommend for others to use the robot ($F(1,98) = 6.91$, $p = 0.010$; $B = 0.323$; $R = 0.257$). The equation was:

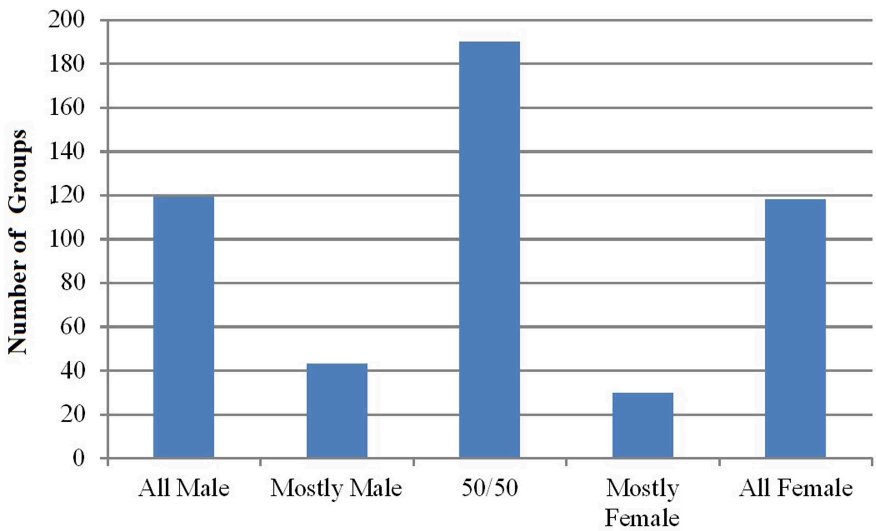


FIGURE 5 | Gender distribution in video-coded groups.

TABLE 5 | Percent of participants, divided by Gender Proportion and Age (ASR). Negative ASR values indicate that the percent was lower than expected.

	Older adult	Adult	Teenager	Child	Baby
All male	1.01 (−2.1)	84.80 (3.5)	6.76 (2.8)	4.39 (−4.3)	3.04 (−1.3)
Mostly male	0.72 (−1.6)	69.78 (−2.3)	2.16 (−1.2)	24.46 (5.2)	2.88 (−0.9)
50/50	4.41 (2.4)	81.90 (2.7)	0.46 (−4.6)	8.58 (−2.2)	4.64 (0.3)
Mostly female	1.00 (−1.1)	68.00 (−2.4)	0.00 (−2.1)	28.00 (5.5)	3.00 (−0.7)
All female	3.91 (1.3)	70.68 (−3.3)	8.47 (4.6)	10.42 (−0.6)	6.51 (2.1)

TABLE 6 | Percent of participants who interacted with the robot (ASR), divided by gender and group type.

	Alone	Age-diverse	Age-similar	Business-dressed	Total
Male	3.42 (−13.3)	35.88 (13.9)	20.12 (3.9)	12.12 (0.4)	11.14 (−6.1)
Female	4.95 (−9.4)	39.33 (11.0)	17.27 (−1.1)	13.33 (−0.9)	19.75 (6.1)
Total	3.83 (−16.9)	37.80 (18.5)	18.49 (2.5)	12.31 (−0.8)	

Likeliness to recommend = 2.89 + 0.323 * (Human group entitativity).

Norms of Interaction Affected Participants (H3), Especially Males (H3a, H3b)

Survey responses related to gender ratio were too few ($N = 69$), so for tests of norms we examined only behavior.

Participants (H3), especially males (H3a) were more likely to *interact* with the robot if others in their group previously interacted.. Overall, a 2 (Gender: Male, Female) × 4 (Seen Previous Interaction: No Previous, Different Group, Different Group and Individual, Own Group) × 2 (Interaction: Yes, No) chi-squared test revealed that participants were more likely to interact than expected with the robot if they saw Own Group ($ASR = 27.3$) interact or currently interacting with the robot, and less likely to interact than expected if they saw No Previous ($ASR = -10.0$) or Different Group ($ASR = -4.1$) interacting ($X^2(3, N = 2{,}466) = 745.96$, $p < 0.001$; **Figure 7**), regardless of gender. There was an interaction effect between Gender and Seen Previous ($X^2(1, N = 2{,}466) = 36.84$, $p < 0.001$). Females were more likely to interact than expected compared to Males when they saw No Previous ($ASR = 2.2$, $p = 0.031$) interaction, Different Group ($ASR = 2.7$, $p = 0.007$), or Own Group, and Individual ($ASR = 2.2$, $p = 0.025$). However, there was no significant difference in Male and Female interaction with the robot when Own Group had previously interacted ($ASR = -1.0$, $p = 0.315$).

Participants (H3) were more likely to *socially gesture* toward the robot if others in their group previously interacted. Gender did not affect the relationship (H3a). When the same test was run on Gesture (Yes, No), the Expected Count for Own Group Male Gesture was too low ($N = 2.4$). However, the effects were similar across gender. Therefore, Gender was collapsed and a 4 (Seen Previous Interaction: No Previous, Group Not, Different Group and Individual, Own Group) × 2 (Gesture: Yes, No) chi-squared test was run ($X^2(3, N = 2{,}466) = 133.31$, $p < 0.001$) indicating that participants gestured at the robot less often than expected when they saw No Previous ($ASR = -3.9$) and more often when they saw Own Group ($ASR = 11.5$) interacting. Because Different Group and Different Group and Individual showed interaction in similar direction, they were combined to find that participants who saw those not in their group interacting were also less likely to make a social gesture toward the robot ($ASR = -2.5$; $X^2(3, N = 2{,}466) = 140.84$, $p < 0.001$).

Males (H3a) *interacted for a longer duration* with the robot if others in their group had previously interacted. The effect was not clear for participants overall (H3). A 2 (Gender: Male,

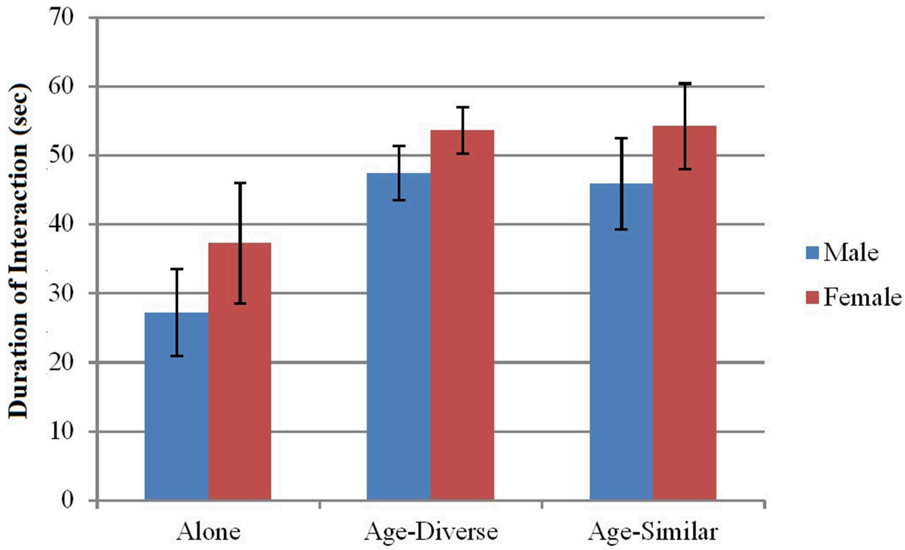


FIGURE 6 | Group type and gender on duration of interaction. Error bars indicate standard error.

TABLE 7 | Percent of participants who socially gestured toward the robot as divided by Gender and Group Type [Adjusted Standardized Residual (ARS)].

	Alone	Age-diverse	Age-similar	Total
Male	0.57 (−10.3)	12.98 (8.5)	10.37 (4.6)	3.86
Female	1.30 (−6.3)	16.46 (7.1)	6.82 (−0.7)	7.94
Total	0.77 (−12.3)	14.92 (11.7)	8.33 (2.7)	5.44

Female) × 4 (Seen Previous Interaction: No previous, Different Group, Different Group and Individual, Own Group) ANOVA was run on interaction time for participants who interacted with the robot. There was an interaction effect between Gender and Seen Previous Interaction ($F(3,359) = 2.69$, $p = 0.046$, $n_p^2 = 0.022$) such that Males interacted for less time than Females unless they saw their Own Group interacting (**Figure 8**).

Males *interacted more* with and made more *social gestures* toward the robot if there were more females in their groups (H3b). We ran logistic regressions for the effects of Gender and Gender Ratio on behavior relating to the robot ($N = 1,270$). On interaction (Y/N), (Nagelkerke's $R^2 = 0.015$, $X^2(5) = 61.89$, $p < 0.001$), only the effect of Gender Ratio was significant (Wald(1) $= 9.23$, $p = 0.002$, $B = 0.749$) such that the greater percentage female in the group, the more likely participants were to interact with the robot. The same was true of gesture (Nagelkerke's $R^2 = 0.015$, $X^2(5) = 16.67$, $p = 0.011$) (Wald(1) $= 4.98$, $p = 0.026$, $B = 0.818$). When Gender Ratio was used as a covariate in the 2 (Gender) × 3 (Group Type: Age-Diverse, Age-Similar, Business-Dressed), no effects were found.

DISCUSSION

In this study, participants interacted with a humanoid robot in a mall setting. Participants who interacted with the robot were given the opportunity to complete a survey. Behavior of those who passed through the area was examined by independent video coders with high accuracy. The main findings of the study were twofold: Groups (H1, H2a, H2b), especially entitative groups (H2, H2ai, H2bii) increased following of group goals, and group norms of interaction increased actual interaction (H3). These findings are described in more depth below.

H1 and H2. Groups, Especially Entitative Groups, Enjoyed the Robot

Groups, especially entitative groups, as compared to individuals, (1) interacted more and for longer with, (2) behaved more socially toward, and (3) were more positive toward a robot in the mall. These results were shown across survey and behavioral measures, supporting H1, H2a, and H2b. These findings indicate that in a naturalistic setting, groups had more positive interaction with the robot than individuals—at least in the positive and friendly mall environment. They also contribute the novel information that the entitativity of pre-existing participant groups can positively affect subjective and behavioral responses toward a robot in a naturalistic setting. These results are useful for HRI because they show that groups do not necessarily turn people against a robot; group effects can also work in favor of robots in situations in which the group members support each other to explore the environment and a robot.

Family, friend, and female groups were highly entitative (H2ai, H2bii) and responded more positively toward the robot than others. These results support research indicating that different types of groups naturally have different levels of entitativity (Lickel et al., 2001). An alternative reason for increased interaction when in groups is because participants in groups could watch their families or friends interact with the robot, which accounted for some of the time and close interaction distance with the robot. However, this would not necessarily account for increased positivity of entitative groups.

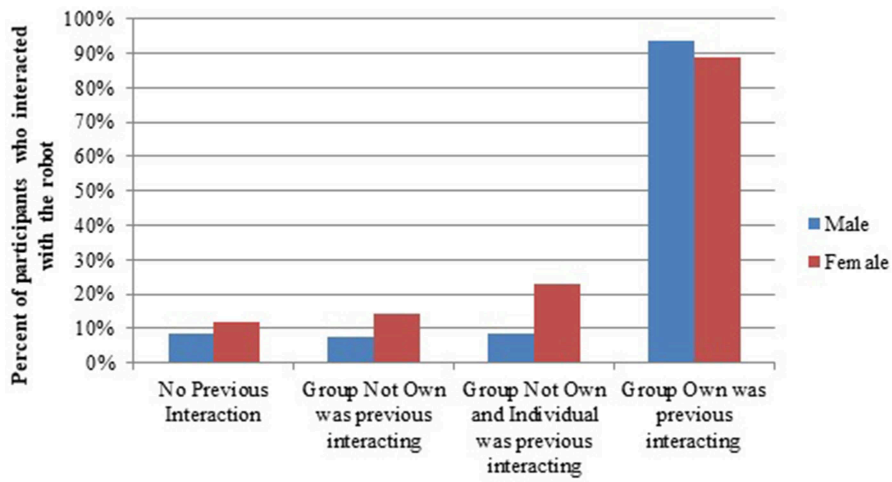


FIGURE 7 | Interaction with the robot based on previous exposure to interactions with it.

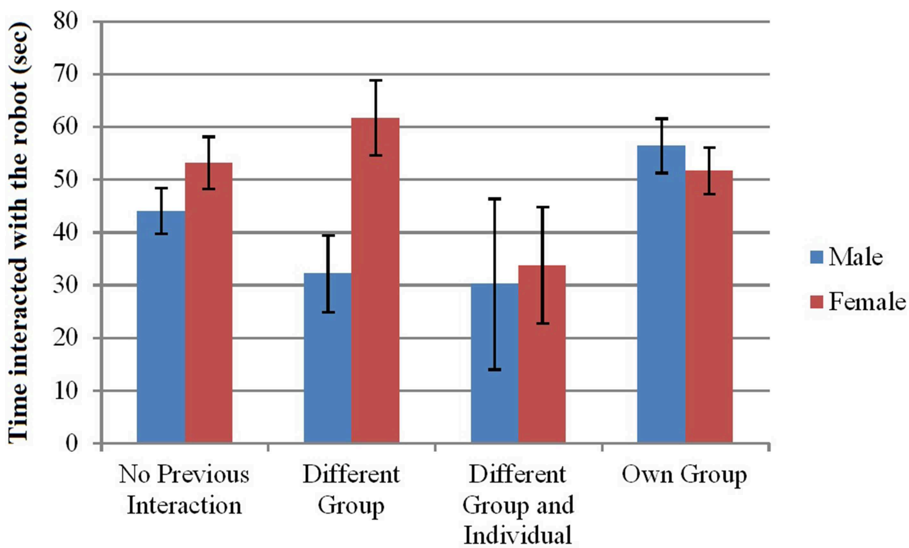


FIGURE 8 | Effect of gender and group type on time interacted with the robot. Error bars indicate standard error.

Robot designers may wish to target entitative groups of users to increase interaction time with their robots. A robot could initially target those in a group who are more likely to want to interact immediately with the robot (e.g., women in this study), and once the group members are present, could switch strategies to appeal to other group members. Indeed, once participants' groups were interacting with the robot, the participants themselves were more likely to interact with it.

H3. Social Norms Affected Interaction

Participants followed group norms of interacting with the robot (H3). That is, when others in participants' group interacted with the robot, participants who did not typically interact were more likely to (H3a). Because males interacted less than females, as in previous observational research of naturalistic interactions with the robot in public spaces (Chang et al., 2014), this was especially pronounced for male participants (H3b). Further, once one's own group was interacting with the robot, males, and females were similarly likely to approach and interact. As suggested above, practitioners trying to increase interaction with the robot might initially target group members that are more likely to interact with the robot (females in this case), and once the group is interacting, engage other group members.

An alternate explanation is that female groups in the mall were more likely to be there for leisure whereas the male groups were more likely to be there for business. Therefore, those groups with more females were more likely to have the group purpose of exploring and therefore, have more positive interactions with the robot, whereas those in groups with more males were more likely to have the purpose of business and therefore, have fewer interactions with the robot. Regardless, these findings show that the group composition and purpose of the group bring participant behavior to be closer to the behavior of the other members of the group.

Limitations, Design Recommendations, and Future Studies

In this study, we examined actual interaction with a robot in a real world setting, giving our results high external reliability. As such, none of the variables examined were experimentally manipulated, meaning that we cannot infer causation. Future studies should confirm that the effects we found can be replicated in more controlled studies during actual manipulation, and in other contexts. For example, in this study, it could be surmised that people at the mall for fun rather than business (i.e., family, friend, females, compared to males, coworkers) were more likely to interact with the robot, rate it positively, and interact with it for longer. This was especially true when they were in groups, especially more entitative groups. Future studies should examine if robots made for a work environment would attract more coworkers and if entitativity of coworker groups would also increase interaction time with the robot.

One confound in the study is that groups that are more entitative (family, friend, female) were also more likely to be in the mall for leisure than lower-entitative groups (coworkers, men). We recommend that scholars take these results with caution. Future studies should directly manipulate entitativity or examine entitative groups in different settings to disentangle these variables.

Further, family groups and female friend groups were not only more entitative, but perceived the robot as more positive than other groups. Future designers might market robots toward women and children for mall settings because these were the typical users in this study. However, the results indicate that it is important to remember that it is the social context (e.g., family outing) that is at least as important as gender in affecting responses toward a robot.

Additionally, this study was conducted in Japan. Findings, especially those related to business people and gender, may differ in different countries and in different social contexts.

In this study, we did not have enough survey data to examine the relationship between behavior toward the robot and survey ratings of the robot. Future studies could employ more surveys, such as by introducing a simple button near the robot to rate the interaction (positive, neutral, negative) for participants to employ after interaction to gain more explicit ratings of the robot.

Another limitation is that, although in some groups we were able to survey multiple participants, in other groups, only one participant would take the survey. This may bias the results if, for example, the person who was most likely to take the survey was the person who responded most positively to the robot. This type of self-selection bias is a limitation of all naturalistic studies that request survey participation. Examining these effects in a laboratory setting would circumnavigate this limitation. However, this is not a major concern because participant actual behavior supports conclusions drawn from surveys.

Finally, the video coders were not accurately able to identify family vs. friend groups. This could be a limitation in that we cannot make strong conclusions about behaviors of family and friends. However, if robots are programmed to identify different groups, they may also not be able to correctly identify family vs. friend groups. The identification of age-similar vs. age-diverse groups that were used in this study could plausibly be employed by robots in the near future, making this research directly applicable to today's HRI.

CONCLUSION

Overall, in this study we sought to find how the presence of a groups, group characteristics, and group norms relate to people's behavior toward a humanoid mall guidance robot. The results indicated that, in this friendly context, groups, and especially entitative groups, were more positive toward a robot. Second, group norms of interacting with a robot influenced participants who would not normally interact with the robot to interact with the robot. Practitioners can apply these results to the design and implementation of public HRI, with robots targeting high-entitative groups if they are looking for longer interactions. Future studies might examine more ways to engage low-entitative groups and others that are less likely to interact with a public robot.

DATA AVAILABILITY

The datasets generated for this study of survey and codes of video behavior are available through the Open Science Foundation (OSF) at https://osf.io/ew8ta/?view_only=df9ef0f0919b48afb64a773ffa0251ac.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of Indiana University, Institutional Review Board with verbal informed consent from all survey subjects. The study was exempt from requiring written informed consent because the study was minimal risk and in a public domain. Video subjects were not required to provide written or verbal consent because they were in a public location; however, a large, prominent sign informed them that video recordings were being made in the location. The protocol was approved by the Indiana University Institutional Review Board.

AUTHOR CONTRIBUTIONS

The ideas for this paper were conceptualized and developed by MF, SŠ, and TK. MF ran the study, data analysis, and wrote paper drafts, MF consulted SŠ throughout the process of data collection, analysis, and paper writing. SŠ and TK significantly edited the paper.

CONTRIBUTION TO THE FIELD

Robots are increasingly being placed in public settings, including malls, museums, and city streets, to help guide and direct people. The results of this study demonstrate how the characteristics of human groups influence people's behavior when interacting with a robot. It presents the novel finding that groups (especially integrative groups), compared to individuals followed the group norm more for interacting with a robot. Researchers and practitioners can use this information when designing robots for public interaction with people, and for engaging people who might not otherwise be interested in interacting with a robot.

FUNDING

Thank you to the National Science Foundation (NSF) Grant #1650681 and JST CREST Grant Number JPMJCR17A2 for funding this research, including researchers on the grant and equipment.

ACKNOWLEDGMENTS

Thank you to research assistants Andrew Buzzelli and Yulu Huang, Ji Hyung, Sara Wang, and Emily Wattimena for video coding. Thank you to our partners from Japan who helped run the study: Jason Lim, Kanako Tomita, and Satoru Satake.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frobt.2019.00048/full#supplementary-material

REFERENCES

Admoni, H., Hayes, B., Feil-Seifer, D., Ullman, D., and Scassellati, B. (2013). "Dancing with myself: The effect of majority group size on perceptions of majority and minority robot group members," in *Proceedings of the Annual Meeting of the Cognitive Science Society* (Berlin).

Al Moubayed, S., Beskow, J., Granström, B., Gustafson, J., Mirning, N., Skantze, G., et al. (2012). "Furhat goes to Robotville: A large-scale multiparty human-robot interaction data collection in a public space," *Paper presented at the Proc of LREC Workshop on Multimodal Corpora* (Istanbul).

Alves-Oliveira, P., Sequeira, P., Melo, F. S., Castellano, G., and Paiva, A. (2019). Empathic robot for group learning: a field study. *ACM Transact. Human-Robot Interact.* 8:3. doi: 10.1145/3300188

Anastasio, P., Bachman, B., Gaertner, S., and Dovidio, J. (1997). Categorization, recategorization and common ingroup identity. *Soc. Psychol. Stereotyp Group Life* 236:256.

Bennewitz, M., Faber, F., Joho, D., Schreiber, M., and Behnke, S. (2005). "Towards a humanoid museum guide robot that interacts with multiple persons," in *5th IEEE-RAS International Conference on Humanoid Robots, 2005* (Tsukuba: IEEE), 418–423.

Brscić, D., Kidokoro, H., Suehiro, Y., and Kanda, T. (2015). "Escaping from children's abuse of social robots," in *Proceedings of the tenth Annual ACM/IEEE International Conference on Human Robot Interaction* (Portland, OR: ACM), 59–66.

Burger, J. M., and Shelton, M. (2011). Changing everyday health behaviors through descriptive norm manipulations. *Soc. Influence* 6, 69–77. doi: 10.1080/15534510.2010.542305

Campbell, D. T. (1958). Common fate, similarity, and other indices of the status of aggregates of persons as social entities. *Behav. Sci.* 3, 14–25. doi: 10.1002/bs.3830030103

Castano, E., Yzerbyt, V., and Bourguignon, D. (2003). We are one and I like it: The impact of ingroup entitativity on ingroup identification. *Eur. J. Soc. Psychol.* 33, 735–754. doi: 10.1002/ejsp.175

Chang, W. L., White, J. P., Park, J., Holm, A., and Šabanović, S. (2012). "The effect of group size on people's attitudes and cooperative behaviors toward robots in interactive gameplay," in *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication* (Paris: IEEE), 845–850.

Chang, W. L., Šabanovic, S., and Huber, L. (2014). "Observational study of naturalistic interactions with the socially assistive robot PARO in a nursing home," in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication* (Edinburgh: IEEE), 294–299.

Cialdini, R. B. (2007). Descriptive social norms as underappreciated sources of social control. *Psychometrika* 72, 263–268. doi: 10.1007/s11336-006-1560-6

Costa Jr., P. T., Terracciano, A., and McCrae, R. R. (2001). Gender differences in personality traits across cultures: robust and surprising findings. *J. Pers. Soc. Psychol.* 81, 322–331. doi: 10.1037//0022-3514.81.2.322

Eagly, A. H., and Wood, W. (1999). The origins of sex differences in human behavior: Evolved dispositions versus social roles. *Am. Psychol.* 54, 408–423. doi: 10.1037/0003-066X.54.6.408

Eyssel, F., De Ruiter, L., Kuchenbrandt, D., Bobinger, S., and Hegel, F. (2012). "'If you sound like me, you must be more human': on the interplay of robot and user features on human-robot acceptance and anthropomorphism," in *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Paris: IEEE), 125–126.

Foster, M. E., Gaschler, A., Giuliani, M., Isard, A., Pateraki, M., and Petrick, R. (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* (Santa Monica, CA: ACM), 3–10.

Fraune, M. R., Kawakami, S., Sabanovic, S., De Silva, R., and Okada, M. (2015). *Three's company, or a crowd?: The effects of robot number and behavior on HRI in Japan and the USA*. Robotics: Science and Systems. doi: 10.15607/RSS.2015.XI.033

Fraune, M. R., Šabanović, S., Smith, E. R., Nishiwaki, Y., and Okada, M. (2017). "Threatening flocks and mindful snowflakes: How group entitativity affects perceptions of robots," in *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Vienna: IEEE), 205–213.

Fraune, M. R., Sherrin, S., Šabanović, S., and Smith, E. R. (2019). "Is human robot interaction more competitive between groups than between individuals?,"in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Daegu: IEEE), 104–113.

Gaertner, L., and Schopler, J. (1998). Perceived ingroup entitativity and intergroup bias: an interconnection of self and others. *Eur. J. Soc. Psychol.* 28, 963–980.

Gaertner, S. L., Dovidio, J. F., Banker, B. S., Houlette, M., Johnson, K. M., and McGlynn, E. A. (2000). Reducing intergroup conflict: From superordinate goals to decategorization, recategorization, and mutual differentiation. *Group Dyn Theory Res Pract.* 4, 98–114. doi: 10.1037//1089-2699.4.1.98

Gergen, K. J., Gergen, M. M., and Barton, W. H. (1973). Deviance In Dark. *Psychol. Today,* 7, 129–130. doi: 10.1037/e400532009-005

Giuliani, M., Petrick, R., Foster, M. E., Gaschler, A., Isard, A., Pateraki, M., et al. (2013). "Comparing task-based and socially intelligent behaviour in a robot bartender," in *Paper presented at the Proceedings of the 15th ACM on International Conference on Multimodal Interaction* (Sydney).

Goldstein, N. J., Cialdini, R. B., and Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *J Consu Res.* 35, 472–482. doi: 10.1086/586910

Gomez, R., Kawahara, T., Nakamura, K., and Nakadai, K. (2012). "Multi-party human-robot interaction with distant-talking speech recognition," *Paper Presented at the Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction* (Paris).

Holthaus, P., Pitsch, K., and Wachsmuth, S. (2011). How can I help? *Inter. J. Soc. Robot.* 3, 383–393. doi: 10.1007/s12369-011-0108-9

Hyde, J. S. (1984). How large are gender differences in aggression?: a developmental meta-analysis. *Dev. Psychol.* 20, 722–736. doi: 10.1037/0012-1649.20.4.722

Insko, C. A., Hoyle, R. H., Pinkley, R. L., Hong, G.-Y., Slim, R. M., Dalton, B., et al. (1988). Individual-group discontinuity: The role of a consensus rule. *J. Exp. Soc. Psychol.,* 24, 505–519. doi: 10.1016/0022-1031(88)90049-2

Insko, C. A., Wildschut, T., and Cohen, T. R. (2013). Interindividual–intergroup discontinuity in the prisoner's dilemma game: How common fate, proximity, and similarity affect intergroup competition. *Organ. Behav. Hum. Decis. Process.* 120, 168–180. doi: 10.1016/j.obhdp.2012.07.004

Johansson, M., and Skantze, G. (2015). "Opportunities and obligations to take turns in collaborative multi-party human-robot interaction," in *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue* (Stockholm), 305–314.

Johansson, M., Skantze, G., and Gustafson, J. (2013). *Head pose patterns in multiparty human-robot team-building interactions*. Paper presented at the International Conference on Social Robotics. doi: 10.1007/978-3-319-02675-6_35

Johnson, R. D., and Downing, L. L. (1979). Deindividuation and valence of cues: Effects on prosocial and antisocial behavior. *J. Pers. Soc. Psychol.* 37, 1532–1538. doi: 10.1037/0022-3514.37.9.1532

Jung, M. F., Martelaro, N., and Hinds, P. J. (2015). "Using robots to moderate team conflict: the case of repairing violations," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (Portland, OR: ACM), 229–236.

Kanda, T., Hirano, T., Eaton, D., and Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. *Hum Comp Interact.* 19, 61–84. doi: 10.1207/s15327051hci1901&2_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 Transact. Robot.* 23, 962–971. doi: 10.1109/TRO.2007.904904

Koike, S., Sugawara, M., Kutsukake, Y., Yamanouchi, S., Sato, K., Fujita, Y., et al. (2009). "Social robot design," in *International Conference on Human Centered Design* (Berlin: Springer), 462–467.

Kumagai, T., and Ohbuchi, K.-I. (2009). The effects of group identification and the unfairness of harm on third party aggression. *Jap. J. Soc. Psychol.* 24, 200–207.

Labor Force. (2018). *Female (% of total labor force)*. The World Bank. Available online at: https://data.worldbank.org/indicator/SL.TLF.TOTL.FE.ZS

Lee, M. K., Kiesler, S., and Forlizzi, J. (2010). "Receptionist or information kiosk: How do people talk with a robot?," in *Proceedings of the 2010 ACM conference on Computer Supported Cooperative Work* (Savannah, GA: ACM), 31–40.

Leite, I., McCoy, M., Lohani, M., Ullman, D., Salomons, N., Stokes, C., et al. (2015). "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* (Portland, OR: ACM), 75–82.

Lickel, B., Hamilton, D. L., and Sherman, S. J. (2001). Elements of a lay theory of groups: types of groups, relational styles, and the perception of group entitativity. *Pers. Soc. Psychol. Rev.* 5, 129–140. doi: 10.1207/S15327957PSPR0502_4

Lickel, B., Hamilton, D. L., Wieczorkowska, G., Lewis, A., Sherman, S. J., and Uhles, A. N. (2000). Varieties of groups and the perception of group entitativity. *J. Pers. Soc. Psychol.* 78, 223–246. doi: 10.1037/0022-3514.78.2.223

Martin, J. S., and Chaney, L. H. (2012). *Global business etiquette: A guide to international communication and customs.* Santa Barbara, CA: ABC-CLIO.

Meier, B. P., and Hinsz, V. B. (2004). A comparison of human aggression committed by groups and individuals: An interindividual–intergroup discontinuity. *J. Exp. Soc. Psychol.,* 40, 551–559. doi: 10.1016/j.jesp.2003.11.002

Nawata, K., and Yamaguchi, H. (2011). The ingroup audience effect on intergroup vicarious retribution. *Research in Soc. Psychol.* 26, 167–177.

Pereira, A. T., Prada, R., and Paiva, A. (2014). "Improving social presence in human-agent interaction," in *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems* (Toronto: ACM), 1449–1458.

Pitsch, K., Gehle, R., and Wrede, S. (2013). "A museum guide robot: dealing with multiple participants in the real-world," *Paper Presented at the Workshop "Robots in Public Spaces: Towards Multi-Party, Short-Term, Dynamic Human-Robotinteraction"* (Bristol, UK: ICSR).

Reicher, S. D., Spears, R., and Postmes, T. (1995). A social identity model of deindividuation phenomena. *Eur. Rev. Soc. Psychol.* 6, 161–198. doi: 10.1080/14792779443000049

Rosario, M., Shinn, M., Mørch, H., and Huckabee, C. B. (1988). Gender differences in coping and social supports: Testing socialization and role constraint theories. *J. Community Psychol.* 16, 55–69.

Sabanovic, S., Michalowski, M. P., and Simmons, R. (2006). "Robots in the wild: observing human-robot social interaction outside the lab," *Paper Presented at the Advanced Motion Control, 2006. 9th IEEE International Workshop on* (Istanbul).

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* (Munich: ACM), 263–270.

Sherif, M. (1936). *The Psychology of Social Norms*. Oxford, England: Harper.

Siegel, M., Breazeal, C., and Norton, M. I. (2009). "Persuasive robotics: the influence of robot gender on human behavior," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems* (St. Louis: IEEE), 2563–2568.

Siino, R. M., and Hinds, P. J. (2005). "Robots, gender & sensemaking: sex segregation's impact on workers making sense of a mobile autonomous robot," in *Proceedings of the 2005 IEEE International Conference on Robotics and Automation* (Barcelona: IEEE), 2773–2778.

Smith, J. R., Hogg, M. A., Martin, R., and Terry, D. J. (2007). Uncertainty and the influence of group norms in the attitude–behaviour relationship. *Br. J. Soc. Psychol.* 46, 769–792. doi: 10.1348/014466606X164439

Steinberg, C., and Nakane, M. M. (2012). *Can Women Save Japan? (No. 12-248)*. International Monetary Fund.

Tajfel, H., Billig, M. G., Bundy, R. P., and Flament, C. (1971). Social categorization and intergroup behaviour. *Eur. J. Soc. Psychol.* 1, 149–178. doi: 10.1002/ejsp.2420010202

Taylor, A., and Riek, L. D. (2016). "Robot perception of human groups in the real world: State of the art," *Paper presented at the AAAI Fall Symposium Series: Artificial Intelligence for Human-Robot Interaction Technical Report FS-16-01* (Palo Alto, CA).

Tseng, S.-H., Chao, Y., Lin, C., and Fu, L.-C. (2016). Service robots: system design for tracking people through data fusion and initiating interaction with the human group by inferring social situations. *Rob. Auton. Syst.* 83, 188–202. doi: 10.1016/j.robot.2016.05.004

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

Wang, Y. (2014). *Gendering human-robot interaction: exploring how a person's gender impacts attitudes toward and interaction with robots (FGS -Electronic Theses & Dissertations (Public) [23278]).* University of Manitoba Libraries MSpace. Retrieved from: http://hdl.handle.net/1993/24446

Weiss, A., Bernhaupt, R., Tscheligi, M., Wollherr, D., Kuhnlenz, K., and Buss, M. (2008). "A methodological variation for acceptance evaluation of human-robot interaction in public places," in *RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication* (Munich: IEEE), 713–718.

Wildschut, T., and Insko, C. A. (2007). Explanations of interindividual–intergroup discontinuity: a review of the evidence. *Eur. Rev. Social. Psychol.* 18, 175–211. doi: 10.1080/10463280701676543

Wildschut, T., Insko, C. A., and Gaertner, L. (2002). Intragroup social influence and intergroup competition. *J. Pers. Soc. Psychol.* 82, 975–992. doi: 10.1037/0022-3514.82.6.975

Wildschut, T., Insko, C. A., and Pinter, B. (2007). Interindividual-intergroup discontinuity as a joint function of acting as a group and interacting with a group. *Eur. J. Soc. Psychol.* 37, 390–399. doi: 10.1002/ejsp.374

Wildschut, T., Pinter, B., Vevea, J. L., Insko, C. A., and Schopler, J. (2003). Beyond the group mind: A quantitative review of the interindividual-intergroup discontinuity effect. *Psychol. Bull.* 129, 698–722. doi: 10.1037/0033-2909.129.5.698

Wright, E. O., Baxter, J., and Birkelund, G. E. (1995). The gender gap in workplace authority: A cross-national study. *Am. Sociol. Rev.* 407–435. doi: 10.2307/2096422

Zanlungo, F., Yücel, Z., Brščić, D., Kanda, T., and Hagita, N. (2017). Intrinsic group behaviour: Dependence of pedestrian dyad dynamics on principal social and personal features. *PLoS ONE.* 12:e0187253. doi: 10.1371/journal.pone.0187253

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.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 03 September 2019
doi: 10.3389/frobt.2019.00081

A Long-Term Study of Young Children's Rapport, Social Emulation, and Language Learning With a Peer-Like Robot Playmate in Preschool

Jacqueline M. Kory-Westlund and Cynthia Breazeal*

MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA, United States

Prior research has demonstrated the importance of children's peers for their learning and development. In particular, peer interaction, especially with more advanced peers, can enhance preschool children's language growth. In this paper, we explore one factor that may modulate children's language learning with a peer-like social robot: rapport. We explore connections between preschool children's learning, rapport, and emulation of the robot's language during a storytelling intervention. We performed a long-term field study in a preschool with 17 children aged 4–6 years. Children played a storytelling game with a social robot for 8 sessions over two months. For some children, the robot matched the level of its stories to the children's language ability, acting as a slightly more advanced peer (*Matched* condition); for the others, the robot did not match the story level (*Unmatched* condition). We examined children's use of target vocabulary words and key phrases used by the robot, children's emulation of the robot's stories during their own storytelling, and children's language style matching (LSM—a measure of overlap in function word use and speaking style associated with rapport and relationship) to see whether they mirrored the robot more over time. We found that not only did children emulate the robot more over time, but also, children who emulated more of the robot's phrases during storytelling scored higher on the vocabulary posttest. Children with higher LSM scores were more likely to emulate the robot's content words in their stories. Furthermore, the robot's personalization in the *Matched* condition led to increases in both children's emulation and their LSM scores. Together, these results suggest first, that interacting with a more advanced peer is beneficial for children, and second, that children's emulation of the robot's language may be related to their rapport and their learning. This is the first study to empirically support that rapport may be a modulating factor in children's peer learning, and furthermore, that a social robot can serve as an effective intervention for language development by leveraging this insight.

Keywords: children, language development, mimicry, peer modeling, rapport, relationship, social robotics, storytelling

Edited by:
Chung Hyuk Park, George Washington University, United States

Reviewed by:
Tony Belpaeme, University of Plymouth, United Kingdom
Sofia Serholt, Department of Applied Information Technology, University of Gothenburg, Sweden

****Correspondence:***
Jacqueline M. Kory-Westlund
jakory@alum.mit.edu

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *17 May 2019*
Accepted: *16 August 2019*
Published: *03 September 2019*

Citation:
Kory-Westlund JM and Breazeal C (2019) A Long-Term Study of Young Children's Rapport, Social Emulation, and Language Learning With a Peer-Like Robot Playmate in Preschool. Front. Robot. AI 6:81. doi: 10.3389/frobt.2019.00081

1. INTRODUCTION

Children's early language development is linked to their academic and overall life success. Numerous studies in the United States, for example, have found that children who are not exposed to rich language learning opportunities as they grow up—such as vocabulary-building curricula, cognitively challenging preschool activities, greater numbers of novel words and total words heard—may be significantly impacted, showing language deficits, lower reading comprehension, and lower vocabulary ability (Huttenlocher et al., 1991, 2002, 2010; Hart and Risley, 1995; Fish and Pinkerman, 2003; Griffin et al., 2004; Paez et al., 2007; Snow et al., 2007; Perkins et al., 2013; Schwab and Lew-Williams, 2016). Numerous interventions have been developed to support children's early language development, such as preschool readiness programs, teacher, and parent resources, and a wide range of language-focused educational apps, games, and computer programs.

One way children's language learning can be supported is through peer interaction. Children's peer relationships provide opportunities for openness, exploration, and discovery. Research from the past several decades shows that children's peers, particularly more advanced peers, can enhance their overall preschool competency and language growth (Fuchs et al., 1997; Mathes et al., 1998; Topping, 2005; Schechter and Bye, 2007; Whitebread et al., 2007; Mashburn et al., 2009; Justice et al., 2011; DeLay et al., 2016; Lin et al., 2016). Mashburn et al. (2009), for example, measured preschool children's receptive and expressive language skills at the start and end of a school year. Children's language growth during the year was positively related to their peers' expressive language abilities, a result later replicated by Justice et al. (2011). Notably, children, particularly children with lower skills, appeared to benefit most from having higher ability peers around them.

This research is in line with various theories about how peer learning occurs, including Vygotsky's theory that a child's more advanced peers can help support or scaffold the child in acquiring and practicing skills that are otherwise beyond their skill level (Vygotsky, 1978; Tudge and Rogoff, 1989; Rubin et al., 1998); Bandura and Walters' social learning theory which argues that children frequently learn through observing and imitating others (e.g., observing and imitating their speech; Bandura and Walters, 1963; Bandura, 1971; Rubin et al., 1998); and Piaget's theories regarding the importance of dialogue and discussion among peers in promoting cognitive development (Piaget, 1932; Tudge and Rogoff, 1989; Rubin et al., 1998; De Lisi and Golbeck, 1999).

Because children's peers can significantly and positively affect their language learning, numerous researchers in human-robot interaction have hypothesized that playing with a peer-like robot companion may lead to similar benefits. For example, some robots have been positioned as slightly advanced peers (e.g., Kanda et al., 2004; Kory and Breazeal, 2014; Gordon et al., 2016; Kory Westlund et al., 2017b); while others have been positioned as younger peers or novices (e.g., Movellan et al., 2009; Tanaka and Kimura, 2009; Tanaka and Matsuzoe, 2012; Gordon and Breazeal, 2015; Hood et al., 2015; Tanaka et al., 2015). Some virtual agents have also been created as peer-like learning companions (Bers et al., 1998; Cassell and Ryokai, 2001; Ryokai et al., 2003; Cassell, 2004; Cassell et al., 2007). In language learning applications, research has focused primarily on children's vocabulary learning, often in English and often with English as a second language, though language production is also a growing area of study (Kanero et al., 2018).

It is also very common for robots to be situated as teachers or tutors (e.g., Robins et al., 2005; You et al., 2006; Chang et al., 2010; Lee et al., 2011; Alemi et al., 2014; Serholt et al., 2014; Deshmukh et al., 2015; Kennedy et al., 2016; Park et al., 2017b; Vogt et al., 2017, 2019; Rintjema et al., 2018). A recent survey of 101 studies of social robots in education revealed that 86% of studies set up robots as teachers or tutors, 4% positioned the robot in a mixed tutor/teacher role, only 9% set up the robot as a peer or novice, and 1% gave the robot another role (Belpaeme et al., 2018). In this survey, nearly 60% of the studies surveyed involved children, and it included studies of many different educational activities, including language, math, and reading.

Given this interest in using social robots to support children's language learning, we should examine more closely what modulates children's learning with peers, and by extension, mechanisms that robots can use to be more effective learning companions. That is: are children's peers approximately equal as sources for promoting language learning, or will children learn more effectively from some peers than from others? What features or behavior might help a social robot better enable children's language learning?

Some work has begun exploring these questions. For example, robots that use nonverbal social cues and nonverbal immediacy behaviors have led to increases in children's engagement, learning, and relationships during educational activities (e.g., Kanda et al., 2004, 2007, 2012; Breazeal et al., 2016; Kennedy et al., 2017; Kory Westlund et al., 2017a,b). These results jibe with literature in psychology and education, where research has linked improved learning outcomes to use of appropriate social cues (e.g., Bloom, 2000; Meltzoff et al., 2009; Sage and Baldwin, 2010; Kuhl, 2011), social interaction and greater numbers of conversational turns (e.g., Hoff, 2006; Romeo et al., 2018a,b), and nonverbal immediacy (Mehrabian, 1968; Christophel, 1990; Witt et al., 2004). Robots that personalize content or behavior to children have also led to increased learning and engagement (e.g., Leite et al., 2012; Kory and Breazeal, 2014; Gordon et al., 2016; Palestra et al., 2016; Scassellati et al., 2018; Park et al., 2019).

Another mechanism that may improve children's learning is rapport, as suggested by two recent studies of children's language learning during storytelling with social peer-like robots (Kory Westlund et al., 2017b; Kory-Westlund, 2019; Kory-Westlund and Breazeal, 2019b). Kory Westlund et al. (2017b) found that playing with a robot with a more expressive voice led to increases in children's engagement and vocabulary learning as well as increased emulation of the robot's language. Kory-Westlund (2019) found that children's language emulation, positive emotion, and acceptance of the robot were positively affected by the robot's use of speech entrainment and an appropriate backstory about its abilities. These studies suggest that children's rapport may be reflected in their language emulation, a result that jibes with related work showing that humans who have greater rapport with each other will mimic each other's language (e.g., Niederhoffer and Pennebaker, 2002; Pennebaker et al., 2003; Huttenlocher et al., 2004; Tausczik and Pennebaker, 2010; Ireland et al., 2011; Babcock et al., 2014) and vocal prosody (e.g., Porzel et al., 2006; Reitter et al., 2011; Borrie and Liss, 2014) more.

Earlier work with adults and robots (Kidd and Breazeal, 2008; Lubold et al., 2016, 2018; Lubold, 2017), as well work in human-human tutoring (Sinha and Cassell, 2015a,b), have also suggested links between learning and rapport. Children's social bonds with their teachers can predict their performance (Wentzel, 1997). Children who have stronger parasocial relationships with media characters may learn more effectively from those characters (Gola et al., 2013; Richards and Calvert, 2017).

Taken together, the research so far suggests that children's rapport with an interlocutor may affect their learning and language behavior. However, these studies were primarily one session; they did not examine children's learning or language behavior over time. As such, one open and important question was whether children would emulate the robot's language long-term, and if they did, whether this would be related to their vocabulary learning or their rapport with the robot. To explore this question, we performed new analyses on an existing dataset from an 8-session study in which children played a storytelling game with a peer-like social robot. The design and early results from this study were presented in (Kory, 2014; Kory and Breazeal, 2014; Kory Westlund and Breazeal, 2015); here we present the full methodology, as well as results and discussion.

2. METHODOLOGY

2.1. Research Questions

We wanted to explore connections between children's learning, their rapport, and their emulation of a peer-like robot's language behavior. We asked whether children would be more likely to emulate language of a robot with whom they had more positive rapport, whether this was correlated with their learning, and furthermore, whether children's emulation or rapport were consistent over time.

2.2. Design

We performed new analyses on an existing dataset that included stories from 14 children, who had played a storytelling game with a robot 1–2 times per week for 8 sessions (**Figure 1**) (Kory, 2014; Kory and Breazeal, 2014; Kory Westlund and Breazeal, 2015).

The original study explored whether a peer-like social robot could facilitate preschool children's oral language development. In addition to being one of the first studies exploring the effectiveness of a long-term, storytelling intervention, this study examined whether personalizing the general language complexity of the robot's stories might increase children's learning of new words and use of more complex language in their own stories. The hypothesis was that presenting stories of an appropriate challenge for the child, slightly ahead of the child's general ability in the zone of proximal development, might promote learning (Vygotsky, 1978; Csikszentmihalyi, 1990). Thus, the study followed a two-condition design.

Two versions of each story told by the robot were created, a harder version and an easier version (for more detail regarding story creation, see Kory, 2014; Kory and Breazeal, 2014). In the first half of the study (sessions 1–4), all children heard the same versions of the stories. In the second half of the study (sessions 5–8), children in the *Matched* condition (12 children—6 female, 6 male) heard stories matched to their language ability (i.e., harder stories for children with higher ability; easier stories for children with lower ability. Children in the *Unmatched* condition (5 children—4 female, 1 male) heard stories that were not matched (e.g., easy stories for children with higher ability).

2.3. Participants

Seventeen children aged 4–6 years (10 female, 7 male) from two Boston-area preschools (9 from the first and 8 from the second) participated in the original study. Children were recruited from two schools in order to recruit sufficient children for the study. There were three 4-year-olds, thirteen 5-year-olds, and one 6-year-old ($M = 4.88$, $SD = 0.49$). The 6-year-old girl did not complete the final session, and one 4-year-old girl completed only the first 4 sessions. Children in this age range were targeted because their expressive language abilities are developed enough to be able to tell stories. They are still in the process of developing their narrative abilities. Younger children, as was discovered during pilot testing, may not tell stories at all and are less likely to understand and follow the rules of the game.

For the purposes of our analyses here, our data included 206 stories from 14 children (8 female, 6 male, two 4-year-olds, twelve 5-year-olds, age $M = 4.86$, $SD = 0.36$) and full transcripts from all 17 children (3 children did not tell stories).

Children's parents gave written informed consent prior to the start of the study, and all children assented to participate. The protocol was approved by the MIT Committee on the Use of Humans as Experimental Subjects.

2.4. Hypotheses

We expected the following:

- **H1:** Children who showed greater rapport with the robot would be more likely to learn the target vocabulary words, with receptive knowledge indexed by vocabulary assessment scores and productive knowledge by use of the words in their stories. We expected this because prior work has shown that rapport can facilitate learning (Sinha and Cassell, 2015a,b), and children have previously mirrored a robot's vocabulary in their stories (Kory Westlund et al., 2017b)
- **H2:** Children who showed greater rapport would be more likely to emulate the robot's language in their stories and throughout the full interaction session. We expected this because people frequently mirror the language and behavior of those with whom they have rapport (e.g., Dijksterhuis and Bargh, 2001; Niederhoffer and Pennebaker, 2002; Huttenlocher et al., 2004; Chartrand and van Baaren, 2009; Tausczik and Pennebaker, 2010; Ireland et al., 2011; Babcock et al., 2014).

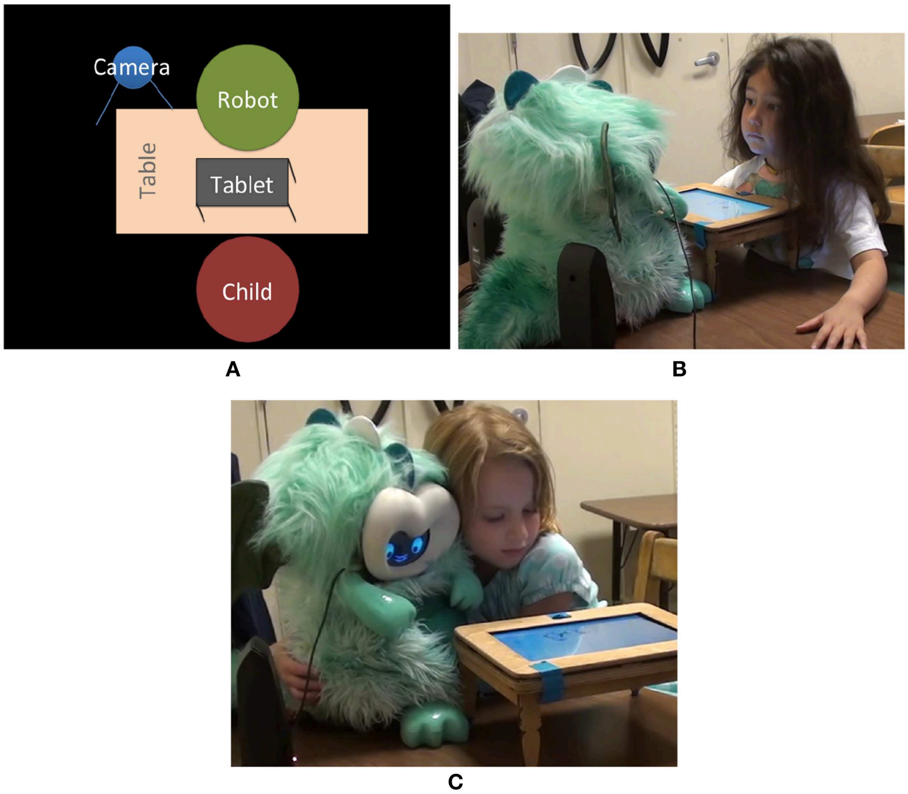


FIGURE 1 | (A) The robot was placed on a table across from the child. The tablet was set in a small table between them. The camera was set up behind the robot to the left. **(B)** A girl listens while the robot tells a story. **(C)** This girl turned the robot and tablet so she could sit beside the robot. Written informed consent was obtained to use these images.

- **H3:** Because of the expected connections between children's rapport and their learning, we also expected that children who emulated the robot's language more would also show more vocabulary learning.
- **H4:** We expected children's rapport and their emulation of the robot's language to increase over time as they became more familiar and comfortable with the robot.
- **H5:** Children who heard personalized stories from the robot would emulate more, learn more words, and have greater rapport. We expected this because of suggested links between a robot's personalization and children's engagement and learning (e.g., Leite et al., 2012; Gordon et al., 2016; Palestra et al., 2016; Scassellati et al., 2018; Park et al., 2019).

2.5. Procedure

Each child participated in a pretest session and 8 sessions with a teleoperated robot, over 10 weeks (Kory, 2014; Kory and Breazeal, 2014; Kory Westlund and Breazeal, 2015). During the pretest, children were given a language assessment, a subset of the Preschool Language Scale, 5th Edition (Zimmerman et al., 2011), to assess aspects of their expressive and receptive language ability. This assessment did not use any of the robot's target words. Children were also given a separate receptive vocabulary pretest for the target words the robot used in its stories. In this test, for each word, children were shown a set of four pictures and were asked to point to the picture showing the target word.

These initial assessments was used to split children into two groups: higher language ability (above the mean), and lower language ability (below the mean). These categorizations were for this study only; "higher/lower language ability" did not mean children were necessarily above or below what might be expected for their age, just that they were divided into two groups for the purposes of the robot's language level personalization. Children were randomly assigned to the *Matched* or *Unmatched* conditions after these assessments; their initial language assessment scores were taken into account in an attempt to balance language ability across conditions.

Each of the 8 sessions with the robot was 10–15 min long (**Figure 1**). The robot briefly engaged the child in conversation (e.g., asking if the child had done anything fun that morning or sharing a fact about itself), then showed a story scene on a tablet and told a short story. Next, the child was invited to tell their own story about the scene. The robot then showed a second story scene and told a second story, and the child was invited to tell a second story. After a brief closing conversation, the interaction ended. In some sessions, the robot showed a story scene but asked the child to tell a story first. If children declined to tell their own story, the robot briefly encouraged them to do so, but if they refused again, the robot moved on.

As mentioned above, in the first half of the study (sessions 1–4), all children heard the same stories. In the second half of the study (sessions 5–8), children in the *Matched* condition heard stories matched to their language ability, while children in the *Unmatched* condition heard stories that were not matched.

A storytelling activity was used to promote language development because storytelling is a socially situated activity that combines play and narrative, which are two important aspects of children's learning and development (Nicolopoulou, 1993; Engel, 1995). Storytelling can allow collaborative, creative conversation and language practice, and can support emergent literacy skills, including metalinguistic knowledge about language patterns, structure, and function; vocabulary; "decontextualized" language that can be understood outside its original context; as well as supporting cognitive, communicative, and linguistic development more broadly (Engel, 1995; Cassell, 2004; Curenton et al., 2008).

Children were interviewed about their perception of the robot and interaction after sessions 4 and 8. The questions were adapted in part from (Jipson and Gelman, 2007; Kahn et al., 2012). Children were invited to answer numerous questions using a verbal 3-point scale ("a lot," "a little bit," or "not very much"). While this methodology presents some challenges due to children's tendency to answer in socially acceptable ways, anecdotally, children's engagement and interest observed during the activities was reflected in their interview responses. Furthermore, many of the interview questions were followed up by asking children to explain their response or to say more, which helped give context to children's ratings. All interview questions and language assessments are available on figshare at https://doi.org/10.6084/m9.figshare.8144456.

2.6. Materials

2.6.1. Robot

This study used the Dragonbot (Setapen, 2012; Kory et al., 2013) as the learning companion. This robot is capable of expressive movement based on "squash and stretch" principles of animation. It can display a variety of facial expressions on the smart phone that also runs its software, as well as play sounds or speech. The robot wore green fur, was named "Green," and was referred to in a distinctly non-gendered way by the experimenter throughout the study.

The robot followed a script of speech, expressions, and movement. Speech was recorded by a human adult female. The pitch of the speech was shifted higher to sound more like a child.

2.6.2. Teleoperation

A human operator used a custom control interface to send action and speech commands to the robot. The teleoperator attended to the child's speech and actions in order to trigger the robot's actions (e.g., playing back speech or showing a facial expression) at appropriate times. Including a human in the loop allowed the robot to appear autonomous while sidestepping technical barriers such as autonomatic speech recognition and natural language understanding. When the robot's actions depended on what the child said or did, such as during the introductory conversation or when asking the child if they wanted to tell a story, the teleoperator selected among a limited set of dialogue options. The robot's gaze was automatically directed to either look up at the child or down at the game, based on data collected during the pilot study regarding where children look during play.

The teleoperator followed several general rules. First, the teleoperator made the robot's behavior as socially contingent as possible—reacting to the child as closely to as a human would in the same circumstance. When the child spoke, the robot would acknowledge through speech, verbal exclamations such as "Ooh!" and "Oh no!," smiles, and short affirmative non-linguistic noises. These acknowledgments were primarily triggered during pauses in the child's speech. The same sounds or animations were not triggered twice in close succession, though the same sounds and animations were often used multiple times per session. Finally, the teleoperator made the robot's behavior as consistent as possible across participants, using the same set of sounds and animations with approximately the same frequency for all children. The same person operated the robot for all participants and had been previously operated this robot in numerous earlier studies.

2.6.3. Storytelling Game

The storytelling game was inspired by the game developed by Ryokai et al. (2003) for their virtual peer, in which the virtual agent which took turns with children telling stories about characters in a toy castle. In this study, the shared game surface was a tablet screen set into a small wooden table. Story scenes showed a background image with several characters and objects that could be dragged around on the screen, much like virtual stick puppets. When the robot told stories, the characters were moved automatically in concert with the robot's speech. These movements were recorded and played back so that they would be consistent for all children. There were no additional animations or sound effects.

The game included eight story scenes (**Figure 2**). Over the course of the study, the robot told two stories using each scene.

The robot's stories were based on stories told by children during pilot testing of the game at the Boston Museum of Science (Kory, 2014). Two versions of each story were crafted for the personalization with the same general content, but with one having greater with greater language complexity ("hard" stories) and one with less ("easy" stories Kory and Breazeal, 2014). For example, part of one easier story included, "George liked to climb up massive icebergs. He liked to slide back down in the snow," while the more complex version was, "George enjoyed climbing up to the very top of massive icebergs, then sliding all the way back down on his belly, beak first".

2.6.4. Vocabulary

Twenty-four target vocabulary words were selected from Andrew Biemiller's "Words Worth Teaching" lists (Biemiller, 2010), including nouns (e.g., structure, clump), verbs (e.g., expect, plunge), and adjectives (e.g., ancient, massive). Three words were used in each of the robot's stories. Because the robot told two stories each session, six words were used each session. After sessions 1–4, all the words had been introduced. During the sessions 5–8, the words were used again in new stories to provide

FIGURE 2 | The eight story scenes used for the storytelling game. Two stories were written for each scene, for a total of 16 stories.

additional opportunities for learning. Children were tested on the vocabulary words using a picture-based assessment before and after the study. In each item on the assessment, children were shown four pictures. They were asked to point to the picture corresponding to the target word.

2.7. Data

Audio and video of the study sessions were recorded with a camera beside the robot (**Figure 1**). Children's responses to the vocabulary assessments and interview questions were recorded on paper and later transferred to a spreadsheet.

2.8. Data Analysis

The recorded audio was used to transcribe children's speech. Children's stories were extracted from the full transcripts. All children spoke during the conversations with the robot, and most told stories as well.

The data we analyzed in this paper included 206 stories from 14 children and full transcripts from 17 children (3 children did not tell stories). In these data, we examined children's use of key vocabulary words and key phrases used by the robot, children's emulation of the robot's stories during their own storytelling, and children's language style matching (LSM). LSM is a measure of overlap in function words and speaking style as opposed to content words. Our phrase matching metrics looked primarily at content words. Research has shown that the more "in sync" two people are, the more they will match function words in their speech; it may reflect rapport and relationship (Niederhoffer and Pennebaker, 2002; Pennebaker et al., 2003; Tausczik and Pennebaker, 2010; Ireland et al., 2011; Babcock et al., 2014). We use LSM here as a measure of rapport.

One limitation of this methodology is that LSM is a linguistic measure of rapport. It would be useful in future work to examine additional ways of measuring children's rapport with the robot, to see whether children's word and phrase use was related to any non-linguistic signs of rapport or relationship as well.

2.8.1. Target Words and Key Phrases

Using automated software tools, we counted the number of times children used each of the target vocabulary words in each session and in their stories. This analysis was performed on the full transcripts of each session. Usage of the words may reflect expressive vocabulary ability, which is often a stronger indicator of knowledge of a word than the receptive knowledge tested with the vocabulary assessment (Bloom, 1974; Ingram, 1974; Sénéchal, 1997), as well as mimicry of the robot. We also counted the number of times children used key phrases that the robot had used (e.g., "Once upon a time," "I'll tell a story about…," "See you later, alligator!"). For these, our goal was to see whether children adopted any of the robot's frequently used phrases, as this mimicry may reflect greater rapport.

2.8.2. Language Style Matching (LSM)

LSM analysis requires a minimum of 50 words per participant in the conversation, but works better with a greater number of words (Pennebaker et al., 2003; Tausczik and Pennebaker, 2010). Thus, to get sufficient data for an LSM analysis, we aggregated all of each child's stories for sessions 1–4 (the first half of the study) and then for sessions 5–8 (the second half). We obtained an LSM score for each set using software tools to access the Receptivity API (Tausczik and Pennebaker, 2010). LSM scores range from 0 to 1.00, but more often range from 0.5 to 1.00. The closer the score is to 1.00, the more matching is present.

2.8.3. Stories and Phrase Matching

We analyzed children's transcribed stories in five ways: length (in seconds), word count, vocabulary word use, and emulation of the robot's phrases. We created an automatic tool to obtain phrase matching scores comparing each child story to each robot story that the child had heard prior to telling the story. For example, a story told by a child in session 2 was compared to the stories the robot told in session 1 as well as any stories the robot told before the child in session 2. The analysis was then threefold: (1) compare each child story to the robot story just prior to it; (2) compare each child story to other stories in the same scene; (3) compare each child story to all stories prior to it. The matching algorithm was as follows:

1. Remove stopwords (i.e., words with no significant information such as "the," "uh," and "an").
2. Stem words, i.e., convert words to their original form (e.g., "running" becomes "run").
3. Find all N-grams in each text, where an N-gram is a continuous sequence of N words from the text.
4. Remove duplicate N-grams from one text.
5. Count how many N-grams are the same in both texts.
6. Return that number as the match score.

This produced a score reflecting the number of exact matches—i.e., words used in the same order by both the child and robot. It also produced a higher match score for texts that have both more matching phrases and longer matching phrases. We also implemented an algorithm for counting similar matches that were close to each other, but not exactly the same. This algorithm followed the same steps listed above, where step 5 (counting matching N-grams) used a fuzzy string matching algorithm to determine if N-grams matched.

For exact matches, we used $N = 3$ because a smaller N may not retain enough information to be considered actual phrase matching, while a larger N may contain more information than would comprise a single phrase. For similar matches, we used $N = 4$, so that when phrases differed by one or two words, they might still match.

For example, one of the robot's stories included the sentences, "But Turtle still couldn't find Squirrel. Eventually, it got dark out and they all got sleepy. So Squirrel had to show his hiding place." After stopword removal and stemming, this was converted to: "turtle still couldn't find squirrel eventually get dark out they all get sleepy squirrel show hiding place." One child's story included the similar section, "But he still couldn't find Squirrel. Then he bumped into him and started playing. And it's getting late out. So Squirrel had not showed his hiding place," which was converted to "he still couldn't find squirrel then he bump into him start play get late squirrel show hiding place." This segment included several exactly matching phrases, e.g., "couldn't find squirrel," as well as several similar matching phrases, e.g., (robot) "squirrel show hiding place" \ (child) "late squirrel show hiding."

3. RESULTS

First, we discuss children's vocabulary learning and information about the kinds of stories children told. Some of these results were previously reported in Kory (2014), Kory Westlund and Breazeal (2015), Kory-Westlund (2019). We also briefly discuss children's responses to the interview questions about their perception of the robot. These interviews are relevant because they showed that nearly all children reported liking the robot, and that children's liking was not identical with our measures of emulation and rapport.

Next, we present our new analyses regarding children's use of the target words and key phrases, emulation of the robot, LSM scores, and correlations among these measures. Because the new analyses were *post-hoc*, we corrected for multiple comparisons using the Benjamini Hochberg method (to control the false discovery rate), which indicated that the results with p <0.011 could be considered significant.

3.1. Interviews

As reported in Kory (2014), most children reported that they liked the game a lot (76.5%), that the robot was their friend (87.5%), that they wanted to play again (87.5%), that they liked the stories (93.6%), and that they thought the stories were interesting (93.6%), and understandable (93.6%) (**Figure 3**). There were no differences by condition.

3.2. Target Vocabulary

Across all the children, children's scores on the vocabulary assessment increased from the pretest (mean words correct = 13.4 of 24, $SD = 3.62$) to the posttest ($M = 18.9$, $SD = 2.84$), $t_{(14)} = 7.21$, p <0.001, $d = 1.7$. Children's scores increased by a mean of 5.7 words ($SD = 3.08$). Children's scores increased more in the *Matched* condition ($M = 6.91$ more words correct at the posttest, $SD = 2.51$) than in the *Unmatched* condition ($M = 2.50$, $SD = 2.08$), $t_{(13)} = 3.13$, $p = 0.008$, $d = 1.9$ (**Figure 4**).

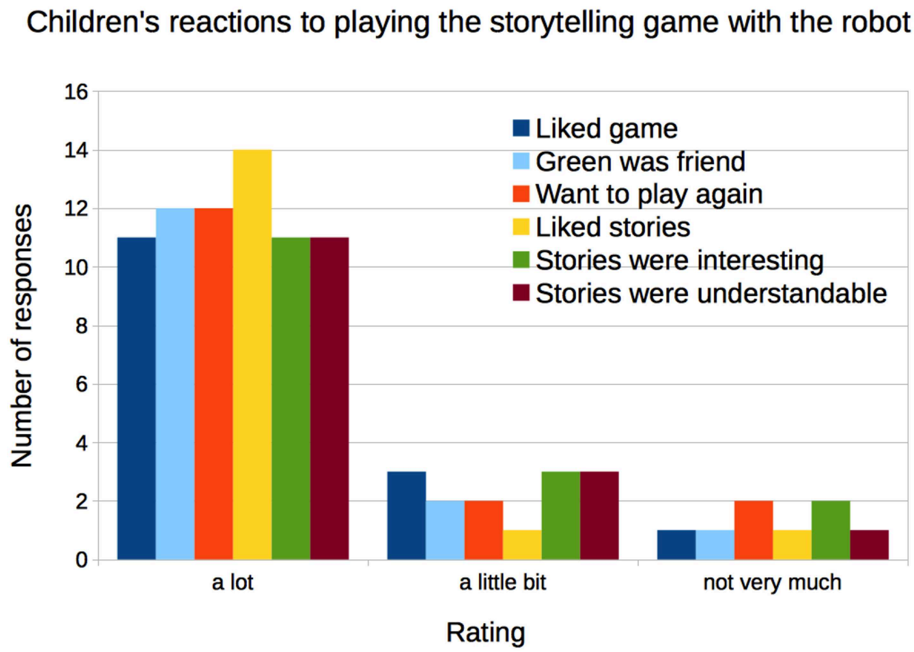


FIGURE 3 | The majority of children reported liking the robot and the storytelling game.

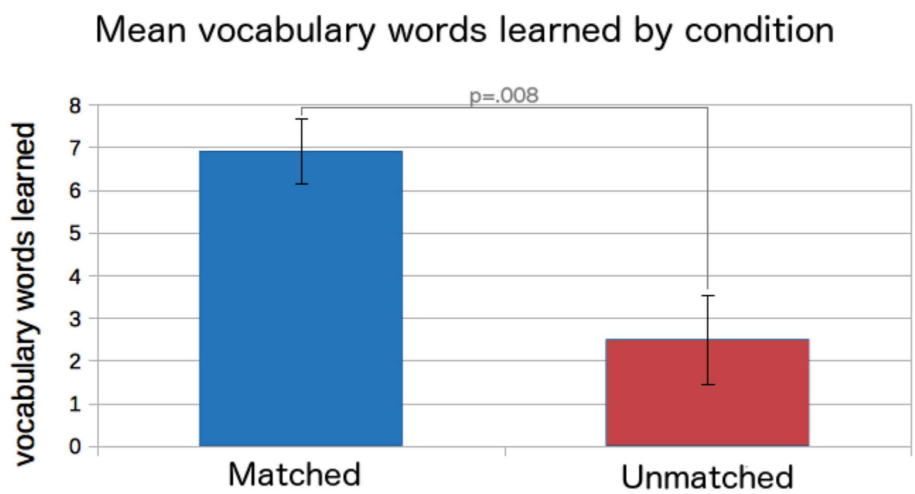


FIGURE 4 | Children's vocabulary scores increased over the study, but more so in the *Matched* condition.

3.3. Stories

Nine children told stories aloud every session. Five children told primarily silent stories, in which they spent time dragging characters on the tablet and sometimes murmuring to themselves, but not speaking aloud very often. Their stories often appeared short because only spoken words were counted. Several of these "silent tellers" began vocalizing their stories more by the final session, telling stories closer in length to the other children. Three children told no stories, though they did talk at other times.

The children who spoke aloud told 206 stories with a mean word count of 81.7 words ($SD = 77.8$). Of these, 141 stories were 20 words or longer; the shorter stories were primarily from the children who only occasionally spoke while playing the storytelling game.

Qualitatively, children covered a range of themes in their stories. We observed that children often borrowed elements from the robot's stories—such as character names and activities characters performed. For example, one of the robot's stories was about a boy named Micah, who played ball with his friends.

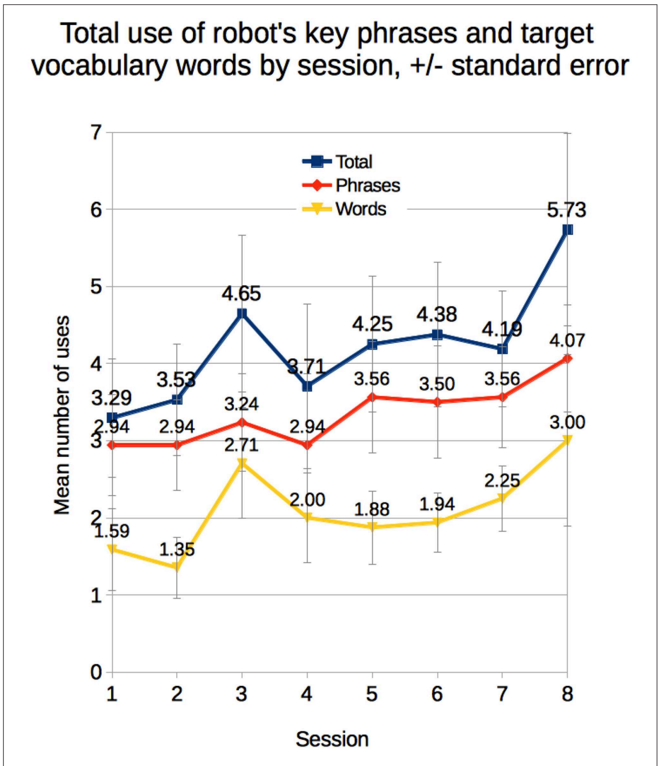


FIGURE 5 | Children's mean use of the robot's key phrases and target vocabulary words by session.

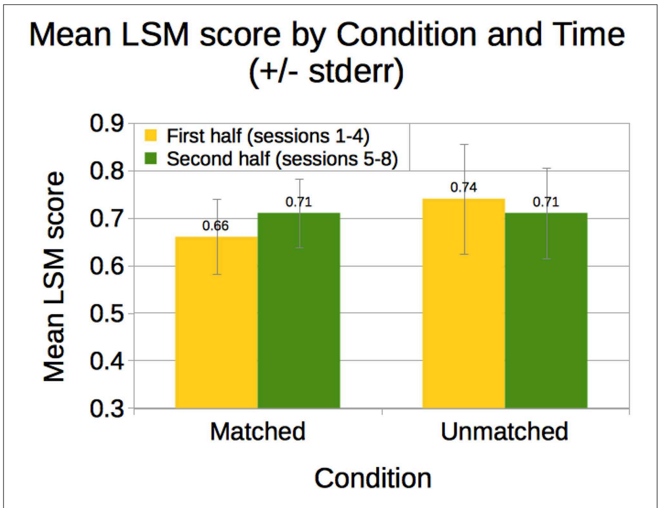


FIGURE 6 | Children's mean LSM scores by condition for the first half vs. second half of the study.

One child continued using this name and theme (XX's indicate inaudible words in the transcript):

"One time there were three friends, XX, Micah and Isabella. Micah liked going on the swings. Isabella liked going on the slide. One time they made a new friend, Daisy. She liked ball. One time she hid behind a bush until nobody saw her. Then both of the kids that were playing, approached and hid. Then, Micah slid down the slide and saw her. She stepped out but landed on the top of the brick tower. So then, they both came down together. The end."

Several children also retold versions of the robot's stories, without prompting (they were merely asked to tell a story and were not prompted with regards to content). For example, after the robot told a story about three animals that played hide-and-seek together, one child told the following story:

"Once upon a time there was a squirrel named, Squirrel, a turtle named Turtle and a rabbit named Rabbit. That particular day they played hide and seek. Squirrel hid in the mud. Turtle hid in the trees while Bunny counted. One, two, three, four. Found you! Found you, Turtle. My turn. XX behind a tree. Squirrel found Turtle. And then they played again and again. The end."

Our observations of these emulations suggested that children were, in fact, emulating the robot's stories, which was revealed quantitatively in our language eumulation results below.

3.4. Keywords and Key Phrases

We performed mixed analysis of variance with condition (between: *Matched* vs. *Unmatched*) and mean of sessions (within: sessions 1–4 vs. sessions 5–8) on children's use of the robot's target vocabulary words and key phrases. We observed a trend toward a main effect of session on the total number of key phrases and target vocabulary words children used from the first half to the second half of the study, $F_{(1,13)} = 2.95$, $p = 0.11$, $d = 0.22$ (**Figure 5**). Children used somewhat more of the key phrases and target words in the second half of the study than in the first half. In particular, children tended to use the phrases "once upon a time" and "See you later, alligator" more in later sessions.

3.5. LSM

We observed LSM scores ranging from 0.063 to 0.892, with a mean of 0.696 ($SD = 0.212$). Only two children had scores below 0.500; in both cases, their scores increased from the first half to second half of the study. A mixed analysis of variance with time (within: first half of the study vs. second half) and condition (between: *Matched* vs. *Unmatched*) revealed a trend toward an interaction of time with condition, $F_{(1,12)} = 4.29$, $p = 0.061$. As shown in **Figure 6**, LSM scores increased slightly for children in the *Matched* condition (first: $M = 0.66$, $SD = 0.25$; second: $M = 0.71$, $SD = 0.23$; $d = 0.21$); the scores decreased slightly for children in the *Unmatched* condition (first: $M = 0.74$, $SD = 0.23$; second: $M = 0.71$, $SD = 0.19$; $d = 0.14$).

3.6. Language Emulation

As described earlier, phrase matching scores were computed against all previously heard stories, only stories from the same story scene, and only the story heard just prior to the child's. We used children's phrase matching scores as a measure of language emulation. We performed mixed analysis of variance with condition (between: *Matched* vs. *Unmatched*) and mean of sessions (within: sessions 1–4 vs. sessions 5–8) for the mean of children's exact and similar phrase matching scores per story and for the sum of children's exact and similar phrase matching scores across all stories.

3.6.1. Compared to All Previously Heard Stories

We observed a trend for main effect of time on the mean number of matching phrases used per story, $F_{(1,12)} = 5.65$, $p = 0.035$, and

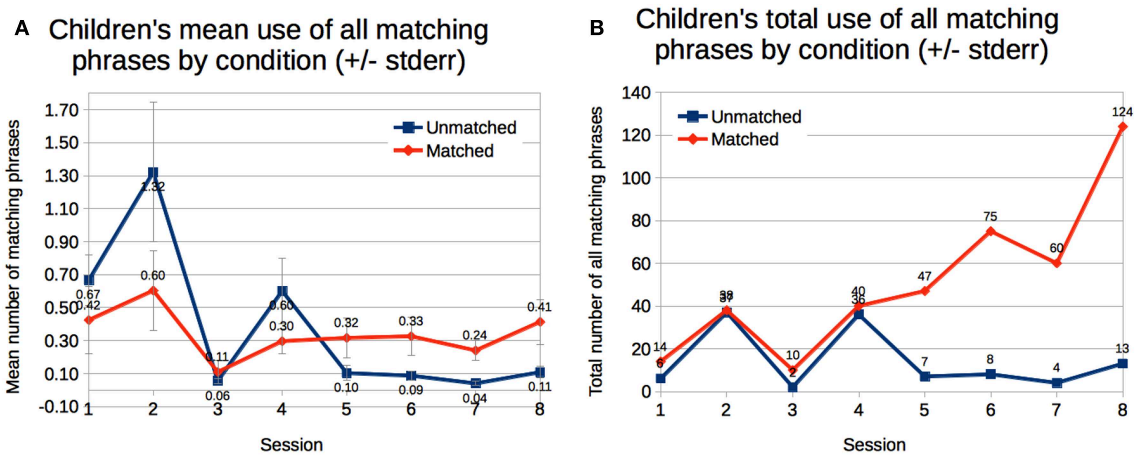


FIGURE 7 | Children emulation the robot's phrases during their storytelling. Their emulation increased during the second half of the study in the *Matched* condition. **(A)** Children's emulation decreased in the *Unmatched* condition in the second half of the study. **(B)** Children's emulation increased in the second half of the study in the *Matched* condition.

a significant interaction of time with condition, $F_{(1,12)} = 10.0$, $p = 0.008$. Children emulated more of the robot's phrases per story in the first half of the study, and children in the *Unmatched* condition decreased usage more (**Figure 7A**). We observed a significant interaction of time with condition when looking at the sum of matching phrases across stories, $F_{(1,12)} = 9.81$, $p = 0.009$. Children in the *Matched* condition increased their usage of matching phrases, while children in the *Unmatched* condition decreased their usage (**Figure 7B**).

3.6.2. Compared to Stories Heard From the Same Story Scene

We observed a significant interaction of time with condition for the mean number of matching phrases used per story, $F_{(1,12)} = 9.10$, $p = 0.011$. Children in the *Unmatched* condition used fewer matching phrases on average in the second half of the study, while children in the *Matched* condition did not change significantly. There were no significant differences for the sum of matching phrases across stories.

3.6.3. Compared to the Story Heard Just Prior

We observed a trend for an interaction of time with condition for the mean number of matching phrases used per story, $F_{(1,12)} = 4.82$, $p = 0.048$. Again, children in the *Unmatched* condition used fewer matching phrases in the second half of the study. There were no significant differences for the sum of matching phrases across stories.

3.7. Correlations

Children who emulated more of the robot's phrases during their storytelling also scored higher on the vocabulary posttest, $r_{s15} = 0.511$, $p = 0.052$ (**Figure 8A**); as did children who used more of the robot's key words and phrases $r_{s15} = 0.532$, $p = 0.041$ (**Figure 8B**). Children who emulated the robot more during storytelling were also more likely to use more of the robot's key words and phrases, $r_{s15} = 0.688$, $p = 0.003$ (**Figure 8C**). This pattern was also apparent when looking at the mean of all children's scores for sessions 1–8 (**Figure 8D**).

Children who had higher LSM scores during sessions 1–4 were more likely to emulate the robot's phrases during storytelling, $r_{s15} = 0.667$, $p = 0.007$; they were also more likely to use the robot's key words and phrases, $r_{s15} = 0.548$, p = 0.034 (**Figures 9A,B**). The same pattern held for children's LSM scores in sessions 5–8 for phrase emulation, $r_{s14} = 0.732$, $p = 0.003$; and for key word and phrase use, $r_{s14} = 0.554$, $p = 0.040$ (**Figures 9C,D**). Children's LSM scores from sessions 1–4 were strongly correlated with their LSM scores from sessions 5–8, $r_{s14} = 0.802$, $p < 0.001$, suggesting little change in children's rapport and style matching over time.

When looking at the mean of all children's scores for sessions 1–8, we observed that children who told longer stories also used more unique words ($r_{s8} = 0.954$, $p < 0.001$) and, as one might expect, spent more time telling their stories ($r_{s8} = 0.715$, $p = 0.046$; **Figure 10**).

4. DISCUSSION

We asked whether children would show greater vocabulary learning and language emulation when they showed greater rapport with a social robot with whom they played a storytelling game over time. We found some evidence supporting our hypotheses.

First, we observed that most children liked the robot, and their LSM scores reflected that liking, being reasonably high overall. We observed that children learned new vocabulary words, as evidenced by higher vocabulary posttest scores and use of the target words in their stories. This result reflects prior work in which children have learned and mirrored new vocabulary words with social robots during storytelling activities (e.g., Kory Westlund et al., 2017b; Park et al., 2017a, 2019). However, because children were exposed to the target words during the

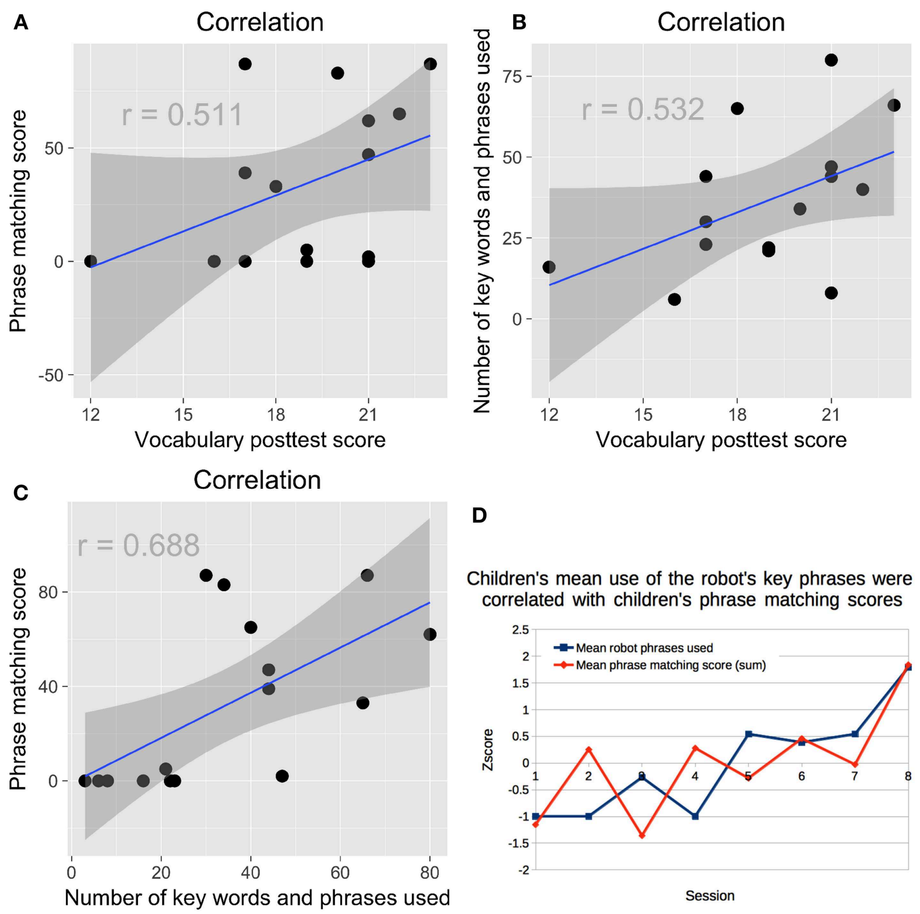


FIGURE 8 | (A) Children who emulated more of the robot's phrases during their storytelling scored higher on the vocabulary posttest. **(B)** Children who used more of the robot's key words and phrases scored higher on the vocabulary posttest. **(C)** Children who emulated more of the robot's phrases were more likely to use the robot's key words and phrases. **(D)** Children's use of the robot's key words and phrases was correlated with their emulation of the robot's language over time.

pretest, it is possible that the pretest posed a first learning opportunity, and that they learned somewhat fewer words with the robot than the posttest indicates.

In partial support of H1, we observed that children's LSM scores were positively related to their use of the robot's key words and phrases. However, contrary to our expectations, LSM scores were not significantly related to children's vocabulary test scores.

This may be for several reasons. First, because the sessions with the robot were fairly short (10–15min) and because not all children told long stories, the amount of conversation between the robot and child was limited. As such, the amount of data used to compute the LSM scores was limited, and the LSM scores should be interpreted with a degree of caution. Second, children's LSM scores may not perfectly reflect rapport. Prior work linked higher LSM scores between two people to higher rapport and a deeper relationship (e.g., Pennebaker et al., 2003; Ireland et al., 2011; Babcock et al., 2014), but this work has primarily been done with adults, not children. Third, we do not know exactly how rapport affects learning, and thus, the causal connection between rapport and learning seen in earlier work in human-human peer tutoring (Sinha and Cassell, 2015a,b) may not appear with younger children in a language learning context. Rapport may not necessarily directly impact learning; it may be, for example, that rapport increases emulation of various behaviors, which in some contexts could increase learning, or that rapport facilitates being in a more positive state of mind, which perhaps leads to more engagement and learning. Furthermore, rapport may play a different role in peer learning with social robots than in other contexts with humans.

In our analyses here, we did observe that children's LSM scores correlated positively with their emulation of the robot during storytelling, as expected (H2). This suggests that rapport is linked to emulation, which is in line with prior work showing that people will mirror a variety of different behaviors in others with whom they have high rapport (e.g., Tickle-Degnen and Rosenthal, 1990; Chisholm and Strayer, 1995; Dijksterhuis and Bargh, 2001; Rotenberg et al., 2003; Dijksterhuis, 2005; Chartrand and van Baaren, 2009; Wiltermuth and Heath, 2009; Lubold, 2017).

In addition, we saw that children's emulation of the robot's language was positively correlated with their vocabulary scores, supporting H3. Children who correctly identified more of the

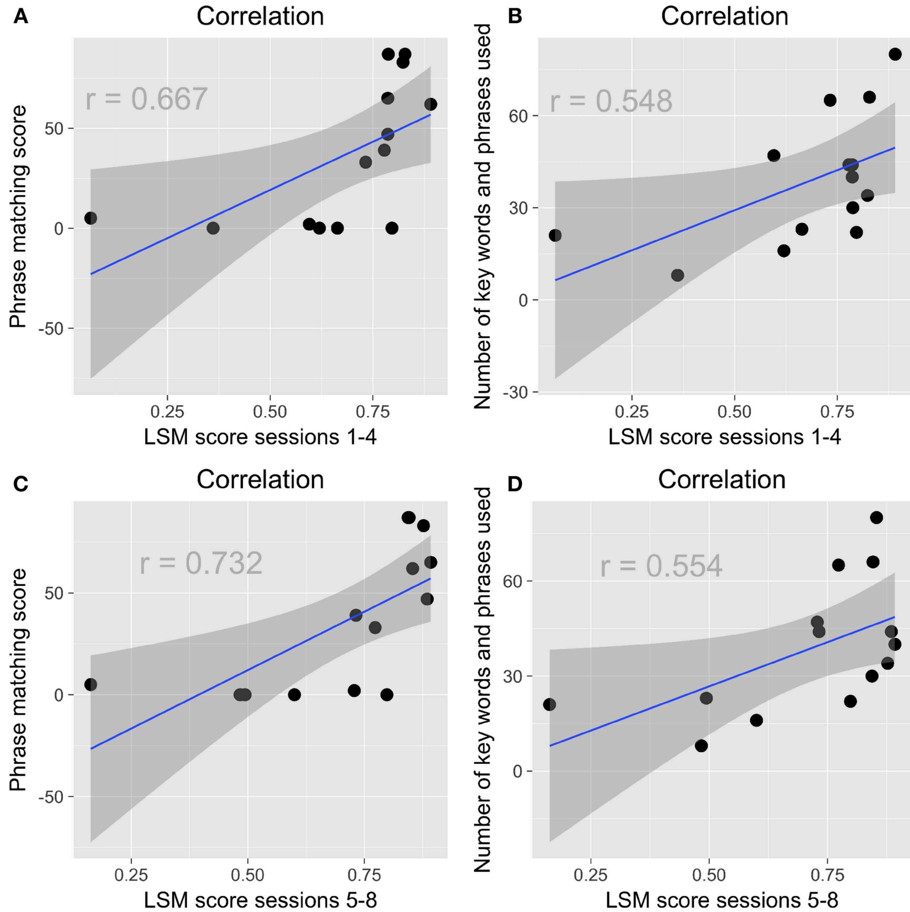


FIGURE 9 | (A) In the first half of the study, children who had higher LSM scores were more likely to emulate the robot's phrases. **(B)** In the first half of the study, children who had higher LSM scores were more likely to use the robot's key words and phrases. **(C)** In the second half of the study, children who had higher LSM scores were more likely to emulate the robot's phrases. **(D)** In the second half of the study, children who had higher LSM scores were more likely to use the robot's key words and phrases.

target words on the receptive vocabulary test were also more likely to expressively use the words in their stories. These results suggest that children's emulation was related to their learning—perhaps their rapport with the robot led to greater emulation, and greater emulation was indicative of greater word learning. This would be worth investigating in a systematic way in follow-up work.

We find partial support for H4: When examining children's behavior over time, we saw that children slightly increased their use of the robot's keywords and phrases from the first half of the study to the second half. However, children's overall emulation decreased over time, while their use of unique words increased. It may be that children were more creative over time when telling stories, making up their own that drew less on the robot's stories for inspiration. The storytelling activity was designed to facilitate language development, so both creatively using language as well as imitating the robot's language were beneficial outcomes. Story re-telling (i.e., intentionally imitating another's storytelling) has often been used as an educational activity for helping children learn stories and vocabulary (e.g., Isbell, 2002; Dunst et al., 2012; Kory Westlund et al., 2017b; Otwinowska et al., 2018; Kory-Westlund and Breazeal, 2019b).

Children's LSM scores, on average, did not show a strong increase over time (there were differences by condition, as discussed further below). This could indicate little increase in rapport, or could mean that LSM is not sufficiently sensitive to capture children's changes in rapport over the study.

Children's LSM scores and phrase emulation during storytelling increased over time for children in the *Matched* condition, but decreased slightly for children in the *Unmatched* condition. Children in the *Matched* scoring also had higher scores on the vocabulary posttest. These results provide some support for H5; however, given the small sample size, these results should be interpreted with caution. The robot's story level personalization appeared to positively impact children's emulation of the robot's language, their rapport as indexed by LSM, and their vocabulary learning. This is in line with prior work showing links between a robot's personalization

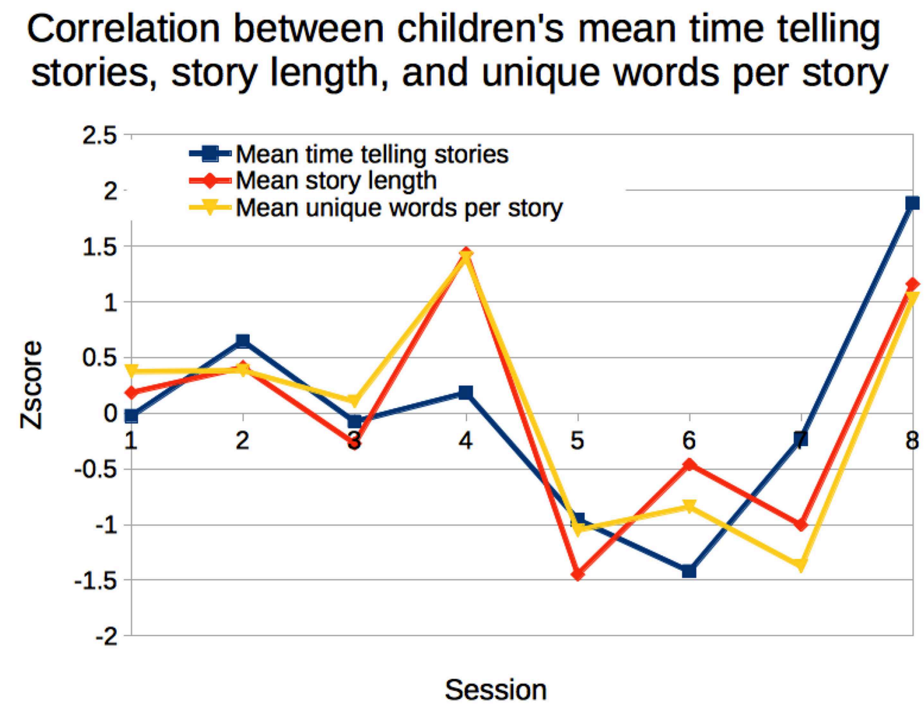


FIGURE 10 | Children who told longer stories also used more unique words and spent more time telling their stories.

and children's engagement and learning (e.g., Leite et al., 2012; Gordon et al., 2016; Palestra et al., 2016; Scassellati et al., 2018; Park et al., 2019)

However, in addition to the small sample size, the two conditions were not fully balanced. There were more children in the *Matched* condition and there was only one boy in the *Unmatched* condition. In addition, although children were assigned to conditions prior to the start of the robot interaction using their initial language assessment scores to attempt to balance language ability across conditions, we did observe somewhat higher scores for children in the *Unmatched* condition across various metrics during the first half of the study (prior to the robot's personalization/matching, which only occurred in the second half of the study). We expect that were the groups more balanced, these initial differences may be smaller or might even disappear, while differences between conditions as a result of the personalization would be larger.

Taken together, our results suggest that first, interacting with a more advanced peer-like social robot can be beneficial for children's language learning. This is in line with work examining children's language learning with human peers (Fuchs et al., 1997; Mathes et al., 1998; Topping, 2005; Schechter and Bye, 2007; Whitebread et al., 2007; Mashburn et al., 2009; Justice et al., 2011; DeLay et al., 2016; Lin et al., 2016). Second, children's emulation of the robot's language may be related to their rapport and to their learning. Earlier work has shown that children will emulate the behavior of social robots—including mirroring expressiveness (Spaulding et al., 2016), curiosity (Gordon et al., 2015), and language (Kory Westlund et al., 2017b)—but had not yet explored mechanisms that might affect children's emulation and peer learning. Our results suggest that rapport may be one such mechanism. This is the first study we know of to empirically support that rapport may indeed be a modulating factor in children's peer learning.

Finally, this study highlights new opportunities we have for using social robots as interventions for early language development, specifically by leveraging this connection between rapport and learning.

4.1. Limitations

This study had several limitations. First, as mentioned earlier, the sample size was fairly small and conditions were unbalanced in number. As such, the statistical power of our analyses are underpowered. In addition, children's individual differences were not controlled for, such as learning ability or socio-economic status. These factors may all influence children's learning and social interactions with the robot. Future work should attempt to recruit a more balanced, homogeneous sample and explore the stability of the results across individual differences.

The target vocabulary words presented in the robot's stories included some words that were known by numerous children at the start of the study (as reported above, children identified a mean of 13.4 of 24 words correctly at the pretest, $SD = 3.62$). The difference between children's vocabulary scores on the pretest vs. the posttest did show that children knew more of the words at the end of the study, but because a set of common words and not nonce words were used, we cannot know for sure that children learned these words as a result of the robot interaction or because of other events that occurred during the two months during which the study took place.

Another limitation of the dataset was the lack of additional assessments of relationship and rapport. We used children's LSM scores as a measure of rapport, since numerous prior studies have linked higher LSM scores between two people to higher rapport and a deeper relationship (e.g., Pennebaker et al., 2003; Ireland et al., 2011; Babcock et al., 2014). However, future work should endeavor to measure children's rapport and relationship with the robot in additional ways, e.g., using measures presented in Kory-Westlund et al. (2018) and Kory-Westlund and Breazeal (2019a).

Finally, this study explored a one-on-one interaction with the robot. However, children often learn with others—friends, siblings, parents, and teachers. Future work should explore group interactions that include multiple children or children with parents, caregivers, and teachers. This could give us insight into how to integrate robots into real-world educational contexts, such as schools and homes.

Despite these limitations, we did see numerous correlations and differences that are suggestive of links between children's learning, rapport, and language emulation. While these results are exploratory and not definitive, they do provide evidence that this in an area that warrants further study.

DATA AVAILABILITY

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

JK-W and CB: the study was conceived, designed, the paper was drafted, written, revised, and approved. JK-W: data analysis was performed.

FUNDING

This research was supported by an MIT Media Lab Learning Innovation Fellowship and the National Science Foundation (NSF) under Grants 122886, CCF-1138986, and Graduate Research Fellowship under Grant No. 1122374. Any opinions, findings and conclusions, or recommendations expressed in this paper are those of the authors and do not represent the views of the NSF.

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. doi: 10.1142/S0219843614500224

Babcock, M. J., Ta, V. P., and Ickes, W. (2014). Latent semantic similarity and language style matching in initial dyadic interactions. *J. Lang. Soc. Psychol.* 33, 78–88. doi: 10.1177/0261927X13499331

Bandura, A. (1971). *Social Learning Theory*. Morristown, NJ: General Learning Press.

Bandura, A., and Walters, H. (1963). *Social Learning and Personality Development*. New York, NY: Holt Rinehart and Winston.

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., and Tanaka, F. (2018). Social robots for education: a review. *Sci. Robot.* 3:eaat5954. doi: 10.1126/scirobotics.aat5954

Bers, M. U., Ackermann, E., Cassell, J., Donegan, B., Gonzalez-Heydrich, J., DeMaso, D. R., et al. (1998). "Interactive storytelling environments: coping with cardiac illness at Boston's Children's Hospital," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Los Angeles, CA: ACM Press/Addison-Wesley Publishing Co.), 603–610. doi: 10.1145/274644.274725

Biemiller, A. (2010). *Words Worth Teaching: Closing the Vocabulary Gap*. Columbus, OH: McGraw-Hill SRA .

Bloom, L. (1974). "Talking, understanding, and thinking: developmental relationship between receptive and expressive language," in *Language Perspectives: Acquisition, Retardation and Intervention*, eds R. Schiefelbusch and L. Lloyd (Baltimore, MD: University Park Press), 285–311.

Bloom, P. (2000). *How Children Learn the Meanings of Words*. Cambridge, MA: MIT press.

Borrie, S. A., and Liss, J. M. (2014). Rhythm as a coordinating device: entrainment with disordered speech. *J. Speech Lang. Hear. Res.* 57, 815–824. doi: 10.1044/2014-JSLHR-S-13-0149

Breazeal, C., Dautenhahn, K., and Kanda, T. (2016). "Social robotics," in *Springer Handbook of Robotics*, eds B. Siciliano and O. Khatib (New York, NY: Springer International Publishing), 1935–1972.

Cassell, J. (2004). Towards a model of technology and literacy development: story listening systems. *J. Appl. Dev. Psychol.* 25, 75–105. doi: 10.1016/j.appdev.2003.11.003

Cassell, J., and Ryokai, K. (2001). Making space for voice: technologies to support children's fantasy and storytelling. *Pers. Ubiquit. Comput.* 5, 169–190. doi: 10.1007/PL00000018

Cassell, J., Tartaro, A., Rankin, Y., Oza, V., Tse, C., and Searle, F. (2007). Virtual peers for literacy learning. *Educ. Technol. Saddle Brook Then Englewood Cliffs NJ*, 47, 39–43. Available online at: https://www.jstor.org/stable/44429376

Chang, C.-W., Lee, J.-H., Chao, P.-Y., Wang, C.-Y., and Chen, G.-D. (2010). Exploring the possibility of using humanoid robots as instructional tools for teaching a second language in primary school. *J. Educ. Techn. Soc.* 13, 13–24. Available online at: https://www.jstor.org/stable/jeductechsoci.13.2.13

Chartrand, T. L., and van Baaren, R. (2009). Human mimicry. *Adv. Exp. Soc. Psychol.* 41, 219–274. doi: 10.1016/S0065-2601(08)00405-X

Chisholm, K., and Strayer, J. (1995). Verbal and facial measures of children′s emotion and empathy. *J. Exp. Child Psychol.* 59, 299–316.

Christophel, D. M. (1990). The relationships among teacher immediacy behaviors, student motivation, and learning. *Commun. Educ.* 39, 323–340.

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. New York, NY: Harper and row.

Curenton, S. M., Craig, M. J., and Flanigan, N. (2008). Use of decontextualized talk across story contexts: how oral storytelling and emergent reading can scaffold children's development. *Early Educ. Dev.* 19, 161–187. doi: 10.1080/10409280701839296

De Lisi, R., and Golbeck, S. L. (1999). "Implications of Piagetian theory for peer learning," in *Cognitive Perspectives on Peer Learning, The Rutgers Invitational Symposium On Education Series*, eds A. M. O, and A. King (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 3–37.

DeLay, D., Hanish, L. D., Martin, C. L., and Fabes, R. A. (2016). Peer effects on Head Start children's preschool competency. *Dev. Psychol.* 52, 58–70. doi: 10.1037/dev0000066

Deshmukh, A., Jones, A., Janarthanam, S., Hastie, H., Ribeiro, T., Aylett, R., et al. (2015). "An empathic robotic tutor in a map application," in *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems* (Istanbul: International Foundation for Autonomous Agents and Multiagent Systems), 1923–1924.

Dijksterhuis, A. (2005). "Why we are social animals: the high road to imitation as social glue," in *Perspectives on Imitation: From Neuroscience to Social Science*, Vol. 2, eds S. Hurley and N. Chater (Cambridge, MA: MIT Press), 207–220.

Dijksterhuis, A., and Bargh, J. A. (2001). The perception-behavior expressway: automatic effects of social perception on social behavior. *Adv. Exp. Soc. Psychol.* 33, 1–40. doi: 10.1016/S0065-2601(01)80003-4

Dunst, C. J., Simkus, A., and Hamby, D. W. (2012). Children's story retelling as a literacy and language enhancement strategy. *Center Early Literacy Learn.* 5, 1–14.

Engel, S. (1995). *The Stories Children Tell: Making Sense of the Narratives of Childhood*. New York, NY: Macmillan.

Fish, M., and Pinkerman, B. (2003). Language skills in low-SES rural Appalachian children: Normative development and individual differences, infancy to preschool. *J. Appl. Dev. Psychol.* 23, 539–565. doi: 10.1016/S0193-3973(02)00141-7

Fuchs, D., Fuchs, L. S., Mathes, P. G., and Simmons, D. C. (1997). Peer-assisted learning strategies: making classrooms more responsive to diversity. *Am. Educ. Res. J.* 34, 174–206.

Gola, A. A. H., Richards, M. N., Lauricella, A. R., and Calvert, S. L. (2013). Building meaningful parasocial relationships between toddlers and media characters to teach early mathematical skills. *Media Psychol.* 16, 390–411. doi: 10.1080/15213269.2013.783774

Gordon, G., and Breazeal, C. (2015). "Bayesian active learning-based robot tutor for children's word-reading skills," in *Proceedings of the 29th AAAI Conference on Artificial Intelligence* (Austin, TX).

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

Gordon, G., Spaulding, S., Kory Westlund, J., Lee, J. J., Plummer, L., Martinez, M., et al. (2016). "Affective personalization of a social robot tutor for children's second language skill," in *Proceedings of the 30th AAAI Conference on Artificial Intelligence* (Palo Alto, CA).

Griffin, T. M., Hemphill, L., Camp, L., and Wolf, D. P. (2004). Oral discourse in the preschool years and later literacy skills. *First Language* 24, 123–147. doi: 10.1177/0142723704042369

Hart, B., and Risley, T. R. (1995). *Meaningful Differences in the Everyday Experience of Young American Children.* Baltimore, MD: ERIC.

Hoff, E. (2006). How social contexts support and shape language development. *Dev. Rev.* 26, 55–88. doi: 10.1016/j.dr.2005.11.002

Hood, D., Lemaignan, S., and Dillenbourg, P. (2015). "When children teach a robot to write: an autonomous teachable humanoid which uses simulated handwriting," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI '15 (New York, NY: ACM), 83–90.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., and Lyons, T. (1991). Early vocabulary growth: relation to language input and gender. *Dev. Psychol.* 27, 236–248.

Huttenlocher, J., Vasilyeva, M., Cymerman, E., and Levine, S. (2002). Language input and child syntax. *Cogn. Psychol.* 45, 337–374. doi: 10.1016/S0010-0285(02)00500-5

Huttenlocher, J., Vasilyeva, M., and Shimpi, P. (2004). Syntactic priming in young children. *J. Memory Language* 50, 182–195. doi: 10.1016/j.jml.2003.09.003

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., and Hedges, L. V. (2010). Sources of variability in children's language growth. *Cogn. Psychol.* 61, 343–365. doi: 10.1016/j.cogpsych.2010.08.002

Ingram, D. (1974). "The relationship between comprehension and production," in *Language Perspectives: Acquisition, Retardation, and Intervention,* eds R. Schiefelbusch and L. Lloyd (Baltimore, MD: University Park Press), 670.

Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., and Pennebaker, J. W. (2011). Language style matching predicts relationship initiation and stability. *Psychol. Sci.* 22, 39–44. doi: 10.1177/0956797610392928

Isbell, R. T. (2002/00/00). Telling and retelling stories: learning language and literacy. Supporting language learning. *Young Children* 57, 26–30. Available online at: https://eric.ed.gov/?id=EJ656302?

Jipson, J. L., and Gelman, S. A. (2007). Robots and rodents: children's inferences about living and nonliving kinds. *Child Dev.* 78, 1675–1688. doi: 10.1111/j.1467-8624.2007.01095.x

Justice, L. M., Petscher, Y., Schatschneider, C., and Mashburn, A. (2011). Peer effects in preschool classrooms: is children's language growth associated with their classmates' skills? *Child Dev.* 82, 1768–1777. doi: 10.1111/j.1467-8624.2011.01665.x

Kahn, P. H., Kanda, T., Ishiguro, H., Freier, N. G., Severson, R. L., Gill, B. T., et al. (2012). "Robovie, you'll have to go into the closet now": children's social and moral relationships with a humanoid robot. *Dev. Psychol.* 48:303. doi: 10.1037/a0027033

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/s15327051hci19012-4

Kanda, T., Sato, R., Saiwaki, N., and Ishiguro, H. (2007). A two-month field trial in an elementary school for long-term Human #x2013;robot interaction. *IEEE Trans. Robot.* 23, 962–971. doi: 10.1109/TRO.2007.904904

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

Kanero, J., Geçkin, V., Oranç, C., Mamus, E., Küntay, A. C., and Göksun, T. (2018). Social Robots for early language learning: current evidence and future directions. *Child Dev. Perspect.* 12, 146–151. doi: 10.1111/cdep.12277

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). "Social robot tutoring for child second language learning," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Christchurch: IEEE), 231–238.

Kidd, C. D., and Breazeal, C. (2008). "Robots at home: understanding long-term human-robot interaction," in *IROS 2008. IEEE/RSJ International Conference On Intelligent Robots and Systems, 2008* (Nice: IEEE), 3230–3235.

Kory Westlund, J. M., and Breazeal, C. (2015). "The interplay of robot language level with children's language learning during storytelling," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, HRI'15 Extended Abstracts (New York, NY: ACM), 65–66.

Kory Westlund, J. M., Dickens, L., Jeong, S., Harris, P. L., DeSteno, D., and Breazeal, C. L. (2017a). Children use non-verbal cues to learn new words from robots as well as people. *Int. J. Child Comput. Interact.* 13, 1–9. doi: 10.1016/j.ijcci.2017.04.001

Kory Westlund, J. M., Jeong, S., Park, H. W., Ronfard, S., Adhikari, A., Harris, P. L., et al. (2017b). Flat versus expressive storytelling: young children's learning and retention of a social robot's narrative. *Front. Hum. Neurosci.* 11:295. doi: 10.3389/fnhum.2017.00295

Kory, J. (2014). *Storytelling with Robots: Effects of Robot Language Level on Children's Language Learning*. Master's Thesis, Cambridge, MA: Massachusetts Institute of Technology,

Kory, J., and Breazeal, C. (2014). "Storytelling with robots: learning companions for preschool children's language development," in *2014 RO-MAN: The 23rd IEEE International Symposium on Robot and Human Interactive Communication* (Edinburgh), 643–648.

Kory, J. M., Jeong, S., and Breazeal, C. L. (2013). "Robotic learning companions for early language development," in *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*, eds J. In, F. Epps, S. Chen, and K. Oviatt Mase (New York, NY: ACM), 71–72.

Kory-Westlund, J. M. (2019). *Relational AI: Creating Long-Term Interpersonal Interaction, Rapport, and Relationships with Social Robots.* Ph.D. Thesis, Cambridge, MA: Massachusetts Institute of Technology.

Kory-Westlund, J. M., and Breazeal, C. (2019a). "Assessing children's perceptions and acceptance of a social robot," in *Proceedings of the 18th ACM Conference on Interaction Design and Children* (Boise, ID: ACM).

Kory-Westlund, J. M., and Breazeal, C. (2019b). Exploring the effects of a social robot's speech entrainment and backstory on young children's emotion, rapport, relationship, and learning. *Front. Robot. AI* 6:54. doi: 10.3389/frobt.2019.00054

Kory-Westlund, J. M., Park, H. W., Williams, R., and Breazeal, C. (2018). "Measuring young children's long-term relationships with social robots," in

Proceedings of the 17th ACM Conference on Interaction Design and Children (Trondheim: ACM), 207–218.

Kuhl, P. K. (2011). "Social mechanisms in early language acquisition : understanding integrated brain systems supporting language," in *The Oxford Handbook of Social Neuroscience*, eds J. Decety and J. T. Cacioppo (Oxford, UK: Oxford University Press), 649–667.

Lee, S., Noh, H., Lee, J., Lee, K., Lee, G. G., Sagong, S., and Kim, M. (2011). On the effectiveness of robot-assisted language learning. *ReCALL* 23, 25–58. doi: 10.1017/S0958344010000273

Leite, I., Pereira, A. A., Castellano, G., Mascarenhas, S., Martinho, C., and Paiva, A. (2012). "Modelling empathy in social robotic companions," in *Advances in User Modeling,* eds L. Ardissono and T. Kuflik (New York, NY: Springer), 135–147.

Lin, T.-J., Justice, L. M., Paul, N., and Mashburn, A. J. (2016). Peer interaction in rural preschool classrooms: contributions of children's learning-related behaviors, language and literacy skills, and problem behaviors. *Early Childhood Res. Quart.* 37, 106–117. doi: 10.1016/j.ecresq.2016.04.001

Lubold, N. (2017). "Building rapport through dynamic models of acoustic-prosodic entrainment," in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17 (New York, NY: ACM), 297–300.

Lubold, N., Walker, E., and Pon-Barry, H. (2016). "Effects of voice-adaptation and social dialogue on perceptions of a robotic learning companion," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, HRI '16 (Piscataway, NJ: IEEE Press), 255–262.

Lubold, N., Walker, E., Pon-Barry, H., and Ogan, A. (2018). "Automated pitch convergence improves learning in a social, teachable robot for middle school mathematics," in *Artificial Intelligence in Education, Lecture Notes in Computer Science*, eds C. Penstein Rosé, R. Martínez-Maldonado, H. U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta (New York, NY: Springer International Publishing), 282–296.

Mashburn, A. J., Justice, L. M., Downer, J. T., and Pianta, R. C. (2009). Peer effects on children's language achievement during pre-kindergarten. *Child Dev.* 80, 686–702. doi: 10.1111/j.1467-8624.2009.01291.x

Mathes, P. G., Howard, J. K., Allen, S. H., and Fuchs, D. (1998). Peer-assisted learning strategies for first-grade readers: responding to the needs of diverse learners. *Read. Res. Quart.* 33, 62–94.

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

Meltzoff, A. N., Kuhl, P. K., Movellan, J., and Sejnowski, T. J. (2009). Foundations for a new science of learning. *Science* 325, 284–288. doi: 10.1126/science.1175626

Movellan, J., Eckhardt, M., Virnes, M., and Rodriguez, A. (2009). "Sociable robot improves toddler vocabulary skills," in *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction* (San Diego, CA: ACM), 307–308.

Nicolopoulou, A. (1993). Play, cognitive development, and the social world: piaget, Vygotsky, and beyond. *Hum. Dev.* 36, 1–23.

Niederhoffer, K. G., and Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *J. Lang. Soc. Psychol.* 21, 337–360. doi: 10.1177/026192702237953

Otwinowska, A., Mieszkowska, K., Bia\lecka-Pikul, M., Opacki, M., and Haman, E. (2018). Retelling a model story improves the narratives of Polish-English bilingual children. *Int. J. Biling. Educ. Biling.* 1–25. doi: 10.1080/13670050.2018.1434124

Páez, M. M., Tabors, P. O., and López, L. M. (2007). Dual language and literacy development of Spanish-speaking preschool children. *J. Appl. Dev. Psychol.* 28, 85–102. doi: 10.1016/j.appdev.2006.12.007

Palestra, G., Varni, G., Chetouani, M., and Esposito, F. (2016). "A multimodal and multilevel system for robotics treatment of autism in children," in *Proceedings of the International Workshop on Social Learning and Multimodal Interaction for Designing Artificial Agents*, DAA '16 (New York, NY: ACM) 3, 1–3:6.

Park, H. W., Gelsomini, M., Lee, J. J., and Breazeal, C. (2017a). "Telling stories to robots: the effect of backchanneling on a child's storytelling," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '17 (New York, NY: ACM), 100–108.

Park, H. W., Grover, I., Spaulding, S., Gomez, L., and Breazeal, C. (2019). "A model-free affective reinforcement learning approach to personalization of an autonomous social robot companion for early literacy education," in *Proceedings of AAAI Conference on Artificial Intelligence,* Vol. 33 (Honolulu, HI: AAAI), 687–694.

Park, H. W., Rosenberg-Kima, R., Rosenberg, M., Gordon, G., and Breazeal, C. (2017b). "Growing growth mindset with a social robot peer," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '17 (New York, NY: ACM), 137–145,

Pennebaker, J. W., Mehl, M. R., and Niederhoffer, K. G. (2003). Psychological aspects of natural language use: our words, our selves. *Ann. Rev. Psychol.* 54, 547–577. doi: 10.1146/annurev.psych.54.101601.145041

Perkins, S. C., Finegood, E. D., and Swain, J. E. (2013). Poverty and language development: roles of parenting and stress. *Innovat. Clin. Neurosci.* 10, 10–19.

Piaget, J. (1932). *The Moral Development of the Child*. London: Kegan Paul.

Porzel, R., Scheffler, A., and Malaka, R. (2006). "How entrainment increases dialogical effectiveness," in *Proceedings of the IUI*, eds C. Paris and C. L. Sidner (Sydney, NSW: Citeseer), 35–42.

Reitter, D., Keller, F., and Moore, J. D. (2011). A computational cognitive model of syntactic priming. *Cogn. Sci.* 35, 587–637. doi: 10.1111/j.1551-6709.2010.01165.x

Richards, M. N., and Calvert, S. L. (2017). "Media characters, parasocial relationships, and the social aspects of children's learning across media platforms," in *Media Exposure During Infancy and Early Childhood: The Effects of Content and Context on Learning and Development*, eds R. Barr and D. N. Linebarger (Cham: Springer International Publishing), 141–163.

Rintjema, E., van den Berghe, R., Kessels, A., de Wit, J., and Vogt, P. (2018). "A robot teaching young children a second language: the Effect of multiple interactions on engagement and performance," in *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '18 (New York, NY: ACM), 219–220.

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

Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., et al. (2018a). Beyond the 30-million-word gap: children's conversational exposure is associated with language-related brain function. *Psychol. Sci.* 29, 700–710. doi: 10.1177/0956797617742725

Romeo, R. R., Segaran, J., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., et al. (2018b). Language exposure relates to structural neural connectivity in childhood. *J. Neurosci.* 38, 7870–7877. doi: 10.1523/JNEUROSCI.0484-18.2018

Rotenberg, K. J., Eisenberg, N., Cumming, C., Smith, A., Singh, M., and Terlicher, E. (2003). The contribution of adults' nonverbal cues and children's shyness to the development of rapport between adults and preschool children. *Int. J. Behav. Dev.* 27, 21–30. doi: 10.1080/01650250143000571

Rubin, K. H., Bukowski, W. M., and Parker, J. G. (1998). "Peer interactions, relationships, and groups," in *Handbook of Child Psychology,* ed N. Eisenberg (Hoboken, NJ: John Wiley & Sons, Inc.)

Ryokai, K., Vaucelle, C., and Cassell, J. (2003). Virtual peers as partners in storytelling and literacy learning. *J. Comput. Assis. Learn.* 19, 195–208. doi: 10.1046/j.0266-4909.2003.00020.x

Sage, K. D., and Baldwin, D. (2010). Social gating and pedagogy: mechanisms for learning and implications for robotics. *Neural Netw.* 23, 1091–1098. doi: 10.1016/j.neunet.2010.09.004

Scassellati, B., Boccanfuso, L., Huang, C.-M., Mademtzi, M., Qin, M., Salomons, N., et al. (2018). Improving social skills in children with ASD using a long-term, in-home social robot. *Sci. Robot.* 3:eaat7544. doi: 10.1126/scirobotics.aat7544

Schechter, C., and Bye, B. (2007). Preliminary evidence for the impact of mixed-income preschools on low-income children's language growth. *Early Childhood Res. Quart.* 22, 137–146. doi: 10.1016/j.ecresq.2006.11.005

Schwab, J. F., and Lew-Williams, C. (2016). Language learning, socioeconomic status, and child-directed speech. *Wiley Interdiscipl. Rev. Cogn. Sci.* 7, 264–275. doi: 10.1002/wcs.1393

Sénéchal, M. (1997). The differential effect of storybook reading on preschoolers' acquisition of expressive and receptive vocabulary. *J. Child Lang.* 24, 123–138.

Serholt, S., Barendregt, W., Leite, I., Hastie, H., Jones, A., Paiva, A., et al. (2014). "Teachers' views on the use of empathic robotic tutors in the classroom," in *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium On* (Edinburgh: IEEE), 955–960.

Setapen, A. M. (2012). *Creating Robotic Characters for Long-Term Interaction*. Master's Thesis, Cambridge, MA: MIT.

Sinha, T., and Cassell, J. (2015a). "Fine-grained analyses of interpersonal processes and their effect on learning," in *Artificial Intelligence in Education,* eds C. P. Rosé, R. Martínez-Maldonado, H. U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren, and B. du Boulay (Cham: Springer), 781–785.

Sinha, T., and Cassell, J. (2015b). "We click, we align, we learn: impact of influence and convergence processes on student learning and rapport building," in *Proceedings of the 1st Workshop on Modeling INTERPERsonal SynchrONy And infLuence*, INTERPERSONAL '15 (New York, NY: ACM), 13–20,

Snow, C. E., Porche, M. V., Tabors, P. O., and Harris, S. R. (2007). *Is Literacy Enough? Pathways to Academic Success for Adolescents.* Baltimore, MD: ERIC.

Spaulding, S., Gordon, G., and Breazeal, C. (2016). "Affect-aware student models for robot tutors," in *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, AAMAS '16 (Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems), 864–872.

Tanaka, F., Isshiki, K., Takahashi, F., Uekusa, M., Sei, R., and Hayashi, K. (2015). 'Pepper learns together with children: development of an educational application," in *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)* (Seoul), 270–275.

Tanaka, F., and Kimura, T. (2009). "The use of robots in early education: a scenario based on ethical consideration," in *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication* (Toyama), 558–560.

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

Tausczik, Y. R., and Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *J. Lang. Soc. Psychol.* 29, 24–54. doi: 10.1177/0261927X09351676

Tickle-Degnen, L., and Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychol. Inquiry* 1, 285–293.

Topping, K. J. (2005). Trends in Peer Learning. *Educ. Psychol.* 25, 631–645. doi: 10.1080/01443410500345172

Tudge, J., and Rogoff, B. (1989). "Peer influences on cognitive development: piagetian and Vygotskian perspectives," in *Interaction in Human Development, Crosscurrents in Contemporary Psychology*, eds M. H. Bornstein and J. S. Bruner (Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.), 17–40.

Vogt, P., de Haas, M., de Jong, C., Baxter, P., and Krahmer, E. (2017). Child-robot interactions for second language tutoring to preschool children. *Front. Hum. Neurosci.* 11:73. doi: 10.3389/fnhum.2017.00073

Vogt, P., van den Berghe, R., de Haas, M., Hoffmann, L., Kanero, J., Mamus, E., et al. (2019). "Second language tutoring using social robots. A large-scale study," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Daegu). doi: 10.1109/HRI.2019.8673077

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Wentzel, K. R. (1997). Student motivation in middle school: the role of perceived pedagogical caring. *J. Educ. Psychol.* 89, 411–19.

Whitebread, D., Bingham, S., Grau, V., Pasternak, D. P., and Sangster, C. (2007). Development of metacognition and self-regulated learning in young children: role of collaborative and peer-assisted learning. *J. Cogn. Educ. Psychol.* 6, 433–455. doi: 10.1891/194589507787382043

Wiltermuth, S. S., and Heath, C. (2009). Synchrony and cooperation. *Psychol. Sci.* 20, 1–5. doi: 10.1111/j.1467-9280.2008.02253.x

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

You, Z.-J., Shen, C.-Y., Chang, C.-W., Liu, B.-J., and Chen, G.-D. (2006). "A robot as a teaching assistant in an English class," in *Sixth International Conference On Advanced Learning Technologies, 2006* (Washington, DC: IEEE), 87–91.

Zimmerman, I. L., Steiner, V. G., and Pond, R. E. (2011). Preschool Language Scales, 5th Edn (PLS-5). Available online at: www.pearsonclinical.com.

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.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 01 November 2019
doi: 10.3389/frobt.2019.00107

Educators' Views on Using Humanoid Robots With Autistic Learners in Special Education Settings in England

Alyssa M. Alcorn[1]*, *Eloise Ainger*[1], *Vicky Charisi*[2], *Stefania Mantinioti*[1], *Sunčica Petrović*[3], *Bob R. Schadenberg*[4], *Teresa Tavassoli*[5] and *Elizabeth Pellicano*[1,6]

[1] Centre for Research in Autism and Education, UCL Institute of Education, University College London, London, United Kingdom, [2] Joint Research Centre, European Commission, Brussels, Belgium, [3] Serbian Society of Autism, Belgrade, Serbia, [4] Department of Human Media Interaction, University of Twente, Enschede, Netherlands, [5] Department of Psychology and Clinical Language Sciences, University of Reading, Reading, United Kingdom, [6] Department of Educational Studies, Macquarie University, Sydney, NSW, Australia

Researchers, industry, and practitioners are increasingly interested in the potential of social robots in education for learners on the autism spectrum. In this study, we conducted semi-structured interviews and focus groups with educators in England to gain their perspectives on the potential use of humanoid robots with autistic pupils, eliciting ideas, and specific examples of potential use. Understanding educator views is essential, because they are key decision-makers for the adoption of robots and would directly facilitate future use with pupils. Educators were provided with several example images (e.g., NAO, KASPAR, Milo), but did not directly interact with robots or receive information on current technical capabilities. The goal was for educators to respond to the general concept of humanoid robots as an educational tool, rather than to focus on the existing uses or behaviour of a particular robot. Thirty-one autism education staff participated, representing a range of special education settings and age groups as well as multiple professional roles (e.g., teachers, teaching assistants, speech, and language therapists). Thematic analysis of the interview transcripts identified four themes: Engagingness of robots, Predictability and consistency, Roles of robots in autism education, and Need for children to interact with people, not robots. Although almost all interviewees were receptive toward using humanoid robots in the classroom, they were not uncritically approving. Rather, they perceived future robot use as likely posing a series of complex cost-benefit trade-offs over time. For example, they felt that a highly motivating, predictable social robot might increase children's readiness to learn in the classroom, but it could also prevent children from engaging fully with other people or activities. Educator views also assumed that skills learned with a robot would generalise, and that robots' predictability is beneficial for autistic children—claims that need further supporting evidence. These interview results offer many points of guidance to the HRI research community about how humanoid robots could meet the specific needs of autistic learners, as well as identifying issues that will need to be resolved for robots to be both acceptable and successfully deployed in special education contexts.

Keywords: education, special education, schools, teachers, autism, children, humanoid robots, social robots

OPEN ACCESS

Edited by:
Raquel Ros, Universitat Ramon Llull, Spain

Reviewed by:
Sofia Serholt, University of Gothenburg, Sweden
Serge Thill, Radboud University, Netherlands

***Correspondence:**
Alyssa M. Alcorn
a.alcorn@ucl.ac.uk

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *31 May 2019*
Accepted: *11 October 2019*
Published: *01 November 2019*

Citation:
Alcorn AM, Ainger E, Charisi V, Mantinioti S, Petrović S, Schadenberg BR, Tavassoli T and Pellicano E (2019) Educators' Views on Using Humanoid Robots With Autistic Learners in Special Education Settings in England. Front. Robot. AI 6:107. doi: 10.3389/frobt.2019.00107

INTRODUCTION

Robotic systems targeted toward people on the autism spectrum, especially children, are a growing subfield of social robotics and human-robot interaction (HRI) research. Autism is a lifelong neurodevelopmental condition or spectrum of related conditions that affects the way a person interacts with others and experiences the world around them (American Psychiatric Association, 2013). Many autistic[1] individuals also have additional difficulties with spoken language and/or intellectual disability, as well as co-occurring mental health problems, especially anxiety, and attentional difficulties—all of which can involve complex, long-term support needs. In England, ∼120,000 children are documented as having autism as their primary form of special educational need and disability [SEND; (Department for Education, 2018)]. Of these, 28% percent of autistic children are educated in special schools and represent over a quarter of the total special school population. The children attending these schools often have complex needs, including an additional intellectual disability and/or limited-to-no spoken communication, and often require much higher levels of support from specialist teaching and allied-health staff than regular, mainstream schools can typically provide. These particular children are frequently overlooked by researchers (Tager-Flusberg and Kasari, 2013) but, along with the specialist staff that support them, represent two sizeable populations of potential robot users in England—and were thus the focus of the current investigation.

Autistic children are thought to be especially interested in and motivated by robots, potentially related to the fact that they are interactive—but programmed and ultimately rule-based—devices. Indeed, robot-based programmes are often cited to be potentially beneficial for this group in particular because they offer the possibility of fairly predictable and consistent interactions (e.g., Dautenhahn, 1999; Dautenhahn and Werry, 2004; Duquette et al., 2008; Rudovic et al., 2017; Straten et al., 2018). These are precisely the sort of interactions that autistic people are often said to favour (Pellicano and Burr, 2012; Lawson et al., 2014). The extant HRI literature suggests that autistic children may be highly engaged during robot interactions (Robins and Dautenhahn, 2006; Straten et al., 2018), and show spontaneous joint attention and other social behaviours that are often challenging for this group (Anzalone et al., 2014; Warren et al., 2015). Yet, existing research on social robotics for autism often constitutes proof-of-concept studies with small samples ($n < 10$), single rather than repeated robot-child interactions, and incomplete information about the autistic participants, making it more difficult to understand the potential applicability of the work as education or therapy [see reviews by (Diehl et al., 2012; Scassellati et al., 2012; Begum et al., 2016), for discussion].

Existing autism and HRI studies have predominantly studied children interacting with robots in lab-based settings (e.g., Salvador et al., 2015; Yun et al., 2016) or closely controlled, researcher-designed procedures that effectively re-create labs in schools (e.g., Kozima et al., 2007; Robins et al., 2012). Although there is much to be learned from studies in controlled lab-like settings, moving robots from the lab into the classroom (or "the wild"), where teachers apply the teaching programme unsupervised, is no straightforward task (Diehl et al., 2012; Huijnen et al., 2016). Embedding robots into existing autism contexts and pedagogical practices requires in-depth understanding of *specific* contexts and practices, and of the adult users who will support robot-based programmes. Understanding the views of these adults is therefore essential, as they are key decision-makers for the adoption of new technologies, and would be the ones to directly facilitate any future use of robots.

Several studies have sought teachers and professionals' views to explore implementing robots within regular educational settings (Fridin and Belokopytov, 2014; Kennedy et al., 2016; Serholt et al., 2017; Cheng et al., 2018) but only a handful have done so within special education settings. Diep et al. (2015) interviewed six teachers from a Canadian school for children with multiple and complex needs about their perceptions of social robots, in relation to an anticipatory governance framework (Guston, 2014). Although their results make some reference to autistic learners, they do not primarily focus on this group. In a larger study, Hughes-Roberts and Brown (2015) conducted interviews and focus groups with 20 teachers in special (though not autism-specific) education settings in the UK, incorporating a demonstration of a humanoid robot, NAO. Teachers stressed sustained engagement as a key indicator of success for many of their SEND pupils, and thus considered facilitating engagement as a key robot requirement. They highlighted three teacher-proposed robot activities, which included adults facilitating one or more children's game-like interactions. Perceived barriers to adoption focused on technical factors, describing the need for simple, fast, versatile, and usable robot controls. The only other limiting factor mentioned was the potential for robots to distract students from learning—at least while the robots were new. It was unclear, however, whether these educators considered, overall, robots to be relevant, appropriate, and feasible for their SEND settings and learners—and, most relevant to the current study, whether they might be especially useful for *autistic* learners.

Huijnen et al. (2017) took a related approach, combining focus groups, and co-creation sessions with autism stakeholders and professionals (including teachers and other school-based roles, all in the Netherlands) to develop 10 specific "intervention templates" for the humanoid robot, KASPAR. These included clear statements of goals, and explicitly mapped out the planned roles and "flow" of an interaction between a child, robot, and professional. This group discussed the role, requirements, and potential impact of the *adult* robot user in far more detail than any other study, ultimately "expect[ing] that the person operating KASPAR is a huge determiner of the success of the interaction and thereby of the intervention" (p. 3085). They also discussed characteristics or subgroups of autistic learners in relation to the suitability of robot use and, in a related paper, identified the potential educational roles that KASPAR could play, including those of a trainer, prompter, or mediator (Huijnen et al., 2019).

[1]We use "identify-first" language ("autistic person") rather than person-first language ("person with autism"), because it is the preferred term of autistic activists (e.g., Sinclair, 1999) and many autistic people and their families (Kenny et al., 2016) and is less associated with stigma (Gernsbacher, 2017).

The findings from Hughes-Roberts and Brown (2015) and Huijnen et al. (2016, 2017) suggest that many educators seem to be broadly receptive—albeit cautious—toward at least some purposes of robots in autism or special education [though see (Diep et al., 2015), for more negative or mixed sentiments]. Educator interviews provide a valuable starting point for understanding whether and how robots might be integrated into existing educational practices, and might transition into being teacher- (not researcher-)managed tools. Yet, these studies only give a partial picture of the information researchers need to know to work toward robot deployment with autistic learners within special education settings. This is for three key reasons. First, these learners' specific needs and the strategies used to support them can be very distinct from those educated within mainstream settings (Eaves and Ho, 1997). Greater knowledge is needed about the utility of robot-based programmes for these particular children in their own specific, specialist contexts. Second, these and other existing studies have frequently asked educators to answer questions or discuss ideas in relation to demonstrations of existing robots (e.g., Hughes-Roberts and Brown, 2015; Coeckelbergh et al., 2016; Huijnen et al., 2016; as in Cheng et al., 2018). This approach can be useful if the goal is to generate or assess applications for those specific robots, but it is necessarily limiting with respect to discussing perceptions and applications of robots as a *category* of tools, or for generating novel use cases, as it primes participants to think of *that specific* robot when developing their ideas. Third, much existing research has either used surveys and questionnaires (e.g., Coeckelbergh et al., 2016; Kennedy et al., 2016; Cheng et al., 2018) to ask educators to *respond to topics and ideas that have been pre-identified by researchers*, or, have effectively leveraged educators' expertise for solving particular design or pedagogical problems (e.g., Huijnen et al., 2016, 2017). Educators' priorities and ideas about robotics might be different than those of researchers, but existing work seems to have given limited opportunities to explore these issues.

The current study is part of the European Union funded DE-ENIGMA project (de-enigma.eu), in which teams with technical and autism education expertise are collaborating to explore the potential of humanoid robots as tools in autism education, particularly with respect to teaching social and emotional skills, and to develop real-time multimodal processing of autistic children's behaviour. One strand of the project sought to better understand current specialist autism education settings in England, i.e., the target users and context of use for DE-ENIGMA outputs. This paper reports Part B of a two-part interview study with autism educators. We focused on educators, rather than a wider range of autism stakeholders, because DE-ENIGMA's focus has been specifically on schools. Part A (reported in Ainger et al., Manuscript in Preparation) investigated autism educators' current goals and pedagogical practices. Part B, reported here, discussed the potential future use of robots.

Our goal in Part B was to elicit educators' views and perspectives on the potential use of humanoid robots with autistic learners in special schools, to better understand the factors perceived to be important for deploying robots in these settings. We also focused on understanding educators' perceptions and suggested applications of humanoid robots as tools for teaching social and emotional skills, due to the focus on this topic within the DE-ENIGMA project. Unlike some previous studies that have asked educators to respond to ideas and topics pre-identified by researchers (e.g., in surveys and questionnaires), we used a semi-structured interview schedule, with researchers exploring participants' ideas in detail, following from fairly open questions.

MATERIALS AND METHODS

Participants and Educational Settings

Thirty-one educators (female: $n = 25$) took part in individual semi-structured interviews or small focus groups, between December 2016 and January 2018. These educators were recruited via convenience sampling through existing community and personal contacts. All of our participants worked in specialist settings in England: 26 in special schools ($n = 7$, autism-specific; $n = 18$, general SEND), five in autism resource bases attached to a mainstream school, and one working across multiple SEND settings.

Autistic children educated in special schools in England usually have a high degree of adult interaction and support throughout the school day. In special schools, classes are small (often 5–10 children), with a highly trained teacher and a team of teaching assistants, who often have less specialist training. There is further input from specialist allied health professionals, including speech and language therapists and occupational therapists. Consistent with this context, our participants reported working with learners on the autism spectrum in a variety of educational roles, including as a primary ($n = 12$) or secondary ($n = 5$) teacher, teachers working across multiple ages and/or school settings ($n = 2$), a teaching assistant ($n = 2$), a headteacher or deputy headteacher ($n = 3$), a speech and language therapist ($n = 3$), or an occupational therapist ($n = 2$). Many participants indicated more than one autism-related role and had worked with multiple age groups over time, from Early Years education (<5 years), up to age 18–19 years. They varied widely in their level of experience, ranging from <1 to 18 years' experience in their current education setting ($M = 4.7$ years, $SD = 4.1$) (see **Supplementary Table 1** for participant details).

Procedure

Fourteen participants (female: $n = 11$) completed individual, semi-structured interviews in a quiet room at the university or school, and 17 participants took part in one of three focus groups (female: $n = 14$) in participating schools (two groups contained six participants, one contained five), facilitated by a researcher (see **Supplementary Table 1**). Part A of the interview study (Ainger et al., Manuscript in Preparation) focused on current educational contexts and practices, including participants' aspirations for their autistic students, their views on how social and emotional skills are currently taught within classrooms, curricula and supports used in their setting, and uses of technology (see **Supplementary Table 2**). To introduce the discussion of humanoid robots in Part B, the focus of the current study, participants viewed six example images of existing robots (Milo, KASPAR, NAO, Flobi, PARLO, and Pepper). They were not given any further information about these particular robots, their current capabilities, or examples of use and were encouraged not to be concerned about issues of technical feasibility. Instead, they were asked to consider the potential uses of humanoid robots for autistic children's learning, including potential benefits and concerns (see **Supplementary Table 2**).

While the interview did not explicitly ask about respondents' prior experience with or knowledge of robotics, almost all educators stated that they had no prior experience or knowledge of robots. The exceptions were one educator working with older students, who reported using commercially available Bee-Bot® robots to teach science and programming, and some educators who had seen previous demonstrations of a humanoid robot (Milo) in connection with the DE-ENIGMA project.

The protocol was approved by UCL Institute of Education Research Ethics Committee (REC857). All participants gave written informed consent to the interviews, including audio recording, in accordance with the Declaration of Helsinki. The total duration of the individual semi-structured interviews lasted 30–54 min ($M = 40$ min) and focus groups lasted 52–78 min ($M = 62$ min). The robotics-focused questions (Part B) lasted 5–12 min in individual interviews, equating to 14–31% of the total time ($M = 8.5$ min, 20%), and the robot section of focus groups lasted 15–18 min, or 24–35% of total discussion time ($M = 17$ min, 29%).

Thematic Analysis

Audio-recordings were transcribed verbatim. The robot interview data were analysed using thematic analysis (Braun and Clarke, 2006), which included familiarisation of the data; generating of initial codes; generating themes, reviewing, defining and naming themes; and compiling this report. We adopted an inductive approach (i.e., without integrating the themes within any pre-existing coding scheme or preconceptions of the researchers) within an essentialist framework (to report the experiences, meanings, and reality of the participants). Two authors (AA and EP) independently familiarised themselves with the data and liaised several times to review the themes and subthemes, focusing on semantic features of the data, resolving discrepancies and deciding the final definitions of themes and subthemes. Analysis was thus iterative and reflexive in nature. Participants' responses to Part A of the interviews, on current educational goals and practices, were analysed and reported separately (Ainger et al., Manuscript in Preparation).

RESULTS

We identified four themes in educators' interviews (see **Figure 1** for summary of themes and subthemes). Throughout, educator quotes are attributed via participant ID numbers.

Theme 1: Autistic Children Are Likely to Find Robots Engaging

Participants stressed the importance of engagement and motivation for learning, and anticipated that the autistic learners in their settings would be "so interested" in and motivated by humanoid robots, potentially more motivated than when interacting with adult educators, or non-technical activities. One explained: "I think if the robot's doing it [modelling behaviours], it's more captivating than just us as a person. This is a toy that plays back essentially, it's engaging" [101]. Educators also felt that this engagement could have a positive impact on their readiness to learn: "They would be really happy to work with it for longer periods of time, much longer than usual because, let's be honest, a piece of paper and a worksheet, it's not as exciting as a humanoid robot can be" [011].

Participants reported that, for some children, the attraction of a humanoid robot might be sufficient to encourage them to engage in otherwise challenging social interactions: "engagement is a big key to the social barriers that children may face, and if they're able to engage and experience some of those interactive activities, which they avoid at all cost in other settings… I really think [a robot] could support the social skills" [004]. Yet, robot attractiveness and engagement were not perceived as wholly positive. Respondents often discussed this characteristic alongside potential drawbacks, including concerns "about the extent we're going to use the robots… when we're talking about autistic children, we need to be very careful with something [not] to become an obsession" [011]. Another educator commented: "particularly with the younger ones with autism, we're trying to make them think that people are amazing… so all the teachers in the sessions try to become the most exciting thing in the room" [105]. For some children, educators further felt that access to a highly attractive robot could conflict with overarching educational goals to help autistic learners attend to and understand other people (see also Theme 4B).

Theme 2: Robots Offer Predictable, Consistent Interactions; Children Know What to Expect

Educators in this study expected humanoid robots to be "consistent" and "obviously predictable" compared to people, who "behave in all sorts of different manners and ways" [015]. One educator summed it up:

> "Robots, unlike humans, they will always be the same. Their tone of voice will always be the same, their inflection will always be the same, the body language is always the same. They're very predictable, like if you say a certain thing, it will say a certain thing back to you. So I think with kids with autism, they love that kind of thing, predictability" [014].

Overwhelmingly, they saw predictability as a potential benefit for their students but, as in Theme 1, they frequently discussed this benefit alongside less-positive implications.

Subtheme 2A: Predictability Is Understandable and Non-threatening

Educators emphasised autistic children's difficulty in making sense of other people's often-unpredictable behaviour: "this is a struggle, they cannot predict people but a robot is quite predictable with its reactions" [017]. Robots were perceived to be "easier" for children because "they know what to expect" [010] and could help them to "predict what might happen" [015].

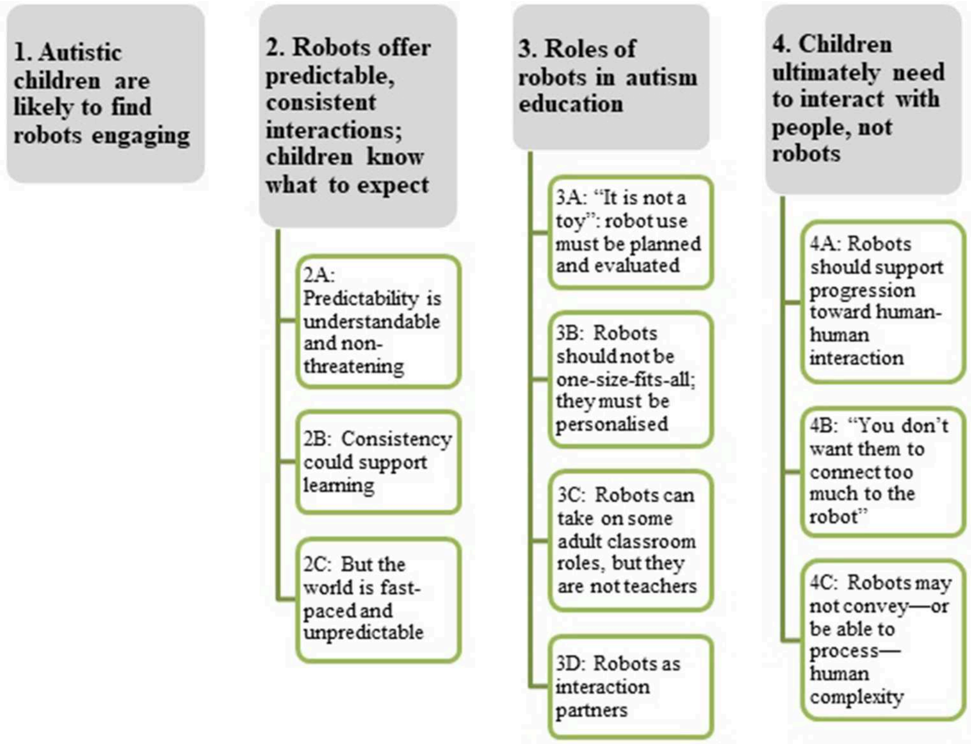


FIGURE 1 | Summary of themes and subthemes.

Educators often talked about the importance of their students feeling safe and secure, and thought that "a robot like that would be something safe for my students, safe to interact, safe to communicate… they wouldn't feel threatened" [008]. One specific, anticipated benefit of a robot's predictability was that children might feel more at ease interacting with robots, relative to how they feel in other school activities or human interactions: "These children might respond to the robots better than the way they respond to other people because they might predict their reaction. So, for example, if they know that when they say 'happy' he smiles, it could be less scary for them" [002]. Some educators also felt that this benefit could have a positive impact on their learning:

> "Many of my students won't push themselves harder because they are afraid of making a mistake. Maybe if a robot like that would exist in my classroom, they wouldn't feel so intimidated or threatened from the teacher's authority and they would be able to try different things and that would help them progress and develop in different aspects" [008].

Subtheme 2B: Consistency Could Support Learning

Respondents also highlighted the possible benefits of a robot's consistency or "sameness," particularly in its visual appearance and manner. One educator remarked:

> "We do have different people coming in as supply teachers or supply TAs [teaching assistants] for the day and, if some of the students do not like the way someone is dressed or smells or talks to them, they won't communicate with them. But a robot like that will have the same specific characteristics every single day and that's something that would be very useful for my students. They will know that this robot would look exactly the same every day and they will be able to build a trust with the robot and communicate more" [008].

Another respondent suggested using the robot for helping to focus their attention on academic learning due to their unchanging manner and appearance: "[autistic students] can only concentrate to the words that the robot says. When we [staff] used to teach them, they could concentrate on everything else on us, like the way we move our hands, the way that our hair is today. So I think a robot could actually attract their interest on a specific thing that we want them to learn" [005].

Educators also used "consistency" to refer to a concept sometimes described in the autism-robotics literature as *repeatability:* a robot could repeat usually-variable social behaviour (e.g., a facial expression) over and over, helping autistic children to begin identifying patterns and associating meanings with the behaviour.

> "The challenges of face-to-face and eye contact and response to facial expression and understanding somebody's facial expressions are so inconsistent that, with a robot, [autistic children] can start to learn what those consistencies are and it becomes much easier for them to respond to them, rather

> than a human facial expression, which could mean all kinds of things. I think with a robot they learn very quickly… [they] may start to associate meaning with some of those facial expressions and recognise those in others and maybe seek some of those communicative responses" [004].

Educators also felt that robotic consistency might be particularly advantageous if applied to classroom interventions that *require* consistency and rule following, such as the Picture Exchange Communication System (PECS; Bondy and Frost, 1994), a widely-used alternative/augmentative communication system. Indeed, they felt that a robot might deliver such an intervention with *more* fidelity than a human teacher:

> "The PECS system is very definite and it's very, very rule based, but as humans, there's distractions and that means the delivery of this rule-based training we often get wrong. Robots would do it consistently so that a child, an autistic child working with a robot that's programmed to deliver training only in that specific way following that specific algorithm, [the child is] going to respond much better because they're getting a consistent response. So I think you'd have better outcomes if robots are teaching autistic kids certain protocols" [013].

Subtheme 2C: But the World Is Fast-Paced and Unpredictable

Educators repeatedly highlighted that, unlike the expected behaviour of robots, both humans and life are *un*predictable, and that one key educational goal was to support children in learning to deal with this uncertainty. Educators were concerned that predictable and consistent robots would potentially hinder children's progress in this regard: "[technology], largely speaking, you know, does what they want it to do. What we want them to understand is that the world is unpredictable and the world has huge variety in it, and we want them to be able to respond quite flexibly to things, as well as follow somebody else's agenda" [201].

Educators noted that, while a robot might not "mind how long it takes for a child to do anything… it could be really deskilling for the child because you don't have all the time in the world with a robot waiting for you when you're an adult, like you do have to just go on the bus and swipe your [bus pass], you do just have to transition" [103]. Transitioning between activities and/or settings can be an area of particular difficulty for autistic children.

Our participants also felt that, while children might learn more easily or feel more comfortable with a highly predictable robot than when learning with a person, that type of learning could be counter-productive in the long run because it does not support skill generalisation: "I don't know, maybe it's going to be too predictable for them, and then how will they generalise when they actually have to interact with actual people. So maybe by teaching them this predictability, it's not that easy to help them generalise it" [010]. Some educators reported that a robot could provide a "good base" for teaching simple social skills but warned, "if our goal is to teach kids social skills and interaction and how to interact into the world and the community, then that's not through robots because at the end of the day, our community and our real world are not made of robots. So it's very important that we phase out a bit and then have more human contact" [014].

Theme 3: Roles of Robots in Autism Education

Educators' examples of how robots could potentially be used varied widely depending on their settings and the profile of their learners. Nevertheless, there were several key commonalities across the interviews.

Subtheme 3A: "It Is Not a Toy": Robot Use Must Be Planned and Evaluated

Educators agreed that robots are "not a toy." Rather, any use of social robots in their settings would need to be planned by teachers, "really thinking carefully, 'How do I use it? Is it appropriate?"' because a robot "might not be appropriate for every single child" [203]. Some framed the need for planning in relation to their past experiences with iPads in class. Like robots, iPads were perceived to be attractive, flexible technologies but, according to educators, were often introduced without clear goals, creating knock-on problems in which autistic learners might "see an iPad or a technological device as something that is mainly a toy. They can develop some obsessive behaviours or they will be repeatedly asking for an iPad without completing the work" [008]. One interviewee neatly summarised: "I don't think [robots] should end up being used like iPads, just for fun and just as a toy. I think they, when you use them, should have a very clear target for why you're using it and for a very clear amount of time and with a purpose" [011].

Indeed, educators emphasised that educational planning would therefore need to consider whether the robot was "appropriate for every single child... really just thinking carefully, like everything we do here, 'Oh is that child ready?' and to really teach something specific, not thinking just putting them in the same room with the robot and then leave and think they'll know everything" [204]. Another focus group agreed: "It wouldn't have to be like, 'okay, now we're learning the social skills next time the robot is coming up' but I would look at each child and see like, 'okay, how am I going to use it with that learner' and then find a time and a setting that feels appropriate" [302].

One respondent further suggested that planning to introduce robots or any new tool must incorporate *evaluation*, perhaps especially if teachers have high expectations and perceive the new tool as a "[scheme] that they believe will work and will fix everything." She noted,

> "We say, 'oh yes, try that, that might work,' and there's nobody assessing as to whether or not it is working. We need a baseline check to start with and then we need to check whether or not it's worked at the end of the intervention. Interventions are incredibly expensive so therefore you have got to have the mindset that you're going to look to see whether that intervention has worked" [007].

Subtheme 3B: Robots Should Not Be One-Size-Fits-All; They Must Be Personalised

In autism education, "personalisation" is a fundamental task in which educators choose, adapt (and often, invent) tools, and strategies "that are catered to that child" [301]. The educators we spoke to clearly expected that the same type of fine-grained child-level personalisation would be necessary and "programmable" with robots, in addition to choices about the types of learning activities different students may do: "[robots have] got to be based on the likes and dislikes of the child… the adjustments would be, you know, that [the robots are] programmed to do a variety of different things" [013].

> "If I have a very verbal student who just needs to practise reciprocal conversation or needs to practise its tone of voice or practise identification of feelings and expressions, then I'd program the robot for that. But then if I have only that one robot but then I want to use it with a different kid, who's non-verbal, doesn't like interacting with people at all, then I would have the robot programmed to not say anything, to not maybe do any sudden movements. I would program it depending on what level the student is or what social skill I want to work on" [014].

One respondent agreed wholeheartedly with the importance of technology personalisation, but questioned how well teacher-implemented robot personalisation would work in practice, based on their current experiences with a dyslexia-focused app, Wordshark (https://www.wordshark.co.uk/). She described how this program "can be tailor made to fit the particular child and quite often teachers don't use that tailor-made bit. They just think, 'oh yeah, Wordshark, Wordshark is supposedly very good so let's use it,' and they're not using it in the way that the manufacturers intended" [007]. She also pointed out that these issues around personalisation and correct use can be exacerbated by school-level decisions around technology and training, in which institutions "invest in one particular member of staff, 'here you are, you're the expert in this' and then either that trend is not cascaded down, or that person then leaves and the technology is left behind and nobody really knows how to use it."

Subtheme 3C: Robots Can Take on Some Adult Classroom Roles, but They Are Not Teachers

In their discussions, interviewees' suggested robot roles reflect the types of routine support that staff members offer autistic children throughout the day, including "to guide them, to give them ideas, and maybe even to prompt them or to praise them" [008], "especially the higher ability ones, who when I leave them to work independently, they lose track of what they're doing" [009]. Others also felt that educators "could use it as a tool as a part of the group, so the robot could almost form part of the group or it could be used as, it might lead the session or the group" [302].

Yet, while the interviewees suggested that robots could usefully offer some types of support and facilitation currently provided by various adult staff, other discussions made clear that the robot was not seen as a potential *teacher*. Educators emphasised that "the adult always needs to be in control with what's happening" [303], especially with regard to planning and goal-setting. Where some respondents indicated that robots could be adaptively responding to children, these comments were always made within the context of supporting educational or social goals already identified by teachers. There was no discussion of future humanoid robots "assessing" or identifying children's needs.

Beyond issues of planning and control, respondents pointed out that special education teachers are trained in a distinct set of skills and strategies that they need to support their learners. Educators were concerned that reliance on a robot may both deprive autistic learners of the benefit of those skills, but also (over time) detract from staff members' ability to exercise those skills. One focus group participant explained:

> "Part of our skills we have as special needs educators is that we're able to empathise, and use lots of creative strategies, to the point where you understand why someone finds it challenging to transition and hopefully don't find it so frustrating anymore. I think it's important to swap around as a team as well, not just leave it to a robot" [102].

This "professional deskilling" concern was shared. Another educator noted their own lack of robot experience and training, explaining that their "main concern is whether I would be able to use it appropriately and I wouldn't lose other aspects of my teaching. For example, I wouldn't want to rely too much on the robot to communicate with my students or to help my students access the knowledge" [008].

Subtheme 3D: Robots as Interaction Partners

Even before being explicitly asked about possible applications of humanoid robots to social and emotional skills teaching, respondents spontaneously suggested social applications and roles for the robot. As they had explained earlier in the interviews (Ainger et al., Manuscript in Preparation), "the most important goal is to help them progress with their social skills" [010]. Teachers believed that attractive robots might act as *social partners*, motivating children to work on inherently challenging social and communication skills that are already targeted in existing class activities, such as turn-taking in activities ("You're waiting for the robot to finish talking and then it's your turn to talk, so it's like turn taking, you know, how to have a conversation with somebody" [011]) and conversations ("like having kids just learn general conversations like teach them to say, 'hi, my name is A, what's your name? How are you feeling today?' Like just have them practise conversations, have them practise answering questions but also having the kids practise coming up with questions themselves" [014]).

Educators specifically highlighted the role a robot could play in understanding how children's own behaviour affects others—one of "the biggest thing[s] for our learners" [302]. Another interviewee concurred that some autistic learners "cannot see how the way that they're behaving affects other people. So this would be a nice thing to use the robots for… [learners] could perhaps see how their behaviour was affecting somebody else" [007]. Other respondents gave specific examples of how they might work on this concept, using the robot, including "a programme for how to make the robot happy today… The programme might ask for some steps that the child has to do like feeding or giving water or going for a walk or holding hands or playing a game, whatever makes a robot happy" [017]. Another offered:

> "I just think of like a robot crying and then having like props of tissues or whatever, you know, and then making my children try to calm him down… care for the robot as well, you know, when he says that he's angry or he's got a cut in his wrist or something, I think they really could connect with that. I think that could be a great tool actually" [015].

Educators also suggested that understanding cause-and-effect with the robot could also be used to go beyond grasping event relationships, to "build that empathy and understanding [of] other people. So, the child is angry and might be pulling or shaking the robot or hitting the robot, that the robot might be able to respond to that in a way that it's communicating to the child how those actions are making him feel" [304].

Respondents were not universally approving of using the robot to teach social communication. One respondent was receptive to the idea of robots in general, saying "with the right software or the right purpose, it could be awesome," but was emphatic that its uses should not include anything "related to emotions or behaviour management or any patronising sort of thing" and "nothing like engaging in social skills or emotional stuff" [016]. This same respondent had expressed particular concern about the robot's capacity to meaningfully render complex human behaviour, and to respond appropriately to autistic children.

Theme 4: Children Ultimately Need to Interact With People, Not Robots

While they expressed interest and cautious optimism about the use of humanoid robots in autism education, interviewees were also very clear that robots were perceived to have potential and acceptability primarily as "stepping stones" to fostering human-human interaction.

Subtheme 4A: Robots Supporting Progression Toward Human-Human Interaction

Respondents either implicitly or explicitly indicated that working with a robot in a school context would be a transitory, middle phase between two different *types* of human-human interaction. Educators felt that they would first need to introduce the robot "in a familiar space, with trust and familiar adults that can say, it's okay" [301]. Many autistic children are highly anxious about all new people and activities, and staff suggested addressing this issue using existing educational strategies such as "a social story about it, [showing] pictures beforehand, [explaining] what's going to happen with the robot, when the robot will be coming" [301] (see Gray, 1994, on social stories). These steps, which can "build up almost the story of this robot, how it's coming here, and when it arrives then the pupils will probably be more—shall we say, prepared for its arrival" [101], are useful for any child's interaction with a robot, but especially so for autistic children, who require additional preparation to adapt to novel objects and events their environment.

Educators then described how children might work with the robot on skills or activities over time, again potentially supported by some degree of adult guidance: "that's one of our targets, especially in my class, is getting kids to talk to one another. So that could be almost the first step, rather than talking to an adult, you're talking to the robot" [305]. At a later point, children might transition away from work with the robot, applying those skills in interaction with peers, adults, or the community: "You can practice having conversations, you can have the robot opposite you and you can set certain rules and you can first practise with robots before you move on to adults" [011].

Respondents suggested that humanoid robots might be particularly successful at supporting social learning and later generalisation, because "the fact that it is human-like might help them to associate the robot with human behaviour." Another explained with reference to the robot image examples provided in the interviews/focus groups: "I prefer the ones that look more like a human. Most importantly, it's going to be like it's a real boy, it's a real-life example. They would consider the rest like a toy but this [humanoid robot] might be actually an example" [010]. Other educators felt the opposite, that human-like robot appearance and behaviour could be confusing and create problems: "I think that will be my main concern, you know, how to explain to the child that this is only a robot, it doesn't have feelings, and it's different than mum and dad or friends and teachers" [015]. Another agreed that "we don't want them to start thinking this is a human, 'this is my friend' or 'It's the same as my peers'" [204]. Others thought children's understanding of robot-human differences would be dependent on their age and cognitive ability, and one respondent flatly dismissed these concerns, maintaining that to "someone who has autism, a robot is a robot, even if it looks like a person" [016].

Subtheme 4B: "You Don't Want Them to Connect Too Much to the Robot"

Educators expressed concern that children might have "too much" interaction with a humanoid robot, in various ways. Some perceived time spent interacting with the robot as directly detracting from time spent with people: "with my kids, you know, [my concern] would just be maybe about the amount of time they would be engaging with it and making sure that they're not always engaging with the robot and they're engaging with other children" [302].

Our participants were also worried about children's emotional investment in the robot. They felt certain that autistic learners could trust and emotionally connect with a robot—perhaps more so than with a person: "You don't want them to connect too much to the robot, that then it's almost like an imaginary friend, like that they rely so heavily on this robot that then they don't socialise" [303]. Another predicted: "they will become too dependent, they will prefer to be with the robot than be with mum or be with sibling and interact with friends. I would be just scared that they will get too attached. I would rather see my children interacting and playing with me or with each other than with the robot" [015].

Suggested applications where children would "build up" from robot interactions to human interactions were repeatedly positioned as a way to balance the potential benefits of supportive, reciprocal robot interactions with the risk of these overshadowing existing relationships. One participant summed this up:

"I feel that a robot will work more or less in the same way as our students. There would be a common ground to communicate and share feelings and emotions, a better way to express those emotions instead of interacting with an adult, or their peers. And I'm not saying necessarily to interact just with the robot because that would lose their communication part with, the other human beings in the classroom, with the adults or with their peers. But I think that would be the first step for them to start expressing their feelings and emotions and then it would be easier for them to involve other human beings in the classroom… [in] their everyday lives and showing their emotions and communicating their needs" [008].

Subtheme 4C: Robots May Not Convey—or Be Able to Process—Human Complexity

Educators repeatedly noted the complexity of human behaviour, and were concerned that humanoid robots' behaviour would lack nuance and variation, particularly for social communication: "You could teach a robot to do this and that but not everyone does it the same way. One person when they're angry might cross their arms but some people might tap their foot. So human behaviour is so erratic and unpredictable and everybody's behaviour for whatever emotion is different" [001]. Educators felt that this lack of variation would limit the robot's potential with regard to what it *could* teach: "With autistic kids, certainly they could mimic [the robots] but because they could mimic them, they would be in risk of learning one expression for one feeling and that's not right 'cause the diversity of emotions is so wide and the way we adjust and the way we process emotions is so different" [016]. As with the mixed implications of robot predictability and consistency (Theme 2), educators felt that a robot that is programmed to—or is physically limited to—showing a social behaviour in only one way might potentially do autistic children a disservice by not preparing them to understand the true range of human behaviour. They also described how a real, two-way exchange of feeling would be missing: "Social interaction is emotional for both sides, so it's something more than you just get with the robot who is just there, he's predictable. Human relationships are much more complex than the robot I think can show" [104].

Other concerns focused on how the underlying technology would not be able to adequately cope with—and adapt to—the diversity and unpredictability of autistic learners' behaviour: "Even if our students are very structured and predictable, they can also be unpredictable and I don't know if a robot could be able to adjust to those things" [013]. Additionally, "I doubt that a robot could recognise the different ways a person with autism could express [the] same emotions. I think it would be hard to design a software for that" [016].

DISCUSSION

In this study, educators were provided with minimal information about what humanoid robots "are like" or their current or future uses to avoid biasing educators' reflections toward specific, existing examples. Educators were therefore free to project their own ideas of whether, and in what ways, future humanoid robots might contribute to autism education. This approach differs from some recent practitioner studies, where participants were introduced to specific robots, or were asked to solve specific problems (e.g., whether KASPAR could add value to a particular learning domain; Huijnen et al., 2016). Overall, the current respondents were open to discussing humanoid robots within autism education contexts. They expressed a willingness to find out more about them, or to try interacting with them for themselves to see what their capabilities might be. These respondents from autism education settings shared many basic perceptions of robots with both the mainstream, UK-based educators in Kennedy et al. (2016), including robots as having "simplistic interactions" and being "primarily seen as a scripted, reactive machine" (p. 5), and with the Canada-based special educators in Diep et al. (2015), who felt that robots might "[provide] structure and repetitiveness in a consistent fashion" (p. 2). Yet, the same qualities that our participants saw as potentially so promising for meeting the needs of autistic learners were perceived as *obstacles* to adoption by the Kennedy et al. (2016) mainstream sample (see also Serholt et al., 2017); an illustration that "educators," "autistic children" and "schools" are not homogenous groups and will have different needs—which need to be fully understood to inform future robotics work.

Our respondents' openness to discussing future robot use did not equate to unqualified endorsement, however. Where educators predicted that robots could benefit their learners, these predictions were both conditional and carefully circumscribed: robots may be beneficial, *if* used in a certain way, and *if* certain measures are in place. These circumscriptions consistently position proposed future robot use within established educational goals and supports. Educator responses also revealed a shared prediction that any future robot use would pose a series of complex cost-benefit trade-offs: if a robot is appealing and motivating, it may become a liability if children engage with it to the exclusion of other interactions; a predictable robot could support short-term learning goals, but might then interfere with children's longer-term capabilities to cope with a mutable world. As part of their initial consideration of whether robots belong in autism education, teachers were already looking at the implication of robots across a child's school career, or their lifespan. Such predicted trade-offs must be addressed by carefully planning robot use, within existing practices and within individual learners' pre-existing goals (subtheme 3A). Autism specialists in Huijnen et al. (2017) made similar comments on the imperativeness of planning robot use, though did not discuss its longer-term implications and trade-offs as did the current participants. These perceived benefits and trade-offs have significant implications for the autism-robotics field, and will be discussed in turn below.

Robots Are Novel, but Not Different From Existing Tools

Across all of the interview prompts, educators discussed humanoid robots in a remarkably similar way. Interviewees proposed robot uses that supported existing curricular goals, and volunteered a range of established educational strategies that could be applied to introduce robots and support their use. Suggested robot activities and roles built on existing classwork (e.g., practicing turn-taking in a small group) and staff roles. Respondents' emphases on cause-and-effect and turn-taking, plus the specification that adults must be present to support robot use, echo the teacher-proposed robot learning activities in Hughes-Roberts and Brown (2015) and indicate that social skills practice with robots has wider relevance for special education populations.

Humanoid robots are a novel technology to autism educators, and one for which they can propose possible applications. However, the current interviewees did not have an expectation of robots affording *completely new* educational goals, but rather, of robots representing a potentially powerful tool to pursue existing goals. Overall, humanoid robots were not perceived as being fundamentally *different* from current, widespread technologies, such as tablets. Autism specialists interviewed on their existing iPad use in King et al. (2017) described comparable patterns of use to those that our respondents envisioned for robots, "attempting to integrate tablets into the standard instructional methods that they were already using" (p. 9). To the current respondents, humanoid robots could be fully compatible with current autism education practices, *if* they can support key longer-term priorities (see Generalisation and Effectiveness: Challenges to Educational Robot Adoption?). This perceived instructional compatibility does not negate the desire for specialist training about robot use, and for that training to be distributed across school staff. Respondents in Huijnen et al. (2017) and King et al. (2017) made similar points about KASPAR and iPads respectively: they wanted training both on how to operate the devices and how to make the most of them pedagogically.

As with any educational tool, educators indicated that humanoid robots should be one component or phase of educational activity that is carefully planned to integrate into wider practices; participants in Huijnen et al. (2017) similarly stressed the need for integration. Lesson planning, introducing the robot, and—eventually—transitioning to human interaction were envisioned as being planned and managed by teachers. At least some teachers also seemed to envision taking responsibility for programming robots, or otherwise adapting them to individual learners (see Personalisation, Content, and Teachers-as-Programmers). Respondents' examples of potential robot use implied that some degree of autonomous behaviour would be acceptable and useful, such as robots being able to respond to children in an ongoing activity, to detect when children need prompting, or to offer praise. In Huijnen et al. (2016), participants suggested similar preferences for "semi-autonomous" robot operation with autistic learners with specific reference to the existing KASPAR platform. However, some current interviewees raised the concern that robotic technology may not be well-equipped to autonomously interpret and respond to autistic children's variable behaviour.

Even if robots do not demand new ways of working, interviewees still identified areas of desired improvement from existing practices around technology use in their schools. They clearly had mixed experiences with iPads in particular, as devices that could be *too* engaging, and specifically referenced them when emphasising the need for careful lesson planning around robot use. Once again, there is close alignment between these respondents' views and those reported in King et al. (2017), in which educators acknowledged "numerous challenges" of iPads such as "perseveration," but yet retained "an overall optimism about tablet use. They were aware of the incredible motivation tablets provided for [autistic children] and realised their potential across several areas" (p. 8).

One area in which the current results *differed* from other teacher studies on robots or iPads was the degree of concern over children becoming too emotionally attached, or robots potentially detracting from children's peer, family, and staff relationships. This is more specific than concerns over the amount of use, and also seems distinct from concerns about technology isolating autistic children (e.g., King et al., 2017). This may be one area in which *humanoid* robots are perceived as special and facilitative of social relationships with autistic children in a way that other devices may not be. However, as with other robot characteristics, human-ness and social capacity were also perceived as pedagogically important (subthemes 3D, 4A). These concerns about overly close and important social relationships with robots are diametrically opposed to some of the Canadian special educators' opinions in Diep et al. (2015), where "face-to-face interaction was seen as an important task they felt the robot could not provide" and that robots "cannot perform the task of providing emotional comfort or communication" (p. 2). These divergent views may indicate both differences of opinion between groups of educators, but also views of robots shifting over time (data from Diep et al. were collected from six teachers in 2012) as technology becomes more sophisticated and is increasingly publicised.

Generalisation and Effectiveness: Challenges to Educational Robot Adoption?

When asked to discuss potential applications of humanoid robots, educators consistently talked about them as a "stepping stone" to learning, between an introduction that is carefully managed by school staff and a supported transition away from the robot, toward applying new skills with human partners. Endorsing this basic three-stage pattern of robot use appeared to counteract some respondents' concerns about the possibility of children becoming overly reliant on robots, or interacting with them at the expense of classmates and families (subtheme 4B), and made them more ethically acceptable. The stepping stone pattern also relies on educators' special skills and knowledge of children. (Huijnen et al., 2017) participants perceived this same factor as critical to the robot's success, and also linked it to the potential for generalisability—especially in Wizard-of-Oz interfaces with direct and fine-grained adult control. A child could practice transfer even *within* robot interactions by working with different staff, or in different locations.

The "stepping stone" strategy (see also Vygotsky, 1978; or "social bridge" in Hughes-Roberts and Brown, 2015; Huijnen et al., 2016) assumes that children would successfully generalise skills from a robot interaction context to a human one, after sufficient practice. Yet, supporting autistic children to generalise, or transfer, their skills from the lab/intervention setting to a more real-world context is notoriously difficult (e.g., Schreibman et al., 2015). Concepts such as the "therapy register" (e.g., Johnston, 1988; Yoder et al., 2006) capture the issue of autistic children successfully learning and applying skills in one setting (e.g., speech and language therapy), but struggling to apply them in other relevant settings and situations (e.g., at home). Several studies that have specifically investigated autistic children generalising skills from technological contexts have not been particularly promising [e.g., see (Wainer and Ingersoll, 2011; Wass and Porayska-Pomsta, 2014; Whyte et al., 2015)]. With respect to technology-based autism tools, McCleery (2015) points out that there has been very limited, direct study of *near transfer* (i.e., skill transfer to another related task), and *far transfer* (i.e., skill transfer to other domains or naturalistic interaction contexts). The existing research has focused predominantly on screen-based technologies, over a wide range of ages and ability profiles, but not on social robots. More research is needed to test specifically whether robot-based activities can support near and far transfer of skills, and for *which* robots, activities, and subgroups of autistic learners (see section Conclusion). Following Huijnen et al. (2017), perhaps the role of adults in robot-based interventions, and in supporting successful transfer, should also be more overtly defined. For educators to see humanoid robots as potentially valuable and ethically acceptable tools, future research should focus on providing evidence of robots consistently supporting skill transfer into "real contexts."

The interviewees' examples of potential future robot use also make a second critical assumption: that robots can actually teach autistic children new skills, particularly through implicit instruction. As with generalisation, this is not a settled question. Numerous social robotics studies have tested the *efficacy* of robots (i.e., whether a process can produce an intended result in a highly controlled setting), teaching autistic children specific, isolated skills such as point-following (e.g., David et al., 2018). Yet there are relatively few—if any—studies of robots' teaching *effectiveness* in non-lab contexts (though see Scassellati et al., 2018) and methodological issues mean many HRI studies do not provide clinically useful evidence (see Begum et al., 2016). Many of the skills that these educators wish to teach are also more complex than those in existing studies, with murkier criteria for success (e.g., a child understanding how her actions affect another person). Assuming that robots *could* facilitate skill transfer and show effectiveness in educational contexts, one outstanding question is whether robots could offer sufficient added value (vs. other technological/educational tools) to compensate for their current expense, fragility, and complexity.

Personalisation, Content, and Teachers-as-Programmers

Strikingly, *none* of the educators made any reference to any kind of "robot app store," or of otherwise buying or accessing pre-packaged curricula for robots, as they may already do with tablets or with some autism interventions. Instead, they repeatedly highlighted that successful robot use would need personalisation or adaptation of teacher-planned activities, especially given the enormous diversity of behaviours, preferences and traits of autistic learners. Directly or indirectly, respondents indicated that they (or people in teaching roles) should be the ones to implement whatever robot personalisation is required, with some explicitly explaining this in terms of programming (subtheme 3B). In both Hughes-Roberts and Brown (2015) and Huijnen et al. (2017), participants also stressed the need to personalise activities and robot behaviours (e.g., speech) to individual learners, suggesting that teachers would have responsibility over personalisation within the classroom, and even during the course of an interaction.

Yet, technical complexity and need for expertise were perceived as significant practical barriers to robot adoption. One participant in a leadership role described existing problems with teachers not using the personalisation capacities of existing technologies, such as apps, due to lack of training or time constraints. Others were concerned that technology expertise and training may be deliberately limited to single "experts," and thus not easily "cascaded" through an entire teaching team. Other participants agree: Hughes-Roberts and Brown (2015) interviewees raised similar requirements for "the teacher [to] manipulate the robot without needing external support," warning that "if it takes too long to set up the robot or deliver a lesson... [teachers] won't use (it)" (p. 52). Participants in Huijnen et al. (2016) cited as a particular strength of KASPAR that they would be able to use software to create interaction scenarios *themselves*, without specialist technical support. These views and concerns highlight a clear deployment challenge for robot developers and for educators: if the type of flexible robots that educators envision require extensive training or technical knowledge, they may struggle to gain traction in schools because of expertise bottlenecks, or overly complex, time-consuming procedures.

What Type of Tools Are Robots? Educator Views vs. Current Research

The current findings suggest that autism educators at special schools in England have notably different expectations and priorities for humanoid robots than many existing HRI research projects, though share many points of agreement with other SEND and autism educators (Hughes-Roberts and Brown, 2015; Huijnen et al., 2017; though see Diep et al., 2015) and autism specialists working with other technologies (King et al., 2017). Educators expected that if they could access humanoid robots in the future, these would be *flexible tools for them and their teams.* They would be able to plan lessons using the same robot to work on different goals with individual learners or small groups, depending on need. This "flexible tool" view also agrees with a recent survey of UK-based teachers in regular, mainstream schools, where the second most popular proposed use of robots in schools was as a "versatile tool for the teacher, used in many situations" (Kennedy et al., 2016, Figure 6).

Yet, many existing autism-robotics and educational robotics research projects do not appear to be working toward a "flexible tools" endpoint. There are some clear practical reasons for that, including the difficulty of demonstrating feasibility and efficacy for a tool that could be used in almost any way, or investigating learning gains when every participant may have unique targets. Existing proof-of-concept and psychological experimentation work with robots (see section Introduction) often have basic science goals that add to the autism-robotics knowledge base and have focused on the needs of child users, rather than the needs of adult users who may operate robot systems. While the KASPAR research programme (e.g., Robins and Dautenhahn, 2017) has worked on iteratively developing and evaluating domain-specific robot-based lessons over time and has created customisation software for end-users to develop new learning scenarios, this capability does not appear to be well-known or well-documented compared to other aspects of the project (though see Huijnen et al., 2017). There are also several examples of packaged robot-based or robot delivered content. US-based Robokind manufactures humanoid robots, but has also developed and sells the "robots4autism" curriculum for autistic learners (https://robots4autism.com/). Scassellati et al. (2018) developed a month-long home-based social communication intervention for school-aged autistic children, using an autonomous robot. While both robots4autism and the Scassellati et al. (2018) system can present content adaptively to different children, neither offers the degree of flexibility and type of personalisation that educators within autism-specific special education settings seem to envision (e.g., programming the robot to use particular phrasing).

At present, the robotics industry may be offering something closer to educators' desired flexible use and to the "single, simple point of control" that Hughes-Roberts and Brown suggested (2015, p. 52). There are several tablet-based controls for commercially available robot NAO, such as the "AskNAO Tablet" app (Softbank and ERM Robotique https://www.asknao-tablet.com/), which offers a range of controls from push-button selection of pre-programmed actions to integrating with a powerful desktop program (Choreographe) for programming new robot behaviours. They also have a companion blocks-based visual programming language, AskNAO Blockly. Also using NAO, the EU-funded DREAM project developed a simplified, tablet-controlled version of their original autonomous system, DREAM Lite (Mazel and Matu, 2019), which therapists in Romania found fairly easy to learn and use, though they also requested further simplification (Cao et al., 2019). In addition to the contributions made by doing controlled robot experiments and developing specific teaching programmes, it would be a much-needed contribution for HRI and Human-Computer Interaction researchers and the commercial robotics industry to collaborate with educators, developing or modifying robot programming/control platforms to be both usable and secure.

LIMITATIONS

This study is not without limitations. First, given the convenience sampling of participants, we cannot be sure that our findings reflect the views of autism educators in all special schools across England, or of educators working with autistic students in mainstream schools (in which the majority of autistic students are educated; Department for Education, 2018). Nevertheless, given the current interviewees' expertise in working with autistic students, particularly those with high support needs, they are likely to have provided particularly informed and nuanced views on the potential of robots as educational tools, as our findings attest.

A second key limitation is that the interviews prioritised the concerns of the larger DE-ENIGMA project in asking specifically about *humanoid* robots. Our respondents may have had different views and suggested other uses for animal-like robots such as Keepon (Kozima et al., 2007), or non-biomimetic robots. It is unclear whether the consensus present in the current dataset, such as using robots as "stepping stones" to human interaction, would also be present if discussing other robots. For the same reason, the interviews also specifically prompted respondents to consider applications for social and emotional skills teaching, but did not prompt them about academic or other applications, somewhat skewing the dataset in terms of the types of educational activities discussed.

CONCLUSION

The findings of this study show multiple, strong points of agreements with how related participant groups (e.g., Hughes-Roberts and Brown, 2015; Huijnen et al., 2016, 2017, 2019) have conceptualised robots as potential tools for autism education. Importantly, our educators were not uncritically approving of the use of robots in the classroom (see also (Serholt et al., 2017), for similar views from mainstream educators). Rather, they carefully outlined specific use-cases and circumstances in which robots were predicted to be beneficial (e.g., as "stepping stones" to social interaction), and conditions that would need to be met to ensure their adoption in the classroom, including integration with educational curricula, and the capacity to personalise robots to meet the specific needs of individual, autistic learners.

The findings suggest several promising avenues for future research. First, educators repeatedly highlighted the idea, prevalent in HRI literature, that robots' predictability and consistency of behaviour should benefit autistic learners in particular (e.g., Rudovic et al., 2017; Straten et al., 2018); it should reduce demands on them, put them at ease, and potentially facilitate learning. These claims are logical based on the diagnostic features of autism and current educational practices that aim to offer children predictability and structure at school (e.g., Mesibov and Shea, 2010), as well as theories of autistic perception and information processing (e.g., Pellicano and Burr, 2012; Lawson et al., 2014). However, they have not been rigorously operationalised and evaluated at a behavioural level. Research is required to test these widely-held beliefs about the benefits of robot predictability and exactly how it may affect children in learning contexts.

Second, the capacity of humanoid robots to support autistic children in developing transferrable, generalisable skills is not currently supported by clear research evidence. Given the centrality of educator views that robots need to be a stepping stone to human-human interaction, investigating skill transfer should be an urgent priority. Further generalisation studies might also test educators' beliefs, as expressed herein, that a humanoid robot might *better* teach, or support transfer of, social skills, than would other robot morphologies. These questions are not only the domain of autism education researchers; they should also concern robotics researchers. Based on this current research, robots for autism education—no matter how appealing or user-friendly—would not meet educators' and children's needs if they did not consistently support skill transfer. Robots that only facilitate learning gains within robot-based activities (i.e., *training effects*) are unlikely to be ethically or financially justifiable for educators or the broader autism community.

Educator interview studies are a valuable source about of information for robotics researchers and industry about the needs of child and adult users, but are not in themselves sufficient to bridge the "deployment gap" between preliminary, lab-based research, and the vision of robots as educational tools. Huijnen et al. describe this gap perfectly, writing:

> "For socially interactive robots to actually make a difference to the lives of children with ASD and their carers, they have to find their way out from case studies with 'standalone' robots in robotics labs to… education environments as part of daily activities/therapies. Being effective in eliciting a certain target behaviour of a particular child in a lab environment, will not automatically ensure… adoption of use by professionals in the field" (2016, p. 446).

Greater engagement with educators—and other key stakeholders, including autistic children themselves—during design, implementation, and evaluation should help to ensure that the resulting robotics systems and programmes are relevant to autistic learners and those who support them, sufficiently tailored to the realities of their everyday learning contexts, and consistent with their values (e.g., Lloyd and White, 2011). Such participatory processes are being championed across autism research (Nicolaidis et al., 2011; Pellicano and Stears, 2011; Fletcher-Watson et al., 2019), but especially within technology-related autism research (Frauenberger et al., 2011; Porayska-Pomsta et al., 2012; Brosnan et al., 2016). The children's interaction design community can offer useful examples and methodological guidance for undertaking participatory technology research with educators and children, including children on the autism spectrum (e.g., Frauenberger et al., 2013).

In advocating for HRI researchers to engage more fully with autism education practitioners while planning, developing, and evaluating robotic tools, we realise that this could pose a substantial change to many established ways of working, and that fully co-produced research might not be possible on many projects. Yet stakeholder participation in research—beyond being a passive participant or subject—can take many forms, including as advisors, as consultants, or as full decision-making partners throughout a project. The risks of designing robots that do not consider stakeholders' views, needs and contexts could be far-reaching for research and industry, especially given the costs of developing and deploying robots.

The current findings highlight that there will be no one-size-fits all design "solution" for robotics in autism education, and that current "solutions" may pose later challenges for autistic children. Such future work therefore needs to involve key stakeholders in the design and implementation process (see also Serholt et al., 2017), designing *with* educators, parents and autistic children, rather than *to, on,* or *for* them, to ensure that this work has a direct and sustained impact on those who need it most. This process will require beginning from a point of rigorously co-investigating the assumed and predicted benefits of robotics for autistic children, and balancing these against potential interpersonal, developmental, and resource costs. We envision that robot design driven by technical innovation will be increasingly combined with—or shaped by—approaches that prioritise the needs and values of users.

DATA AVAILABILITY STATEMENT

The datasets generated for this study will not be made publicly available because participants did not consent to future re-use of their interview data by other researchers.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The UCL Institute of Education Research Ethics Committee (REC857). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

VC, SP, BS, TT, and EP devised and piloted the interview schedule. EA, SM, and AA recruited participants. EA, SM, and AA interviewed participants. AA and EP analysed the data. AA and EP drafted the manuscript. All authors commented on and edited the manuscript prior to submission.

FUNDING

This work was financially supported by the European Council under the European Council's Horizon 2020 Work Programme (H2020-ICT-2015-688835)/Grant Agreement No. 688835 (DE-ENIGMA: Multi-Modal Human-Robot Interaction for Teaching and Expanding Social Imagination in Autistic Children).

ACKNOWLEDGMENTS

The authors and the DE-ENIGMA project would like to thank the educators who generously gave their time to this study, and the three schools who hosted focus groups. We also thank Rebecca Sealy, Demi Eades, and the DE-ENIGMA student volunteers who supported the timing, note-taking, and recording for many interviews.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frobt.2019.00107/full#supplementary-material

REFERENCES

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. Washington, DC: American Psychiatric Publishing. doi: 10.1176/appi.books.9780890425596

Anzalone, S. M., Tilmont, E., Boucenna, S., Xavier, J., Jouen, A. L., Bodeau, N., et al. (2014). How children with autism spectrum disorder behave and explore the 4-dimensional (spatial 3D+ time) environment during a joint attention induction task with a robot. *Res. Autism Spectr. Disord.* 8, 814–826. doi: 10.1016/j.rasd.2014.03.002

Begum, M., Serna, R. W., and Yanco, H. A. (2016). Are robots ready to deliver autism interventions? A comprehensive review. *Int. J. Soc. Robot.* 8, 157–181. doi: 10.1007/s12369-016-0346-y

Bondy, A. S., and Frost, L. A. (1994). The picture exchange communication system. *Focus Autistic Behav.* 9, 1–19. doi: 10.1177/108835769400900301

Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qual. Res. Psychol.* 3, 77–101. doi: 10.1191/1478088706qp063oa

Brosnan, M., Parsons, S., Good, J., and Yuill, N. (2016). How can participatory design inform the design and development of innovative technologies for autistic communities? *J. Assist. Technol.* 10, 115–120. doi: 10.1108/JAT-12-2015-0033

Cao, H. L., Esteban, P., Bartlett, M., Baxter, P. E., Belpaeme, T., Billing, E., et al. (2019). Robot-enhanced therapy: development and validation of a supervised autonomous robotic system for autism spectrum disorders therapy. *IEEE Robot Autom Mag.* 26, 49–58. doi: 10.1109/MRA.2019.2904121

Cheng, Y. W., Sun, P. C., and Chen, N. S. (2018). The essential applications of educational robot: requirement analysis from the perspectives of experts, researchers and instructors. *Comput. Educ.* 126, 399–416. doi: 10.1016/j.compedu.2018.07.020

Coeckelbergh, M., Pop, C., Simut, R., Peca, A., Pintea, S., David, D., et al. (2016). A survey of expectations about the role of robots in robot-assisted therapy for children with ASD: Ethical acceptability, trust, sociability, appearance, and attachment. *Sci. Eng. Ethics*, 22, 47–65. doi: 10.1007/s11948-015-9649-x

Dautenhahn, K. (1999). "Robots as social actors: aurora and the case of autism," in *Proceedings CT99, the 3rd International Cognitive Technology Conference* (San-Francisco, CA), 359–374.

Dautenhahn, K., and Werry, I. (2004). Towards interactive robots in autism therapy: Background, motivation and challenges. *Prag. Cognit.* 12, 1–35. doi: 10.1075/pc.12.1.03dau

David, D. O., Costescu, C. A., Matu, S., Szentagotai, A., and Dobrean, A. (2018). Developing joint attention for children with autism in robot-enhanced therapy. *Int. J. Soc. Robot.* 10, 595–605. doi: 10.1007/s12369-017-0457-0

Department for Education (2018). *Data From: Special Educational Needs in England - January 2018: National Tables*. Available online at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/729209/SEN_2018_National_tables.xlsx

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

Diep, L., Cabibihan, J. J., and Wolbring, G. (2015). Social Robots: Views of special education teachers," in *Proceedings of the 3rd 2015 Workshop on ICTs for Improving Patients Rehabilitation Research Techniques* (Lisbon: ACM), 160–163. doi: 10.1145/2838944.2838983

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. *Autonom. Robots* 24, 147–157. doi: 10.1007/s10514-007-9056-5

Eaves, L. C., and Ho, H. H. (1997). School placement and academic achievement in children with autistic spectrum disorders. *J. Dev. Phys. Disabil.* 9, 277–291. doi: 10.1023/A:1024944226971

Fletcher-Watson, S., Adams, J., Brook, K., Charman, T., Cusack, J., Crane, L., et al. (2019). Making the future together: Shaping autism research through meaningful participation. *Autism* 23, 943–953. doi: 10.1177/1362361318786721

Frauenberger, C., Good, J., Alcorn, A., and Pain, H. (2013). Conversing through and about technologies: design critique as an opportunity to engage children with autism and broaden research(er) perspectives. *Int. J. Child-Comput. Int.* 1, 38–49. doi: 10.1016/j.ijcci.2013.02.001

Frauenberger, C., Good, J., and Keay-Bright, W. (2011). Designing technology for children with special needs: bridging perspectives through participatory design. *CoDesign*, 7, 1–28. doi: 10.1080/15710882.2011.587013

Fridin, M., and Belokopytov, M. (2014). Acceptance of socially assistive humanoid robot by preschool and elementary school teachers. *Comput. Hum. Behav.* 33, 23–31. doi: 10.1016/j.chb.2013.12.016

Gernsbacher, M. A. (2017). Editorial perspective: the use of person-first language in scholarly writing may accentuate stigma. *J. Child Psychol. Psychiatry*, 58, 859–861. doi: 10.1111/jcpp.12706

Gray, C. (1994). *Social Stories.* Arlington, TX: Future Horizons.

Guston, D. H. (2014). Understanding 'anticipatory governance'. *Soc. Stud. Sci.* 44, 218–242. doi: 10.1177/0306312713508669

Hughes-Roberts, T., and Brown, D. (2015). "Implementing a robot-based pedagogy in the classroom: initial results from stakeholder interviews," in *2015 International Conference on Interactive Technologies and Games* (Nottingham: IEEE), 49–54. doi: 10.1109/iTAG.2015.18

Huijnen, C. A., Lexis, M. A., and de Witte, L. P. (2016). Matching robot KASPAR to autism spectrum disorder (ASD) therapy and educational goals. *Int. J. Soc. Robot.* 8, 445–455. doi: 10.1007/s12369-016-0369-4

Huijnen, C. A., Lexis, M. A., Jansens, R., and de Witte, L. P. (2017). How to implement robots in interventions for children with autism? A co-creation study involving people with autism, parents and professionals. *J. Autism Dev. Disord.* 47, 3079–3096. doi: 10.1007/s10803-017-3235-9

Huijnen, C. A., Lexis, M. A., Jansens, R., and de Witte, L. P. (2019). Roles, strengths and challenges of using robots in interventions for children with autism spectrum disorder (ASD). *J. Autism Dev. Disord.* 49, 11–21. doi: 10.1007/s10803-018-3683-x

Johnston, J. R. (1988). Generalization: the nature of change. *Lang. Speech Hear. Serv. Sch.* 19, 314–329. doi: 10.1044/0161-1461.1903.314

Kennedy, J., Lemaignan, S., and Belpaeme, T. (2016). "The cautious attitude of teachers towards social robots in schools," in *Robots 4 Learning Workshop at IEEE RO-MAN 2016*, (New York, NY).

Kenny, L., Hattersley, C., Molins, B., Buckley, C., Povey, C., and Pellicano, E. (2016). Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism* 20, 442–462. doi: 10.1177/1362361315588200

King, A. M., Brady, K. W., and Voreis, G. (2017). "It's a blessing and a curse": Perspectives on tablet use in children with autism spectrum disorder. *Autism Dev. Lang. Impair.* 2:2396941516683183. doi: 10.1177/2396941516683183

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

Lawson, R. P., Rees, G., and Friston, K. J. (2014). An aberrant precision account of autism. *Front. Hum. Neurosci.* 8:302. doi: 10.3389/fnhum.2014.00302

Lloyd, K., and White, J. (2011). Democratizing clinical research. *Nature* 474, 277–278. doi: 10.1038/474277a

Mazel, A., and Matu, S. (2019). "DREAM lite: simplifying robot assisted therapy for ASD," in *The Second Workshop of Social Robots in Therapy and Care, 14th ACM/IEEE International Conference on Human-Robot Interaction* (Daegu), 11–14.

McCleery, J. P. (2015). Comment on technology-based intervention research for individuals on the autism spectrum. *J. Autism Dev. Disord.* 45, 3832–3835. doi: 10.1007/s10803-015-2627-y

Mesibov, G. B., and Shea, V. (2010). The TEACCH program in the era of evidence-based practice. *J. Autism Dev. Disord.* 40, 570–579. doi: 10.1007/s10803-009-0901-6

Nicolaidis, C., Raymaker, D., McDonald, K., et al. (2011). Collaboration strategies in non-traditional community-based participatory research partnerships: lessons from an academic–community partnership with autistic self-advocates. *Prog. Commun. Health Partnersh.* 5:143. doi: 10.1353/cpr.2011.0022

Pellicano, E., and Burr, D. (2012). When the world becomes 'too real': a Bayesian explanation of autistic perception. *Trends Cognit. Sci.* 16, 504–510. doi: 10.1016/j.tics.2012.08.009

Pellicano, E., and Stears, M. (2011). Bridging autism, science and society: moving towards an ethically informed approach to autism research. *Autism Res.* 4, 271–282. doi: 10.1002/aur.201

Porayska-Pomsta, K., Frauenberger, C., Pain, H., Rajendran, G., Smith, T., Menzies, R., et al. (2012). Developing technology for autism: an interdisciplinary approach. *Pers. Ubiquitous Comput.* 16, 117–127. doi: 10.1007/s00779-011-0384-2

Robins, B., and Dautenhahn, K. (2006). "The role of the experimenter in HRI research-a case study evaluation of children with autism interacting with a robotic toy," in *ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication* (Hatfield: IEEE), 646–651.

Robins, B., and Dautenhahn, K. (2017). "The iterative development of the humanoid robot kaspar: an assistive robot for children with autism," in *Social Robotics: 9th International Conference, ICSR 2017* (Tsukuba: Springer).

Robins, B., Dautenhahn, K., Ferrari, E., Kronreif, G., Prazak-Aram, B., Marti, P., et al. (2012). Scenarios of robot-assisted play for children with cognitive and physical disabilities. *Int. Stud.* 13, 189–234. doi: 10.1075/is.13.2.03rob

Rudovic, O., Lee, J., Mascarell-Maricic, L., Schuller, B. W., and Picard, R. W. (2017). Measuring engagement in robot-assisted autism therapy: a cross-cultural study. *Front. Robot. AI* 4:36. doi: 10.3389/frobt.2017.00036

Salvador, M. J., Silver, S., and Mahoor, M. H. (2015). "An emotion recognition comparative study of autistic and typically-developing children using the zeno robot" in *2015 IEEE International Conference on Robotics and Automation (ICRA) (*Seattle, WA), 6128–6133. doi: 10.1109/ICRA.2015.7140059

Scassellati, B., Admoni, H., and Matarić, M. (2012). Robots for use in autism research. *Ann. Rev. Biomed. Eng.* 14, 275–294. doi: 10.1146/annurev-bioeng-071811-150036

Scassellati, B., Boccanfuso, L., Huang, C. M., Mademtzi, M., Qin, M., Salomons, N., et al. (2018). Improving social skills in children with ASD using a long-term, in-home social robot. *Sci. Robot.* 3:eaat7544. doi: 10.1126/scirobotics.aat7544

Schreibman, L., Dawson, G., Stahmer, A. C., Landa, R., Rogers, S. J., McGee, G. G., et al. (2015). Naturalistic developmental behavioural interventions: empirically validated treatments for autism spectrum disorder. *J. Autism Dev. Disord.* 45, 2411–2428. doi: 10.1007/s10803-015-2407-8

Serholt, S., Barendregt, W., Vasalou, A., Alves-Oliveira, P., Jones, A., Petisca, S., et al. (2017). The case of classroom robots: Teachers' deliberations on the ethical tensions. *AI Soc.* 32, 613–631. doi: 10.1007/s00146-016-0667-2

Sinclair, J. (1999). *Why I Dislike 'Person-First' Language*. Jim Sinclair's website. Available online at http://web.archive.org/web/20090210190652/http://web.syr.edu/~jisincla/person_first.htm (accessed August 8, 2019).

Straten, C. L., Smeekens, I., Barakova, E., Glennon, J., Buitelaar, J., and Chen, A. (2018). Effects of robots' intonation and bodily appearance on robot-mediated communicative treatment outcomes for children with autism spectrum disorder. *Person. Ubiquit. Comput.* 22, 379–390. doi: 10.1007/s00779-017-1060-y

Tager-Flusberg, H., and Kasari, C. (2013). Minimally verbal school-aged children with autism spectrum disorder: The neglected end of the spectrum. *Autism Res.* 6, 468–478. doi: 10.1002/aur.1329

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Wainer, A. L., and Ingersoll, B. R. (2011). The use of innovative computer technology for teaching social communication to individuals with autism spectrum disorders. *Res. Autism Spectr. Disord.* 5, 96–107. doi: 10.1016/j.rasd.2010.08.002

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? *Journal of Autism Dev. Disord.* 45, 3726–3734. doi: 10.1007/s10803-013-1918-4

Wass, S. V., and Porayska-Pomsta, K. (2014). The uses of cognitive training technologies in the treatment of autism spectrum disorders. *Autism* 18, 851–871. doi: 10.1177/1362361313499827

Whyte, E. M., Smyth, J. M., and Scherf, K. S. (2015). Designing serious game interventions for individuals with autism. *J. Autism Dev. Disord.* 45, 3820–3831. doi: 10.1007/s10803-014-2333-1

Yoder, P. J., McDuffie, A. S., Charman, T., and Stone, W. (2006). "Treatment of responding to and initiating joint attention," in *Social and Communication Development in Autism Spectrum Disorders: Early Identification, Diagnosis, and Intervention*, eds T. Charman and W. Stone (New York, NY: The Guilford Press), 117–142.

Yun, S.-S., Kim, H., Choi, J., and Park, S.-K. (2016). A robot-assisted behavioral intervention system for children with autism spectrum disorders. *Robot. Autonom. Syst.* 76, 58–67. doi: 10.1016/j.robot.2015.11.004

Conflict of Interest: 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.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 06 November 2019
doi: 10.3389/frobt.2019.00110

Check for updates

Long-Term Personalization of an In-Home Socially Assistive Robot for Children With Autism Spectrum Disorders

Caitlyn Clabaugh[1†], Kartik Mahajan[1*†], Shomik Jain[1], Roxanna Pakkar[1], David Becerra[1], Zhonghao Shi[1], Eric Deng[1], Rhianna Lee[1], Gisele Ragusa[2] and Maja Matarić[1]

[1] Interaction Lab, Department of Computer Science, Viterbi School of Engineering, University of Southern California, Los Angeles, CA, United States, [2] STEM Education Research Group, Division of Engineering Education, Viterbi School of Engineering, University of Southern California, Los Angeles, CA, United States

Socially assistive robots (SAR) have shown great potential to augment the social and educational development of children with autism spectrum disorders (ASD). As SAR continues to substantiate itself as an effective enhancement to human intervention, researchers have sought to study its longitudinal impacts in real-world environments, including the home. Computational personalization stands out as a central computational challenge as it is necessary to enable SAR systems to adapt to each child's unique and changing needs. Toward that end, we formalized personalization as a hierarchical human robot learning framework (hHRL) consisting of five controllers (disclosure, promise, instruction, feedback, and inquiry) mediated by a meta-controller that utilized reinforcement learning to personalize instruction challenge levels and robot feedback based on each user's unique learning patterns. We instantiated and evaluated the approach in a study with 17 children with ASD, aged 3–7 years old, over month-long interventions in their homes. Our findings demonstrate that the fully autonomous SAR system was able to personalize its instruction and feedback over time to each child's proficiency. As a result, every child participant showed improvements in targeted skills and long-term retention of intervention content. Moreover, all child users were engaged for a majority of the intervention, and their families reported the SAR system to be useful and adaptable. In summary, our results show that autonomous, personalized SAR interventions are both feasible and effective in providing long-term in-home developmental support for children with diverse learning needs.

Keywords: long-term human-robot interaction, personalization, socially assistive robotics, reinforcement learning, home robot, autism spectrum disorders, early childhood

Edited by:
Raquel Ros, Universitat Ramon Llull, Spain

Reviewed by:
Tony Belpaeme, University of Plymouth, United Kingdom
Mehdi Khamassi, Centre National de la Recherche Scientifique (CNRS), France

****Correspondence:***
Kartik Mahajan
kmahajan@usc.edu

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *19 June 2019*
Accepted: *15 October 2019*
Published: *06 November 2019*

Citation:
Clabaugh C, Mahajan K, Jain S, Pakkar R, Becerra D, Shi Z, Deng E, Lee R, Ragusa G and Matarić M (2019) Long-Term Personalization of an In-Home Socially Assistive Robot for Children With Autism Spectrum Disorders. Front. Robot. AI 6:110. doi: 10.3389/frobt.2019.00110

1. INTRODUCTION

Human development follows non-linear trajectories unique to each individual (Vygotsky, 1978). Therefore, socially assistive interventions need to be tailored toward the specific needs and preferences of each participant over time. In a long-term setting, this means interventions must continuously and rapidly adapt toward the user's unique personality. Given the complexity, unpredictability, and uniqueness of each user's progress, intervention strategies must be adapted *in situ* via untrained human feedback. Creating autonomous long-term personalized adaptation poses many computational and engineering challenges.

Benefits of personalization are well-established across the domains of education (Bloom, 1984; Anderson et al., 2001) and healthcare (Artz and Armour-Thomas, 1992; Beevers and McGeary, 2012; Cesuroglu et al., 2012; Swan, 2012). While personalized services are paramount, they are neither universally nor equitably affordable. This provides motivation for human-machine interaction research that seeks to develop personalized assistance via socially assistive agents, whether disembodied, virtually embodied (DeVault et al., 2014), or physically embodied (Matarić, 2017).

Socially assistive robotics (SAR) combines robotics and computational methods to broaden access to personalized, socially situated, and physically co-present interventions (Feil-Seifer and Matarić, 2011). A large body of work has supported the importance of physical embodiment (Deng et al., 2019), including its role in increasing compliance (Bainbridge et al., 2008), social engagement (Lee et al., 2006; Wainer et al., 2006), and cognitive learning gains (Leyzberg et al., 2012). Correspondingly, there has been a significant body of work using various types of robots for children with autism spectrum disorders (ASD) in short-term studies (Diehl et al., 2012; Scassellati et al., 2012; Begum et al., 2016), and one long-term study (Scassellati et al., 2018).

The majority of past work with SAR for ASD has been related to social skills. However, it is well-established that learning in general is impacted by social factors; this is particularly important for young learners, because their learning is most often socially mediated (Durlak, 2011). Social difficulties often interfere with children's learning; therefore embedding social contexts in learning environments presents a developmentally appropriate practice that is preferable over isolating social behaviors from cognitive activities (Zins et al., 2004). Consequently, this work addresses the social and cognitive learning domains in tandem, in an intervention that is specifically designed for such learning by children with ASD (White et al., 2007; Guadalupe, 2016).

Personalizing the learning process is especially important in ASD. Given sufficient domain knowledge, personalization of SAR can be achieved through human-in-the-loop or Wizard of Oz (WoZ) frameworks, wherein intervention strategies are mapped to individuals *a priori* or *in situ* via human input (Riek, 2012). However, in practice, considering diverse individual needs and the noise of real-world environments, and the scale of need in ASD, non-autonomous personalization of SAR is infeasible. Reinforcement learning (RL) methods have been successfully applied to adapting to a user's learning habits over time, particularly in early child development studies (Ros et al., 2011). Moreover, recent long-term SAR studies have demonstrated success in maintaining persistent co-present support for educators, students, and caregivers (Bongaarts, 2004). There is therefore an opportunity to develop RL-based personalized long-term learning SAR systems, especially when teaching abstract concepts, such as mathematics (Clabaugh et al., 2015).

In this work, we propose a *personalized* SAR intervention framework that can provide accessible and effective long-term, in-home support for children with ASD. To accommodate the variable nature of ASD, our framework personalizes to each user's individual needs. To that end, we introduce a hierarchical framework for Human Robot Learning (hHRL) that decomposes SAR interventions into computationally tractable state-action subspaces contained with a meta-controller. The meta-controller consists of disclosure, promise, instruction, feedback, and inquiry controllers that personalize instruction challenge levels and robot feedback based on each child's unique learning patterns. The framework is implemented and evaluated in a fully autonomous SAR system deployed in homes for session-based, single-subject interventions with 17 child participants diagnosed with ASD aged 3–7 years old. Using space-themed mathematics problems, the system combined tenets of educational robotics and computational personalization to maximize each child participant's cognitive gains. Our findings show that the SAR system successfully personalized its instruction and feedback to each participant over time. Furthermore, most families reported the SAR system to be useful and adaptable, and correspondingly, all users were engaged for a majority of the in-home intervention. As a result, all participants showed improvements in math skills and long-term retention of intervention content. These outcomes demonstrate that computational personalization methods can be successfully incorporated in long-term personalized SAR deployments to support children with diverse learning needs.

This paper is organized as follows. *Background* overviews SAR in the relevant contexts of learning, ASD, and personalization. *Formalizing Personalization in SAR* describes the hierarchical human robot learning framework, with a focus on personalization of the challenge level and robot feedback. *Personalized SAR Intervention Design* details the study design, data collection, and outcome measures. The *Results* section details the adaptation performance of the SAR system, its influence on user engagement, participating families' perspectives, and cognitive learning gains over the long-term interaction. *Discussion* and *Conclusion* summarize key insights and recommendations for future work.

2. BACKGROUND

Socially Assistive Robotics (SAR) lies at the intersection of socially interactive robotics and assistive robotics, and focuses on developing intelligent, socially interactive robots that provide assistance through social interaction, with measurable outcomes (Feil-Seifer and Matarić, 2005; Matarić and Scassellati, 2016). We review the relevant background in the main contribution areas of this work: SAR for *learning* (section 2.1) and SAR for *personalization* (section 2.2), both with a particular emphasis on the ASD context, given particular challenges and opportunities for SAR.

2.1. SAR for Learning

A large body of evidence across multiple disciplines supports personalized instruction as a means of positively impacting development and motivation of *individual* learners. Examples include personalized tutoring systems in human-computer interaction research (Wenger, 2014), personalized robot tutors in HRI and SAR research (Leyzberg et al., 2014) and optimal challenge points (Guadagnoli and Lee, 2004), and the Zone of Proximal Development methodologies in education research (Chaiklin, 2003).

A significant body of SAR research has focused on user learning, with a specific focus on developing personalized robot tutors for young children (Clabaugh and Matarić, 2019). Many SAR and HRI studies have found a robot's embodiment to augment learning in a variety of settings (Gallese and Goldman, 1998; Lee et al., 2006; Gazzola et al., 2007; Wainer et al., 2007; Bainbridge et al., 2008; Leyzberg et al., 2012; Fridin and Belokopytov, 2014). Additionally, several studies on intelligent tutoring systems (ITS) have involved computational models of student learning patterns; however, in contrast to SAR, these works have predominately focused on university students in highly controlled environments (Anderson, 1985; Murray, 1999). From that body of past work, key principles about SAR for learning have been grounded in theories of embodied cognition, situated learning, and user engagement.

Embodied cognition research has shown that knowledge is directly tied to perceptual, somatosensory, and motoric experience, and that a robot's physical embodiment can help contextualize a user's ideas (Niedenthal, 2007; Deng et al., 2019). For example, SAR has helped participants develop motor (Goldin-Meadow and Beilock, 2010), behavioral (Fong et al., 2003), and cognitive skills (Toh et al., 2016). SAR has also shown success in helping users learn abstract concepts; for example, Clabaugh et al. (2015) implemented a SAR system that used deictic gestures to help preschoolers learn number concepts.

Situated learning refers to the importance of the social and physical environment on the learning process and outcomes (McLellan, 1996). Cognitive gains are dependent on context and are enhanced by social interaction (Anderson et al., 1996). Therefore, SAR intervention efficacy must be analyzed in real-world learning settings, involving user learning in various spatial and social contexts (Sabanovic et al., 2006). Environmental conditions impact the quality of SAR interactions and the resulting assistive outcomes. However, real-world scenarios are inherently noisier and less predictable, requiring more complex experimental designs and robust robot platforms (Ros et al., 2011).

User engagement is an important measure of SAR's effectiveness and is inherently tied to learning. In the context of HRI, engagement is widely accepted as a combination of behavioral, affective, and cognitive constructs. Specifically, engagement involves on-task behavior, interest in the robot and task at hand, and a willingness to remain focused (Scassellati et al., 2012). Rudovic et al. (2018) successfully modeled users' engagement with a personalized deep learning framework, however the model was developed *post-hoc*, not in real time. As discussed in Kidd (2008), maintaining user engagement in real time is a major challenge for real-world, long-term studies, as are overcoming technological difficulties and accounting for external human actors.

All of the challenges of SAR for learning are significantly amplified in the ASD context, but ASD is also the context where the success of SAR in supporting learning is especially promising. ASD is a complex developmental disorder that is often marked by delays in language skills and social skills, including turn-taking, perspective-taking, and joint attention (White et al., 2007). Personalized therapeutic and learning interventions are critical for individuals with ASD, but the substantial time and financial resources required for such services make them inaccessible to many (Ospina et al., 2008; Lavelle et al., 2014), creating an opportunity for SAR support.

There is a large and growing body of research on using SAR for a variety of ASD interventions, as reviewed in Diehl et al. (2012), Scassellati et al. (2012), and Begum et al. (2016). SAR has been shown to help children with ASD develop behavioral and cognitive skills, specifically increased attention (Duquette et al., 2008), turn-taking (Baxter et al., 2013), social interaction (Robins et al., 2005), and many other skills. SAR's ability to perceive, respond, and adapt to user behavior is especially critical in the ASD context (Clabaugh and Matarić, 2019), as users with ASD vary greatly in symptoms and severities, underscoring the need for personalization, as our work also demonstrates.

2.2. Personalization in SAR

SAR systems have shown great potential for providing long-term situated support for meeting individual learning needs. Autonomous or computational personalization in SAR often seeks to maximize the participants' focus and performance, using rule-, model-, or goal-based approaches to personalization.

Rule-based approaches to personalization have been successful in both short-term and long-term SAR interventions. For example, Ramachandran et al. (2018) designed single session interventions where the robot encouraged participants to think out loud. Scenarios were presented based on whether a participant successfully answered a question, and this simple rule-based method resulted in learning gains across all users. Additional studies have expanded rule-based approaches for sequential interactions using hierarchical decision trees (Kidd and Breazeal, 2008; Reardon et al., 2015). Furthermore, in a study setup similar to ours, Scassellati et al. (2018) developed a personalized SAR system for month-long interventions with children with ASD. The system adapted the challenge level of activities using past performance and fixed thresholds. As a result, participants showed increased engagement to the robot and improved attention skills with adults when not in the presence of the robot. In contrast, this work personalizes feedback and challenge level using a goal-based approach, discussed below.

Model-based approaches use models to evaluate the user's success and make optimal decisions. Bayesian Knowledge Tracing (BKT), a domain-specific form of Hidden Markov Models (HMMs), is a common *model-based* approach to personalization in SAR where the hidden state is based on the user's performance and loosely represents their knowledge (Desmarais and Baker, 2012; van De Sande, 2013). For example, BKT can assess how well a participant understands a concept, such as basic addition by examining the sequence of the user's correct and incorrect responses. To represent the variability present in most learning interactions, BKT uses two domain-specific parameters: the probability that a participant will slip and the probability that they will guess. These parameters are dependent on the interaction context; students with ASD may have difficulty concentrating for extended periods and thus may slip more frequently than typically-developing users (Schiller, 1996). BKT has been successfully applied in SAR; Gordon and Breazeal (2015) and Schodde et al. (2017) used it to adapt to user age and experience, leading to increased learning gains. Leyzberg et al. (2014) applied BKT to training a SAR system to help users solve challenging puzzles more quickly. While outside of ASD, these studies demonstrate the value of BKT in adapting SAR to varying learner needs.

Goal-based methods help the SAR system to select actions that maximize the user's progress toward an assistive outcome. Reinforcement Learning (RL) is a popular goal-based approach, where each user action produces some reward representing progress toward the goal. Throughout the interaction, RL develops a unique, personalized strategy for each participant based on reward-favoring paths. Within HRI, RL has been used to maximize the user's affective state, leading to more effective interactions (Conn et al., 2008; Chan and Nejat, 2011; Castellano et al., 2012; Gordon et al., 2016). Prior studies have shown RL to require deep datasets given the noise of real-world environments. In a single-session context, Gordon et al. (2016) showed RL to successfully adapt in an average of three out of seven sessions. Castellano et al. (2012) also utilized RL to increase engagement; the model was trained on a 15 min interaction and was no better at adapting than a randomized empathetic policy. Conn et al. (2008) showed a RL which was able to adapt quickly, but simplified the robot state to three distinct behaviors. To enable a broader range of behaviors, Chan and Nejat (2011) implemented a hierarchical RL model to personalize feedback within a memory-based SAR interaction. As demonstrated by past work, long-term studies provide the datasets needed for effective RL-based personalization.

Related work has addressed improving social skills of children with ASD. To manage noisy environments and the unpredictable nature of ASD, two studies are particularly relevant as they used RL to parameterize action spaces and speed up robot learning. Velentzas and Khamassi (2018) used RL to personalize the robot's actions to maximize a child's engagement; the robot guided children through a Tower of Hanoi puzzle and used RL to effectively identify non-verbal cues and teach at the learning rate of the participant. That work parameterized the robot's action space to enable efficient decision making and learn single moves in the absence of traditional hierarchical models. Khamassi and Tzafestas (2018) utilized a parameterized action space to select the appropriate robot reaction that maximizes a child's engagement. That work also used RL to maximize the participant's engagement when interacting with a robot. By using the participant's gaze and past variations in engagement, their Q-Learning algorithm became more robust over time. The work used a parameterized environment to simultaneously explore a discrete action space (e.g., moving an object) and a continuous stream of movement features (e.g., expressivity, strength, velocity). These two studies provide insight into maximizing engagement in the absence of hierarchical models, especially when encouraging social interaction (e.g., talking, moving).

The work described in this paper is complementary but different from past work in that it analyzes a long-term SAR intervention for abstract concept learning, specifically helping children with ASD learn mathematics skills. As the next section details, a goal-based RL approach was developed to personalize the instruction and feedback provided to each child by the SAR system.

3. FORMALIZING PERSONALIZATION IN SAR

To address the challenge of long-term personalization in SAR in a principled way, we present a solution to the problem as a controller-based environment which we define as hierarchical human-robot learning (hHRL).

3.1. Human-Robot Learning

Past work has explored methods for computational personalization, with the objective of finding an optimal sequence of actions that steers the user toward a desired goal. While this problem has been studied in the contexts of user modeling in HCI (Fischer, 2001), machine teaching (Chen et al., 2018), as well as active (Cohn et al., 1996), and interactive machine learning (ML) (Amershi et al., 2014; Dudley and Kristensson, 2018), computational personalization is yet to be formalized in the context of SAR.

We define and formalize *Human-Robot Learning (HRL)* as the interactive and co-adaptive process of personalizing SAR. At the highest level, the quality of a SAR intervention can be assessed relative to some goal G. Since SAR contexts often involve long-term goals, success is better assessed via intermediate measures of progress toward G. Hence, it is important to design and represent SAR intervention interactions in a manner that maximizes observability. In this work, HRL is framed from the perspective of the robot, so that optimization is limited to the robot's actions and not those of the human, in contrast to human-robot collaboration and multi-agent learning (Littman, 1994; Nikolaidis et al., 2016).

3.2. Hierarchical HRL

We introduce *a hierarchical framework for HRL (hHRL)* as one that decomposes SAR interventions into computationally tractable state-action subspaces. Building on the work of Kulkarni et al. (2016), the hHRL framework is structured as a two-level hierarchy, shown in **Figure 1**. At the top level, a meta-controller considers high-level information about the intervention state and activates some lower-level controller. SAR-specific controllers wait for activation to select the robot's action based on a simplified state representation. The hHRL framework assumes that SAR interventions can be characterized by five abstract action categories: (1) instructions I, (2) promises P, (3) feedback F, (4) disclosures D, and (5) inquiries Q. Each category is modeled by a separate controller activated by the overarching meta-controller.

Each controller is responsible for a theoretical subset of SAR actions, henceforth referred to as *SAR acts*. In this work, SAR acts are formalized based on the directive, commissive, and representative illocutionary speech acts, or simply illocutions,

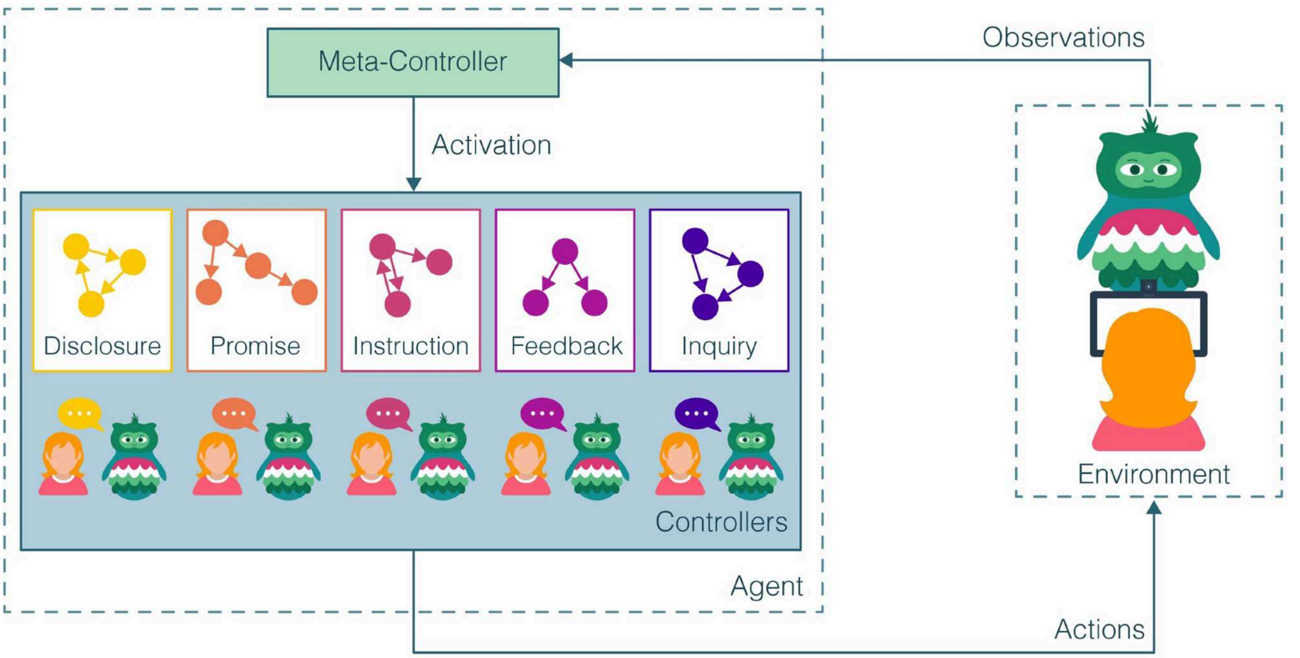


FIGURE 1 | The hierarchical framework for human-robot learning (hHRL) comprises of a two-level hierarchy: (1) The *meta-controller* takes high-level information about the current state of the intervention and activates a lower-level controller. (2) The lower-level controllers await activation to select the robot's action based on a simplified state representation, reward, and action category: instructions, promises, feedback, disclosures, and inquiries.

originally defined in linguistic semantics (Searle and Searle, 1969; Austin, 1975). Searle (1976) defined *illocutions* in terms of speaker, hearer, sincerity condition, psychological state, propositional content, and direction of fit. In the SAR context, the speaker is the robot, the hearer is the human user, and the direction of fit is either action-to-state, where the objective is to make the robot's action match the state of the intervention, or state-to-action, where the objective is to make the state of the intervention match what is expressed through the robot's action. Illocutions uniquely manifest themselves through other modes of communication, such as gestures (Mehrabian, 2017), pictures (Danesi, 2016), music (Kohn et al., 2004), and other multimodal signals (Horn, 1998; Forceville and Urios-Aparisi, 2009). These alternative signals are particularly relevant to SAR because robots have inherently expressive embodiments (Fong et al., 2003). Additionally, SAR interventions target special populations (Feil-Seifer and Matarić, 2005), such as linguistic minorities [e.g., American Sign Language (Stokoe et al., 1976)] or personals with disabilities that involve speech and language difficulties or delays [e.g., Dysarthria (Darley et al., 1969) or autism spectrum disorder (Kasari et al., 2012)]. Therefore, SAR acts are defined to be illocutions irrespective of communicative modality.

3.3. Abstract Controllers

Within the hierarchical model, *instructions* are defined as attempts by the robot to get the user to do something that might generate progress toward the intervention goal. Within the instruction controller, there may be some predefined or learned ordering among instructions, such as the level of challenge or specificity.

Feedback is defined as beliefs expressed to the user by the robot about their past and current interactions. The *direction of fit* is action-to-state, the *sincerity condition* is belief B, and the *propositional content* is that some past or current state s had or has some property p. In this way, feedback F is defined as a specific form of representatives. *Representatives* were defined by Searle (1976) to commit the speaker to the truth of the expressed proposition. The *propositional content* is information about the state relative to some instruction or goal. The feedback controller is responsible for selecting the information or assistance given to the user by the robot. Feedback can be modeled in a variety of ways, the impacts of which have been studied in psychology and human-machine interaction. Specifically, feedback can be adapted to match individual proficiency or independence, as in scaffolded (Finn and Metcalfe, 2010) or graded cueing models (Feil-Seifer and Matarić, 2012; Greczek et al., 2013). It can also be modeled to increase self-efficacy, as in the growth mindset (O'Rourke et al., 2014; Park et al., 2017) and constructive feedback models (Ovando, 1994). Additionally, feedback timing has also been studied (Kulik and Kulik, 1988), such as feedback in response to help-seeking (Roll et al., 2011) and disengagement (Leite et al., 2015).

Extrinsic motivation is a well-studied driver of behavior, explored in educational (Vallerand et al., 1992), professional (Amabile, 1993), and personal settings (Sansone and Harackiewicz, 2000), as well as a common measure in evaluating the effectiveness of human-robot interaction (Breazeal, 1998;

Dautenhahn, 2007; Fasola and Matarić, 2012). In the hHRL framework, *promises* are defined as commitments made by the robot for performing future actions that aim to motivate the user through the promise controller. Promises also relay information critical to collaboration and transparency, expressed via verbal or non-verbal signals, such as gross motion (Dragan et al., 2013). Although they are not directly tied to quantitative measures, promises help to make the robot more personable and consistent over a long-term study period.

Disclosures are defined as beliefs expressed to the user by the robot about its past or current self. The disclosure controller selects internal information for the robot to share with the user as a means of fostering human-robot reciprocity and solidarity. Robot transparency has shown to increase trust (Hancock et al., 2011; Yagoda and Gillan, 2012), improve collaboration (Breazeal et al., 2005; Kim and Hinds, 2006), and build empathetic relationships (Leite et al., 2013). Past work has also shown that non-verbal signals can be particularly effective in disclosing internal states, such as emotion (Bruce et al., 2002).

Inquiries are defined as attempts by the robot to get the user to express some truth. The *inquiry controller* selects what information the robot should attempt to elicit. Inquires may be posed for a variety of interaction benefits, such as improving engagement, relationship, and trust (Hancock et al., 2011). Inquires may also be used to gather feedback about the robot or information about the user, as in interactive machine learning (Amershi et al., 2014).

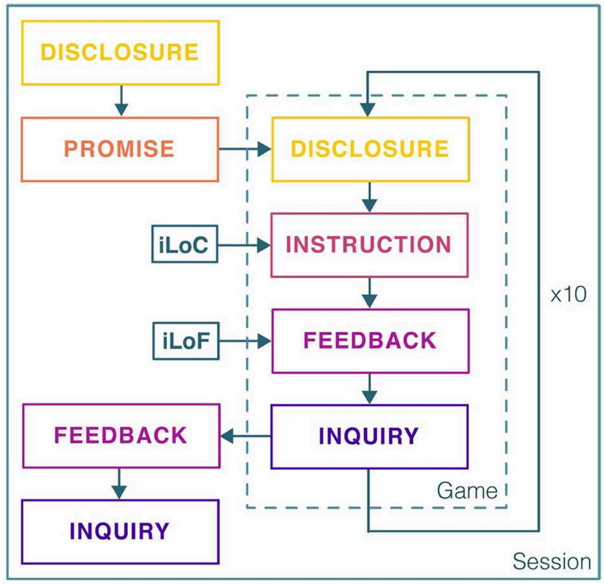


FIGURE 2 | The real-world long-term SAR intervention for early childhood math learning used a meta-controller that sequentially executed each controller. At the beginning of a session, the robot *disclosed* that it needed the child's help to reach a specific planet. It then *promised* that they would reach the planet if they completed all the games they needed to do that day. The child and robot played 10 games whose challenge (pLoC) and feedback (pLoF) levels were set by the computational personalization methods. At the end of each session, the robot congratulated the child on completing the games and reaching the promised planet. It then asked the child some open-ended questions about their day.

3.4. Computational Personalization

To personalize the SAR system, the proposed hHRL controller was instantiated as a group of domain controllers, based on the abstract controllers defined in section 3.3. **Figure 2** represents how the abstract controllers were contained within a domain-specific *meta-controller*. The meta-controller activated one controller at a time. We used insights and data from our prior work, reported in Clabaugh et al. (2015), to inform the design of the controllers for SAR personalization. Our prior study collected data from 31 typically developing preschool children who interacted with a SAR tutor in a single session at their child development center preschool. The data were used to develop a model that predicted a child's performance on the game. We used this performance model to bootstrap the prediction of the children's performance in our study. Specifically, the instruction controller in the personalization framework optimized the level of challenge (pLoC) and the level of feedback (pLoF) to match each child's performance, as described next.

3.4.1. Personalization of the Level of Challenge

Personalization was partially accomplished within the instruction controller. Learning games g were randomly sampled without replacement from all games G and parameterized by some personalized *level of challenge (LoC)* $c \in [1, 5]$. The instruction controller was designed to optimize LoC to match individual proficiency. This optimization problem was based on the concepts of *optimal challenge* from the *Challenge Point Framework* by Guadagnoli and Lee (2004) and from the research on the *Zone of Proximal Development* (ZPD) by Chaiklin (2003); both define the goal as challenging individuals enough that they are presented with new information, but not so much that there is too much new information to interpret.

Since the goal is long-term adaptation, *personalized LoC (pLoC)* was framed as a RL problem, trained using Q-learning (Watkins and Dayan, 1992). Within the instruction controller, a reward function was used to quantify the intervention state and supply Q-learning. The intervention, at time t, was defined by:

1. the current game g_t,
2. the current LoC c_t, and
3. the current number of mistakes m_t

More formally, the state space was defined as G and the action space was defined as C, for a total of $G \times C = 10 \times 5 = 50$ (state, action) couples. As previously explained, the next game g was randomly sampled without replacement from all games G. Therefore, the RL seeks to find a policy with the optimal LoC $c \in [1, 5]$ per game for the individual child.

Given the formulation above, the RL would select and evaluate different LoCs for each child. If some LoC in some game was too difficult or too easy for a child, the RL would learn to select a different LoC for that game, over time. This was accomplished through a reward function designed to maximize LoC without pushing the learner to make too many mistakes. Formally, at time t, let m_t be the number of mistakes a learner has made and M be the pre-defined threshold of maximum mistakes [we used a threshold of five, based on our empirical findings from prior research (Clabaugh et al., 2015)]. The reward function $R(t)$ returns a value equivalent to the LoC c_t, unless the $m_t > M$; then, $R(t)$ returns the inverse of LoC.

$$R(t) = c_t \cdot MC(t), \tag{1}$$

where

$$MC(t) = \begin{cases} 1, & \text{if } m_t \leq M \\ -1, & \text{otherwise.} \end{cases} \tag{2}$$

3.4.2. Personalization of the Level of Feedback

Feil-Seifer and Matarić (2012) and Greczek et al. (2014) applied the concept of *graded cueing* to adapt feedback in the context of SAR interventions for children with ASD. A similar approach was taken in this work to instantiate the feedback controller, as mentioned in section 3.4. Analogous to the instruction controller, the feedback controller was modeled as a MDP, wherein the decision was to select one of five levels of feedback (LoF) $f \in [1, 5]$ to match individual need. The feedback actions were specific to early mathematics learning.

Personalized LoF (pLoF) was framed as a RL problem, trained using Q-learning (Watkins and Dayan, 1992) over many repeated interactions. Within the feedback controller, at time t, the intervention was represented by four parameters:

1. the current game g_t,
2. the current LoF f_t,
3. the current number of mistakes m_t, and
4. the current number of help requests h_t

Similar to LoC, the state space for the feedback controller consisted of the $G = 10$ game states. The action space consisted of the four LoFs $f \in [1, 4]$. The final LoF $f = 5$ was not included as part of the personalization problem. The final feedback level was selected if and only if the child made more than the five allotted mistakes, and the meta-controller would move on to the next interaction. Therefore, the feedback controller included a total of $G \times F = 10 \times 4 = 40$ (state, action) couples.

The reward function was designed to minimize LoF without pushing the learner to make too many M mistakes (where M was the predefined threshold of maximum mistakes) or penalizing them too heavily for making help requests.

$$R(t) = \left(-1 \cdot \frac{f_t}{m_t + h_t + 1}\right) + MC(t), \tag{3}$$

where

$$MC(t) = \begin{cases} 5, & \text{if } m_t \leq M \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

4. PERSONALIZED SAR INTERVENTION DESIGN

The SAR personalization framework was instantiated in a SAR systems designed for and evaluated in a month-long, in-home SAR intervention in the homes of children with ASD, and approved under USC IRB UP-16-00755. The details of the SAR system, study design, data collection, and outcomes measures are described next.

4.1. System Design

The physical robot was designed to be a near-peer learning assistant, intended to act as the child's companion rather than tutor. Toward that end, it was given a neutral, non-threatening character that presented educational games on a tablet and provided personalized feedback.

4.1.1. Physical Design

To enable long-term in-home deployments, including ensuring the protection of the system's sensitive components, we designed a self-contained and portable system, shown in **Figure 3**, consisting of the robot, and a container that encompassed the robot's power supply, speakers, and tablet. The container was approximately the same width as the robot to minimize the overall system footprint.

The robot platform we designed was modified the Stewart Platform Robot for Interactive Tabletop Engagement (SPRITE) with the Kiwi skin (Short et al., 2017). SPRITE used CoR-Dial, also known as the Co-Robot Dialogue system, the software stack that controls the robot's physical movements and virtual face. The SPRITE consists of a 3D printed base, housing electronic components and threaded rods that support a laser-cut platform with six degrees of freedom. Within the exterior skin, a small display was used to animate the robot's face that included two eyes, eyebrows, and a mouth, all of which were controlled using Facial Action Coding System (FACS) coding in CoR-Dial.

The Kiwi skin and character were designed to appeal to the target user population. Children with ASD are often overwhelmed by sensory input, so Kiwi was designed to be non-threatening and simple in its affective displays. It was also gender-neutral in its appearance, allowing each child to assign the robot's gender if and as desired.

4.1.2. Game Design

The design of the SAR intervention was conceptualized by our multidisciplinary team of researchers, leveraging established game design principles, including iterative prototyping (Adams, 2013). Through these processes, Clabaugh et al. (2018) designed an intervention that balanced the needs of the domain with the limitations of SAR technology. The initial game prototype was presented to a focus group of early childhood educators who served as subject matter experts and provided formative feedback on what was developmentally appropriate for children with ASD diagnoses. This informed the second generation of the game design, which was then piloted in a preschool classroom. Following these pilot studies, further adjustments were made

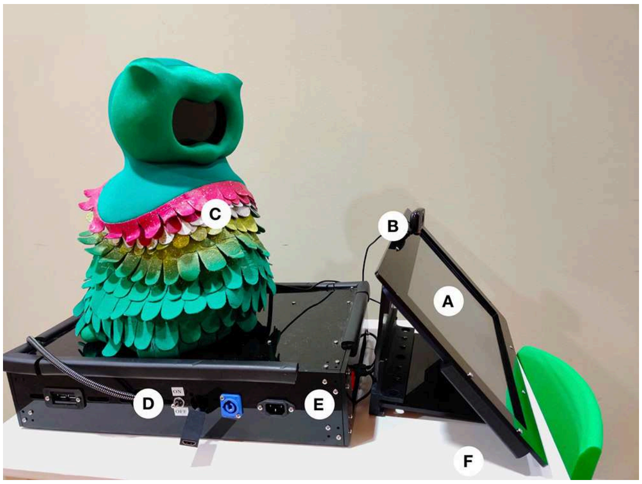


FIGURE 3 | The physical in-home setup included the SPRITE robot with the Kiwi skin (C) mounted on top of the container encasing a computer, power supply, and speakers (E), with an easy-access power switch (D), a camera (B), and touchscreen monitor (A), all located on a standard child-sized table (F).

to accommodate specific needs of children with ASD before the system was iteratively deployed and validated over multiple, long-term, in-home interventions for 17 children with ASD.

The game types within the system were tailored specifically for children with ASD, based on previous case studies, developmentally appropriate practices in working with children with ASD (Copple and Bredekamp, 2009), and standards recognized by the National Association of the Education of Young Children. More specifically, they were developed in concert with developmentally appropriate practices for young children ages 3–8 and informed by contemporary learning theory Omrod et al. (2017). Each game employed a scaffolded approach to gradually increasing difficulty level as the child navigated through successful completion of a particular game level (Sweller et al., 2007). Both the content and difficulty levels were also aligned both to best practices in child development standards of the National Association of Education of Young Children (NAEYC) and the National Common Core Mathematics Standards [for more advanced levels; CCMS (Copple and Bredekamp, 2009)].

The games were also aligned with the Wechsler Individual Achievement Test (WIAT), a developmental-level standard assessment (Wechsler, 2005), used as a pre-post measure of the impact of the game, as described in section 4.3.1. Numerical operation and math reasoning were selected as pre-academic content for the games because they are the early math skills needed by children in preschool and kindergarten. As a result, they are also areas that control for potential social biases found in many early childhood games.

Figure 4 illustrates an example of the different challenge levels of one of the games; game challenge levels were personalized to each child participant as described in section 3.4.1.

4.1.3. Child-Robot Interaction

The Kiwi character described itself as a space explorer and a peer to the child user that continuously needed help from the child in order to return to its home planet. Users were told they could help Kiwi by playing the provided tablet-based games. The games tested a variety of preschool and kindergarten math skills, including addition, counting, and pattern matching. The SAR system offered ten different types of games based on five different levels of challenge (LoC). Child participants were encouraged to play at least one game during each interaction; the games involved the user performing the following on-screen tasks:

1. Pack Moon-Rocks: Drag 1–10 moon-rocks into a box.
2. Select Galaxy: Select the galaxy with more or fewer stars.
3. Select Planet: Select the planet with a particular number.
4. Feed Space Pets: Evenly divide a set of stars between two "alien pets."
5. Pets on a Spaceship: Drag numbered "alien pets" into a spaceship in increasing or decreasing order.
6. Organize Moon-Rocks: Separate and organize moon-rocks based on sprite and number.
7. Organize Space Objects: Separate and organize various space-themed objects based on sprite and number.
8. Pattern Completion: Complete a pattern with the provided space objects.
9. Identify Alien Emotion: Determine the emotion of one or more "alien friends" based on their facial expressions.

The graphics in the game used an age-appropriate comic book design style, with colorful aliens guiding the user through the games. Each game allowed up to five mistakes; every mistake was followed by a verbal hint delivered by Kiwi paired with child-like body movements that signaled whether the user was struggling or excelling. The feedback actions were specific to the game context of early mathematics learning. For example, if the child was presented with the instruction "Put five energy crystals into a box" but used too few crystals, the feedback controller executed one of the following actions:

1. "We need to have a total of five energy crystals inside the box."
2. "Try counting out loud as you drag each crystal one by one."
3. "You have too few energy crystals. Try adding some to the box."
4. "We currently have three energy crystals. So we need two more energy crystals. Can you drag two more crystals into the box?"
5. "Let's try something else."

4.2. Study Participants

Seventeen children with ASD were included in this research, hereafter referred to as P1-P17. Families were recruited through regional centers within the state's Department of Developmental Services and through local school districts. Together, these two recruitment venues provide services for >10,000 children and adults with ASD, with ∼1/3 of the population under the age of ten.

Study recruitment flyers were provided to service coordinators, school district administrators, and family research center coordinators who are employed in regional centers and the schools. Families who were interested in participating in the research contacted our research team and provided written information about their child with ASD. A licensed psychologist

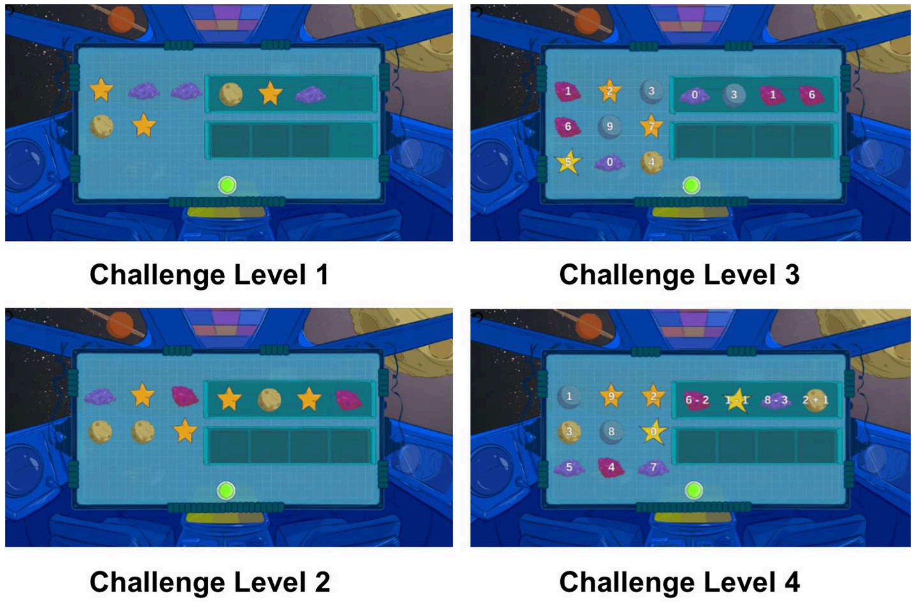


FIGURE 4 | The child-robot interaction was designed around Kiwi as a robot space explorer. The following diagram displays varying challenge levels of the Pack Moon-Rocks game, with more challenging problems combining math reasoning and numerical operation concepts.

on our team reviewed each child's developmental and health information for a match with the study's inclusion criteria:

1. Age between 3 and 8 years old
2. Stable physical, sensory (hearing, vision), and medical health
3. English as a primary language spoken in the family
4. Clinical diagnosis of ASD in mild to moderate ranges as described in the Diagnostic and Statistical Manual of Mental Disorders–Version 5 (Van Bourgondien et al., 1992; Baird et al., 2003; Dover and Le Couteur, 2007; Kanne et al., 2008).

Of the 17 children in the study, 2 were female and 15 male. They were between 3 years, 4 months and 7 years, 8 months of age. Additionally there were 3 sets of sibling pairs (P3 and P4, P5 and P6, and P16 and P17). More information about each participants living situation, education level, and age can be found in the **Supplementary Materials**.

Due to the challenges of ASD and real-world studies, there were some exceptions among the participants. Specifically, there are no personalization data for P1 and P2, as the system was not yet fully developed for those first two deployments. Additionally, P3 did not complete the post-study assessments for personal reasons, but did participate in the study for over a month and provided all other study data. Besides these exceptions, the rest of the participants participated in the entire study.

There is no control condition in this study, as is common in ASD studies, because individuals on the autism spectrum present an extremely broad range of symptoms, symptom combinations, and symptom severities. Consequently, work with ASD participants typically follows a single-case study model rather than the randomized trial model. The single-case study model relies on pre/post-comparisons, as was done in this paper (Lobo et al., 2017). The pre/post-WIAT Interventions in section 4.3.1 serve as a sample baseline to evaluate participant improvement over the course of the study.

All child participants in the study were enrolled in full-time educational and therapeutic interventions that were consistent with the state's educational and developmental services standards and statutes. These services varied based on child needs and family preferences. All child participants had intelligence scores within "normal" limits levels (scores>70) based either on the Leiter International Performance Scale-3 (Roid et al., 2013) or the by the Differential Ability Scales (Elliott, 2012).

The child participants' ASD diagnoses were obtained via clinical best estimate (CBE) by trained psychologists or psychiatrists who had >10 years of experience in diagnosing children with ASD and other developmental disabilities. The tools used to diagnose ASD varied across clinician and referring agency. In each case, multiple measures were used to determine the diagnosis and level of ASD. Common measures used for the ASD diagnoses were the Autism Diagnostic Interview-Revised (Wing et al., 2002; Tadevosyan-Leyfer et al., 2003), Autism Diagnostic Observation Schedule (Lord et al., 2000; Gotham et al., 2008), and Child Autism Rating Scale (Van Bourgondien et al., 1992). All children with ASD diagnoses in the study had diagnoses in the mild to moderate range (Van Bourgondien et al., 1992).

4.3. Procedure

The SAR intervention was deployed in the home of each participating family for at least 30 days. The duration of each deployment was determined by when the minimum number of 20 child-robot interactions was completed; the average duration of deployment was 41 days, with a standard deviation of 5.92 days. On the day of deployment for each family, all system equipment was provided and assembled by the research team; the only requirement from participating families was a power outlet and sufficient space. During system setup, child participants were assessed by an educational psychologist using the measures described in section 4.2. After the system was set up, the research team conducted a system tutorial with the child participant and family.

To capture natural in-home interactions, the SAR system was fully autonomous and could be turned on and off whenever the family desired. The child participants were encouraged but not required to complete five sessions per week. Similarly, during each session, they were encouraged but not required to play each of the 10 games at least once.

4.3.1. Objective Measures

A large corpus of multi-modal data was collected, including video, audio, and performance on the games. The USB camera mounted at the top of the game tablet recorded a front view of the child participant. A second camera recorded the child-robot interaction from a side view. All interactions with the tablet were recorded, including help requests and answers to game questions.

User engagement was annotated by analyzing the camera data. A participant was considered to be engaged when paying full attention to the interaction, immediately responding to the robot's prompts, or seeking further guidance or feedback from others in the room.

Due to numerous technological challenges common in noisy real-world studies, we were able to analyze sufficient video and audio data from seven participants (P5, P7, P9, P11, P12, P16, P17). A primary expert coder annotated whether a participant was engaged or disengaged for those seven participants. To verify the absence of bias, two additional annotators independently annotated 20% of data for each participant; inter-rater reliability was measured using Fleiss' kappa, and a reliability of $k = 0.84$ was achieved between the primary and verifying annotators.

The primary quantitative measure of cognitive skills gained throughout the study were the pre- and post-assessments, inspired by the standardized Wechsler Individual Achievement Test (WIAT II) (Wechsler, 2005) used to assess the academic achievement of children, adolescents, college students, and adults, aged 4–85. The test evaluates a broad range of academics skills using four basic scales: Reading, Math, Writing, and Oral Language. Within those, there are nine subtest scores, including two math subtests, *numerical operations* (NO) and *math reasoning* (MR), which were the most relevant to the SAR intervention content. For young children, NO refers to early math calculations, number discrimination, and related skills; MR refers to concepts of quantity and order, early word problems, patterning, and other skills that require reasoning to solve problems. WIAT II was selected over the WIAT III because the timing of math fluency in version III presents a potential bias for children with ASD diagnoses.

The WIAT II provides raw and composite scores. Standard scores and percentile ranking are computed by comparing an individual assessment to large national samples of typically developing individuals aged 3 to adult (i.e., 2015 US normative sample $N = 2,950$). A standard score of 100-110 is considered an "average achievement score" by national standards. The percentile ranking indicates how an individual compares to the national sample on which the tests were normed. The WIAT-II was used as a pre-post comparison measure to determine achievement gains over the SAR intervention. Procedurally, the pre-assessment was conducted during the first few days of the intervention and the post-assessment was conducted at the end of the intervention for each child.

4.3.2. Subjective Measures

We conducted biweekly interviews with participating families throughout the deployments to evaluate the system in terms of its usefulness and relationship with the participating child, rating responses on a 7-point Likert scale with 1 being least likable and 7 being most likable. Given the variable nature of in-home studies and different degrees of ASD across the participants, the surveys used a single-subject design (Horner et al., 2005) wherein each child served as their own unique baseline. The semi-structured interviews contained similar questions, each tailored for a specific evaluation criterion, as follows.

Based on prior work by Moon and Kim (2001), these were the questions about Kiwi's usefulness:

- Does Kiwi help your child do better on the tasks? Why or why not?
- How could Kiwi be more useful?
- How involved do you have to be while your child is playing with Kiwi?

Based on prior work by Lee et al. (2005) and Rau et al. (2009), these were the questions about the child-robot relationship:

- Do you think Kiwi is your friend?
- Do you think Kiwi listens to you?
- Do you feel like Kiwi knows you?

5. RESULTS

The presented month-long in-home deployments produced a large set of results. Sections 5.1 and 5.2 describe the patterns and quantitative results, respectively, of the hHRL framework instantiation. Section 5.4 discusses how the adaptive system influenced the engagement of the child participants. Section 5.5 reports on how the adaptive SAR system influenced cognitive skills gains across all participants, as measured by the pre-post intervention assessments.

5.1. Personalized Level of Challenge

As illustrated by the learning curve in **Figure 5**, the personalized level of challenge (pLoC) changed over time and varied by participant. Since the goal of the adaptation was to find the optimal LoC for each participant, this learning curve cannot be interpreted in a traditional sense. For instance, if a child was not proficient at math, the learning system may not have been able to reach higher reward values, because the reward is based on both LoC and child performance.

Therefore, other factors must also be considered in interpreting the pLoC results. First, more than 100 episodes or

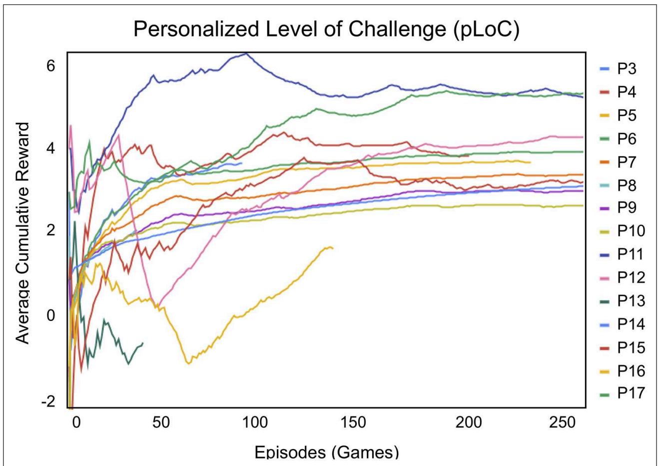


FIGURE 5 | The reinforcement learning reward for the personalized level of challenge (pLoC) ranged between −5 and 5, where 5 indicated that the user was completing games at the highest challenge level. The average cumulative reward of pLoC matched each participant's pre-intervention scores of the WIAT II subtests for numerical operations and math reasoning. Therefore, the pLoC adapted to each participant over the month-long intervention (there is no pLoC data for P1 or P2, as explained in section 4.2).

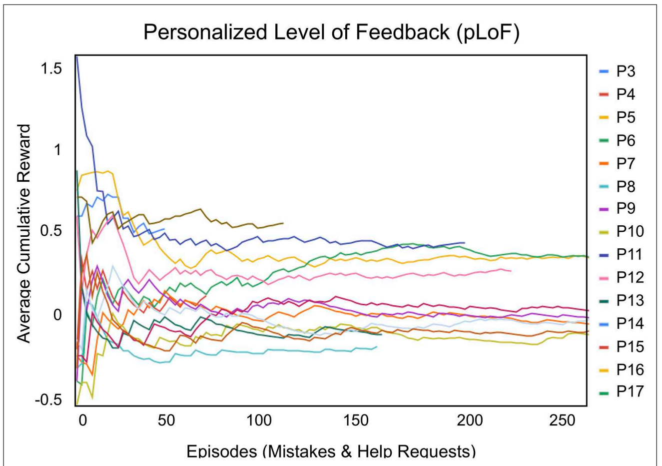


FIGURE 6 | The reinforcement learning reward for personalized level of feedback (pLoF) ranged between 0 and 1, where 1 indicated that a child completed games with the least amount of support or feedback. The average cumulative reward of pLoF converged over 25–50 episodes (i.e., mistakes and help requests) and is correlated with the intervention length and the average number of mistakes made by child participants per game (there is no pLoF data for P1 or P2, as explained in section 4.2).

games were required for pLoC to begin to converge. For example, for participants P3 and P8, the pLoC curve did not have a chance to converge over the few games these participants played. On the other hand, for P6, P11, and P15, the system was able to smoothly adapt given the long interaction periods.

If a child played 10-20 games in a session, consistent with the 13.27 average of the study, then 10 sessions were required before the pLoC began to converge, totalling to ∼132 episodes. This is a reasonable requirement given that the participants completed an average of 14.10 sessions with the robot. Excluding participants P3 and P8, who, as noted above, played significantly fewer games per session, we find an average of 17.57 sessions with the robot.

Consequently, we can conclude that the SAR system was able to adapt and personalize to each child over time. Specifically, the pLoC implementation of the instruction controller did personalize to each child, but required a minimum number of episodes and interaction consistency to do so.

5.2. Personalized Level of Feedback

The learning curve for the personalized level of feedback (pLoF) model, shown in **Figure 6**, adapted the level of feedback to each participant more rapidly than pLoC. Analogous to pLoC, the pLoF learning curve cannot be interpreted in the traditional sense of simply maximizing cumulative reward; it is meant to match each child's need.

Participants with high mistake totals and long interventions usually had the longest feedback curves and, subsequently, allowed the system to adapt to their needs. The pLoF tail was longest for P10, who had the third highest mistake average, balanced with overall intervention length. Although P3 and P8 had the highest mistake averages, they also had the shortest interventions. This can be compared to the pLoF curves for P5 and P6, who had the lowest mistake averages and longest intervention lengths; the cumulative reward is higher and tails are shorter for both P5 and P6 compared to those of P10. Subsequently, P10 stands out as the longest and flattest among the three, demonstrating the value of longer interactions. Overall, the pLoF model successfully adapted to each child participant over time.

5.3. Participant SAR Evaluation

SAR survey results, utilizing a seven-point Likert scale, assessed the average adaptability and usefulness of the system throughout the study.

Participants' responses about the adaptability of the SAR system, seen in **Figure 7**, correlate with the challenges encountered in adapting to the individual needs of each participant. P17 reported the lowest average score for adaptability and usefulness. The result for P17 is likely due to the participant's age, as this was the second oldest and highest performing student in the study, so even the maximum difficulty was too easy for the participant.

The reported scores for usefulness were similar to those of adaptability, as seen in **Figure 7**, since the two measures are related: a system that is more adaptive to a participant is more useful. P1 and P17 were once again outliers with the lowest reported scores for usefulness.

P1 and P7 had personal similarities: they were less than a year apart in age and had parents with the same levels of education (high school). Consequently, one would expect the system to adapt relatively similarly to both participants. Their reported scores for usefulness (5 vs. 3) and adaptability (5 vs. 4) were similar, thus supporting the consistency of the system across participants.

Sibling pairs (P3 and P4, P5 and P6, and P16 and P17) showed discrepancies that can be explained by the fact that the system was

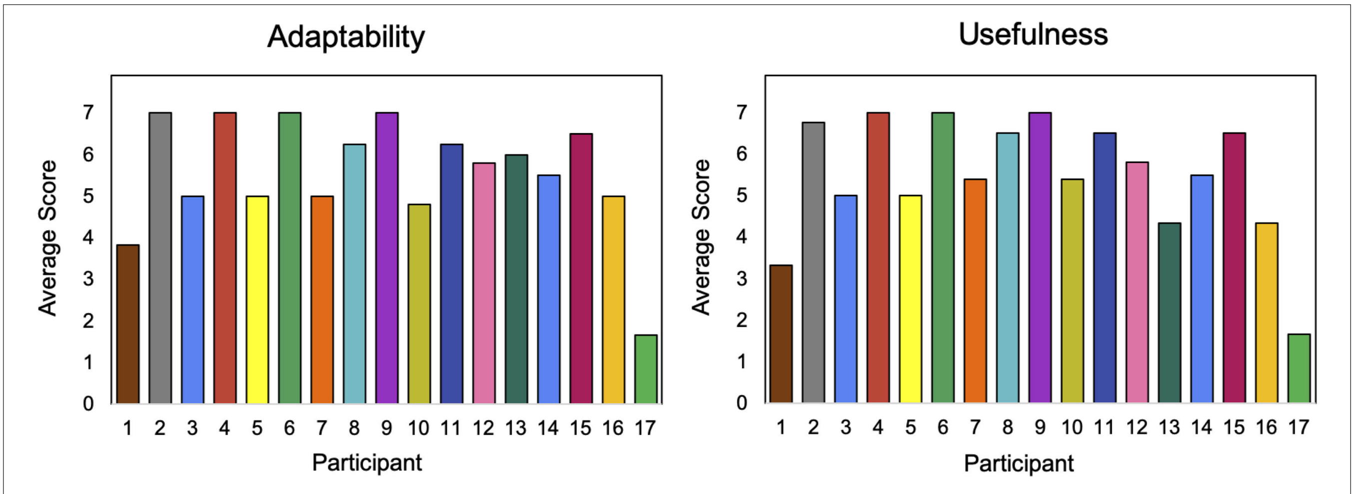


FIGURE 7 | Participant survey results for adaptability **(left)** and usefulness **(right)**.

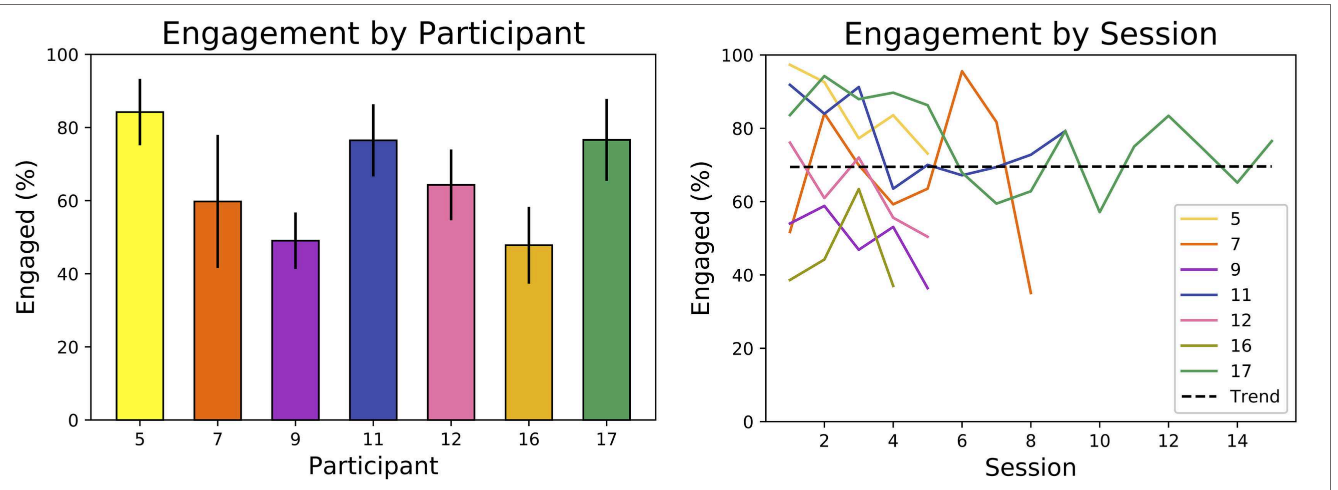


FIGURE 8 | Overall engagement for each analyzed participant, with error bars denoting standard deviation across sessions (left). High variance but no linear trend ($p = 0.99$) in engagement is observed across sessions (right) (engagement was only analyzed for participants with adequate video and audio data).

better suited to the needs of one sibling than the other, likely due to their age. For example, P3 was younger than P4, and therefore was not able to engage with the games as well, resulting in the lower adaptability and usefulness scores. Similarly, P6, the older sibling, reported higher scores for adaptability and usefulness than P5. The higher scores mean that the child liked the robot more and found it more adaptable and useful.

5.4. Effect of Personalization on Engagement

We found that our SAR system elicited and maintained participants' engagement throughout the month-long intervention, an important measure of effectiveness. As mentioned in section 4.3.1, we analyzed the seven participants (P5, P7, P9, P11, P12, P16, P17) with adequate video and audio data to analyze measures of engagement.

5.4.1. Short-Term and Long-Term Engagement

The SAR system maintained reasonable levels of participant engagement during individual sessions and over the month-long intervention. As shown in **Figure 8**, all participants were engaged on average 65% of the intervention. Across sessions, participants had an average engagement range of 32% and standard deviation of 11%. However, there was no statistically significant ($p = 0.99$) increase or decrease in engagement over the study, as determined by a regression t-test and shown by the plotted trend line. In addition, the median duration of continuous engagement over all participants was higher than the median duration of continuous disengagement: 13–5 s on average, respectively.

Furthermore, the robot was able to elicit and maintain user engagement during each game. Engagement was higher shortly after the robot had spoken; participants were engaged about 70% of the time when the robot had spoken in the previous minute, but <50% of the time when the robot had not spoken for over

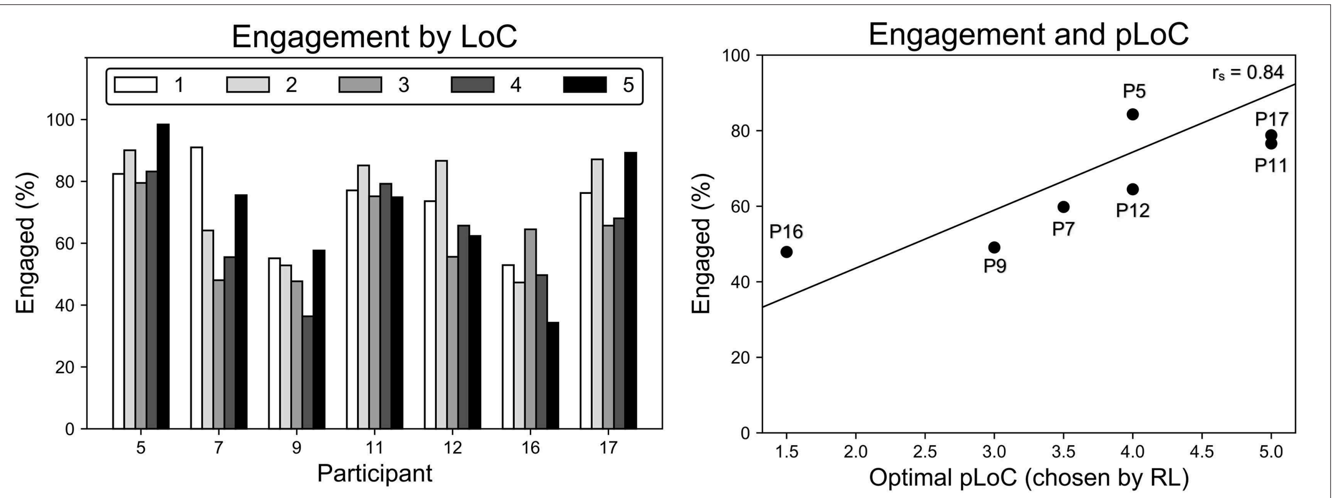


FIGURE 9 | Variance in engagement is higher across participants than across LoC **(left)**. Participants with high optimal LoC were more engaged ($r_s = 0.84, p = 0.018$) **(right)**. Engagement was only analyzed for participants with adequate video and audio data.

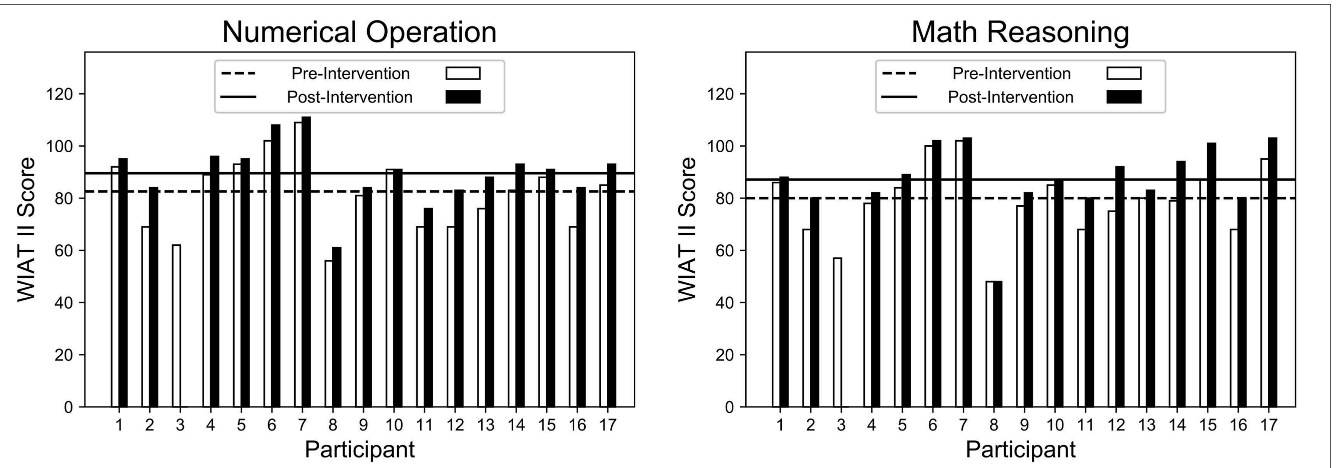


FIGURE 10 | All participants (excluding P3 who did not complete the study) showed significant improvements ($d = 0.54, p < 0.01$) on the WIAT II subtests for numerical operations **(left)**, and math reasoning **(right)** between the pre-intervention and post-intervention assessments.

a minute. Participants also remained engaged after 5 min of starting a game nearly 60% of the time.

5.4.2. Engagement and the Level of Challenge

Engagement varied significantly across participants and their level of challenge (LoC), as shown in **Figure 9**. A two-way analysis of variance (ANOVA) showed ($p < 0.01$) that average engagement for each participant varied significantly and that average engagement under each LoC also varied significantly. The variance across participants accounted for 91% of the total variance, indicating the importance of personalization in SAR.

The personalized level of challenge (pLoC) did not necessarily maximize engagement. As discussed above, pLoC eventually converged to an optimal LoC for each participant. But, as shown in **Figure 9**, participants whose optimal LoC was low were less engaged ($r_s = 0.84, p = 0.018$). We hypothesize that this effect is due to the time required for the learning system to adapt to each user; it took >100 games for the pLoC to begin to converge, and thus participants with a lower LoC were presented with many games of higher challenge level before convergence. This further supports the importance of personalization for increasing engagement, especially with a sufficiently fast convergence rate.

5.5. Impact on Math Learning

Overall, this study observed positive gains in math learning for all participants, excluding P3 who did not complete the study. As seen in **Figure 10**, participants' pre- and post-intervention scores on the WIAT II subtests increased significantly for numerical operations (NO) ($p = 0.002$) and math reasoning (MR) ($p < 0.001$), as determined by a t-test. In addition, both NO and MR scores had a significant effect size of $d = 0.53$ and $d = 0.54$, respectively, as calculated using Cohen's d.

The result reveal that NO and MR both increased even when there was a large discrepancy between the initial assessment of certain participants. For example, P17 scored much higher on MR than on NO on the pre-assessment and even with such different starting points, both NO and MR increased at the post-assessment. On the other hand, P11 started with the same MR and NO scores, and both scores improved after the intervention.

When observing total cognitive gains, it is important to consider developmental factors: the age and subsequent skill level of each participant. Where older students generally had smaller net gains, they started near or above average. On the other hand, younger students started far below average, and thus had much room to improve. P8's pre-intervention scores (MR = 48; NO = 56) were significantly below the national average. Given P8's age (3.75 years), the scores are cautiously computed in terms of what they represent nationally. In another case, P16's pre-intervention scores (MR = 68; NO = 69) were far below the national average. P16 was the youngest participant (3.11 years) and still made significant progress, improving by over 10 points in both categories (MR = 80; NO = 84). On the other hand, P17 was tied second oldest (7.2 years) and only made marginal gains, despite making few mistakes and performing at the highest challenge level.

6. DISCUSSION

The results of the long-term in-home deployment provide several insights for personalization in SAR.

We found that both the personalized levels of challenge (pLoC) and feedback (pLoF) converged for almost all participants. After ∼100 games, the feedback and challenge curves stabilized, showing that the system adapted to an appropriate LoC for each student. Therefore, the long-term nature of the study was important for successful personalization. The participants with the longest episodes in the pLoC were P6, P16, and P11, with 715, 592, and 520 episodes (games played), respectively. In contrast, pLoF interacted most with for P10, P7, and P9, who had 353, 237, and 228 episodes (mistakes and help requests), respectively. The SAR system adapted to the participants in both cases.

Regardless of the difference in sessions, participants who yielded a consistent score by the end of the interaction in the pLoC had similar success with pLoF, and vice versa. This happened for participants who interacted with equal or above average 113.4 and 302.5 episodes for pLoF and pLoC, respectively. On the other hand, P16 illustrated the negative impact of minimal interaction, as both the pLoC and pLoF failed to standardize given only 129 total interactions both on pLoC and pLoF, reaching over 171 episodes below average for pLoC. Within the interaction, the pLoC reward for P16 fluctuated by 2.6 points between the 61st and 129th episode. For reference, the second highest fluctuation in this same interval was 1.02 points by P16, whose system ultimately converged after 592 games.

Overall, the pLoF and pLoC demonstrate the ability to adapt to each user's preferences given their willingness to interact with the robot and provide the system opportunities to learn. The participant surveys support this conclusion and provide the user's perspective on the SAR's ability to adapt.

P6, P7, P9, and P11, who both the pLoC and pLoF had adapted to, reported in their post-interaction interviews an average rating of 6.25, showing a shared appreciation for the system's adaptiveness. The only study participant who believed the system did not adapt was P17, who likely felt this way because of limited success with the feedback model; P17 had only 112 feedback episodes over 298 total games. Aside from this outlier, the survey results supported the effectiveness of pLoC and pLoFs. P9 was an ideal participant, who believed the system adapted and had above average episodes while stabilizing both pLoC and pLoF.

Usefulness questionnaire data provide additional insights into the value of creating an adaptive system. All participants reported very similar scores for usefulness and adaptiveness, implying that the usefulness of the system is related to its adaptiveness. The pre-post assessments supported this finding while providing quantitative data about the learning gains of each participant as a result of SAR personalization.

Participants whose optimal LoC was lower were less engaged, as shown in **Figure 9**. For example, the system converged to the lowest pLoC for P16, who also had the second lowest engagement. This is likely because P17 was presented with games of higher challenge before the system began to converge to an optimal LoC. When also considering that P16 had a below average number of episodes, it is likely the robot failed to adapt quickly ultimately discouraging the participant from interacting further.

The analysis of the objective and subjective outcome measures supports the success of the system as a whole, with all participants improving in math skills over the course of the long-term in-home interaction. Regardless of whether the system was able to adapt to an optimal LoC, all participants demonstrated cognitive gains. The participants gained an average of 7.0 points on numerical operation (NO) and 7.125 points on math reasoning (MR). Although for P16 the system was unable to adapt both in pLoF and pLoC personalization, that participant was still in the top five in both NO and MO gains, with an increase of 15 and 12 points, respectively. This is likely due to the participant's initially low scores that allowed much room for improvement. All participants who had at or above average number of episodes (either in pLoC and pLoF) showed strong positive gains. P8 illustrated the disadvantages of insufficient interaction time, being the participant with the least episodes in both pLoF and pLoC and resulting with below average gains in both NO and MR assessments.

7. CONCLUSION

Socially assistive robotics (SAR) has demonstrated tremendous potential for use in high impact domains, such as personalized learning for special needs populations. This work considered the problem of computational personalization in the context of long-term real-world SAR interventions. At the intersection of HRI and machine learning, computational personalization seeks to autonomously adapt robot interaction to meet the unique needs and preferences of individual users, providing a foundation for personalization.

This work presented a formalized framework for human-robot learning as a hierarchical decision-making problem (hHRL) that decomposes a SAR intervention for tractable computational personalization, and utilized a reinforcement learning approach to personalize the level of challenge and feedback for each user. The approach was instantiated within the interactive games and tested in month-long in-home deployments with children with ASD. The SAR system was able to personalize to the children with ASD who demonstrated cognitive gains, supporting the effectiveness of the approach.

The body of results of the presented study demonstrate that the hHRL framework and its instantiation can engage and adapt to children with diverse needs in math learning over multiple weeks. These findings highlight the tremendous potential of in-home personalized SAR interventions.

DATA AVAILABILITY STATEMENT

The dataset analyzed in this study includes identifiable video and audio data of children with autism spectrum disorders and their families, along with video information and images representing their homes. Consequently, the University IRB prohibits distribution of the dataset to protect the privacy of the research participants.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board (IRB). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

CC developed the earning framework and its instantiation and parts of the system software, oversaw the implementation of the rest, most of the deployments, data annotation, and provided source text for the paper. KM outlined, wrote, and edited the paper while managing various stakeholders throughout the writing process. SJ designed and modeled the engagement process, and also detailed the Background and Engagement Results. RP facilitated the bi-weekly surveys and edited the Results section. DB contributed to the game performance model development, and led the first set of in-home deployments and data collections. ZS designed the graphs used throughout the paper and served as a secondary editor. ED contributed to the design of the Kiwi robot hardware and the overall SAR system for time-extended in-home deployments. RL rated the survey responses and transcribed all survey questions. GR provided domain expertise in autism and early child learning, assessment methods and tools, lead the participant recruitment, and administered the pre- and post-study assessments and interviews. MM was the project lead. she advised all students, coordinated the robot and study designs, oversaw data analysis, and extensively edited the paper.

FUNDING

This research was supported by the National Science Foundation Expedition in Computing Grant NSF IIS-1139148.

ACKNOWLEDGMENTS

The authors thank National Science Foundation Expedition Grant for funding this study along with the valuable contributions of various annotators including Balasubramanian Thiagarajan, Kun Peng, Leena Mathur, and Julianna Keller.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frobt.2019.00110/full#supplementary-material

REFERENCES

Adams, E. (2013). *Fundamentals of Game Design. 3rd Edn.* Pearson, GA: New Riders.

Amabile, T. M. (1993). Motivational synergy: toward new conceptualizations of intrinsic and extrinsic motivation in the workplace. *Hum. Resour. Manage. Rev.* 3, 185–201.

Amershi, S., Cakmak, M., Knox, W. B., and Kulesza, T. (2014). Power to the people: the role of humans in interactive machine learning. *AI Mag.* 35, 105–120. doi: 10.1609/aimag.v35i4.2513

Anderson, J. R. (1985). *Cognitive Psychology and Its Implications*. New York, NY: WH Freeman; Times Books; Henry Holt & Co.

Anderson, J. R., Reder, L. M., and Simon, H. A. (1996). Situated learning and education. *Educ. Res.* 25, 5–11.

Anderson, L. W., Krathwohl, D. R., Airasian, P., Cruikshank, K., Mayer, R., Pintrich, P., et al. (2001). A taxonomy for learning, teaching and assessing: a revision of bloom's taxonomy. New York, NY: Longman Publishing.

Artz, A. F., and Armour-Thomas, E. (1992). Development of a cognitive-metacognitive framework for protocol analysis of mathematical problem solving in small groups. *Cogn. Instr.* 9, 137–175. doi: 10.1207/s1532690xci0902_3

Austin, J. L. (1975). *How To Do Things With Words, Vol. 88*. London: Oxford University Press.

Bainbridge, W. A., Hart, J., Kim, E. S., and Scassellati, B. (2008). "The effect of presence on human-robot interaction," in *The 17th IEEE International Symposium on Robot and Human Interactive Communication, 2008. RO-MAN 2008* (Munich: IEEE), 701–706.

Baird, G., Cass, H., and Slonims, V. (2003). Diagnosis of autism. *BMJ* 327, 488–493. doi: 10.1136/bmj.327.7413.488

Baxter, P., Kennedy, J., Belpaeme, T., Wood, R., Baroni, I., and Nalin, M. (2013). "Emergence of turn-taking in unstructured child-robot social interactions," in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Christchurch: IEEE), 77–78.

Beevers, C. G., and McGeary, J. E. (2012). Therapygenetics: moving towards personalized psychotherapy treatment. *Trends Cogn. Sci.* 16, 11–12. doi: 10.1016/j.tics.2011.11.004

Begum, M., Serna, R. W., and Yanco, H. A. (2016). Are robots ready to deliver autism interventions? A comprehensive review. *Int. J. Soc. Robot.* 8, 157–181. doi: 10.1007/s12369-016-0346-y

Bloom, B. S. (1984). The 2 sigma problem: the search for methods of group instruction as effective as one-to-one tutoring. *Educ. Res.* 13, 4–16.

Bongaarts, J. (2004). Population aging and the rising cost of public pensions. *Popul. Dev. Rev.* 30, 1–23. doi: 10.31899/pgy1.1012

Breazeal, C. (1998). "Early experiments using motivations to regulate human-robot interaction," in *AAAI Fall Symposium on Emotional and Intelligent: The Tangled Knot of Cognition, Technical Report FS-98-03* (Cambridge, MA), 31–36.

Breazeal, C., Kidd, C. D., Thomaz, A. L., Hoffman, G., and Berlin, M. (2005). "Effects of nonverbal communication on efficiency and robustness in human-robot teamwork," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2005 (IROS 2005)* (Cambridge, MA: IEEE), 708–713.

Bruce, A., Nourbakhsh, I., and Simmons, R. (2002). "The role of expressiveness and attention in human-robot interaction," in *Proceedings ICRA'02. IEEE International Conference on Robotics and Automation, 2002, Vol. 4* (Pittsburgh, PA: IEEE), 4138–4142.

Castellano, G., Martinho, C., Paiva, A., Leite, I., and Pereira, A. (2012). "Modelling empathic behaviour in a robotic game companion for children: an ethnographic study in real-world settings," in *International Conference on Human-Robot Interaction (HRI)* (Boston, MA: ACM), 367–374.

Cesuroglu, T., van Ommen, B., Malats, N., Sudbrak, R., Lehrach, H., and Brand, A. (2012). Public health perspective: from personalized medicine to personal health. *Pers. Med.* 9, 115–119. doi: 10.2217/pme.12.16

Chaiklin, S. (2003). The zone of proximal development in vygotsky's analysis of learning and instruction. *Vygotskys Educ. Theory Cult. Context* 1, 39–64. doi: 10.1017/CBO9780511840975.004

Chan, J., and Nejat, G. (2011). "A learning-based control architecture for an assistive robot providing social engagement during cognitively stimulating activities," in *International Conference on Robotics and Automation (ICRA)* (Shanghai: IEEE), 3928–3933.

Chen, Y., Singla, A., Mac Aodha, O., Perona, P., and Yue, Y. (2018). Understanding the role of adaptivity in machine teaching: The case of version space learners. *arXiv:1802.05190.*

Clabaugh, C., Jain, S., Thiagarajan, B., Shi, Z., Mathur, L., Mahajan, K., et al. (2018). "Attentiveness of children with diverse needs to a socially assistive robot in the home," in *2018 International Symposium on Experimental Robotics* (Buenos Aires: University of Southern California).

Clabaugh, C., and Matarić, M. (2019). Escaping oz: Autonomy in socially assistive robotics. *Annu. Rev. Control Robot. Auton. Syst.* 2, 33–61. doi: 10.1146/annurev-control-060117-104911

Clabaugh, C., Ragusa, G., Sha, F., and Matarić, M. (2015). "Designing a socially assistive robot for personalized number concepts learning in preschool children," in *2015 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)* (Seattle, WA: IEEE), 314–319.

Cohn, D. A., Ghahramani, Z., and Jordan, M. I. (1996). Active learning with statistical models. *J. Artif. Intell. Res.* 4, 129–145.

Conn, K., Liu, C., Sarkar, N., Stone, W., and Warren, Z. (2008). "Affect-sensitive assistive intervention technologies for children with autism: an individual-specific approach," in *International Conference on Robot and Human Interactive Communication (RO-MAN)* (Munich: IEEE), 442–447.

Copple, C., and Bredekamp, S. (2009). *Developmentally Appropriate Practice in Early Childhood Programs: Serving Children From Birth Through Age 8*. Washington, DC: National Association for the Education of Young Children, 3.

Danesi, M. (2016). *The Semiotics of Emoji: The Rise of Visual Language in the Age of the Internet*. London: Bloomsbury Publishing.

Darley, F. L., Aronson, A. E., and Brown, J. R. (1969). Differential diagnostic patterns of dysarthria. *J. Speech Lang. Hear. Res.* 12, 246–269.

Dautenhahn, K. (2007). Socially intelligent robots: dimensions of human–robot interaction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 679–704. doi: 10.1098/rstb.2006.2004

Deng, E., Mutlu, B., and Matarić, M. J. (2019). Embodiment in socially interactive robots. *Found. Trends Robot.* 7, 251–356. doi: 10.1561/2300000056

Desmarais, M. C., and Baker, R. S. (2012). A review of recent advances in learner and skill modeling in intelligent learning environments. *User Model. User Adapt. Interact.* 22, 9–38. doi: 10.1007/s11257-011-9106-8

DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., et al. (2014). "Simsensei kiosk: a virtual human interviewer for healthcare decision support," in *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems* (Paris: International Foundation for Autonomous Agents and Multiagent Systems), 1061–1068.

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

Dover, C. J., and Le Couteur, A. (2007). How to diagnose autism. *Arch. Dis. Child.* 92, 540–545. doi: 10.1136/adc.2005.086280

Dragan, A. D., Lee, K. C., and Srinivasa, S. S. (2013). "Legibility and predictability of robot motion," in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Budapest: IEEE), 301–308.

Dudley, J. J., and Kristensson, P. O. (2018). A review of user interface design for interactive machine learning. *ACM Trans. Interact. Intell. Syst.* 8:8. doi: 10.1145/3185517

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

Durlak, J. (2011). The impact of enhancing students' social and emotional learning: a meta-analysis of school-based universal interventions: social and emotional learning. *Child Dev.* 82, 405–432. doi: 10.1111/j.1467-8624.2010.01564.x

Elliott, C. D. (2012). "The differential ability scales," in *Contemporary Intellectual Assessment: Theories, Tests, and Issues. 2nd Edn.* Chapter 13, eds D. P. Flanagan, and P. L. Harrison (New York, NY: Guilford Press), 336–356.

Fasola, J., and Matarić, M. J. (2012). Using socially assistive human–robot interaction to motivate physical exercise for older adults. *Proc. IEEE* 100, 2512–2526. doi: 10.1109/JPROC.2012.2200539

Feil-Seifer, D., and Matarić, M. (2012). "A simon-says robot providing autonomous imitation feedback using graded cueing," in *Poster Paper in International Meeting for Autism Research (IMFAR)* (Toronto, ON).

Feil-Seifer, D., and Matarić, M. J. (2005). "Defining socially assistive robotics," in *9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005* (Chicago, IL: IEEE), 465–468.

Feil-Seifer, D., and Matarić, M. J. (2011). Socially assistive robotics. *IEEE Robot. Autom. Mag.* 18, 24–31. doi: 10.1109/MRA.2010.940150

Finn, B., and Metcalfe, J. (2010). Scaffolding feedback to maximize long-term error correction. *Mem. Cogn.* 38, 951–961. doi: 10.3758/MC.38.7.951

Fischer, G. (2001). User modeling in human–computer interaction. *User Model. User Adapt. Interact.* 11, 65–86. doi: 10.1023/A:1011145532042

Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2003). A survey of socially interactive robots. *Robot. Auton. Syst.* 42, 143–166. doi: 10.1016/S0921-8890(02)00372-X

Forceville, C., and Urios-Aparisi, E. (2009). *Multimodal Metaphor, Vol. 11*. Berlin: Walter de Gruyter.

Fridin, M., and Belokopytov, M. (2014). Embodied robot versus virtual agent: involvement of preschool children in motor task performance. *Int. J. Hum. Comput. Interact.* 30, 459–469. doi: 10.1080/10447318.2014.888500

Gallese, V., and Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends Cogn. Sci.* 2, 493–501.

Gazzola, V., Rizzolatti, G., Wicker, B., and Keysers, C. (2007). The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. *Neuroimage* 35, 1674–1684. doi: 10.1016/j.neuroimage.2007.02.003

Goldin-Meadow, S., and Beilock, S. L. (2010). Action's influence on thought: the case of gesture. *Perspect. Psychol. Sci.* 5, 664–674. doi: 10.1177/1745691610388764

Gordon, G., and Breazeal, C. (2015). "Bayesian active learning-based robot tutor for children's word-reading skills," in *AAAI* (Portland, OR), 1343–1349.

Gordon, G., Spaulding, S., Westlund, J. K., Lee, J. J., Plummer, L., Martinez, M., et al. (2016). "Affective personalization of a social robot tutor for children's second language skills," in *AAAI* (Phoenix, AR), 3951–3957.

Gotham, K., Risi, S., Dawson, G., Tager-Flusberg, H., Joseph, R., Carter, A., et al. (2008). A replication of the autism diagnostic observation schedule

(ADOS) revised algorithms. *J. Am. Acad. Child Adolesc. Psychiatry* 47, 642–651. doi: 10.1097/CHI.0b013e31816bffb7

Greczek, J., Atrash, A., and Matarić, M. (2013). "A computational model of graded cueing: robots encouraging behavior change," in *International Conference on Human-Computer Interaction* (Las Vegas, NV: Springer), 582–586.

Greczek, J., Kaszubski, E., Atrash, A., and Matarić, M. (2014). "Graded cueing feedback in robot-mediated imitation practice for children with autism spectrum disorders," in *2014 RO-MAN: The 23rd IEEE International Symposium onRobot and Human Interactive Communication* (Edinburgh: IEEE), 561–566.

Guadagnoli, M. A., and Lee, T. D. (2004). Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning. *J. Motor Behav.* 36, 212–224. doi: 10.3200/JMBR.36.2.212-224

Guadalupe, F. (2016). Social skills training for autistic children: a comparison study between inclusion and mainstreaming education. *Acad. J. Pediatr. Neonatol.* 1:555. doi: 10.19080/AJPN.2016.01.555564

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

Horn, R. E. (1998). *Visual Language*. Washington, DC: MacroVu Inc.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., and Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Except. Children* 71, 165–179. doi: 10.1177/001440290507100203

Kanne, S. M., Randolph, J. K., and Farmer, J. E. (2008). Diagnostic and assessment findings: a bridge to academic planning for children with autism spectrum disorders. *Neuropsychol. Rev.* 18, 367–384. doi: 10.1007/s11065-008-9072-z

Kasari, C., Rotheram-Fuller, E., Locke, J., and Gulsrud, A. (2012). Making the connection: randomized controlled trial of social skills at school for children with autism spectrum disorders. *J. Child Psychol. Psychiatry* 53, 431–439. doi: 10.1111/j.1469-7610.2011.02493.x

Khamassi, M., V. G. T. T., and Tzafestas, C. (2018). Robot fast adaptation to changes in human engagement during simulated dynamic social interaction with active exploration in parameterized reinforcement learning. *IEEE Trans. Cogn. Dev. Syst.* 10, 881–893. doi: 10.1109/TCDS.2018.2843122

Kidd, C. D. (2008). *Designing for long-term human-robot interaction and application to weight loss* (Ph.D. thesis), AAI0819995, Cambridge, MA, United States.

Kidd, C. D., and Breazeal, C. (2008). "Robots at home: understanding long-term human-robot interaction," in *International Conference on Intelligent Robots and Systems (IROS)* (Nice: IEEE), 3230–3235.

Kim, T., and Hinds, P. (2006). "Who should I blame? Effects of autonomy and transparency on attributions in human-robot interaction," in *The 15th IEEE International Symposium on Robot and Human Interactive Communication, 2006. ROMAN 2006* (Munich: IEEE), 80–85.

Kohn, R., Saxena, S., Levav, I., and Saraceno, B. (2004). The treatment gap in mental health care. *Bull. World Health Organ.* 82, 858–866. doi: 10.1590/S0042-96862004001100011

Kulik, J. A., and Kulik, C.-L. C. (1988). Timing of feedback and verbal learning. *Rev. Educ. Res.* 58, 79–97.

Kulkarni, T. D., Narasimhan, K., Saeedi, A., and Tenenbaum, J. (2016). "Hierarchical deep reinforcement learning: integrating temporal abstraction and intrinsic motivation," in *Advances in Neural Information Processing Systems* (Barcelona), 3675–3683.

Lavelle, T. A., Weinstein, M. C., Newhouse, J. P., Munir, K., Kuhlthau, K. A., and Prosser, L. A. (2014). Economic burden of childhood autism spectrum disorders. *Pediatrics* 133, e520–e529. doi: 10.1542/peds.2013-0763

Lee, K. M., Jung, Y., Kim, J., and Kim, S. R. (2006). Are physically embodied social agents better than disembodied social agents? The effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction. *Int. J. Hum. Comput. Stud.* 64, 962–973. doi: 10.1016/j.ijhcs.2006.05.002

Lee, K. M., Park, N., and Song, H. (2005). Can a robot be perceived as a developing creature? *Hum. Commun. Res.* 31, 538–563. doi: 10.1111/j.1468-2958.2005.tb00882.x

Leite, I., McCoy, M., Ullman, D., Salomons, N., and Scassellati, B. (2015). "Comparing models of disengagement in individual and group interactions," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (Portland, OR: ACM), 99–105.

Leite, I., Pereira, A., Mascarenhas, S., Martinho, C., Prada, R., and Paiva, A. (2013). The influence of empathy in human–robot relations. *Int. J. Hum. Comput. Stud.* 71, 250–260. doi: 10.1016/j.ijhcs.2012.09.005

Leyzberg, D., Spaulding, S., and Scassellati, B. (2014b). "Personalizing robot tutors to individuals' learning differences," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY: ACM), 423–430.

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 Cognitive Science Society, Vol. 34* (Sapporo).

Littman, M. L. (1994). "Markov games as a framework for multi-agent reinforcement learning," in *Machine Learning Proceedings 1994* (New Brunswick, NJ: Elsevier), 157–163.

Lobo, M., Moeyaert, M., Baraldi, C. A., and Babik, I. (2017). Single-case design, analysis, and quality assessment for intervention research. *J. Neurol. Phys. Ther.* 31:187. doi: 10.1097/NPT.0000000000000187

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., et al. (2000). The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *J. Autism Dev. Disord.* 30, 205–223. doi: 10.1023/A:1005592401947

Matarić, M. J. (2017). Socially assistive robotics: human augmentation versus automation. *Sci. Robot.* 2:eaam5410. doi: 10.1126/scirobotics.aam5410

Matarić, M. J., and Scassellati, B. (2016). *Socially Assistive Robotics*, Chapter 73. Cham: Springer International Publishing, 1973–1994.

McLellan, H. (1996). *Situated Learning Perspectives*. Englewood, NJ: Educational Technology.

Mehrabian, A. (2017). *Nonverbal Communication*. New York, NY: Routledge.

Moon, J.-W., and Kim, Y.-G. (2001). Extending the tam for a world-wide-web context. *Inform. Manage.* 38, 217–230. doi: 10.1016/S0378-7206(00)00061-6

Murray, T. (1999). Authoring intelligent tutoring systems: an analysis of the state of the art. *Int. J. Artif. Intell. Educ.* 10, 98–129.

Niedenthal, P. M. (2007). Embodying emotion. *science* 316, 1002–1005. doi: 10.1126/science.1136930

Nikolaidis, S., Kuznetsov, A., Hsu, D., and Srinivasa, S. (2016). "Formalizing human-robot mutual adaptation: a bounded memory model," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction* (New York, NY: IEEE Press), 75–82.

Omrod, J., Anderman, E., and L.H., A. (2017). *Educational Psychology, Developing Learners* (Upper Saddle River, NJ: Merrill Prentice Hall), 4.

O'Rourke, E., Haimovitz, K., Ballweber, C., Dweck, C., and Popović, Z. (2014). "Brain points: a growth mindset incentive structure boosts persistence in an educational game," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, ON: ACM), 3339–3348.

Ospina, M. B., Seida, J. K., Clark, B., Karkhaneh, M., Hartling, L., Tjosvold, L., et al. (2008). Behavioural and developmental interventions for autism spectrum disorder: a clinical systematic review. *PLoS ONE* 3:e3755. doi: 10.1371/journal.pone.0003755

Ovando, M. N. (1994). Constructive feedback: a key to successful teaching and learning. *Int. J. Educ. Manage.* 8, 19–22.

Park, H. W., Rosenberg-Kima, R., Rosenberg, M., Gordon, G., and Breazeal, C. (2017). "Growing growth mindset with a social robot peer," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (Vienna: ACM), 137–145.

Ramachandran, A., Huang, C.-M., Gartland, E., Scassellati, B., and Scas, B. (2018). "Thinking aloud with a tutoring robot to enhance learning," in *International Conference on Human-Robot Interaction (HRI)* (Chicago, IL: ACM), 59–68.

Rau, P. P., Li, Y., and Li, D. (2009). Effects of communication style and culture on ability to accept recommendations from robots. *Comput. Hum. Behav.* 25, 587–595. doi: 10.1016/j.chb.2008.12.025

Reardon, C., Zhang, H., Wright, R., and Parker, L. E. (2015). "Response prompting for intelligent robot instruction of students with intellectual disabilities," in *International Conference on Robot and Human Interactive Communication (RO-MAN)* (Kobe), 784–790.

Riek, L. D. (2012). Wizard of Oz studies in HRI: a systematic review and new reporting guidelines. *J. Hum. Robot Interact.* 1, 119–136. doi: 10.5898/JHRI.1.1.Riek

Robins, B., Dautenhahn, K., Te Boekhorst, R., 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

Roid, G. H., Miller, L. J., Pomplun, M., and Koch, C. (2013). *Leiter International Performance Scale (Leiter-3)*. Los Angeles, CA: Western Psychological Services.

Roll, I., Aleven, V., McLaren, B. M., and Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learn. Instr.* 21, 267–280. doi: 10.1016/j.learninstruc.2010.07.004

Ros, R., Nalin, M., Wood, R., Baxter, P., Looije, R., Demiris, Y., et al. (2011). "Child-robot interaction in the wild: advice to the aspiring experimenter," in *Proceedings of the 13th International Conference on Multimodal Interfaces* (Alicante: ACM), 335–342.

Rudovic, O., Lee, J., Dai, M., Schuller, B., and Picard, R. W. (2018). Personalized machine learning for robot perception of affect and engagement in autism therapy. *Sci. Robot.* 3:eaao6760. doi: 10.1126/scirobotics.aao6760

Sabanovic, S., Michalowski, M. P., and Simmons, R. (2006). "Robots in the wild: observing human-robot social interaction outside the lab," in *9th IEEE International Workshop on Advanced Motion Control, 2006* (Istanbul: IEEE), 596–601.

Sansone, C., and Harackiewicz, J. M. (2000). *Intrinsic and Extrinsic Motivation: The Search for Optimal Motivation and Performance*. Woodbine, NJ: Elsevier.

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

Scassellati, B., Boccanfuso, L., Huang, C.-M., Mademtzi, M., Qin, M., Salomons, N., et al. (2018). Improving social skills in children with ASD using a long-term, in-home social robot. *Sci. Robot.* 3:eaat7544. doi: 10.1126/scirobotics.aat7544

Schiller, E. (1996). Educating children with attention deficit disorder. *Our Children* 22, 32–33.

Schodde, T., Bergmann, K., and Kopp, S. (2017). "Adaptive robot language tutoring based on Bayesian knowledge tracing and predictive decision-making," in *International Conference on Human-Robot Interaction (HRI)* (Vienna: ACM), 128–136.

Searle, J. R. (1976). A classification of illocutionary acts. *Lang. Soc.* 5, 1–23.

Searle, J. R., and Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language, Vol. 626*. Cambridge, UK: Cambridge University Press.

Short, Y. F. E., Short, D., and Matarić, M. J. (2017). *Sprite: Stewart Platform Robot for Interactive Tabletop Engagement*. Department of Computer Science, University of Southern California, Tech Report.

Stokoe, W. C., Casterline, D. C., and Croneberg, C. G. (1976). *A Dictionary of American Sign Language on Linguistic Principles*. Washington, DC: Linstok Press.

Swan, M. (2012). Health 2050: the realization of personalized medicine through crowdsourcing, the quantified self, and the participatory biocitizen. *J. Pers. Med.* 2, 93–118. doi: 10.3390/jpm2030093

Sweller, J., Kirschner, P., and Clark, R. (2007). Why minimally guided teaching techniques do not work: a reply to commentaries. *Educ. Psychol.* 42, 115–121. doi: 10.1080/00461520701263426

Tadevosyan-Leyfer, O., Dowd, M., Mankoski, R., Winklosky, B., Putnam, S., McGrath, L., et al. (2003). A principal components analysis of the autism diagnostic interview-revised. *J. Am. Acad. Child Adolesc. Psychiatry* 42, 864–872. doi: 10.1097/01.CHI.0000046870.56865.90

Toh, L. P. E., Causo, A., Tzuo, P.-W., Chen, I., Yeo, S. H. (2016). A review on the use of robots in education and young children. *J. Educ. Technol. Soc.* 19:148.

Vallerand, R. J., Pelletier, L. G., Blais, M. R., Briere, N. M., Senecal, C., and Vallieres, E. F. (1992). The academic motivation scale: a measure of intrinsic, extrinsic, and amotivation in education. *Educ. Psychol. Meas.* 52, 1003–1017.

Van Bourgondien, M. E., Marcus, L. M., and Schopler, E. (1992). Comparison of dsm-iii-r and childhood autism rating scale diagnoses of autism. *J. Autism Dev. Disord.* 22, 493–506.

van De Sande, B. (2013). Properties of the bayesian knowledge tracing model. *J. Educ. Data Mining* 5:1.

Velentzas, G., Tsitsimis, T., Rañó, I., Tzafestas, C., and Khamassi, M. (2018). Adaptive reinforcement learning with active state-specific exploration for engagement maximization during simulated child-robot interaction. *Paladyn J. Behav. Robot.* 9, 235–253. doi: 10.1515/pjbr-2018-0016

Vygotsky, L. (1978). Interaction between learning and development. *Read. Dev. Children* 23, 34–41.

Wainer, J., Feil-Seifer, D. J., Shell, D. A., and Matarić, M. J. (2006). "The role of physical embodiment in human-robot interaction," in *The 15th IEEE International Symposium on Robot and Human Interactive Communication, 2006. ROMAN 2006* (Hatfield: IEEE), 117–122.

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

Watkins, C. J., and Dayan, P. (1992). Q-learning. *Mach. Learn.* 8, 279–292.

Wechsler, D. (2005). *Wechsler Individual Achievement Test. 2nd Edn. (WIAT II)*. London: The Psychological Corporation.

Wenger, E. (2014). *Artificial Intelligence and Tutoring Systems: Computational and Cognitive Approaches to the Communication of Knowledge*. San Francisco, CA: Morgan Kaufmann.

White, S. W., Keonig, K., and Scahill, L. (2007). Social skills development in children with autism spectrum disorders: a review of the intervention research. *J. Autism Dev. Disord.* 37, 1858–1868. doi: 10.1007/s10803-006-0320-x

Wing, L., Leekam, S. R., Libby, S. J., Gould, J., and Larcombe, M. (2002). The diagnostic interview for social and communication disorders: background, inter-rater reliability and clinical use. *J. Child Psychol. Psychiatry* 43, 307–325. doi: 10.1111/1469-7610.00023

Yagoda, R. E., and Gillan, D. J. (2012). You want me to trust a robot? The development of a human–robot interaction trust scale. *Int. J. Soc. Robot.* 4, 235–248. doi: 10.1007/s12369-012-0144-0

Zins, J. E., Weissberg, R. P., Wang, M. C., and Walberg, H. (2004). *Building Academic Success on Social and Emotional Learning: What Does the Research Say?* New York, NY: Teachers College Press.

Conflict of Interest: MM was co-founder of Embodied, Inc., but was no longer involved with the company. CC was a full-time employee of Embodied, Inc., but was not involved with the company while the reported work was done.

The remaining 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.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 15 November 2019
doi: 10.3389/frobt.2019.00118

Socio-Cognitive Engineering of a Robotic Partner for Child's Diabetes Self-Management

Mark A. Neerincx[1,2]*, Willeke van Vught[1], Olivier Blanson Henkemans[1], Elettra Oleari[3], Joost Broekens[2], Rifca Peters[2], Frank Kaptein[2], Yiannis Demiris[4], Bernd Kiefer[5], Diego Fumagalli[6] and Bert Bierman[7]

[1] TNO Perceptual and Cognitive Systems, Soesterberg, Netherlands, [2] Department of Intelligent Systems, Delft University of Technology, Interactive Intelligence, Delft, Netherlands, [3] IRCSS Ospedale San Raffaele, Center for Advanced Technology in Health and Wellbeing, Milan, Italy, [4] Department of Electrical and Electronic Engineering, Imperial College, London, United Kingdom, [5] DFKI, Saarbrucken, Germany, [6] Mixel, Milan, Italy, [7] Produxi, Zeist, Netherlands

OPEN ACCESS

Edited by:
Séverin Lemaignan, Bristol Robotics Laboratory, United Kingdom

Reviewed by:
Nicole Salomons, Yale University, United States
Soheil Keshmiri, Advanced Telecommunications Research Institute International (ATR), Japan

****Correspondence:***
Mark A. Neerincx
mark.neerincx@tno.nl

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *15 June 2019*
Accepted: *28 October 2019*
Published: *15 November 2019*

Citation:
Neerincx MA, van Vught W, Blanson Henkemans O, Oleari E, Broekens J, Peters R, Kaptein F, Demiris Y, Kiefer B, Fumagalli D and Bierman B (2019) Socio-Cognitive Engineering of a Robotic Partner for Child's Diabetes Self-Management. Front. Robot. AI 6:118. doi: 10.3389/frobt.2019.00118

Social or humanoid robots do hardly show up in "the wild," aiming at pervasive and enduring human benefits such as child health. This paper presents a socio-cognitive engineering (SCE) methodology that guides the ongoing research & development for an evolving, longer-lasting human-robot partnership in practice. The SCE methodology has been applied in a large European project to develop a robotic partner that supports the daily diabetes management processes of children, aged between 7 and 14 years (i.e., Personal Assistant for a healthy Lifestyle, PAL). Four partnership functions were identified and worked out (joint objectives, agreements, experience sharing, and feedback & explanation) together with a common knowledge-base and interaction design for child's prolonged disease self-management. In an iterative refinement process of three cycles, these functions, knowledge base and interactions were built, integrated, tested, refined, and extended so that the PAL robot could more and more act as an effective partner for diabetes management. The SCE methodology helped to integrate into the human-agent/robot system: (a) theories, models, and methods from different scientific disciplines, (b) technologies from different fields, (c) varying diabetes management practices, and (d) last but not least, the diverse individual and context-dependent needs of the patients and caregivers. The resulting robotic partner proved to support the children on the three basic needs of the Self-Determination Theory: autonomy, competence, and relatedness. This paper presents the R&D methodology and the human-robot partnership framework for prolonged "blended" care of children with a chronic disease (children could use it up to 6 months; the robot in the hospitals and diabetes camps, and its avatar at home). It represents a new type of human-agent/robot systems with an evolving collective intelligence. The underlying ontology and design rationale can be used as foundation for further developments of long-duration human-robot partnerships "in the wild."

Keywords: child-robot interaction, conversational agent, human-robot partnership, socio-cognitive engineering, diabetes management, personal health, pervasive lifestyle support

1. INTRODUCTION

Despite substantial progress in AI, robotics, conversational agents, and related technologies (Klopfenstein et al., 2017; Wu et al., 2017; Anjomshoae et al., 2019; Montenegro et al., 2019), social or humanoid robots do hardly show up in sound long-term field studies for pervasive human benefits such as child health (Moerman et al., 2018; Dawe et al., 2019; Robinson et al., 2019). The studies with a prolonged deployment and long-term behavior support ambition have been more exploratory, covering "only" a few robot functions and interactions in a time span of a couple of weeks, for example to explore child-robot relationship development (Looije et al., 2016; Westlund et al., 2018). To make this ambition reality within the foreseeable future, the research & development approach has to change substantially: We have to take an integrative socio-cognitive approach in which robots are researched and developed as part of a human-robot collective that has collaborative intelligence (Epstein, 2015; Johnson and Vera, 2019; Rahwan et al., 2019).

This paper presents such an approach, the socio-cognitive engineering (SCE) methodology that aims at such a developing collective: The building of human-robot partnerships for prolonged performance and well-being. In an extensive case study, the European Personal Assistant for a healthy Lifestyle (PAL) project, this methodology has been applied to develop a robotic partner and human-robot activities that support the daily diabetes management processes of children, aged between 7 and 14 years (i.e., supporting a healthy lifestyle).

Type 1 Diabetes Mellitus (T1DM) is one of the main chronic diseases in childhood with severe consequences for physical and mental well-being. The disease prevalence is rising substantially, doubling every 20 years. T1DM is often diagnosed in early or middle childhood (age between 1 and 11 years) based on symptoms of high or low blood glucose (i.e., hyper- or hypoglycemia) (Betts et al., 1996; Boyer and Paharia, 2008; Jin et al., 2017). Symptoms of a hyper can be headaches, fatigue, thirst, and nausea, while a hypo can start with tremors, sweating and palpitations, and eventually can continue in confusion, impaired thinking, and even seizures. The long-term health consequences of T1DM can be serious, damaging the eyes (retinopathy), peripheral nerves (neuropathy), or kidney (nephropathy) (Centers for Disease Control and Prevention, 2011). Managing T1DM requires strict lifestyle adjustments, which proves to be complex and demanding (Iannotti et al., 2006). Daily management behaviors are, for example, monitoring blood glucose (at least 4 times a day), counting carbohydrates before every meal or snack, anticipating physical exercises, and calculating and administering insulin (Boyer and Paharia, 2008). When children enter puberty, the management challenges are increasing: Bodily changes (e.g., hormones) bring about new dynamics in the blood glucose regulation processes, socio-emotional changes bring about different (possibly negative) appraisals, and autonomy development can bring about resistance to parents and caregivers advises. Unfortunately, most children are not, in advance, well-prepared or -trained to deal with these challenges, as the parents can take care of them well. The result is a decrease in glycemic control and regimen adherence when children enter puberty (Ellis et al., 2007; Pai and Ostendorf, 2011).

We started the PAL project to develop a social robot that supports the child in learning to correctly manage T1DM and, this way, prevents serious consequences to appear at the age of puberty. The envisioned robot acts as partner in a (small) diabetes management team, primary for the child (as a "pal"), but also for the Health Care Professional (HCP) and parent (as a "mediator," e.g., for responsibility transfer from parent to child). It is a conversational agent that is integrated into a distributed behavior change support system, embodied as a humanoid robot in hospitals and diabetes camps, and as an avatar on a tablet at home. Via a mobile timeline and dashboards, the diabetes management activities, information processes and outcomes are visible, accessible, and manageable at all locations. The collective Human-PAL intelligence is evolving over time based on (1) the incremental additions and refinements of robot capabilities in the successive development cycles and (2) the intrinsic learning capabilities of the humans and robots (e.g., based on experiences and feedback).

This paper provides an overview of the PAL research & development activities and outcomes, focusing on three general research questions. The first question is: "How to develop human-agent partnerships for long-term lifestyle support?" The second question concerns the design outcome: "How can a robotic partner support the daily diabetes management of children over a longer period?" The third evaluative question is: "Does this partnership improve child's diabetes-control and well-being?" Section 2 argues that the SCE-methodology provides an answer to the first question, and provides an overview of this methodology. Section 3, 4, and 5 describe the application and results for each SCE-component: the foundation, specification and evaluation of the PAL system. Taken together, they present the evolving knowledge base and partnership behaviors of the human-robot collective to be applied and further developed in practice. Section 6 contains the general discussion and conclusions.

2. SOCIO-COGNITIVE ENGINEERING

In the eighties, cognitive (system) engineering was proposed to integrate social sciences, like cognitive psychology, into the design of human-machine systems or so-called *joint* cognitive systems (Norman, 1986; Woods and Roth, 1988; Rasmussen et al., 1994; Hollnagel and Woods, 2005). Subsequently, this approach was refined to facilitate re-usability and theory building (generalization) by the construction of a design rationale that explicates the contextual dependencies, calling the methodology *situated* cognitive engineering (SCE) (Neerincx and Lindenberg, 2008; Neerincx, 2011). This rationale describes the design solution with its theoretical and empirical foundation in a coherent and concise format and structure. Core is the specification of claims ("hypotheses") on the effects of machine (e.g., robot) functions in specific use cases, and the development of design patterns for the corresponding machine behaviors (Neerincx et al., 2016b; Looije et al., 2017). To better address the technological progress on Artificial Intelligence (AI), Robotics, Conversational Agents and Connectivity, with their capabilities to transform social processes and human-technology relationships, SCE toke a more principled focus on human-agent/robot teamwork and patterns, and got its corresponding new first syllable: Socio-Cognitive Engineering (Sharples et al., 2002; Bradshaw et al., 2012; Van Diggelen et al., 2018). In our view, SCE can contribute to the research and development of the robotic systems by supporting the acquisition, modeling, sharing, and extension of the evolving social intelligence.

Long-term interaction "in the wild" is an important research and development challenge for socially assistive and educational robots (SAR). For example, Coninx et al. (2016) stated that, to pursue learning and therapeutic goals through child-robot interaction, it is important to ensure the child remains engaged in the relationship and that the child experiences progress in achieving educational goals. To establish such engagement and to accommodate individual differences, they developed an adaptive social robot with which children can perform various activities. This robot was evaluated in three 1 h hospital sessions (with about 2 weeks in between each session), showing positive effects on engagement and bonding (Looije et al., 2016). The design rationale was well-explicated by Looije et al. (2017), but did hardly include formal specifications of robot's social intelligence and did not inform *how* to extend. As a second example, Jones and Castellano (2018) used an open learner model (OLM) for a robotic tutor that promotes self-regulated learning (SRL) in a personalized scaffolding process. Based on this model, the robot shows skill meters for each competency, prompts the learner to reflect on their developing skills, and can suggest to work on an activity of an appropriate difficulty level for learning. The robot was evaluated at a primary school during 4 sessions (1 session per week) with positive results. As far as we know, the underlying OLM- and domain-models are not formalized in a way that enables (automatic) reasoning on causes and effects of the robot feedback (as an evolving "social educative intelligence"). Gordon et al. (2016) provide a third example of long-term human-robot interaction in which children play a second-language learning game with a "social robotic learning companion." An affective policy was developed to provide appropriate affective responses when the child finished a task or was not active for a while. The robot was evaluated in preschool classrooms for a duration of 2 months (each child interacted from 3 to 7 sessions). Personalization of the affective response had a positive effect on child's emotional state (valence). This study is a good example of the design of model-based social responses, but the scope is still rather limited and does not (yet) address robot's role in the class room (e.g., its relation with the teacher). As a last example, Clabaugh et al. (2018) presented preliminary results of a 30-day, in-home case experiment with a robot for children with autism. Their findings underline the importance of personalization of robots and show the relevance of research in realistic long-term, family-situated contexts. For example, parents were more comfortable to let the children interact with the robot independently and reported that the robot gave them more time for other things. How the robot could systematically support such situated social processes is not yet clear, however. Findings of these studies underline the relevance of applying a comprehensive socio-cognitive methodology that systematically addresses the social context, the building of a shared human-robot knowledge base and the opportunities to improve and learn continuously.

Figure 1 presents an overview of the Socio-Cognitive Engineering (SCE) methodology, distinguishing the foundation, specification, and evaluation. To establish the *foundation*, i.e., the operational demands, technology and human factors, a selection of established human-computer interaction and human factors methods can be applied, e.g., from the People, Activity, Context and Technology (PACT) analyses (Benyon, 2019) or Cognitive Work Analyses (Vicente, 1999; Naikar, 2017). SCE puts

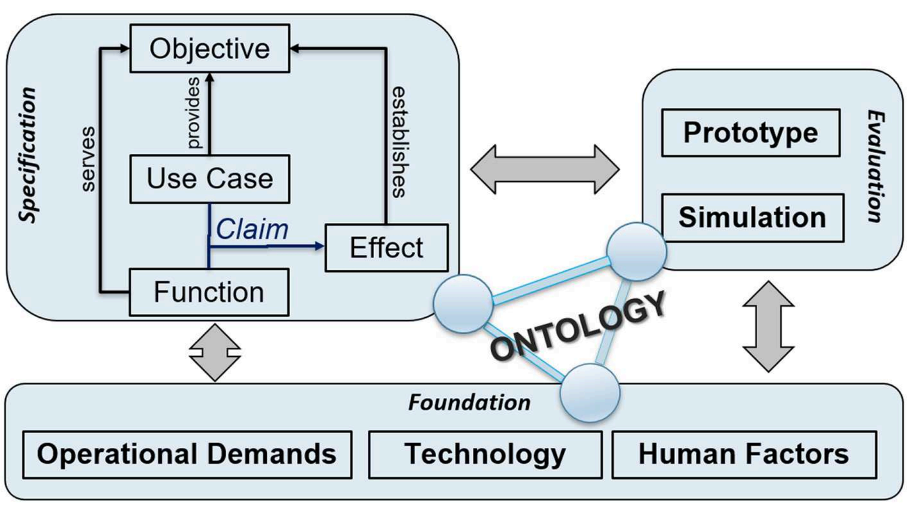


FIGURE 1 | Overview of the socio-cognitive engineering methodology.

specific emphasis on the identification of expert knowledge and cognitive theories that are relevant and can be formalized for implementation in the human-robot knowledge-base. See, for example, the "situated design rationale" method for formalizing and contextualizing behavior change support techniques of Looije et al. (2017). From the foundation, a design *specification* is derived that defines "what" the system shall do (function) in a set of use cases ("when") to bring about a desired effect (i.e., the claim, "why"). In the *evaluation*, the claims are tested via prototyping or simulations, in order to validate and refine the foundation and design specification. It is an iterative, incremental development process, aiming at a sound, theoretically and empirically grounded, prototype with a coherent description of its design rationale. Each design-test cycle will advance (a) the prototype, (b) its foundation in the human factors, technology and operational demands, and (c) the design specification. For the building, maintaining and re-using of design knowledge, SCE distinguishes the following development principles. First, creating human-centered AI and robots is viewed as an inter-disciplinary collaborative activity with active stakeholder involvement during the complete development process (cf. Riek, 2017). Second, functional modules are defined and tested incrementally in an iterative refinement process. As learning and adaptation are key characteristics of human-AI systems, this process of iterations should continue during the complete life-cycle of these systems. Third, design decisions are explicitly based on claims analyses, explicating the up-downside trade-offs. Fourth, keeping and sharing the design rationale is key for progress and coherence in the development of AI and social robots. Fifth, a common ontology should be developed and implemented, which defines the core concepts, with their relationships, for human-robot collaboration (e.g., tasks) and communication (e.g., style).

The PAL project applied these five principles in the three design-test cycles in Italy and the Netherlands (in a period of 4 years). In each cycle, we constructed, extended and refined the foundation, design specification and prototype of the PAL system. It should be noted that the direct stakeholders (children, parents, and Health-Care Professionals), the designers & engineers and the researchers (from computer science, AI, Psychology, educational science, health-care, human-computer interaction) were actively involved in the PAL research & development team from the start to the end of the project. The next sections provide more information on the SCE theories, models and methods that were applied in the PAL project.

3. FOUNDATION OF ROBOTIC LIFESTYLE PARTNER

3.1. Human Factors

Human Factors theories and methods should be used in the development of robotic lifestyle partners. The *Self-Determination Theory* (SDT) provides a coherent and well-founded starting point to support the behavior change that disease management requires. It distinguishes three human basic needs that affect the development and habituation of human behaviors in a social environment: The needs for competence, autonomy and relatedness (Legault, 2017; Ryan and Deci, 2017). By supporting these needs, as important sub-objectives, PAL is expected to achieve the main objective of enhanced self-management. For each basic need, a support strategy for a social robot (NAO) has been designed and tested successfully for children with diabetes (Blanson Henkemans et al., 2017a).

First, autonomy proves to be supported by providing choice and rationale for the (educative) activities, acknowledging children's feelings and minimizing pressure and control. It is expected that personalizing the learning objectives and providing explanations improves the responsibility transfer further. The difficulty of the learning tasks should be attuned to the skill level of the learner for an optimal learning experience and outcome. The Zone of Proximal Development (ZPD) theory states that adaptive support (or "scaffolding") can establish the required balance, encouraging and advancing the individual learning processes (Vygotsky, 1980; Chaiklin, 2003; Charisi et al., 2015). Such a balance will also help to develop an adequate level self-efficacy (Bandura, 1977).

Second, competence proves to be enhanced by providing effectance-based (instead of norm-based), reinforcing and challenging feedback. Applying motivational interviewing techniques can help to improve this feedback, i.e., providing appropriate informative feedback (corrective, descriptive, evaluative, or confirmatory responses) and motivational feedback (encouragement, praise, remark, or mood matching) (Schunk and Lilly, 1984; Tudge et al., 1996).

Third, relatedness proves to evolve positively by approaching the child in a personal, positive and respectful way. Via experience sharing in the form of reciprocal disclosures, relatedness can be further enhanced (see the Social Penetration Theory, Cohn and Strassberg, 1983; Altman and Taylor, 1987; Rotenberg and Chase, 1992; Burger et al., 2017).

Gamification principles have been studied, proposed, and applied for diverse behavior change support systems, to enhance users motivation, for example for child's diabetes self-management (Blanson Henkemans et al., 2017b). For PAL, we worked out these principles in the following educational games. A quiz is used to learn and test knowledge. A break-and-sort game is used to train the players to rapidly recognize the content of box (e.g., the categories of certain foods), challenging player's reflexes. A memory game provides a relatively relaxed and slow-paced experience for thinking and reflection. The general gamification approach entails an activity-based reward system which enables the "players" to unlock additional features for personalized engaging tasks in the PAL timeline. An achievement dashboard (Peters et al., 2019b) shows the personal achievements and (learning) goals, progress toward attainment and the possible activities for further advancement. The achievements and goals are chosen collaboratively between the child and health care professional (HCP) and selected via the HCP-dashboard (inspired by ability trees). Goal attainments are rewarded with coins. **Figures 2**, **3** show, respectively, a screenshot of the achievement dashboard and the goal tree. Coins can be earned and used to unlock new desirable content. In PAL four categories were implemented: Floor images, background images,

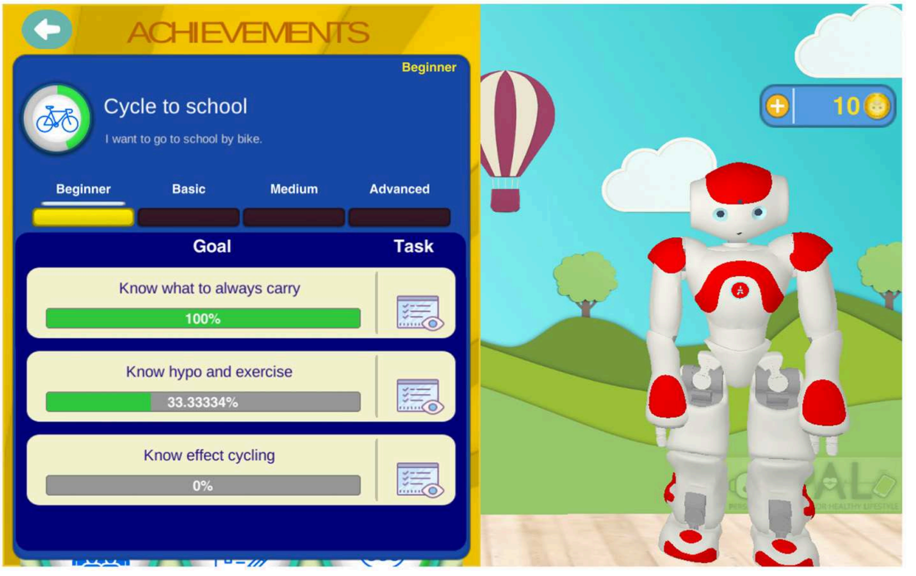


FIGURE 2 | Screenshot of the achievement dashboard in the MyPAL application to follow own goal attainment (this example shows child's progress on "Cycle to school").

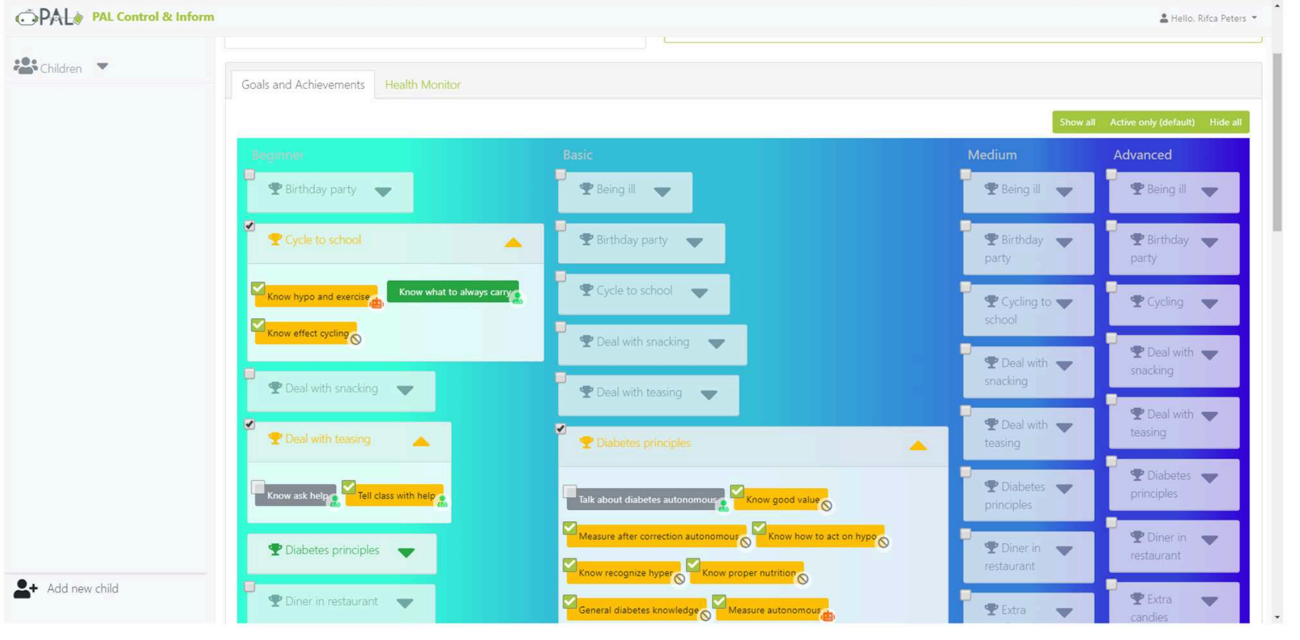


FIGURE 3 | Screenshot of the goal-tree in the PAL control & inform application to select personal goals.

color of the avatar, and dance moves (i.e., features to design dances that can be shown by the avatar and the robot). In a shop, these features can be unlocked and activated. See **Figure 4** for a screenshot of the shop in the MyPAL application.

Children are frequent users of interactive technologies for different kinds of purposes, but have hardly been involved in the design process itself to provide their specific needs and ideas (Druin, 1999, 2002; Davis, 2010). A coherent and concise set of *co-design* methods is needed, which (a) allows children to choose their own way of expression and communication and (b) provides complementary insights in their values, needs and situations (Darbyshire et al., 2005). To fulfill this need, we developed the Co-design for Child-Computer Companionship (4C) suite (Blanson Henkemans et al., 2016), consisting of

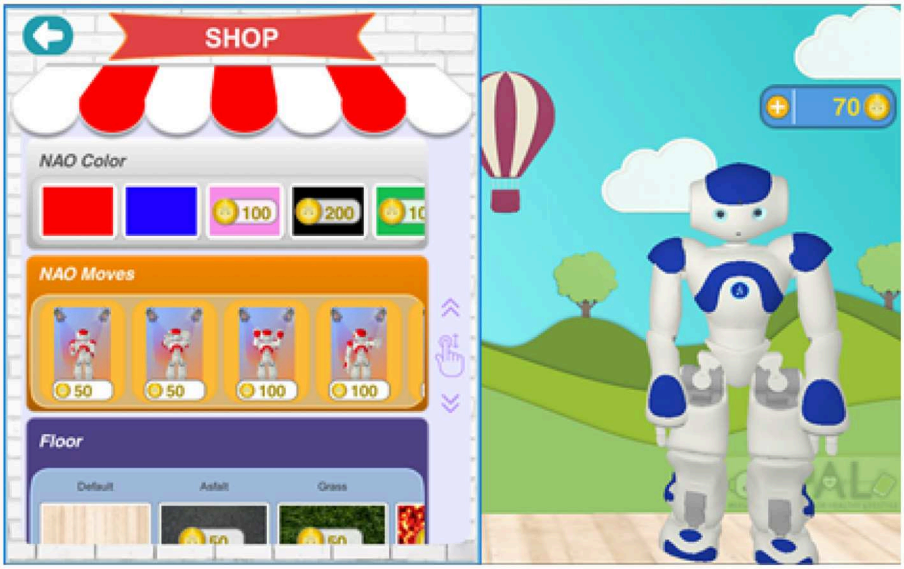


FIGURE 4 | Screenshot of the shop in the MyPAL application where one can buy nice skins and robot dance features with coins earned by doing diabetes related activities in the MyPAL applications.

two methods for eliciting daily experiences, needs and values regarding T1DM (i.e., photo-elicitation and user journey map), and three methods for collecting envisioned interactions and requirements for the PAL system (draw-write-tell, story telling, and image theater). The 4C suite has been developed by a multi-disciplinary team, involving robotics researchers, service designers, psychologists and ethicists, to establish a comprehensive, responsible and practical approach for (a) value, need and context analyses, and (b) generation of design ideas. Blanson Henkemans et al. (submitted) provide more background information on the C4 suite and its development.

3.2. Operational Demands

The direct stakeholders, particularly the children, parents, and health care professionals (HCPs), were intensively involved in both the design and test activities. At the start of the project, focus group sessions with the HCPs and diabetes organizations provided information on diabetes management and the child, family and context factors, *and* on the support needs of the HCPs themselves. Similarities and differences between the nations and hospitals were identified, and explicated in (a) flow charts of the care processes, (b) descriptions of personas and (b) journey maps for these personas ("disease management related activities of a child and his or her caregivers during the week").

Every year, the national patient organizations set-up so-called diabetes camps in Italy and the Netherlands, among other things to acquire further insight in children's values, needs and ideas for PAL support, and to assess interim designs and prototypes. As parents were partially present, their values, needs, ideas, and assessments could be acquired too. Follow-up focus group sessions with the HCPs and diabetes organizations provided further information on diabetes management and the corresponding child, family and context factors, *and* on the support needs of the HCPs themselves.

In these sessions, we acquired so-called *value* stories for each direct stakeholder (i.e., child, parent, and HCP), as a first step of the requirement analysis. Value stories have the following format: As STAKEHOLDER I want/need REQUIREMENT to support VALUE in a certain SITUATION. An example is: As A CHILD I need A PERSONAL ROBOT THAT SHARES EXPERIENCES BY ACTIVE LISTENING AND TELLING ABOUT ITSELF IN A SIMILAR WAY to support RELATEDNESS in THE PAL-ACTIVITIES AT THE CAMP, HOSPITAL AND HOME. In addition, possible value tensions were identified (such as the tension between privacy and health for the sharing of information about diabetes regime adherence). To address these tensions adequately, we formulated a general requirement on the creation, activation, and adaptability of agreements (permissions, prohibitions, and obligations to share information, e.g., when the child is staying with a friend; cf. Kayal et al., 2018a,b).

3.3. Technology

The PAL project uses a state-of-the-art *humanoid robot*, the NAO of Softbank Robotics, which has four microphones, two speakers and two video cameras. The robot is present at the hospital and diabetes camps (and might sometimes visit the child at home or school). As an avatar, a virtual 3D robot model (i.e., a "copy") was developed in the Unity environment, which has the same appearance, movement and interaction characteristics (there is one "expression model" for these two embodiments). The avatar is developed for Android mobile devices, particularly for a tablet that is used at home.

Cloud computing is used to establish an evolving modular and distributed intelligence that facilitates long term interaction

(Kehoe et al., 2015). It enables (a) accessing external libraries for enriching interaction such as dialogues, (b) relatively heavy computations such as statistical analyses of previous behaviors and their outcomes, (c) collective human-agent learning (the human and the robot can, in real-time, learn from each others' interactions by means of data sharing), and (d) monitoring the robot's interactions and adapting the decision making where and when needed. The PAL "brain" is set-up in a modular manner to support incremental development [easy addition and updating of (sub-)modules]. All messages go through a common messaging board called the *nexus*. Each module sends messages of a particular type and decides itself to subscribe to specific message types of other modules.

A *hybrid AI* approach was chosen that combines symbolic reasoning methods [like Belief-Desire-Intention (BDI) agent frameworks] with Machine Learning methods. The symbolic reasoning frameworks allow to implement expert knowledge into the system, and to provide meaningful control and interpretable output for the human. The machine learning methods allow to leverage the available data for potentially continuous performance improvements. For example, estimating child's knowledge level is an important continuous process ("user modeling") for the planning of the next (learning) tasks. Concerning machine learning, a combination of collaborative filtering, Gaussian processing, and covariance matrices is used to track child's knowledge level in PAL (see Cully and Demiris, 2019), and a deep learning Gated Recurrent Unit (GRU) model for aspect extraction to track child's emotional state on the topic of a textual expression (Haanstra and de Boer, 2019). Concerning symbolic reasoning, a Cognitive Agent Architecture Framework is used to provide adaptive—goal-, belief-, emotion-based—explanations (Kaptein et al., 2016; Neerincx et al., 2018), and a dialogue management framework for the human-agent conversations in general.

The knowledge base of the symbolic reasoning framework of PAL entails a *federated ontology*. Ontologies provide explicit, formal descriptions of objects and concepts (their properties), and of the relations among them (Gruber, 1993). In SCE (see **Figure 1**), the ontology covers concepts from the foundation, specification and evaluation, and functions as an evolving knowledge base that: (1) provides an unambiguous vocabulary and communication between stakeholders, (2) supports system implementation of knowledge-based reasoning functionalities, and (3) serves as a basis for interoperability in human-agent interaction (as they contain human expert knowledge and have an inbuilt logic that machines can process and interpret). The PAL ontology integrates individual ontologies ("models") via one top-level ontology. These models are high-level building blocks that contain smaller, more specific areas of interest ("frames") (Neerincx et al., 2016a). When useful, existing ontological frames can be rather easily included in the evolving ontology. The PAL system uses an extended Resource Description Framework (RDF) storage component and reasoner (HFC) to process the knowledge models (classes) and running instances in conjunction (van Bekkum et al., 2016).

Kaptein et al. (in review) provide more background information on the PAL system architecture and technology.

4. SPECIFICATION OF SITUATED HUMAN-ROBOT PARTNERSHIPS

Based on the human factors, operational and technological analyses of section 3, we worked out the core functions (4.1) and knowledge-base (4.2) of the robotic partner with the corresponding interaction design (use cases, requirements, and claims; 4.3) of the PAL system.

4.1. Partnership Functions

Five high-level ("core") functions of a robotic lifestyle partner like PAL are expected to enhance the disease self-management, distinguishing 4 partnership functions in italics (see **Figure 5**):

1. Providing personal, reliable and reinforcing assistance on diabetes management via learn-by-playing activities.
2. Planning and pursuing joint *objectives* for the disease management. These objectives (like enhanced diabetes management) drive robot activities in a consistent and transparent way, and are compliant with stakeholder values. Furthermore, the style of communication is harmonized with the joint objectives (e.g., showing "warmth," "competence," and "dominance," Peters et al., 2015, 2017b, 2019a).
3. Proposing and committing to *agreements* for value-sensitive information sharing. To address value trade-offs adequately, information sharing might be permitted, prohibited or obliged for specific stakeholders, situations and periods (such as keeping emotional statements private in specific situations).
4. Sharing *experiences* via disclosures that match the disclosures of its human partner. For long-term lifestyle partnerships, mutual understanding and relationship building is crucial (such as learning to cope with the effects of specific stress events and sport activities on the personal blood glucose regulation).
5. Providing *feedback* on partner's behaviors, learning progress, and *explanations* of own behaviors. These responsive and pro-active communications should be constructive and personalized to establish prolonged motivation, learning and trust.

4.2. Partnership Knowledge-Base

We worked closely together with health care professionals to obtain and implement an ontology that contains the relevant knowledge and content for these core partner functions. The PAL ontology integrates individual ontologies ("models") via one top-level ontology. Relevant existing ontological frames were identified and included in the PAL ontology. For some, only parts of the frame were relevant, and therefore partially included (e.g., the self-management activities of diabetes, but *not* the entire professional medical diagnosis and treatment model of diabetes). Other frames had to be extended with additional concepts into a PAL model [e.g., the well-known task ontology Van Welie et al. (1998) in the PAL Objective Model]. The PAL ontology contains models that capture mutually different knowledge; no direct dependencies have to be specified for the concepts of one model to the concepts of another model. The independence of the models has as advantage that it provides

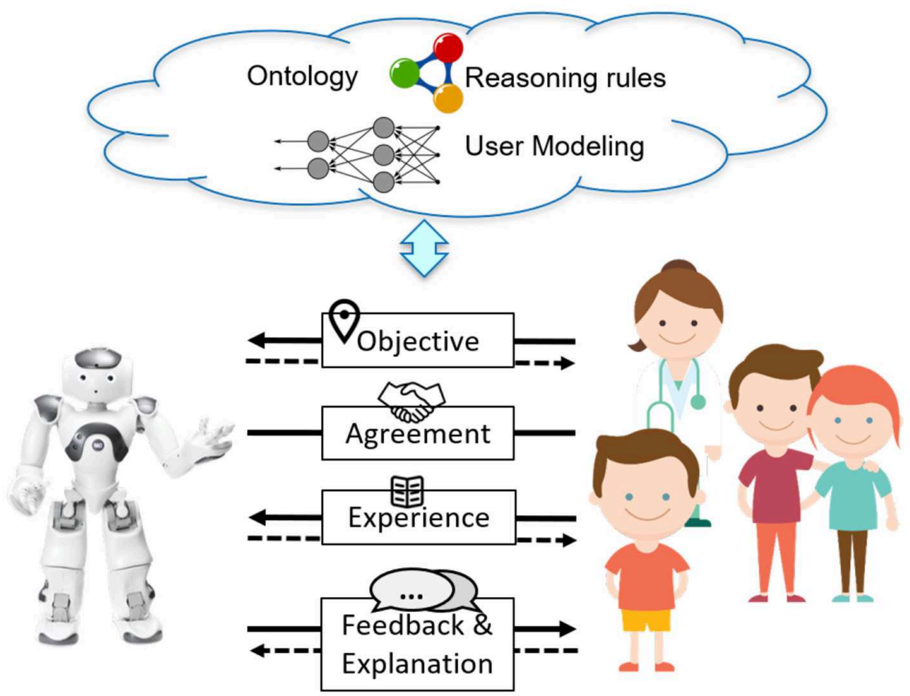


FIGURE 5 | The PAL actor (robot and avatar), its core functions and knowledge-base to act as partner in the diabetes management of a child and his or her diabetes care team.

clean sub-ontologies which can be reusable in other projects and/or domains. Besides that, this structure had a practical advantage that different project partners could work on the sub-ontologies simultaneously, without interfering with each other. Currently, the PAL ontology consists of ten models, the top-level ontology and nine models that capture different knowledge (most are available as separate files using the Web Ontology Language, OWL, which can be directly re-used in other hybrid AI systems):

- The *PAL Objective Model (POM)* entails a decomposition of achievements into (learning) goals, which are further decomposed into (learning) tasks (Peters et al., 2017a, 2019b). When the underlying tasks are completed, the goal is attained (it can be that *either* task A *or* task B has to be completed). When the underlying goals are attained, the achievement is gained. For example, to gain the achievement of competence for a sleepover, the child has to attain the learning goals to know "how and when to measure blood glucose" and "what to do when I am experiencing some tremors" by completing the corresponding tasks of the diabetes quiz and memory game.
- The *Domain* model describes characteristics of diabetes, the PAL system, the direct stakeholders (end-users) and locations (such as hospital, diabetes camp, home). The classes could be core domain concepts (e.g., actor, activity, food, ..) or relate to other classes (e.g., "pen" and "pump" are sub of "device"). Specific information for the dialogue modeling has been included in the domain model (i.e., the classes and properties that constitute the information state of the dialogue components).
- The *Episodic Memory* model combines the Ontology-based Unified Robot Knowledge (OUR-K) with a temporal episode ontology that models the 5W1H (When, Where, Who, Why, What, and How) (Lim et al., 2011, 2013; Han et al., 2013). This model is used to record and activate memory episodes. Based on rules, each episode is tagged with possible triggers (e.g., child login) that, for example, can activate a corresponding speech act.
- The *Agreement* model specifies the underlying concepts with their relations: Type (permission, prohibition, obligation), creditor, debtor, antecedent, consequent, condition, and adaptability (Kayal et al., 2018a,b; Mioch et al., 2018). In the current PAL system, the focus is on the sharing of child's data. For indications of serious health risks (e.g., very high or low blood glucose values, or a long-lasting negative emotional feeling), policies to inform parent and HCP were specified and implemented. The default setting for other information is: Sharing information with PAL is permitted, whereas it is prohibited with other stakeholders (e.g., parent and HCP). Agreements can be set-up to change this information sharing. PAL will act according to these agreements and provide the information as agreed upon.
- The *Semantics* model is tailored to the specific needs of the PAL games (quiz, break & sort, and memory game). We developed a simple frame semantics that is oriented along thematic roles, and deviated from the FrameNet Frame

semantics (which would require heavy modification and extension as it is very general). Among other things, it aims to underpin natural language generation and interpretation, and to support multilingualism (i.e., linking concrete realizations in the different languages to the abstract concept as, e.g., Multilingual WordNet does).

- The *Affect* model is based on James Russell's Circumplex Model of Emotions, the Schachter-Singer theory of emotion and Joseph Forgas' Affect Infusion Model (Schachter and Singer, 1962; Forgas, 1995). It describes how Mood and Emotion continuously influence each other.
- The *Interaction Style* model is based on Leary's Interpersonal Circumplex, the Model of Interpersonal Teacher Behavior, and Grasha's theory on teaching styles (Leary, 1958; Grasha, 1994; Wubbels et al., 2012). It describes how (teaching) styles are constructed from dominance, friendliness and competence expressions, and the (learning) activities for which they are appropriate.
- The *Feedback* model is based on motivational interviewing techniques that distinguish four informative feedback styles (corrective, descriptive, evaluative or confirmatory responses) and four motivational feedback styles (encouragement, praise, remark or mood matching) (Schunk and Lilly, 1984; Tudge et al., 1996). It specifies the events and states that trigger the corresponding feedback style, and the speech acts for each style.
- The *Explanation* model describes characteristics of the explanations and the agents involved (Neerincx et al., 2018). It distinguishes Roles (such as student and teacher), Explanation Types (such as contrastive and BDI-based), Interpretation, Explanandum, Explanans, One or more statements provided through some medium (e.g., sentences) that are offered to explain a phenomenon or an argument.
- The *Small Talk* model provides the data structure specifications for all kinds of small talk dialogues. It distinguishes Starters, Prompts, Disclosures (with topic and intimacy level) and Closure parts to conduct such dialogues. Other concepts have been added to enrich the conversation, like parameters concerning Intimacy level, Topic, Valence, and Liking (Burger et al., 2017).

This ontology represents an important part of the human-robot *collective intelligence*: Knowledge that the robotic system and the humans share and use for their reasoning and conversations (e.g., the feedback and conversations). For example, **Figure 6** presents the "Health Monitor" tab of the Health Care Professional's dashboard that is based on the Domain Model, sharing information on child's glucose level (hype / hyper), insulin administration, carbohydrates in consumed nutrition, activities and emotions (The "Goal and Achievements" tab, not shown, contains child's plan and progress in achievements, goals and tasks, based on the PAL Objective Model).

4.3. Use Cases, Requirements, and Claims

Following the partnership functions and knowledge-base, we specified more in detail a set of use cases *with* the required PAL functions (i.e., the functional requirements) and expected effects (i.e., the claims; see **Figure 1**). Each use case refers explicitly to an objective, its pre- and post-conditions and the actors involved. It specifies the sequence of actions and dialogues with an explicit reference to the corresponding requirement and claim. For example, the *use case* "Managing child's objectives" contains an action and dialogue "HCP monitors child's progression at his or her work place," referring to a *requirement* "PAL shall provide an interactive overview of the realized and active objectives" and a *claim* "HCP identifies progress successes and delays effectively and efficiently." Use cases have been defined for the hospital and home settings. Requirements have been derived for the overall system, the actor (robot and avatar) behaviors, the timeline, the dashboards, and the tasks. Claims concerned diabetes management behaviors (e.g., the working on the learning goals) and outcomes (e.g., the HbA1c as measure of blood glucose regulation in the last period), well-being indicators (e.g., child vulnerability), PAL system usages (e.g., usage time), and social effects (e.g., responsibility transfer). In total, 19 measuring instruments were selected or constructed to test the claims. Looije et al. (2017) give a detailed description of the method (and tool) to specify and test the use cases, requirements and claims in a coherent way.

5. EVALUATIONS OF THE PROTOTYPES

As mentioned before, the 4-year PAL-project entailed three design-test cycles. A summary of the SCE specification per design cycle can be found in **Table 1**. After the first year, we established the first integrated system that was tested in Italy and the Netherlands. Following the incremental development approach, the system kept on running, being available for all the tests and being updated when appropriate (development was taking place on a test environment, a "copy" of the system in use). This way, the development and evaluation activities could continue in parallel and prototypes could be always assessed at all locations. A diverse set of complementing, formative and summative, evaluations was conducted during each cycle. For the formative studies, we developed the 4C-suite that has been described in section 3.1. Dedicated usability tests were performed, e.g., focusing on (a) the games usage at home, the hospital (e.g., see **Figure 7**) or diabetes camp, (b) the comprehensibility and use of the objective model, (c) the ease-of-use in general, and (d) the dashboards (e.g., Peters et al., 2019b).

In *Cycle 1*, a first version of the human-robot partnership framework was worked out, built and tested. The Self-Determination Theory (SDT) is an important foundation of the objectives that are being served, provided and established by, respectively, the functions, use cases and (expected) effects in the design specification (see **Figure 1**). For the "gamified" quiz use case with an empathetic robot and avatar (among other things), three claims were specified that fulfill the human basic needs: Increased knowledge for competence (e.g., being able to recognize symptoms of a hypo), liking for relatedness and positive experiences for autonomy. These concepts were

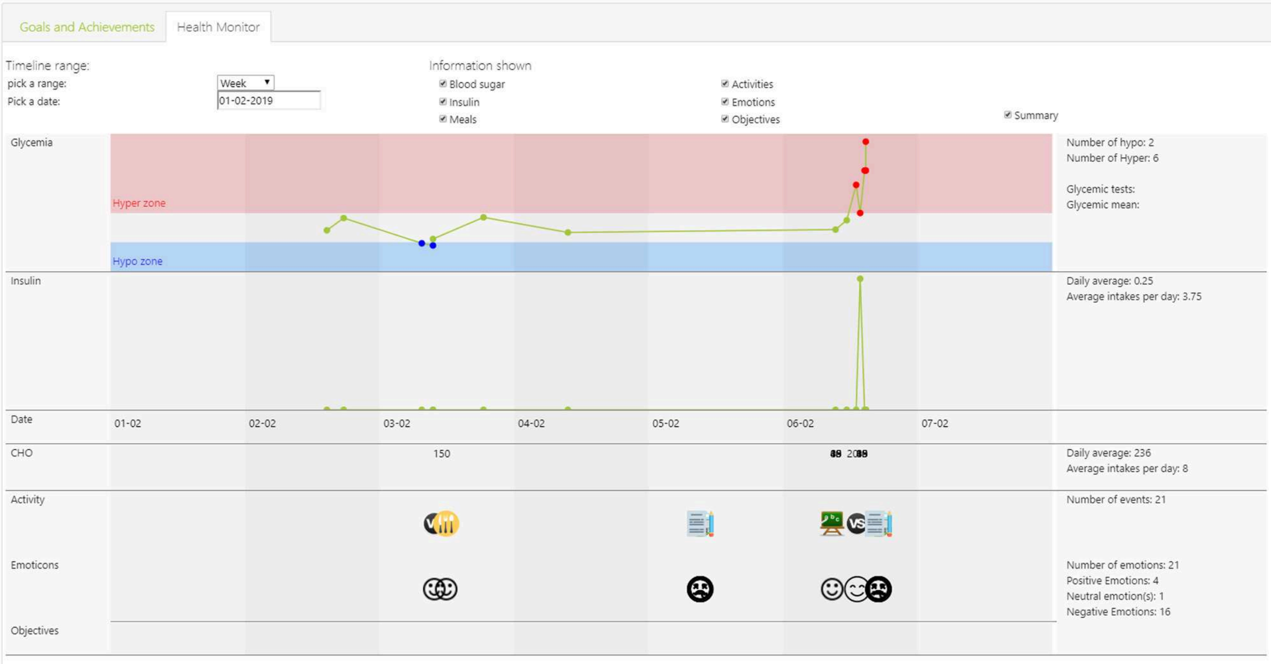


FIGURE 6 | Screenshot of the "Health Monitor" dashboard for Health Care Professionals, showing over a time period: Child's glycemia level (e.g., hypo / hyper), insulin administration, carbohydrates (CHO) in consumed nutrition, activities and emotions.

TABLE 1 | Explanatory selection of PAL's foundation, core functions, use case (UC) implementations, and claims that were evaluated per design-test cycle.

Cycle	Foundation	Core functions	UC implementation	Claims
1	Self-determination theory, zone of proximal development, gamification, ALIZ-E design rationale. Value stories, journey maps, co-designed scenarios. Cloud computing, hybrid AI and federated ontology.	*R1*: PAL shall provide learn-by-playing activities with personal, reliable, and reinforcing assistance on diabetes management. *R2*: PAL actor shall show empathic partnership. *R3*: PAL shall support joint planning and pursuing personalized objectives.	*Robot interaction*: Acquaintance, quiz. *MyPAL environment*: Avatar, timeline and quiz. *Dashboards*: PAL control and inform.	*C1*: Child has increased knowledge on T1DM. *C2*: Child likes the PAL actor (robot and its avatar). *C3*: Child experiences diabetes-related activities more positively.
2	Social penetration theory, motivational interviewing, folk psychology. New co-designed scenarios. System reliability, usability engineering for children.	*R4*: PAL actor shall share experiences via mutual self-disclosure. *R5*: PAL actor shall provide feedback and explanations on behavior. *R6*: PAL actor shall show personalized learning styles	*Robot interaction*: Break and sort game. *MyPAL environment*: Dialogues, reward system (earn coins) and a shop.	*C4*: Child bonds with the PAL actor via the robot *and* its avatar. *C5*: Child is motivated to work on his or her personal objectives with PAL.
3	Expert knowledge on child's learning processes for diabetes management with culture- and hospital dependencies. New co-designed scenarios. Game-based learning.	*R1.1*: PAL's support for planning and pursuing objectives shall be personalized and harmonized to child's daily life. *R7*: PAL shall propose and commit on agreements for information sharing.	*PAL actor*: Small talk, dancing designed by child. *MyPAL environment*: Tip of the day, memory games (3), videos, real world tasks, high score board, interactive overview of objectives. *Dashboards*: Making agreements about information sharing.	*C1.1*: Child has increased situated knowledge on T1DM. *C6*: Child is aware of T1DM state and causes and develops self-efficacy *C7*: Child has a higher Quality of Life concerning T1DM *C8*: Children seamlessly follow culture- and hospital-dependent diabetes management processes. *C9*: Child pursues relatively difficult goals.

The successive rows (cycles) show the increments on the previous row (i.e., the extensions of the foundation, functions, implementations, and claims).

applied and tested at diabetes camps and hospitals with children aged 7–10 years. First, during 1-week diabetes camps in Italy and the Netherlands, a user needs assessment was conducted and PAL mock-ups were tested ($N = 55$). Second, an initial version of the PAL system (PAL 1.0) was evaluated in a 1-month test at Italian and Dutch hospitals ($N = 21$). The

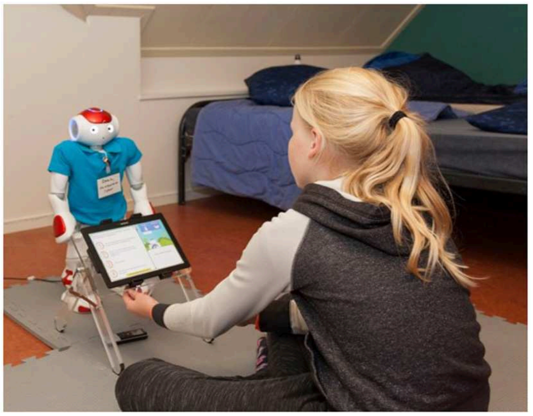


FIGURE 7 | Child-robot interaction during the quiz.

claims were tested in these different evaluations, for usage periods between 1 and 4 weeks. In general, positive effects were recorded on the SDT-related claims. Children enjoyed to interact with the PAL robot and avatar (which made diabetes-related activities more positive) and showed an increased diabetes knowledge when using the PAL system (i.e., functions *R1 and R2* and claims *C1, C2, and C3* in **Table 1**). However, for meaningful benefits over a longer period, the PAL system needed substantial improvement. Particularly, shortcomings of the reliability, usability, and goal structure hindered the acceptance and trust of health care professionals. Further, enhanced personalization proved to be needed to establish adherence of the children.

In *Cycle 2*, the SCE-activities continued, building on the results of the first cycle. After establishing the general PAL framework in cycle 1, innovative PAL functions were identified for which, first, the specific module had to be developed and tested, before it would be integrated in the overall system (actually, already in cycle 1 and continuing in cycle 3, this modular approach was taken). For example, Burger et al. (2017) tested the experience sharing function with 11 children over the course of $\sim$ 2 weeks at home (i.e., function *R4* and claim *C4* in **Table 1**). The number of child disclosures proved to be an indication of their perceived relatedness at the end of the experiment. The higher the relatedness, the better the system usage. Subsequently, this function was implemented in the PAL system. In a similar way, the feedback & explanation functions were tested (i.e., function *R5* and claim *C1* in **Table 1**). For example, Kaptein et al. (2017) tested robot's self-explanation with 19 children and 19 adults in which the robot performed actions to support type 1 diabetes mellitus management. Adults showed a higher preference for goal-based explanations than children, providing a foundation for personalizing the explanation. The explanations have been integrated in the PAL system. As a third example, Peters et al. (2017b) developed a model of non-verbal warmth and competence robot behaviors, which is expected to improve robot's teaching style (i.e., function *R6* and claim *C1* in **Table 1**). A perception experiment at primary schools and a diabetes camp showed that even subtle behavior manipulations affect children's warmth-competence perceptions of the robot. This model has been implemented in the PAL system. The last example of a focused experiment concerns the avatar function that was tested at a diabetes camp (Sinoo et al., 2018). The bonding with the physical robot was higher, but this effect reduced when children perceived the physical robot and its avatar more as the same agency. The stronger friendships, the higher the motivation to perform the tasks to do. Therefore, we improved the similarity and consistency between robot and avatar in the next version of PAL. Finally, a study in Italian and Dutch hospitals was conducted with children aged 7–12 years ($N = 35$). The primary aim of the experiment was to refine, further develop and evaluate the second release of the PAL System (2.0), during a longer period of use (i.e., from 3 to 4 months). Main results were that the children bonded with the PAL actor (robot and avatar), and perceived the robot and avatar somewhat as similar. During the experiment, they perceived it increasingly as a buddy who was supporting and making them happy (i.e., functions *R4, R5, and R6* and claim *C4* in **Table 1**). Notably, the majority of the children stopped using MyPAL App some weeks after the beginning of the study. A large number of children already had participated in cycle 1 and as can be expected, the novelty effect disappeared. They felt there was insufficient new interesting content (i.e., amount and variety of activities) and rather limited child-actor interactions to maintain motivation to use the PAL system for such a long period.

In *Cycle 3*, based on the results of the two first cycles, to work toward ongoing and impacting use of the PAL system by children in regard to T1DM self-management, we introduced new and improved existing functionalities in the PAL system, which were discussed earlier in this paper: (1) General usability of the MyPAL app, goal setting, enriched interaction, additional educational material, gamification, and monitoring for parents (PAL dashboard). This last design-test cycle, contained a randomized controlled trial: A summative evaluation that compared child's self-management with the PAL-system 3.0 vs. "care as usual," for a period of twice 3 months (with 49 children aged 7–14 years, in the Netherlands and Italy). Phase 1 (the first 3 months) consisted of the effect study, and phase 2 of an implementation study in which both the children who used PAL and the children who got care-as-usual ("waiting-list") could chose to use a further improved version of PAL (PAL-system 3.5). In total, 14 children interacted with the MyPAL application for 6 months while 26 children participated with the MyPAL application for 3 months (16 were in the waiting-list group and 10 participated in phase 1 only). Each phase started and ended in the hospital. In the intervention condition at the hospital, the child, parent and health care professional set or reflected on the objectives and made agreements about information sharing using the PAL dashboard. Further, during the first visit to the hospital, the interaction with the robot consisted of introduction, acquaintance and play of a game (quiz or break and sort). During the last visit to the hospital, the robot also did a dance choreographed by the child through MyPAL at home (with the avatar). In between the hospital visits, over a 3 month period, the children could play with the avatar via the MyPAL application at home. It is a "real" avatar that continues the activities and interactions of the robot; i.e., the robot and avatar have "only" a different embodiment but act as the same actor (based on the same models and memory). The children were free in deciding how often they wanted to play with the system. Interaction with the avatar at home consisted of saying "hello," reviewing personal goals, and performing tasks contributing to goals. Human-robot interactions entailed, for example, one of the educative games (quiz, break & sort, memory), watching a video, keeping a diabetes diary, a real life activity, dialogue acts (task suggestions, tips, feedback & explanations) or small talk.

PAL proved to partially support the three human basic needs that affect the development and habituation of human behaviors in a social environment, such as disease self-management (see Self-Determination Theory, section 3.1). Children liked the PAL-robot and were motivated to continue the robot-mediated tasks (*relatedness*). This is consistent with the results of a previous experiment at the diabetes camp, presented in section 5 (Sinoo et al., 2018). The tasks to pursue differed between Italian and Dutch children, reflecting cultural differences on diabetes management (function R1.1 and claim *C8*). In regard to diabetes knowledge, children in the intervention group, using the PAL 3.0, in comparison to the control group, showed a stronger increase after 3 months, than children in the wait-list group [$F_{(30)} = 4.17$, $p = 0.05$]. Moreover, we found a correlation between time playing with the MyPAL app and children's knowledge. Also, children in the intervention groups had a stronger increase in self-care score [$F_{(30)} = 6.60$, $p = 0.01$], as an indication of improved *autonomy*. Furthermore, younger children in the intervention group showed a stronger increase in self-care score, in comparison with their older peers in the intervention group ($p = 0.03$). We did not find an effect of PAL on parental stress and child's glucose regulation (including HbA1c and percentage of measures in healthy range). However, we did find an effect on diabetes related quality of life in children [$F_{(30)} = 6.14$, $p = 0.02$] (i.e., functions *R1 through R7* and claims *C1.1 and C5 through C9* in **Table 1**). Blanson Henkemans et al. (submitted) provide a detailed description of the randomized controlled trial.

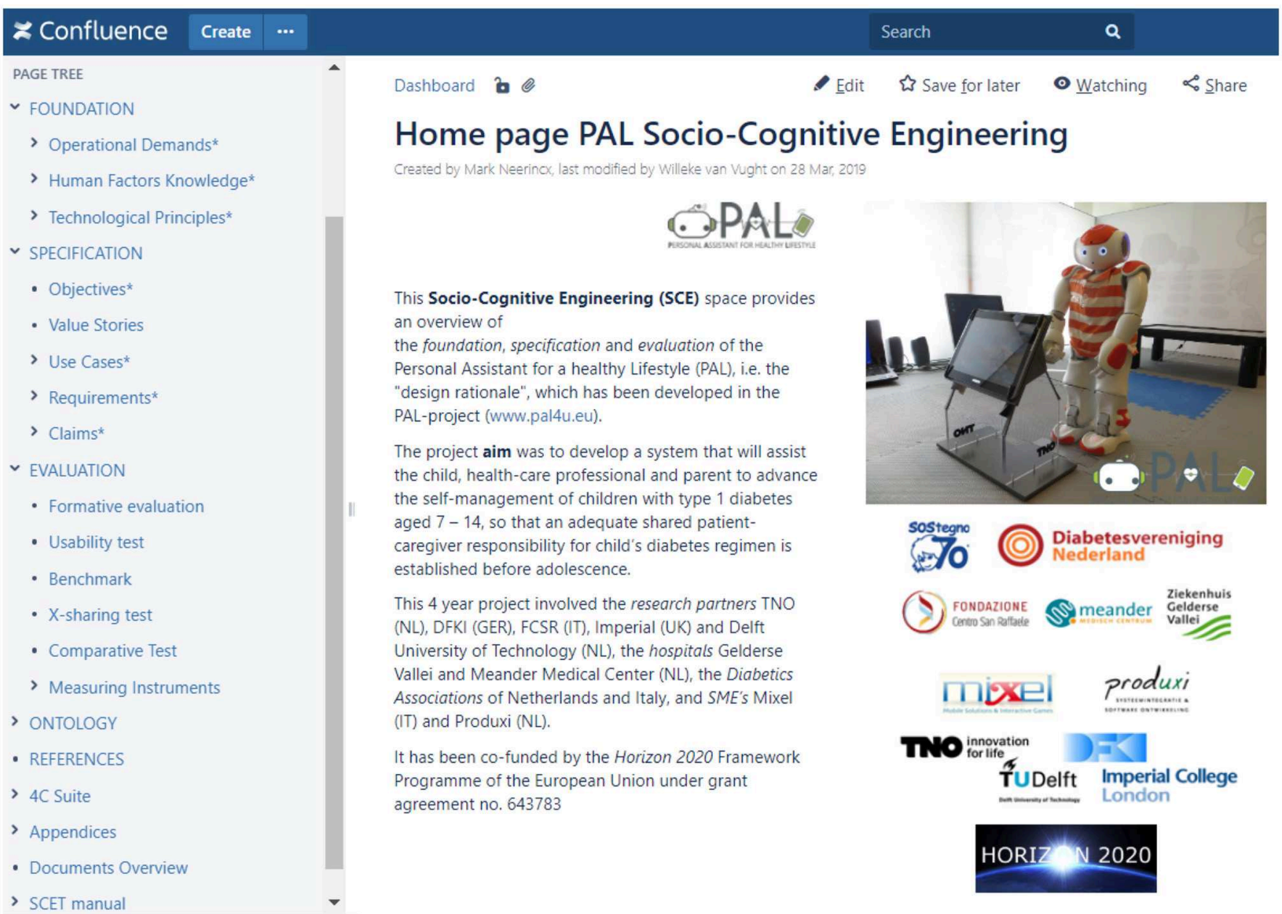


FIGURE 8 | Screenshot of the menu and PAL start page of the Socio-Cognitive Engineering Tool (SCET).

6. CONCLUSIONS

This paper presented an overview of the SCE-methodology and its application for the PAL research & development activities to develop a robotic partner. As a first contribution to the field of social robotics, it shows how to progressively integrate domain and human factors knowledge into social robots via co-design, modeling and evaluations. The models and design rationale, integrated in the robots, are constructed for re-use and further development. As a second contribution, the paper presents a social robot with dedicated partnership functions and a corresponding knowledge-base that is constructed and shared with the human stakeholders. This robotic system has been evaluated "in the wild," i.e., at hospitals, diabetes camps and home, in Italy and the Netherlands.

The introduction of this paper distinguished three research questions. Section 2 proposed the Socio-Cognitive Engineering (SCE) methodology as answer to the first question: "How to develop human-agent partnerships for long-term lifestyle support?" We succeeded to integrate into the PAL-system: (a) theories, models and methods from different scientific disciplines, (b) technologies from different fields, (c) diabetes management practices from different nations and hospitals, and (d) last but not least, the diverse individual and context-dependent needs of the children and their caregivers. Our PAL experiences underpin the argumentation for SCE in section 2, but it needs further grounding in usages by others. It should be mentioned that the re-usable PAL design rationale, ontological models and Co-design for Child-Computer Companionship (C4) suite are maintained and accessible in the Socio-Cognitive Engineering Tool (SCET), which is built and maintained in Atlassian Confluence (a wiki content tool for teams to collaborate and share knowledge efficiently; currently within the PAL consortium, but we are exploring ways to share it with other research & development communities). **Figure 8** shows a screen shot of the SCET with PAL content (Note that the menu left is consistent with **Figure 1**).

The second research question was addressed in section 4 and illustrated by **Figure 5**: "How can a robot partner support the daily diabetes management of children over a longer period?" This section described the 4 partnership functions, the knowledge-base and interaction design of the situated human-robot partnerships for the development of child's disease self-management. In our view, it is one of the first examples of prolonged human-agent/robot teamwork for a healthy lifestyle that has been researched, developed and tested in the field. It represents a new type of evolving human-robot systems with collective intelligence. Both the robot and the human stakeholders acquired more knowledge about child's diabetes management (e.g., recorded in the ontology, like the PAL Objective Model).

The third evaluative question can be answered positively: "Does this partnership improve child's diabetes-control and well-being?" Section 5 provided a brief overview of the evaluation results. PAL proved to support the children on the three basic needs of the Self-Determination Theory: autonomy, competence, and relatedness. To our knowledge, PAL provided the first field study of prolonged "blended" care with a robot for children with a chronic disease, showing positive results in a 3 month evaluation period.

In the next steps of the research and development, we recommend to improve the team aspects concerning responsibility transfer and caregiver involvement. For this, explicit responsibility (transfer) objectives should be included in the PAL Objective Model, and the PAL dashboards should be integrated with the hospital information system (i.e., the work environment of the HCPs). Further, the children would profit substantially from better (technical) integration of their diabetic measurement and administration devices with the PAL system.

Another direction is to apply the models and methods for the management of other diseases of children, such as asthma, and patient or client groups, such as older adults with Type2 Diabetes. Concerning scientific progress, we are researching hybrid AI models that can provide enhanced personalized predictions on patient's health condition (such as hypo or hyper) and can explain these predictions to humans in a way that the human can understand and use (e.g., for the child, the parent, and the HCP).

DATA AVAILABILITY STATEMENT

The datasets generated for this study will not be made publicly available for the Protection of privacy.

ETHICS STATEMENT

Each study involving human participants has been, from case to case, reviewed and approved by the reference ethical committees of the concerning participating organizations (i.e., the hospitals and research partners in the Netherlands and Italy). Written informed consent to participate in this study was provided by the participants and, when appropriate, participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

The 4 year PAL-project involved the *research partners* TNO (NL), DFKI (GER), FCSR (IT), Imperial (UK), and Delft University of Technology (NL), the *hospitals* Gelderse Vallei and Meander Medical Center (NL), the *Diabetics Associations* of Netherlands and Italy, and *SMEs* Mixel (IT) and Produxi (NL). PAL has been co-funded by the *Horizon 2020* Framework Program of the European Union under grant agreement no. 643783.

ACKNOWLEDGMENTS

We thank the children, parents, and health-care professionals who participated in the PAL project. Further, we acknowledge our colleagues and students who substantially contributed to the project outcomes.

REFERENCES

Altman, I., and Taylor, D. (1987). "Communication in interpersonal relationships: Social penetration theory," in *Interpersonal Processes: New Directions in Communication Research*, ed M. E. Roloff (Newbury Park, CA: SAGA Publications, Inc), 257–277.

Anjomshoae, S., Najjar, A., Calvaresi, D., and Främling, K. (2019). "Explainable agents and robots: results from a systematic literature review," in *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems* (Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems), 1078–1088.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychol. Rev.* 84, 191–215. doi: 10.1037//0033-295X.84.2.191

Benyon, D. (2019). *Designing User Experience: A Guide to HCI, UX and Interaction Design*. Amsterdam: Pearson.

Betts, P., Buckley, M., Davies, R., McEvilly, E., and Swift, P. (1996). The care of young people with diabetes. *Diabet. Med.* 13(9 Suppl. 4):S54.

Blanson Henkemans, O., Neerincx, M., Pal, S., van Dam, R., Shin Hong, J., Oleari, E., et al. (2016). "Co-design of the pal robot and avatar that perform joint activities with children for improved diabetes self-management," in *The 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (IEEE Press).

Blanson Henkemans, O. A., Bierman, B. P., Janssen, J., Looije, R., Neerincx, M. A., van Dooren, M. M., et al. (2017a). Design and evaluation of a personal robot playing a self-management education game with children with diabetes type 1. *Int. J. Hum. Comput. Stud.* 106, 63–76. doi: 10.1016/j.ijhcs.2017.06.001

Blanson Henkemans, O. A., Van der Pal, S., Werner, I., Neerincx, M. A., and Looije, R. (2017b). "Learning with Charlie: a robot buddy for children with diabetes," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY: ACM), 406.

Boyer, B. A., and Paharia, M. I. (2008). *Comprehensive Handbook of Clinical Health Psychology*. Hoboken, NJ: John Wiley & Sons.

Bradshaw, J. M., Dignum, V., Jonker, C., and Sierhuis, M. (2012). Human-agent-robot teamwork. *IEEE Intell. Syst.* 27, 8–13. doi: 10.1109/MIS.2012.37

Burger, F., Broekens, J., and Neerincx, M. (2017). "Fostering relatedness between children and virtual agents through reciprocal self-disclosure," in *28th Benelux Conference on Artificial Intelligence, BNAIC 2016. 10 November 2016 through 11 November 2016, 765, 137-154*, eds B. Bredeweg and T. Bosse (New York, NY: Springer Verlag).

Centers for Disease Control and Prevention (2011). *National Diabetes Fact Sheet: National Estimates and General Information on Diabetes and Prediabetes in the United States, 2011*, Vol. 201. Atlanta, GA: US department of health and human services, centers for disease control and prevention, 2568–2569.

Chaiklin, S. (2003). "The zone of proximal development in Vygotskys analysis of learning and instruction," in *Vygotskys Educational Theory in Cultural Context, Vol. 1*, eds A. Kozulin, B. Gindis, V. S. Ageye, and S. M. Miller (Cambridge, UK: Cambridge University Press), 39–64.

Charisi, V., Davison, D., Wijnen, F., Van Der Meij, J., Reidsma, D., Prescott, T., et al. (2015). "Towards a child-robot symbiotic co-development: a theoretical approach," in *AISB Convention 2015* (Canterburry, UK: The Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB)).

Clabaugh, C., Becerra, D., Deng, E., Ragusa, G., and Matarić, M. (2018). "Month-long, in-home case study of a socially assistive robot for children with autism spectrum disorder," in *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY: ACM), 87–88.

Cohn, N. B., and Strassberg, D. S. (1983). Self-disclosure reciprocity among preadolescents. *Pers. Soc. Psychol. Bull.* 9, 97–102. doi: 10.1177/0146167283091013

Coninx, A., Baxter, P., Oleari, E., Bellini, S., Bierman, B., Henkemans, O. B., et al. (2016). Towards long-term social child-robot interaction: using multi-activity switching to engage young users. *J. Hum. Robot Interact.* 5, 32–67. doi: 10.5898/JHRI.5.1.Coninx

Cully, A., and Demiris, Y. (2019). Online knowledge level tracking with data-driven student models and collaborative filtering. *IEEE Trans. Knowl. Data Eng.* doi: 10.1109/TKDE.2019.2912367

Darbyshire, P., MacDougall, C., and Schiller, W. (2005). Multiple methods in qualitative research with children: more insight or just more? *Qual. Res.* 5, 417–436. doi: 10.1177/1468794105056921

Davis, B. E. (2010). Parental wartime deployment and the use of mental health services among young military children. *Pediatrics* 126, 1215–1216. doi: 10.1542/peds.2010-2543

Dawe, J., Sutherland, C., Barco, A., and Broadbent, E. (2019). Can social robots help children in healthcare contexts? A scoping review. *BMJ Paediatr. Open* 3:e000371. doi: 10.1136/bmjpo-2018-000371

Druin, A. (1999). *The Design of Children's Technology*. San Francisco, CA: Morgan Kaufmann Publishers.

Druin, A. (2002). The role of children in the design of new technology. *Behav. Inform. Technol.* 21, 1–25. doi: 10.1080/01449290110108659

Ellis, D. A., Podolski, C.-L., Frey, M., Naar-King, S., Wang, B., and Moltz, K. (2007). The role of parental monitoring in adolescent health outcomes: impact on regimen adherence in youth with type 1 diabetes. *J. Pediatr. Psychol.* 32, 907–917. doi: 10.1093/jpepsy/jsm009

Epstein, S. L. (2015). Wanted: collaborative intelligence. *Artif. Intell.* 221, 36–45. doi: 10.1016/j.artint.2014.12.006

Forgas, J. P. (1995). Mood and judgements: the affect infusion model (AIM). *Psychol. Bull.* 117, 39–66. doi: 10.1037//0033-2909.117.1.39

Gordon, G., Spaulding, S., Westlund, J. K., Lee, J. J., Plummer, L., Martinez, M., et al. (2016). "Affective personalization of a social robot tutor for children's second language skills," in *Thirtieth AAAI Conference on Artificial Intelligence* (Palo Alto, CA: AAAI Press), 3951–3957.

Grasha, A. F. (1994). A matter of style: the teacher as expert, formal authority, personal model, facilitator, and delegator. *College Teach.* 42, 142–149. doi: 10.1080/87567555.1994.9926845

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowl. Acquis.* 5, 199–220. doi: 10.1006/knac.1993.1008

Haanstra, H., and de Boer, M. H. (2019). Opinion aspect extraction in Dutch childrens diary entries. *arXiv:1910.10502 [Preprint]*.

Han, S., Lee, K., Lee, D., and Lee, G. G. (2013). "Counseling dialog system with 5w1h extraction," in *Proceedings of the SIGDIAL 2013 Conference* (Metz), 349–353.

Hollnagel, E., and Woods, D. D. (2005). *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*. Boca Raton, FL: CRC Press.

Iannotti, R. J., Schneider, S., Nansel, T. R., Haynie, D. L., Plotnick, L. P., Clark, L. M., et al. (2006). Self-efficacy, outcome expectations, and diabetes self-management in adolescents with type 1 diabetes. *J. Dev. Behav. Pediatr.* 27, 98–105. doi: 10.1097/00004703-200604000-00003

Jin, M., An, Q., and Wang, L. (2017). Chronic conditions in adolescents. *Exp. Ther. Med.* 14, 478–482. doi: 10.3892/etm.2017.4526

Johnson, M., and Vera, A. (2019). No AI is an island: the case for teaming intelligence. *AI Magaz.* 40, 16–28. doi: 10.1609/aimag.v40i1.2842

Jones, A., and Castellano, G. (2018). Adaptive robotic tutors that support self-regulated learning: a longer-term investigation with primary school children. *Int. J. Soc. Robot.* 10, 357–370. doi: 10.1007/s12369-017-0458-z

Kaptein, F., Broekens, J., Hindriks, K., and Neerincx, M. (2017). "Personalised self-explanation by robots: the role of goals versus beliefs in robot-action explanation for children and adults," in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (IEEE Xplore Digital Library), 676–682.

Kaptein, F., Broekens, J., Hindriks, K. V., and Neerincx, M. (2016). "CAAF: a cognitive affective agent programming framework," in *International Conference on Intelligent Virtual Agents* (New York, NY: Springer), 317–330.

Kayal, A., Brinkman, W.-P., Neerincx, M. A., and Riemsdijk, M. (2018a). Automatic resolution of normative conflicts in supportive technology based on user values. *ACM Trans. Intern. Technol.* 18, 41–62. doi: 10.1145/3158371

Kayal, A., van Riemsdijk, M. B., Neerincx, M. A., and Brinkman, W.-P. (2018b). Socially adaptive electronic partners for improved support of children's values: an empirical study with a location-sharing mobile app. *Int. J. Child Comput. Interact.* 18, 79–89. doi: 10.1016/j.ijcci.2018.09.001

Kehoe, B., Patil, S., Abbeel, P., and Goldberg, K. (2015). A survey of research on cloud robotics and automation. *IEEE Trans. Autom. Sci. Eng.* 12, 398–409. doi: 10.1109/TASE.2014.2376492

Klopfenstein, L. C., Delpriori, S., Malatini, S., and Bogliolo, A. (2017). "The rise of bots: a survey of conversational interfaces, patterns, and paradigms," in *Proceedings of the 2017 Conference on Designing Interactive Systems* (New York, NY: ACM), 555–565.

Leary, T. (1958). Interpersonal diagnosis of personality. *Am. J. Phys. Med. Rehabil.* 37:331.

Legault, L. (2017). "Self-determination theory," in *Encyclopedia of Personality and Individual Differences*, eds V. Zeigler-Hill and T. K. Shackelfod, 1–9. doi: 10.1007/978-3-319-28099-8_1162-1

Lim, G. H., Hong, S. W., Lee, I., Suh, I. H., and Beetz, M. (2013). "Robot recommender system using affection-based episode ontology for personalization," in *2013 IEEE RO-MAN* (IEEE Xplore Digital Library), 155–160.

Lim, G. H., Suh, I. H., and Suh, H. (2011). Ontology-based unified robot knowledge for service robots in indoor environments. *IEEE Trans. Syst. Man Cybern. A Syst. Hum.* 41, 492–509. doi: 10.1109/TSMCA.2010.2076404

Looije, R., Neerincx, M. A., and Hindriks, K. V. (2017). Specifying and testing the design rationale of social robots for behavior change in children. *Cogn. Syst. Res.* 43, 250–265. doi: 10.1016/j.cogsys.2016.07.002

Looije, R., Neerincx, M. A., Peters, J. K., and Blanson Henkemans, O. A. (2016). Integrating robot support functions into varied activities at returning hospital visits. *Int. J. Soc. Robot.* 8, 483–497. doi: 10.1007/s12369-016-0365-8

Mioch, T., Peeters, M. M., and Nccrincx, M. A. (2018). "Improving adaptive human-robot cooperation through work agreements," in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (IEEE Xplore Digital Library), 1105–1110.

Moerman, C. J., van der Heide, L., and Heerink, M. (2018). Social robots to support children's well-being under medical treatment: a systematic state-of-the-art review. *J. Child Health Care*. doi: 10.1177/1367493518803031. [Epub ahead of print].

Montenegro, J. L. Z., da Costa, C. A., and da Rosa Righi, R. (2019). Survey of conversational agents in health. *Expert Syst. Appl.* 129, 56–67. doi: 10.1016/j.eswa.2019.03.054

Naikar, N. (2017). Cognitive work analysis: an influential legacy extending beyond human factors and engineering. *Appl. Ergon.* 59, 528–540. doi: 10.1016/j.apergo.2016.06.001

Neerincx, M. A. (2011). Situated cognitive engineering for crew support in space. *Pers. Ubiq. Comput.* 15, 445–456. doi: 10.1007/s00779-010-0319-3

Neerincx, M. A., Kaptein, F., van Bekkum, M., Krieger, H., Kiefer, B., Peters, R., et al. (2016a). "Ontologies for social, cognitive and affective agent-based support of child's diabetes self management," in *ECAI 2016, 1st ECAI Workshop on Artificial Intelligence for Diabetes at the 22nd European Conference on Artificial Intelligence* (IOS Press Ebooks), 35–38.

Neerincx, M. A., and Lindenberg, J. (2008). "Situated cognitive engineering for complex task environments," in *Naturalistic Decision Making and Macrocognition,* eds J. M. C. Schraagen, L. Militello, T. Ormerod, and R. Lipshitz (Aldershot, UK: Ashgate Publishing, Ltd.), 373–390.

Neerincx, M. A., van der Waa, J., Kaptein, F., and van Diggelen, J. (2018). "Using perceptual and cognitive explanations for enhanced human-agent team performance," in *International Conference on Engineering Psychology and Cognitive Ergonomics* (New York, NY: Springer), 204–214.

Neerincx, M. A., van Diggelen, J., and van Breda, L. (2016b). "Interaction design patterns for adaptive human-agent-robot teamwork in high-risk domains," in *International Conference on Engineering Psychology and Cognitive Ergonomics* (New York, NY: Springer), 211–220.

Norman, D. A. (1986). "Cognitive engineering," in *User Centered System Design: New Perspectives on Human-Computer Interaction*, eds D. A. Norman and S. W. Draper (Hillsdale, NJ: Erlbaum Associates Inc.), 31–64.

Pai, A. L., and Ostendorf, H. M. (2011). Treatment adherence in adolescents and young adults affected by chronic illness during the health care transition from pediatric to adult health care: a literature review. *Childrens Health Care* 40, 16–33.

Peters, R., Broekens, J., Li, K., and Neerincx, M. A. (2019a). "Robots expressing dominance: Effects of behaviours and modulation," in *Proceedings of the 8th International Conference on Affective Computing & Intelligent Interaction* (IEEE Xplore Digital Library).

Peters, R., Broekens, J., and Neerincx, M. A. (2015). "Let me guide you! pedagogical interaction style for a robot in children's education," in *Proceedings of the First International Workshop on Educational Robots - Satellite event of the 7th International Conference on Social Robotics* (IEEE Xplore Digital Library).

Peters, R., Broekens, J., and Neerincx, M. A. (2017a). "Guidelines for tree-based learning goal structuring," in *Proceedings of the 22nd International Conference on Intelligent User Interfaces (IUI117)* (New York, NY: ACM), 401–405.

Peters, R., Broekens, J., and Neerincx, M. A. (2017b). "Robots educate in style: the effect of context and non-verbal behaviour on children's perceptions of warmth and competence," in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (IEEE Xplore Digital Library), 449–455.

Peters, R., Oleari, E., Sardu, F., and Neerincx, M. A. (2019b). "Usability of the PAL Objectives dashboard for children's diabetes self-management education," in *Proceedings of the 2019 the 5th International Conference on e-Society, e-Learning and e-Technologies* (New York, NY: ACM), 22–28.

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., et al. (2019). Machine behaviour. *Nature* 568:477. doi: 10.1038/s41586-019-1138-y

Rasmussen, J., Pejtersen, A. M., and Goodstein, L. P. (1994). *Cognitive systems engineering.* New York, NY: Wiley.

Riek, L. D. (2017). Healthcare robotics. *Commun. ACM* 60, 68–78. doi: 10.1145/3127874

Robinson, N. L., Cottier, T. V., and Kavanagh, D. J. (2019). Psychosocial health interventions by social robots: systematic review of randomized controlled trials. *J. Med. Intern. Res.* 21:e13203. doi: 10.2196/13203

Rotenberg, K. J., and Chase, N. (1992). Development of the reciprocity of self-disclosure. *J. Genet. Psychol.* 153, 75–86. doi: 10.1080/00221325.1992.10753703

Ryan, R. M., and Deci, E. L. (2017). *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*. New York, NY: Guilford Publications.

Schachter, S., and Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychol. Rev*. 69:379. doi: 10.1037/h0 046234

Schunk, D. H., and Lilly, M. W. (1984). Sex differences in self-efficacy and attributions: influence of performance feedback. *J. Early Adolesc.* 4, 203–213. doi: 10.1177/0272431684043004

Sharples, M., Jeffery, N., Du Boulay, J., Teather, D., Teather, B., and Du Boulay, G. (2002). Socio-cognitive engineering: a methodology for the design of human-centred technology. *Eur. J. Operat. Res.* 136, 310–323. doi: 10.1016/S0377-2217(01)00118-7

Sinoo, C., van der Pal, S., Henkemans, O. A. B., Keizer, A., Bierman, B. P., Looije, R., et al. (2018). Friendship with a robot: children's perception of similarity between a robot's physical and virtual embodiment that supports diabetes self-management. *Patient Educ. Counsel.* 101, 1248–1255. doi: 10.1016/j.pec.2018.02.008

Tudge, J. R., Winterhoff, P. A., and Hogan, D. M. (1996). The cognitive consequences of collaborative problem solving with and without feedback. *Child Dev*. 67, 2892–2909. doi: 10.1111/j.1467-8624.1996.tb01894.x

van Bekkum, M. A., Krieger, H.-U., Neerincx, M. A., Kaptein, F., Kiefer, B., Peters, R., et al. (2016). "Ontology engineering for the design and implementation of personal pervasive lifestyle support," in *SEMANTiCS (Posters, Demos, SuCCESS), CEUR Workshop Proceedings, 1613-0073, 1695,* eds M. Martin, M. Cuquet, and E. Folmer.

Van Diggelen, J., Neerincx, M., Peeters, M., and Schraagen, J. M. (2018). Developing effective and resilient human-agent teamwork using team design patterns. *IEEE intell. Syst.* 34, 15–24. doi: 10.1109/MIS.2018.2 886671

Van Welie, M., Van der Veer, G. C., and Eliëns, A. (1998). "An ontology for task models," in *Design, Specification and Verification of Interactive Systems 98* (Springer), 81–94.

Vicente, K. J. (1999). *Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work*. London; New Jersey, NJ: CRC Press.

Vygotsky, L. S. (1980). *Mind in Society: The Development of Higher Psychological Processes*. Vienna: Harvard University Press.

Westlund, J. M. K., Park, H. W., Williams, R., and Breazeal, C. (2018). "Measuring young children's long-term relationships with social robots," in *Proceedings of the 17th ACM Conference on Interaction Design and Children* (New York, NY: ACM), 207–218.

Woods, D. D., and Roth, E. M. (1988). Cognitive engineering: human problem solving with tools. *Hum. Factors* 30, 415–430. doi: 10.1177/001872088803000404

Wu, Y., Yao, X., Vespasiani, G., Nicolucci, A., Dong, Y., Kwong, J., et al. (2017). Mobile app-based interventions to support diabetes self-management: a systematic review of randomized controlled trials to identify functions associated with glycemic efficacy. *JMIR mHealth uHealth* 5:e35. doi: 10.2196/mhealth.6522

Wubbels, T., Brekelmans, M., Brok, P. D., Levy, J., Mainhard, T., and Tartwijk, J. V. (eds). (2012). "Chapter 14: Let's Make Things Better - Developments in Research on Interpersonal Relationships in Education," in *Interpersonal Relationships in Education* (Rotterdam: Sense Publishing), 225–249.

Conflict of Interest: MN, WV, and OB were employed by the non-profit organization TNO. BK was employed by the non-profit organization DFKI. DF was employed by company Mixel, Italy. BB was employed by company Produxi, The Netherlands.

The remaining 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.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 24 January 2020
doi: 10.3389/frobt.2020.00001

Designing Ethical Social Robots—A Longitudinal Field Study With Older Adults

Anouk van Maris[1]*, *Nancy Zook*[2], *Praminda Caleb-Solly*[1], *Matthew Studley*[1], *Alan Winfield*[1] and *Sanja Dogramadzi*[1,3]

[1] Bristol Robotics Laboratory, University of the West of England, Bristol, United Kingdom, [2] Department of Health and Social Sciences, University of the West of England, Bristol, United Kingdom, [3] Department of Automatic Control and Systems Engineering, The University of Sheffield, Sheffield, United Kingdom

Emotional deception and emotional attachment are regarded as ethical concerns in human-robot interaction. Considering these concerns is essential, particularly as little is known about longitudinal effects of interactions with social robots. We ran a longitudinal user study with older adults in two retirement villages, where people interacted with a robot in a didactic setting for eight sessions over a period of 4 weeks. The robot would show either non-emotive or emotive behavior during these interactions in order to investigate emotional deception. Questionnaires were given to investigate participants' acceptance of the robot, perception of the social interactions with the robot and attachment to the robot. Results show that the robot's behavior did not seem to influence participants' acceptance of the robot, perception of the interaction or attachment to the robot. Time did not appear to influence participants' level of attachment to the robot, which ranged from low to medium. The perceived ease of using the robot significantly increased over time. These findings indicate that a robot showing emotions—and perhaps resulting in users being deceived—in a didactic setting may not by default negatively influence participants' acceptance and perception of the robot, and that older adults may not become distressed if the robot would break or be taken away from them, as attachment to the robot in this didactic setting was not high. However, more research is required as there may be other factors influencing these ethical concerns, and support through other measurements than questionnaires is required to be able to draw conclusions regarding these concerns.

Keywords: social robots, older adults, longitudinal study, ethics, deception, attachment

Edited by:
Chien-Ming Huang, Johns Hopkins University, United States

Reviewed by:
Alan R. Wagner, Pennsylvania State University (PSU), United States
Elin A. Björling, University of Washington, United States

***Correspondence:**
Anouk van Maris
anouk.vanmaris@uwe.ac.uk

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *15 July 2019*
Accepted: *06 January 2020*
Published: *24 January 2020*

Citation:
van Maris A, Zook N, Caleb-Solly P, Studley M, Winfield A and Dogramadzi S (2020) Designing Ethical Social Robots—A Longitudinal Field Study With Older Adults. Front. Robot. AI 7:1. doi: 10.3389/frobt.2020.00001

1. INTRODUCTION

Awareness of, and a growing interest in, ethical considerations for the development of social robots is increasing due to the predicted increasing likelihood of robots being a part of our everyday lives in the future (Malle et al., 2015; Esposito et al., 2016; Li et al., 2019). This is evident through the emergence of relatively new conferences like the International Conference on Robot Ethics and Standards[1], and new ethical standards in robotics and AI (Winfield, 2019). Socially assistive robots can provide psycho-social, physical and/or cognitive support while interacting with their users (Robinson et al., 2014). Therefore, potential ethical concerns of prolonged use of social assistive

[1] https://www.icres2019.org/

robots needs to be considered while these systems are still being developed. This will help to ensure that appropriate safeguards are considered and built into systems as an integral part of their design. In addition, this will facilitate clear guidelines and regulations for safe deployment. One area for investigation that has been identified is the use of emotional expression in the robot, which can lead to emotional deception (Sharkey and Sharkey, 2012). Emotional deception could occur when the user believes that the robot really experiences these emotions, leading to unrealistic expectations that can possibly result in the user prioritizing the robot's well-being over other people's or their own well-being, as well as over-relying on the robot as a social assistant without exerting one's own critical judgment (Fulmer et al., 2009). Another ethical concern is the possible development of emotional attachment to the robot (Sullins, 2012), which may cause distress in the user when the robot breaks or is taken away. Whilst these issues are important to consider in all human-robot interactions, the current study focuses on self-reported healthy older adults. This group was selected due to the emergence of social robots as a way to support caregivers and care homes as they meet a growing demand for care for the aging population (Unies, 2015). Safe and responsible introduction of social robots to this target group is essential, as a potential lack of knowledge of and experience with new technologies may lead to situations that potentially affect psychological and/or physical safety (Borenstein et al., 2017). Moreover, the step between utilizing robots for cognitively healthy older adults to vulnerable older adults that suffer from e.g., dementia is small, and baseline requirements found through studies with healthy older adults are essential to ensure that it is ethically safe for vulnerable older adults to interact with the robot.

Frennert and Östlund described several matters of concern that arise during the development of social robots for older adults (Frennert and Östlund, 2014). Some of these entail the role that the robot will play in the older adults' lives, factors that can influence social robot acceptance, methodology used in robotic research and ethical implications. These matters were addressed in this study. A specific role for the robot was determined and communicated to the participants, namely that of a didactic learning companion. Factors that may influence acceptance were investigated, and whether these could have ethical consequences. We addressed methodology concerns and even though the number of participants is still low, we did use a comparison group (Bemelmans et al., 2012) and ran the study in a naturalistic setting. As the goal is to ensure that social interactions are ethically safe and acceptable, the concern that social interactions are driven by technological determinism has been addressed as well.

The aims of this study are to establish whether the ethical concerns of emotional deception and emotional attachment that have been established in the literature are reflected in practice. More specifically, this study investigates whether older adults are emotionally deceived by a robot when it shows emotional expressions during didactic interaction in a naturalistic setting, and whether they will become emotionally attached to the robot over time. Suggestions for how the social robot could be adapted to address ethical and acceptability concerns are considered. As no similar study has been previously conducted in the literature, this work could provide useful insights into conducting longitudinal field studies and lay the foundation for future work with vulnerable populations of older adults, such as those with dementia.

The following hypotheses are investigated in this study: It is expected that effects of emotional deception will be minimal, as the level of deception was designed to be low. Furthermore, any effects that occur will decrease over time, once participants become familiar with the displayed emotions of the robot. Additionally, it is hypothesized that emotional attachment will initially be low but increase over time, and that attachment will be higher for participants interacting with the emotive robot, as the display of emotions by the robot will increase people's perception of the robot being a social entity.

2. SOCIAL ROBOTS AND HUMAN-ROBOT INTERACTION

One issue relating to human-robot interaction and social robotics relates to the lack of a common definition for a social robot. For example, Dautenhahn and Billard state that a social robot is an embodied agent that is part of a society of robots and/or humans (Dautenhahn and Billard, 1999). This statement is followed by the notion that these agents can recognize one another, join a social interaction, and learn from each other. An alternative definition is provided by Fong, who describe a social robot as an agent for which interaction is important (Fong et al., 2003). This lack of consensus regarding a definition presents challenges when developing a framework for the investigation of social robots, as it is difficult to ensure all parties involved envisage the same outcome without a common definition to refer to. Combining research by Breazeal and Fong, we can distinguish seven different classes of social robots: socially evocative, social interface, socially receptive, sociable, socially situated, socially embedded and socially intelligent (Fong et al., 2003; Breazeal, 2004). In this research, we will focus on robots from the first two categories, socially evocative and social interface, since these require little social cognition and will be easier to use in real-world settings in less time, but will, due to the small amount of social cognition required, raise several possible ethical concerns.

2.1. Ethical Concerns in HRI

The population of older adults is growing, and the demand for care is growing with them (Unies, 2015). However, the capacity to supply this call for care is limited. This is one of the reasons why research in robotics is so attractive (Sparrow and Sparrow, 2006). Use of social robots by older adults will require ongoing care and health education. The fact that social robots are not designed to be influenced by an emotional state, nor judge people (Breazeal, 2011) might make them less stigmatizing to use in this context (Breazeal, 2011).

Several ethical implications of using robots for older adults have been established in the literature. Example implications are reduced human contact, loss of control, loss of personal liberty, loss of privacy, matters regarding responsibility, infantilization, emotional deception and emotional attachment (Sharkey and Sharkey, 2012; Sullins, 2012; Kolling et al., 2013). Even though these are all valid concerns, there are counterarguments against several of these as well. Some of these counterarguments have been raised by Sharkey and Sharkey themselves (Sharkey and Sharkey, 2012). For example, the use of social robots may reduce people's contact with others, but it can also reduce isolation and increase conversation opportunities both with the robot and other users (Sharkey and Sharkey, 2012). Other ethical issues are loss of control and loss of personal liberty, but robots can also give older people the opportunity to self-manage their well-being and the ability to reduce risks (Callén et al., 2009). Privacy issues are equally important, but do not apply solely to social robotics and are being investigated for many other technologies deployed in human environment and, therefore, will not be discussed in this paper, as well as matters regarding responsibility, which are being researched through e.g., autonomous cars. Encouraging people to interact with robots, that can sometimes have a toy like appearance, might give them the feeling they are being infantilized. However, this can be addressed by taking into account the aesthetics of the robot and including the older adults in the development process, which have been identified as a matters of concern when developing robots for older adults (Frennert and Östlund, 2014). This leaves two ethical concerns: emotional deception and emotional attachment. These two concerns were investigated in more depth in this study.

2.1.1. Emotional Deception

Deception occurs when false information is communicated to benefit the communicator (Arkin et al., 2012); it implies that an agent acts in a way that it induces a false belief in another agent (Hyman, 1989). This can also mean that no information is communicated at all (Dragan et al., 2015). Deception can be approached through different perspectives like philosophy, economics, and biology (Shim and Arkin, 2013). Both the perspective of philosophy and biology discuss the division of deception into either unintentional or intentional (Dragan et al., 2015). Unintentional deception takes place when some feature of the (unintentional) deceiver evokes unforeseen expectations in the agent being deceived. Intentional deception takes place when the deceiver is aware that these features will raise false expectations in the agent being deceived. This distinction suggests that emotional deception is not a binary materialization, but a spectrum with different gradations (Winkle and Van Maris, 2019).

Deception is generally perceived as bad. However, this is not necessarily the case. An example where deception benefits the agent being deceived is the use of placebo-effect. This beneficial form of deception is called benevolent deception (Adar et al., 2013). Benevolent deception has always been part of medical care (Jackson, 1991), and may even be required to act morally (Arkin et al., 2012).

Deception is created when robots are used in assistive settings (Sharkey and Sharkey, 2011), since the robot's social behavior often does not correspond with its actual capabilities. This could be a risk, since users may perceive robots differently than intended and raise expectations that cannot be met by the robot. For a robot to be able to successfully perform a deceptive action, it requires specific knowledge about the person that it intends to deceive (Wagner and Arkin, 2011). Also, the robot should convey its intentions and have a theory of mind for the person being deceived to be able to manipulate their beliefs (Dragan et al., 2015). However, an area that is often not discussed is the fact that a robot can also unintentionally deceive through its appearance and/or behavior.

When objects, in this case the robot, provide social support, interactions may be more effective (Kidd et al., 2006). More effective interactions will lead to a better interaction quality and improvement of quality of life, which is the main reason why robots are used in care for older adults. Providing a robot with emotive behavior is a way to improve its capability to communicate with a person (Kirby et al., 2010). People are capable of recognizing facial expressions in robots (Kirby et al., 2010), and perceive the emotion they recognize themselves as well.

Displays of robot-human affection would be an appearance of affection from the robot toward the human, as real affection requires emotions, which are difficult to implement in robots (Weijers, 2013). Being able to convey emotions is a requirement for a successful companion robot (Breazeal and Scassellati, 1999). However, one might argue this is a form of deception, as the robot does not actually experience emotions. Especially, as emotional deception is stated as the misrepresentation of one's emotional state (Fulmer et al., 2009), and the robot provides incorrect information about its internal state when displaying emotive behavior. As the perception of emotional deception is a subjective response it is measured indirectly through other variables in this study. Older people may benefit from emotive robot behavior, as receiving little affection can have negative consequences for people that are feeling lonely like cardiovascular function (Cacioppo and Patrick, 2008). Whether emotional deception by a robot is benevolent and thus ethically acceptable, especially when interacting with vulnerable users, has to be researched more thoroughly, which was one of the aims of this study. The results provided insights in effects of emotional deception in a didactic setting, which have not been investigated yet.

2.1.2. Emotional Attachment

Attachment can be described as the sum of cohesion episodes that a person has made with other persons or objects (Huber et al., 2016). A cohesion episode entails joint experiences with these other people or objects, in which cohesion factors are present. Cohesion factors can be defined as shared factors like values and preferences, charisma factors like attractiveness and sympathies, personal factors like expressed openness and social factors like non-situation-specific reciprocity).

Research on attachment and robots has either focused on robots showing attachment to the user or eliciting attachment in the user through its behavior (e.g., Hiolle et al., 2009, 2012). It is possible to become attached to a robot, as people are capable of becoming attached to objects (Keefer et al., 2012). Scheutz has highlighted that there is very little needed for people to become attached to robots, even if these robots do not show behavior that elicits attachment (Scheutz, 2011). Therefore, it is particularly important to explore attachment in socially assistive robots interacting with older adults. There are four attachment styles that distinguish how easily a person becomes attached to someone or something: secure attachment, fearful attachment, preoccupied attachment and dismissive attachment (Brennan et al., 1998). Since social robots (and other assistive technologies) become more advanced, the likeliness of users forming attachment-like bonds to them increases (Collins et al., 2013). Opinions on whether it is acceptable for a robot to elicit attachment in its users are divided. On the one hand, eliciting attachment will support the process and goals of its use (Coeckelbergh et al., 2016). Some even say that eliciting attachment in its users is a necessity for the robot to be fully effective in a care providing context (Birnbaum et al., 2016). However, once users have become attached to the robot, taking it away may cause emotional distress (Sharkey and Sharkey, 2010; Coeckelbergh et al., 2016).

Emotional deception and emotional attachment have been raised as ethical concerns in the literature. However, this has never been investigated in practice, which was one aim of this study. Whether these concerns are reflected in practice was investigated through a longitudinal human-robot interaction study, where people's acceptance of the robot, perception of the social interaction and attachment to the robot were measured over time. The robot's behavior was manipulated to investigate emotional deception. This study investigated how emotional deception and emotional attachment may relate to acceptance of the robot and perception of the social interaction, as these will be indicators for the future development of ethically safe socially assistive robots.

3. METHODS

3.1. Overview

The aim of this study was to investigate emotional deception and emotional attachment. Participants' responses to a robot displaying either emotive or non-emotive behavior, and their level of attachment to the robot over time were investigated. Questionnaires were administered several times during the experiment to study participants' attachment to the robot, their acceptance of the robot, and their perception of the social interactions with the robot. Most participants interacted with the emotive and the non-emotive robot, except for a small control group that interacted with the non-emotive robot only. Other data regarding the participants' affect, physiological state and behavior were also measured using sensors and video recordings, however this paper only presents the qualitative aspects of the study from the participants' perspective.

3.2. Participants

In total 17 older adults participated in this experiment. Participants were recruited from two retirement villages where residents have their own apartments and live independently; however if they need support they can call the village manager for assistance. Participants were offered a gift card to compensate them for their time. Ten participants were recruited through one retirement village and seven from a second retirement village. As this study was directed toward typical aging, prior to scheduling sessions, participants were asked to self-report health issues/diagnoses (i.e., dementia, etc.) that could affect their ability to complete measures or limit their capacity to consent. No participants were excluded based on this criteria. In addition, participants from one retirement village had their capacity for informed consent monitored by a locksmith (an individual who monitors residents). As part of the procedure, participants of the retirement village that did not have a locksmith available were administered using the Montreal Cognitive Assessment test (MOCA; Nasreddine et al., 2005) for overall cognitive function. Based on these scores, data from two participants was excluded as they scored below 15 (out of 30) where all other participants scored between 26 and 30. One participant completed four sessions but was unable to complete the study. As such, data from only 14 participants was included in the analyses. The ages of the participants that completed the experiment (9 male, 5 female) ranged from 61 to 90 years old ($M = 76.29$, $SD = 8.50$). Twelve participants reported being generally unfamiliar with social robots and two participants (both male) indicated that they were somewhat familiar with them [$M = 1.35$, $SD = 0.74$ on a scale from 1 (unfamiliar) to 5 (familiar)]. Participants reported being familiar with technological devices (e.g., smart phones, tablets, laptops, desktops) and using them on a daily ($N = 13$) or weekly ($N = 1$) basis. Participant characteristics can be found in **Table 1**. This table also provides attachment style and level of attachment, which will be discussed in the next section.

TABLE 1 | Case characteristics of the user trials.

Participant	Group	Gender	Age	Familiarity	Attachment style	Level of attachment
1	Test	M	74	Somewhat	Secure	Low
2	Test	M	72	Low	Secure	Medium
3	Test	F	72	Low	Fearful	Low
4	Test	F	77	Low	Dismissive	Medium
5	Test	M	82	Low	Dismissive	Medium
6	Test	M	72	Low	Fearful	High
7	Test	M	61	Low	Secure	Medium
8	Test	F	76	Low	Dismissive	Low
9	Test	M	90	Low	Fearful	Medium
10	Test	F	85	Low	Secure	High
11	Control	M	68	Somewhat	Fearful	Medium
12	Control	M	90	Low	Secure	Low
13	Control	M	68	Low	Secure	Low
14	Control	F	81	Low	Dismissive	Medium

3.3. Materials

The robot used for this study was a Pepper robot, developed by Soft Bank Robotics[2]. The software "Choregraph," provided by Soft Bank, was used to create the robot behaviors and run the experiments.

3.3.1. Questionnaires

The order in which the questionnaires were administered and all items in each questionnaire were randomized. Several existing questionnaires were used. Some of these use a five-point scale, where others use a seven-point scale. For consistency and to make it easier for our participants, it was decided to use a five-point scale for all questionnaires.

- **Demographics:** Age, gender and level of education were collected. Interestingly, several participants did not provide their level of education and gave insufficient answers like "not high" and therefore, this question was not used for data analysis.
- **Montreal Cognitive Assessment (MOCA):** This brief cognitive assessment measures performance in executive functioning, memory, language, attention and visuo-spatial perceptual skills.
- **Acceptance of the robot:** Several constructs of the Almere model of technology acceptance (Heerink et al., 2010) were used to determine participants' acceptance of the robot and whether this changed over time. Used constructs were anxiety to use the robot, attitude toward the robot, perceived enjoyment, perceived ease of use, perceived sociability, perceived usefulness, social influence, social presence and trust.
- **Perception of the social interaction with the robot:** Most constructs of the Godspeed questionnaire (Bartneck et al., 2009) were used to determine participants' perception of the robot, and whether it changed over time. Items used included anthropomorphism, likability of the robot, perceived intelligence of the robot and perceived safety during the interactions.
- **Attachment to the robot:** Unlike the acceptance and perception questionnaires, there is no existing questionnaire for attachment in HRI that has been established in previous work. Therefore, a questionnaire for object attachment (Schifferstein and Zwartkruis-Pelgrim, 2008) was adapted to fit this study. This consisted of nine statements, and the average of these statements was used to get an overall number for attachment.
- **Attachment style:** In order to assess attachment types, participants were asked to fill in an adapted version of the Experiences in Close Relationships Inventory to determine their attachment style (Brennan et al., 1998). Statements involving "(romantic) partners" were adapted to a more general variation with "people that are dear to me."
- **Debrief interview:** Several questions were asked after participants were debriefed to gather their opinion on the ethical concerns. The questions asked in this interview are: "Will you miss Pepper?," "Do you think Pepper had an influence on your mood?," "Do you think Pepper was emotionally deceptive and if yes, do you think this was acceptable?," "Do you think you would get bored of Pepper, if you could use it whenever you want?" and "What role would you like Pepper to play in your life?" Note that for the question regarding emotional deception participants were first given the definition of emotional deception used in this research: a robot is deceiving its user when it displays emotions, which may result in the users building an incorrect mental model of the robot's abilities.

As emotional deception is said to occur when an agent falsely displays feelings of emotions (Fulmer et al., 2009) and therefore is a subjective response, it can best be measured indirectly through other variables. In this study, emotional deception was investigated by looking at participants' acceptance of the robot and perception of the social interaction. For example, if the perceived intelligence or perception of the robot as a social entity are higher for the emotive robot, this might indicate that participants are deceived by its behavior.

3.4. Study Procedure

Ethical approval was obtained from the University of the West of England ethics committee prior to recruitment. Informed consent was gathered for all participants before any data was collected. Participants interacted with the robot for eight sessions: two interactions per week for 4 weeks. During these interactions, the robot informed participants about the Seven Wonders of the Modern World and the Seven Wonders of the Ancient World. Interactions lasted between 5 and 8 min. Ten out of the fourteen participants that completed the experiment interacted with the non-emotive robot during the first four interactions, and the emotive robot during the last four interactions, or vice versa. The order of robot behaviors was counterbalanced between participants. The remaining four participants interacted with the non-emotive robot during all eight sessions as control group. Besides these eight interactions with the robot, there was one introductory session before the first interaction and one debrief session after the last interaction with the experimenter only. The robot was present during the introductory session and would introduce itself briefly so participants would get a first impression of the robot's voice and behavior, so they felt more familiar with it once the interactions started. The robot was not present in the room during the debrief session.

An example part of an interaction between the robot and a participant about the Statue of Zeus at Olympia; R = robot, P = Participant:

- **R:** "As mentioned before the statue depicts Zeus sitting on a wooden throne. However, did you know that the whole statue and not only the throne was made of wood?"
  - *If **P** said "no"*: **R:** "Yes, the whole statue was sculpted in wood. After that, Zeus was covered with ivory and gold plates."

[2]https://www.softbankrobotics.com/emea/en/pepper

  - *If **P** said "yes"*: **R:** "Indeed, the whole statue was sculpted in wood. After that, Zeus was covered with ivory and gold plates."

- **R:**"Have you ever been to Olympia, or other places in Greece?"
  - *If **P** said "no"*: **R:** "Now let us continue with..."
  - *If **P** said "yes"*: **R:** "Would you like to tell me about it?"
    - *If **P** says "no"*: **R:** "Ok, now let us continue with...."
    - *If **P** talks about positive experience:* **R:** "That sounds nice. Now let us continue with...."
    - *If **P** talks about negative experience:* **R:** "Sorry to hear that. Now let us continue with...."

The protocol for a participant who would become upset after mentioning a negative experience was for the robot to not reply to the experience at all. However, no participants became upset during the experiment.

Participants were seated opposite the robot. The distance between the chair and the robot was approximately 1.5 m, which falls within the social space of Hall's proxemics categories (Hall et al., 1968), but approaches the personal zone as well, as the threshold between these two zones is at 1.2 m. The social space represents the distance between two strangers having a conversation, where the personal space represents the distance where two friends have a conversation. **Figures 1**, **2** show the experiment room for the two retirement villages.

The emotive and non-emotive behaviors displayed by the robot have been established in earlier research (Van Maris et al., 2018). In the non-emotive condition, the robot would show "neutral" behavior. In the emotive condition, the robot would show context-appropriate emotions. Differences in emotive and non-emotive behavior were identified by a different pitch in voice (higher pitch when happy, lower pitch when sad), different talking speed (faster when happy, slower when sad), changing head position (chin up when happy, chin down when sad) and arm movements (larger movement when happy, smaller movement when sad). These factors were based on existing literature (Kwon et al., 2007; Beck et al., 2013). The emotive behavior and possible emotional deception were designed to be low in this study. As mentioned earlier, emotional deception can be both intentional and unintentional, and the goal of this research is to investigate unintentional deception. This occurs when emotions are displayed to create a more pleasant interaction experience for the user and not to elicit certain reactions from them. The emotive behaviors in this study may have some influence on people's perception of the robot, but the deception will be much lower compared to when the emotive behaviors elicit reactions from the users. Therefore, the emotional deception in this study is intended to be low.

The Wizard of Oz strategy was used for this experiment. Interactions were pre-programmed, but the experimenter would manually prompt the robot to continue with the interaction to ensure the robot would continue at the appropriate times. This was necessary as speech recognition is not optimal yet, and the need for participants to focus on speaking loudly and clearly could have distracted them from the robot's displayed behavior. In one retirement village, the experimenter was located in an adjacent room with all doors open. In the other retirement village, only one experiment room was available so the experimenter was located behind the participants to be out of sight during the interactions. Participants received a photograph of themselves with the robot (taken after the final interaction with the robot), and a £20 gift card for their participation during the debrief session. Contact details of the experimenter were provided in case they would want to see the robot again. This exit strategy was essential, as it would have been unethical to investigate people's emotional attachment to the robot and not provide them with the possibility to see it again if they wanted to.

When each questionnaire was administered is shown in **Table 2**. The questionnaires given at each session varied as participants could not comment on their attachment and acceptance of the robot prior to interactions.

A more detailed explanation for each measuring point is as follows:

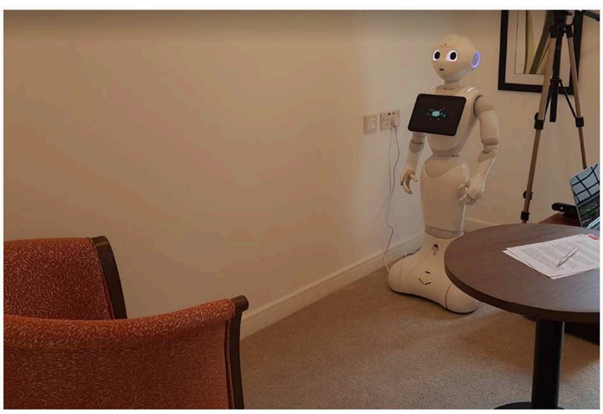


FIGURE 1 | Experimental set-up retirement village.

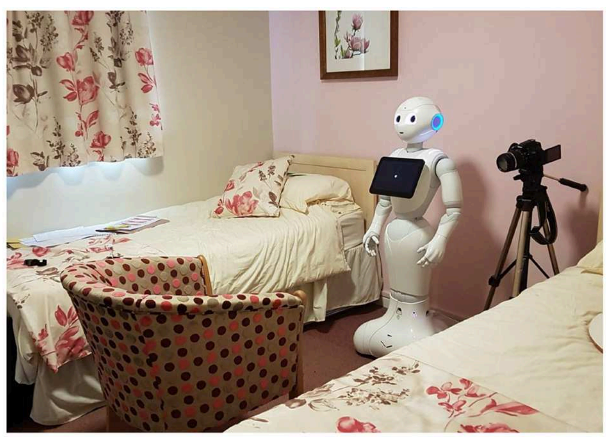


FIGURE 2 | Experimental set-up second retirement village.

- **T1:** Introductory session. For the retirement village that did not have a locksmith available, participants started with MOCA and demographics, followed by an explanation of what would happen during the following sessions. The people that were excluded from data gathering after the MOCA received the explanation without demographics being taken. Other questionnaires taken during this session can be found in **Table 2**. To get a first impression of the robot, it would briefly introduce itself. There was no interaction with the robot.
- **T2:** After finishing interaction 4. Participants would interact with the non-emotive robot the first four sessions and the emotive robot the last four sessions or vice versa. At the end of each condition, they had to fill in questionnaires which can be found in **Table 2**.
- **T3:** After finishing interaction 8, once participants had finished all interactions with the robot.
- **T4:** Debrief session. There would be no interaction with the robot this session, and it would not be present in the experiment room. Participants filled in the attachment, acceptance and perception questionnaires one more time, to investigate whether there was an influence of time and the robot no longer being physically present in the room on their responses. After filling in these questionnaires, they were debriefed by the experimenter. Finally, the participants were asked some final questions in an interview by the experimenter.

TABLE 2 | Questionnaires given at different times.

T1 (introduction)	T2 (after interaction 4)	T3 (after interaction 8)	T4 (debrief)
MOCA (one village only)	Attachment	Attachment	Attachment
Demographics	Almere	Almere	Almere
Attachment Style	Godspeed	Godspeed	Godspeed
Godspeed			Interview

Almere was used to measure acceptance of the robot, Godspeed was used to measure perception of the interaction and MOCA measured cognitive performance.

Figure 3 provides an overview of what questionnaires were performed when, and what behavior the robot displayed during the interactions.

The timeline below provides an overview of the whole experiment. Participants in the control group interacted with the non-emotive robot at all times. Participants in the test group interacted with the non-emotive robot first and the emotive robot later or vice versa. It was aimed for the time between sessions to be as consistent as possible, so most interactions were planned every 3–4 days. If not possible, the minimum would be 2 days between sessions and the maximum 7 days.

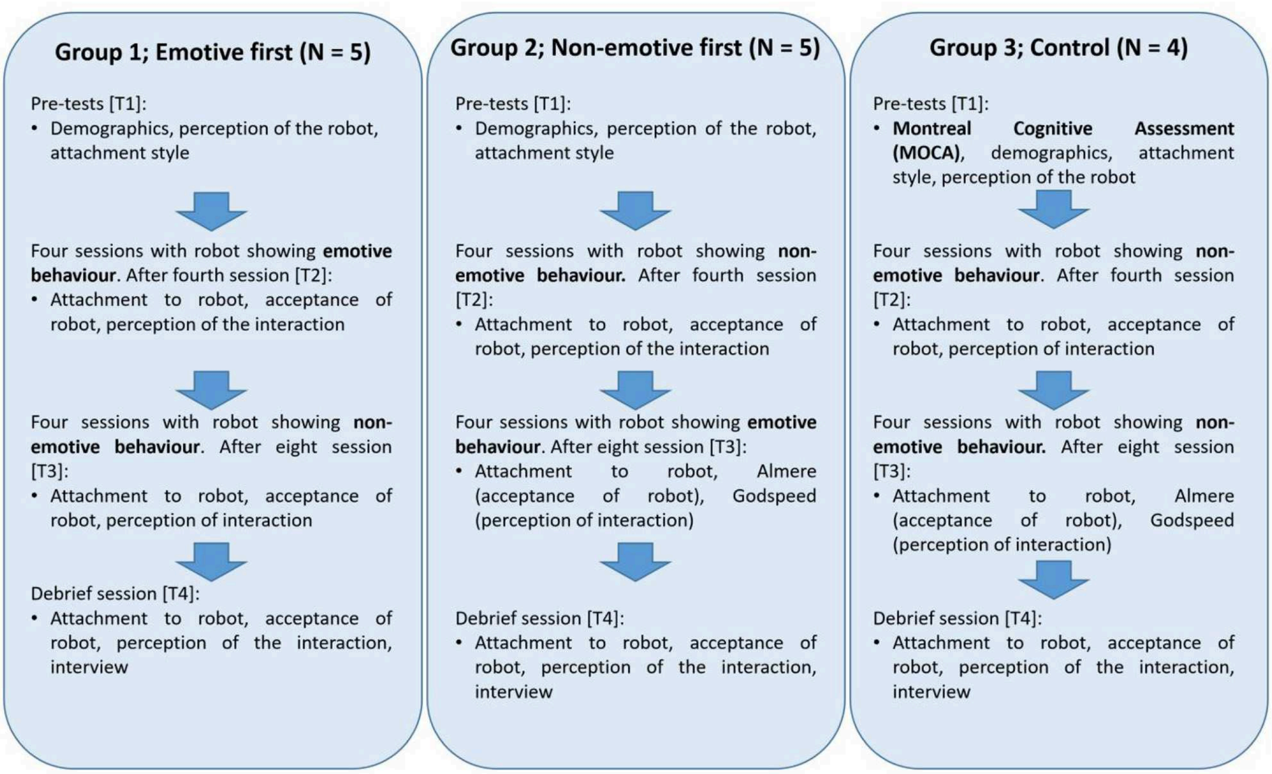
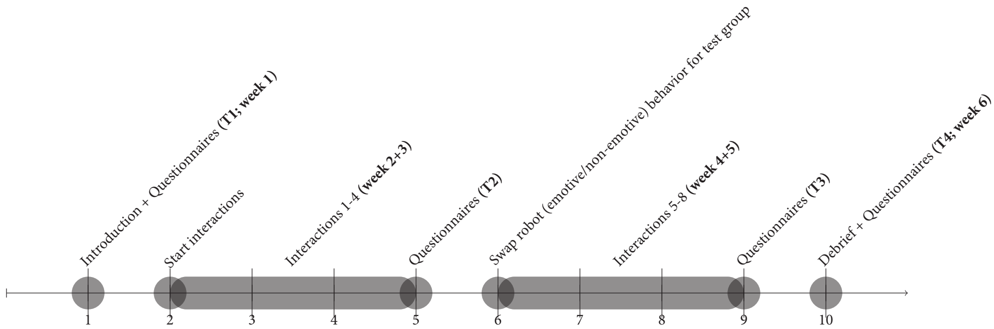


FIGURE 3 | Experimental procedure for the test groups 1 and 2 and control group 3.



4. RESULTS

First, we tested the reliability of the questionnaires we used. The questionnaires measuring acceptance of the robot, perception of the social interaction and attachment questionnaire all showed high internal reliability ($\alpha = 0.85$, $\alpha = 0.83$, and $\alpha = 0.91$, respectively). After participants were debriefed, they were interviewed on their experience with the robot and their opinions regarding emotional deception and emotional attachment.

A mixed between-within subjects design was used for this study. Out of 14 participants that completed the experiment, ten participants (6 male, 4 female, age $M = 76.75$, $SD = 10.75$) were assigned to the test group and interacted with the emotive robot during the first four interactions and the non-emotive robot during the last four interactions, or vice versa. The remaining four participants acted as control group and interacted with the non-emotive robot only (3 male, 1 female, age $M = 76.10$, $SD = 8.10$). This design allowed for between-subjects comparisons after four sessions as well as within-subjects comparisons across the entire study.

4.1. Emotional Deception

A mixed between-within subjects analysis of variance was conducted in order to assess the impact of emotional deception over time (T2, T3, T4) between the two groups (test × control). Examining the acceptance questionnaire, a main effect of group on the perceived social presence of the robot was found, with participants in the test group that interacted with both the emotive and non-emotive robot perceiving the robot as more of a social entity than participants in the control group that only interacted with the non-emotive robot [$F_{(1,12)} = 4.93$, $p = 0.046$, $\eta_p^2 = 0.29$]. This difference can be found in **Figure 4**. No other significant differences were found for any of the constructs in either the acceptance or perception questionnaires.

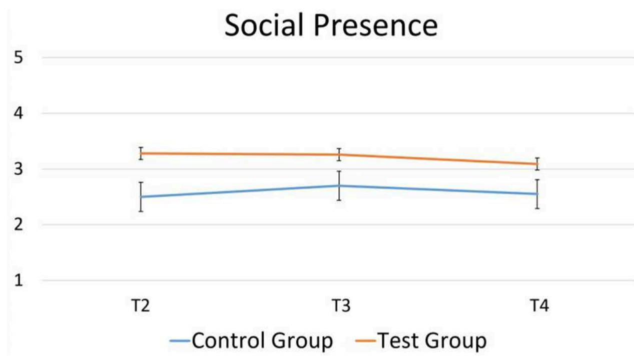


FIGURE 4 | Perceived social presence of the robot over time for control group and test group as measured by the acceptance questionnaire.

To investigate the effect of the robot's displayed behavior, acceptance and perception questionnaire scores from participants from the test group after interacting with the emotive robot were compared to their responses after interacting with the non-emotive robot. One-way ANOVA showed no significant differences for any construct in either of the questionnaires. The averages for the constructs in both questionnaires depending on the robot's displayed behavior can be found in **Figures 5**, **6**.

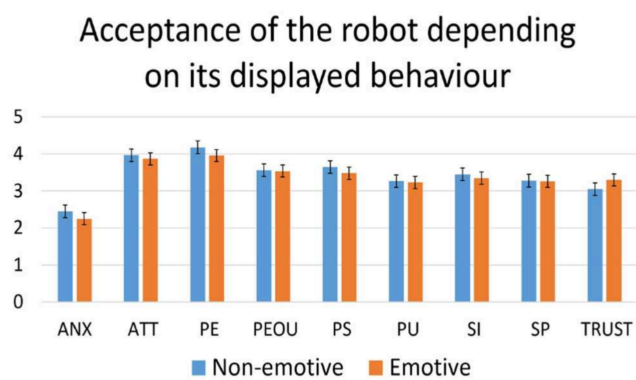


FIGURE 5 | Acceptance of the robot be the test groups depending on its displayed behavior as measured by the acceptance questionnaire. ANX, anxiety to use the robot; ATT, attitude toward technology; PE, perceived enjoyment; PEOU, perceived ease of use; PS, perceived sociability; PU, perceived usefulness; SI, social influence; SP, social presence; TRUST, trust in the robot.

Between-within subjects analysis of variance was conducted in order to compare acceptance and perception ratings by group and over time (T2, T3, T4). There was no significant interaction of time by group or a main effect of group. Investigating the acceptance questionnaire, there was a main effect of time over perceived ease of use [$F_{(2,24)} = 4.22$, $p = 0.03$, $\eta_p^2 = 0.26$], as shown in **Figure 7**. *Post-hoc* tests using the Bonferroni correction showed a significant difference between T2 and T4 [$p = 0.042$], and T3 and T4 [$p = 0.046$], as indicated by the asterisks in **Figure 7**. There was no significant difference between T2 and T3

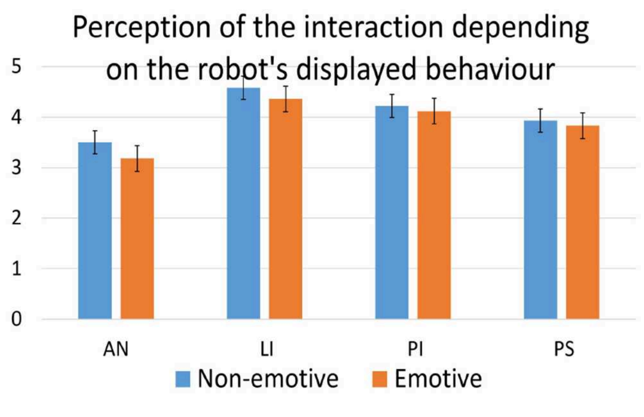


FIGURE 6 | Perception of the social interaction by the test group depending on the robot's displayed behavior as measured by the perception questionnaire. AN, anthropomorphism; LI, likeability; PI, perceived intelligence; PS, perceived safety.

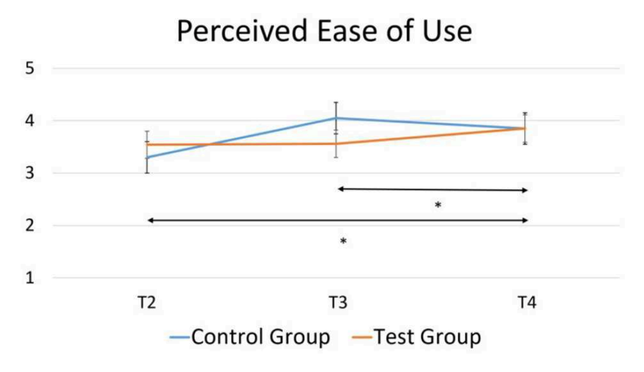


FIGURE 7 | A significant increase in perceived ease of use over time for control group and test group as measured in the acceptance questionnaire. *$p < 0.05$.

[$p = 0.82$]. No other constructs of the acceptance questionnaire, nor any constructs of the perception questionnaire significantly changed over time.

4.2. Emotional Attachment

The acceptance and perception questionnaires were used again to investigate emotional attachment, as trust and anthropomorphism can be indicators of emotional attachment. The attachment questionnaire was included to investigate whether participants became emotionally attached to the robot. Participants' attachment styles can be found in **Table 1**. Six participants were categorized with a secure attachment style, four with a fearful attachment style, four with a dismissive attachment style, and none with a preoccupied attachment style. Acceptance of the robot, perception of social interaction and attachment to the robot were not significantly influenced by participants' attachment style, nor was there an influence of attachment style on any of these factors over time.

Comparing participants' attachment to the robot with constructs of the acceptance questionnaire, Pearson correlation

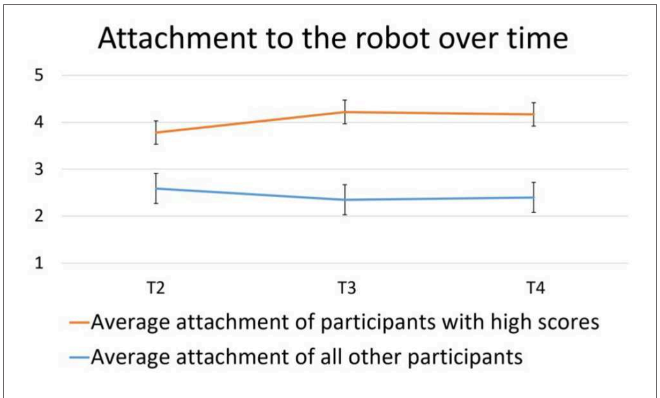


FIGURE 8 | Average attachment to the robot over time as measured by the attachment questionnaire.

analyses showed a positive correlation between participants' attachment to the robot and their perceived ease of using the robot [$r_{(14)} = 0.42$, $p = 0.04$], and the extent to which they perceived the robot as a social entity [$r_{(14)} = 0.42$, $p = 0.04$]. Pearson correlation analyses were run for the constructs of the perception questionnaire and attachment as well, as high scores for anthropomorphism and likability can be an indicator for participants becoming attached to the robot. These analyses showed strong positive correlations between participants' attachment to the robot and the constructs anthropomorphism [$r_{(14)} = 0.66$, $p < 0.01$], likability [$r_{(14)} = 0.51$, $p = 0.01$] and perceived intelligence [$r_{(14)} = 0.51$, $p = 0.01$].

Attachment to the robot fell in the low to medium range, as can be seen in **Figure 8**. However, two participants (one male, one female) scored high on attachment [$M = 4.06$, $SD = 0.24$]. Their attachment to the robot was high when it was first measured at T2 and remained high during T3 and T4. These participants belonged to the test group and interacted with the emotive robot during the experiment.

A mixed between-within subjects analysis of variance was conducted in order to assess the impact of emotional attachment over time (T2, T3, T4) between the two groups (test × control). No significant difference was found for overall attachment [$p = 0.34$]. Looking at the attachment questionnaire items separately, a significant influence of the robot's displayed behavior was found for the statement 'I have feelings for Pepper' that was rated significantly higher by participants from the test group with respect to participants from the control group. [$F_{(1,12)} = 5.33$, $p = 0.04$, $\eta_p^2 = 0.31$]. No other significant differences for attachment between the control group and test group was found.

To investigate the effect of the robot's displayed behavior on attachment to the robot, attachment scores from participants from the test group after interacting with the emotive robot were compared to their responses after interacting with the non-emotive robot, which were taken at T2 and T3. One-way ANOVA showed no significant differences for participants' level of attachment to the robot [$p = 0.55$].

One-way repeated measures ANOVA showed that time did not significantly influence participants' attachment to the robot [$p = 0.61$]. Looking at the nine statements in the attachment questionnaire separately, the only factor that significantly changed was participants reporting that they felt less emotionally affected by Pepper over time [$F_{(2,24)} = 5.88, p < 0.01, \eta_p^2 = 0.33$].

Based on participants' answers to the attachment statements, they were categorized in one of three groups: low, medium, or high attachment. One-way ANOVA was used to compare participants' attachment categories with the constructs of the acceptance questionnaire; there was a significant difference between participants' attachment category and the robot's perceived ease of use [$F_{(2,21)} = 4.76, p = 0.02, \eta_p^2 = 0.31$], social influence [$F_{(2,21)} = 4.97, p = 0.02, \eta_p^2 = 0.32$], social presence [$F_{(2,21)} = 7.82, p < 0.01, \eta_p^2 = 0.43$] and trust [$F_{(2,21)} = 4.25, p = 0.03, \eta_p^2 = 0.29$]. The mean and standard deviations for these constructs and their significance can be found in **Table 3**. Social influence, social presence and trust were significantly lower for participants with low levels of attachment to the robot with respect to participants with high levels of attachment. Perceived intelligence and social influence were significantly lower for participants with low levels of attachment with respect to participants with medium levels of attachment to the robot. There were no significant differences for any of the constructs of the perception questionnaire based on participants' level of attachment to the robot.

Results from T3 and T4 were compared to investigate whether participants' felt differently when some time had passed since their last interaction with the robot and it was not physically present in the room. Paired sample *t*-tests showed that there was no significant difference between the answers given to any of the questionnaires at T3 and T4. As there were no significant differences between T3 and T4 and participants' experience was still fresh at T3 when they finished all interactions, the results of the test group and control group at T3 were compared to investigate the effect of the robot's emotive behavior. One-way ANOVA showed a significant difference for the construct anxiety to use the robot of the acceptance questionnaire [$F_{(1,12)} = 5.52, p = 0.04, \eta_p^2 = 0.32$], where the participants of the control group reported being less afraid to use the robot [$M = 1.56, SD = 0.52$] than the participants of the test group [$M = 2.23, SD = 0.46$]. No other significant differences were found and this difference was not significant for T4.

TABLE 3 | Mean and standard deviation for constructs of the acceptance questionnaire that showed a significant difference between different levels of attachment.

Acceptance construct	Level of attachment	*M*	*SD*	*N*
Perceived ease of use	Low*	3.13	0.50	5
	Medium*	3.78	0.45	7
	High	3.45	0.47	2
Social influence	Low*+	2.75	0.46	5
	Medium*	3.67	0.91	7
	High+	3.88	0.25	2
Social presence	Low**	2.70	0.48	5
	Medium	3.22	0.51	7
	High**	3.80	0.16	2
Trust	Low*	2.75	0.53	5
	Medium	3.25	0.62	7
	High*	3.75	0.50	2

*p <0.05, +p <0.05, **p <0.01.

4.3. Interview

After the participants were debriefed, they were asked some final questions regarding their experience with, and opinion of, the robot. Nine participants (six male, three female) indicated they would want to use the robot on a daily basis in the future, these participants also reported they did not think they would get bored of the robot over time. Four participants (two male, two female) declared they would want to use it on a weekly basis. One participant (male) would not want to use the robot at all. This participant stated that, although he liked interacting with the robot, he found that it was not sufficient for his needs as he preferred a robot that was capable of physical assistance.

Four (three female, one male) participants reported that they would miss Pepper, ranging from "Yes, I guess I will" to "Oh yes, definitely!." Three of these participants were from the test group and one was from the control group. Two of them (both female) scored low on attachment to the robot.

Three participants could not imagine the robot ever playing a role in their life ("do not need a companion," "happy talking to myself when I feel the need to"). Four participants would like to have a robot as a companion, and eight participants thought it would be useful as a helper. This could either be in the sense of helping with tasks, providing useful information or monitoring people's health.

After participants were debriefed about the need to investigate emotional deception, some participants reported finding that the robot was indeed deceptive ("I guess it was deceptive, as it showed some form of emotions"). These participants interacted with the emotive robot and either scored medium or high on attachment. The other participants did not find the robot deceptive, mainly because they thought of it as a machine and/or tool ("I take it for what it is: a distraction for when you are lonely," "I realize it is a machine, therefore I do not find it deceptive"). All participants that found the robot deceptive, all reported finding this deception acceptable, as otherwise the robot would have appeared too machinelike and not pleasant to interact with. Interestingly, two of these three participants were highly attached to the robot.

5. DISCUSSION

This study measured participants' acceptance of a Pepper robot, perception of the social interaction with the robot and attachment to the robot to gain insight into the extent to which emotional deception and emotional attachment are ethical concerns in human-robot interaction. The study consisted of several didactic interactions with the robot spread over several weeks, as time is an essential factor for both emotional deception and emotional attachment. It was expected that effects of emotional deception would be minimal, as the level of deception was designed to be low. Furthermore, any effects that occurred were expected to decrease over time once participants became familiar with the displayed emotions of the robot. Additionally, it was anticipated that emotional attachment would be low but increased over time, and that attachment would be higher for participants interacting with the emotive robot, as the display of emotions by the robot would increase people's perception of the robot being a social entity.

5.1. Emotional Deception

As emotional deception is a subjective response it can best be measured indirectly through other variables. In this study, it was measured through the acceptance and perception questionnaires. Participants from the test group perceived the robot significantly more as a social entity than participants from the control group, indicating that some level of deception may have occurred. However, as expected, no other significant effects were found. This suggests that the effects of the emotional deception used in this study were limited. Emotional deception was designed to be low in this study. It was argued that deception is not a binary value but a spectrum with different gradients, and emotional deception in this study was intended to be low, to find a baseline for acceptable emotive robot behavior. Constructs of the acceptance and perception questionnaires were used to investigate whether emotional deception occurred and how it impacted participants. However, even though it was anticipated that potential effects would decrease over time, indicating that the robot would be perceived as less of a social entity over time, this was not found in the results. During the interview, participants were asked whether they found the robot emotionally deceptive. 21% of the participants indeed thought it was deceptive, where the other 79% did not. It is interesting that all participants that thought the robot was deceptive scored medium or high on attachment. It is possible that these participants thought of the robot more as a social entity than the participants that scored low on attachment and did not think the robot was deceptive as they perceived it as a tool. The participants that did think the robot was deceptive reported they thought it was acceptable as otherwise it would not be pleasant to interact with. As reported in the results, some of these participants were highly attached to the robot. The risks for vulnerable users are supported by research from Klamer and Allouch (2010), who also investigated acceptance of the robot and perception of the social interaction with the robot. They ran a longitudinal study with participants with mild cognitive impairments using similar measurements, but the participants scored higher for nearly all constructs of the acceptance and deception questionnaires than the participants from this study. This may be due to different factors like the use of a different robot and different experiment scenarios. However, as the participants of the study from Klamer and Allouch (2010) were more vulnerable due to mild cognitive impairments with respect to healthy older adults in this study, the additional risks for vulnerable users should be investigated further. These findings indicate that even though the effects of emotional deception appear limited in a didactic setting, deception still occurs and is therefore an ethical concern in practice as well as in the literature. Further investigation into additional measures of perception are necessary before conclusions can be drawn.

5.2. Emotional Attachment

Emotional attachment to the robot ranged from low to medium. This was expected due to the didactic nature of the interactions. However, there was no significant change in attachment over time. This was surprising, as it was hypothesized that attachment would increase when participants were exposed to the robot more often. From an ethical point of view, low attachment to the robot is positive as it suggests that the robot's behavior did not elicit attachment and decreases potential ethical risks, at least for didactic interactions as used in this experiment. However, no change over time indicates that attachment remains high for participants that are attached to the robot from the start. There were two participants who became highly attached to the robot and remained highly attached to the robot for the duration of the experiment. These participants are potentially at risk of experiencing negative consequences of their attachment to the robot such as over-trusting it, having too high expectations of it and relying on it too much. Even though there was not a significant influence of the robot's emotive behavior on participants' overall attachment, the two participants that became highly attached to the robot were both from the test group and therefore interacted with the robot showing emotive behavior as well as non-emotive behavior. As the low number of participants may be a cause for the absence of a significant difference, this is something that needs further investigation. However, even with a low number of participants the findings of this study are still crucial for the development of ethically safe robots.

Participants' attachment category significantly influenced the extent to which the robot was perceived as a social entity and how much the robot was trusted. This shows that participants with a higher level of attachment were more likely to be emotionally deceived by the robot, and may be more at risk of over-trusting the robot and becoming dependent on it. This is especially important for older adults that are more vulnerable due to for example loneliness, as they may become more easily attached to the robot than other users. Thirteen of the participants in this study were either married or in a relationship and reported they did not feel lonely.

No significant influence of participants' attachment style on their level of attachment to the robot was found. However, it is likely that effects of attachment style were not found due to the small sample size, as almost half of the participants had a secure attachment style, where it is expected that people with an insecure attachment style are more vulnerable with respect to emotional deception and emotional attachment. As higher levels of attachment provide more ethical concerns to be aware of, and results from other studies indicated that different attachment styles require different approaches (Dziergwa et al., 2018), attachment style should be regarded as a useful metric for emotional attachment.

During the interview, 36% participants reported they would miss the robot. This may be an indicator that they became attached to the robot; however, 50% of these participants scored low on attachment. A possible explanation for this finding is that participants would miss the whole social experience and the novelty of interacting with a robot, and not necessarily the robot itself. Their willingness to use the robot in the future was high, with nine of them declaring they would like to use it every day. These participants also did not think they would get bored of the robot. Future work will include behavior analysis, speech synthesis and physiological analyses to investigate whether they are useful additional metrics for understanding attachment. Overall, it can be concluded that emotional attachment to the robot may occur in practice and should be investigated in more detail.

5.3. Limitations

One drawback of this study was that, whilst every effort was made to recruit participants for this study, the number of participants was low. Future studies may want to explore why there was resistance to participate and whether or not it was due to the use of robots. However, due to the novelty of this study and its importance for HRI as a research field the findings from this experiment are still deemed to be valuable. This low number of participants and the discrepancy between reported statements from participants and their replies to the attachment questionnaire make it hard to draw conclusions from the results. Future work that includes analyzing participants' behavior and their speech prosody data will hopefully provide clarification. Furthermore, a disadvantage of running field studies is that it is hard to control the experimental environment. In one retirement village, the experiment room was small and the experimenter was located behind the participant, as there was no other place for the experimenter to be. Even though the experimenter tried to limit the interactions between themselves and the participants, it is possible that participants' answers to the questionnaires were influenced by their close proximity to, and therefore bond with the experimenter. Besides that, participants talked to one another about the experiment and possibly influenced each other's opinion of the robot. Additionally, the study was conducted in an environment where other people were working, who accidentally interrupted the experiment by walking into the experiment room. In this study, this happened while participants were filling in questionnaires, not during interactions with the robot. Also, even though the experiment was run in the field and not a laboratory setting, it was still a controlled experiment with a limited number of interactions. The freedom that participants had during these interactions was limited, as the interaction was pre-programmed and although the participants were given the opportunity to provide some personal input, this was limited and may have influenced people's opinion of the robot. This limitation was introduced deliberately to ensure interactions were as similar as possible between participants which made it easier to compare results as all participants would have the same impression of the robot's abilities. However, the nature of the interaction (didactic rather than personal conversations) may possibly have influenced the results. Lastly, as ethical concerns have not been studied in real-life settings as much as other aspects, there are few results to compare this work with which makes it harder to discuss to what extent the findings can be generalized and how easy it is to reproduce the study.

6. CONCLUSION

The likelihood of older adults interacting with social robots is ever increasing (Esposito et al., 2016; Li et al., 2019), and with it ethical concerns regarding these interactions are raised. Some of these concerns are emotional deception and emotional attachment, which have been raised as ethical concerns in the literature (e.g., Sharkey and Sharkey, 2012; Sullins, 2012; Kolling et al., 2013). The aims of this study were to establish whether these concerns are reflected in practice, and investigate what factors influence these concerns. It was found that both concerns may arise in practice and therefore need further investigation. This research is important for HRI as a research field, as it will help develop robots which comply with the principles of ethical design. Moreover, as social robots are also used with vulnerable users like older adults suffering from dementia it is essential to have guidelines on what human-robot interactions are ethically safe and acceptable. Even though the number of participants in this study was low and it is difficult to draw clear conclusions from the findings, this work does provide useful insights into conducting longitudinal field studies and specific directions for future research. Lastly, knowing to what extent ethical concerns raised in the literature have an impact in practice is essential for HRI development as -important as they are- ethical considerations can limit the deployment of these technologies. Speculation about the consequences of a technology can inform research directions, but may carry more weight if proven through experimental study. Speculation may not only be a weak discouragement of poor practice, but may also constrain useful development and study if the worry about a putative ill proves to be unfounded.

Future socially assistive robots should be ethically safe to interact with. Therefore, solely using questionnaires to investigate ethical concerns is a useful starting point to find trends, but not sufficient for stronger claims regarding these concerns. Future work will involve finding additional metrics for emotional attachment, analysis of people's behavior through video recordings and speech patterns, and analyzing people's physiological reactions to the robot's behavior. Once the boundaries for emotive robot behavior with respect to emotional deception and emotional attachment are clear for didactic interactions, the guidelines can be extended to apply to other settings with more personal human-robot interactions.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of the West of England Faculty Research Ethics Committee—Faculty of Environment and Technology. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AM, SD, PC-S, NZ, AW, and MS devised the project and main conceptual ideas. AM designed the study, that was verified by

NZ and PC-S. AM ran the experiments and analyzed the data. All authors contributed to and/or commented on the writing of the manuscript.

FUNDING

This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 721619 for the SOCRATES project.

ACKNOWLEDGMENTS

First and foremost we would like to thank all participants that took part in this study. We also would like to thank the care managers of the retirement villages who gave us the opportunity to run our experiments at their premises.

REFERENCES

Adar, E., Tan, D. S., and Teevan, J. (2013). "Benevolent deception in human computer interaction," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris: ACM), 1863–1872.

Arkin, R. C., Ulam, P., and Wagner, A. R. (2012). Moral decision making in autonomous systems: enforcement, moral emotions, dignity, trust, and deception. *Proc. IEEE* 100, 571–589. doi: 10.1109/JPROC.2011.2173265

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

Beck, A., Cañamero, L., Hiolle, A., Damiano, L., Cosi, P., Tesser, F., et al. (2013). Interpretation of emotional body language displayed by a humanoid robot: a case study with children. *Int. J. Soc. Robot.* 5, 325–334. doi: 10.1007/s12369-013-0193-z

Bemelmans, R., Gelderblom, G. J., Jonker, P., and De Witte, L. (2012). Socially assistive robots in elderly care: a systematic review into effects and effectiveness. *J. Am. Med. Direct. Assoc.* 13, 114–120. doi: 10.1016/j.jamda.2010.10.002

Birnbaum, G. E., Mizrahi, M., Hoffman, G., Reis, H. T., Finkel, E. J., and Sass, O. (2016). What robots can teach us about intimacy: the reassuring effects of robot responsiveness to human disclosure. *Comput. Hum. Behav.* 63, 416–423. doi: 10.1016/j.chb.2016.05.064

Borenstein, J., Wagner, A., and Howard, A. (2017). "A case study in caregiver overtrust of pediatric healthcare robots," in *RSS Workshop on Morality and Social Trust in Autonomous Robots* (Cambridge).

Breazeal, C. (2011). "Social robots for health applications," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE* (Boston, MA: IEEE), 5368–5371.

Breazeal, C., and Scassellati, B. (1999). "How to build robots that make friends and influence people," in *Intelligent Robots and Systems, 1999. IROS'99. Proceedings. 1999 IEEE/RSJ International Conference on*, Vol. 2 (Kyongju: IEEE), 858–863.

Breazeal, C. L. (2004). *Designing Sociable Robots*. Cambridge, MA: MIT Press.

Brennan, K. A., Clark, C. L., and Shaver, P. R. (1998). "Self-report measurement of adult attachment: an integrative overview," in *Attachment Theory and Close Relationships,* eds J. A Simpson and W. S. Rholes (New York, NY: Guilford Press), 46–76.

Cacioppo, J. T., and Patrick, W. (2008). *Loneliness: Human Nature and the Need for Social Connection*. New York, NY: WW Norton & Company.

Callén, B., Domènech, M., López, D., and Tirado, F. (2009). Telecare research:(cosmo) politicizing methodology. *ALTER Eur. J. Disabil. Rese.* 3, 110–122. doi: 10.1016/j.alter.2009.02.001

Coeckelbergh, M., Pop, C., Simut, R., Peca, A., Pintea, S., David, D., et al. (2016). A survey of expectations about the role of robots in robot-assisted therapy for children with asd: ethical acceptability, trust, sociability, appearance, and attachment. *Sci. Eng. Ethics* 22, 47–65. doi: 10.1007/s11948-015-9649-x

Collins, E. C., Millings, A., and Prescott, T. J. (2013). "Attachment to assistive technology: a new conceptualisation," in *Proceedings of the 12th European AAATE Conference (Association for the Advancement of Assistive Technology in Europe)* (Algarve), 823–828.

Dautenhahn, K., and Billard, A. (1999). "Bringing up robots or—the psychology of socially intelligent robots: from theory to implementation," in *Proceedings of the Third Annual Conference on Autonomous Agents* (Seatlle, WA: ACM), 366–367.

Dragan, A., Holladay, R., and Srinivasa, S. (2015). Deceptive robot motion: synthesis, analysis and experiments. *Auton. Robots* 39, 331–345. doi: 10.1007/s10514-015-9458-8

Dziergwa, M., Kaczmarek, M., Kaczmarek, P., Kędzierski, J., and Wadas-Szydłowska, K. (2018). Long-term cohabitation with a social robot: a case study of the influence of human attachment patterns. *Int. J. Soc. Robot.* 10, 163–176. doi: 10.1007/s12369-017-0439-2

Esposito, R., Fiorini, L., Limosani, R., Bonaccorsi, M., Manzi, A., Cavallo, F., et al. (2016). "Supporting active and healthy aging with advanced robotics integrated in smart environment," in *Optimizing Assistive Technologies for Aging Populations,* eds Y. S. Morsi, A. Shukla, and C. P. Rathore (Hershey, PA: IGI Global), 46–77

Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2003). A survey of socially interactive robots. *Robot. Auton. Syst.* 42, 143–166. doi: 10.1016/S0921-8890(02)00372-X

Frennert, S., and Östlund, B. (2014). Seven matters of concern of social robots and older people. *Int. J. Soc. Robot.* 6, 299–310. doi: 10.1007/s12369-013-0225-8

Fulmer, I. S., Barry, B., and Long, D. A. (2009). Lying and smiling: informational and emotional deception in negotiation. *J. Bus. Ethics* 88, 691–709. doi: 10.1007/s10551-008-9975-x

Hall, E. T., Birdwhistell, R. L., Bock, B., Bohannan, P., Diebold A. R. Jr., Durbin, M., et al. (1968). Proxemics [and comments and replies]. *Curr. Anthropol.* 9, 83–108. doi: 10.1086/200975

Heerink, M., Kröse, B., Evers, V., and Wielinga, B. (2010). Assessing acceptance of assistive social agent technology by older adults: the almere model. *Int. J. Soc. Robot.* 2, 361–375. doi: 10.1007/s12369-010-0068-5

Hiolle, A., Bard, K. A., and Canamero, L. (2009). "Assessing human reactions to different robot attachment profiles," in *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication* (Toyama: IEEE), 251–256.

Hiolle, A., Cañamero, L., Davila-Ross, M., and Bard, K. A. (2012). Eliciting caregiving behavior in dyadic human-robot attachment-like interactions. *ACM Trans. Interact. Intell. Syst.* 2:3. doi: 10.1145/2133366.2133369

Huber, A., Weiss, A., and Rauhala, M. (2016). "The ethical risk of attachment how to identify, investigate and predict potential ethical risks in the development of social companion robots," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Christchurch: IEEE), 367–374.

Hyman, R. (1989). The psychology of deception. *Annu. Rev. Psychol.* 40, 133–154. doi: 10.1146/annurev.ps.40.020189.001025

Jackson, J. (1991). Telling the truth. *J. Med. Ethics* 17, 5–9. doi: 10.1136/jme.17.1.5

Keefer, L. A., Landau, M. J., Rothschild, Z. K., and Sullivan, D. (2012). Attachment to objects as compensation for close others' perceived unreliability. *J. Exp. Soc. Psychol.* 48, 912–917. doi: 10.1016/j.jesp.2012.02.007

Kidd, C. D., Taggart, W., and Turkle, S. (2006). "A sociable robot to encourage social interaction among the elderly," in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on* (Orlando, FL: IEEE), 3972–3976.

Kirby, R., Forlizzi, J., and Simmons, R. (2010). Affective social robots. *Robot. Auton. Syst.* 58, 322–332. doi: 10.1016/j.robot.2009.09.015

Klamer, T., and Allouch, S. B. (2010). *Acceptance and Use of a Zoomorphic Robot in a Domestic Setting*. Freyung: Austrian Society for Cybernetic Studies.

Kolling, T., Haberstroh, J., Kaspar, R., Pantel, J., Oswald, F., and Knopf, M. (2013). Evidence and deployment-based research into care for the elderly using emotional robots. *J. Gerontopsychol. Geriatr. Psychiatry* 26:83. doi: 10.1024/1662-9647/a000084

Kwon, D.-S., Kwak, Y. K., Park, J. C., Chung, M. J., Jee, E.-S., Park, K.-S., et al. (2007). "Emotion interaction system for a service robot," in *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on* (Jeju: IEEE), 351–356. doi: 10.1109/ROMAN.2007.4415108

Li, H., Milani, S., Krishnamoorthy, V., Lewis, M., and Sycara, K. (2019). "Perceptions of domestic robots normative behavior across cultures," in *AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society* (Honolulu, HI). doi: 10.1145/3306618.3314251

Malle, B. F., Scheutz, M., Arnold, T., Voiklis, J., and Cusimano, C. (2015). "Sacrifice one for the good of many? People apply different moral norms to human and robot agents," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (Portland, OR: ACM), 117–124.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al. (2005). The montreal cognitive assessment, MOCA: a brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* 53, 695–699. doi: 10.1111/j.1532-5415.2005.53221.x

Robinson, H., MacDonald, B., and Broadbent, E. (2014). The role of healthcare robots for older people at home: a review. *Int. J. Soc. Robot.* 6, 575–591. doi: 10.1007/s12369-014-0242-2

Scheutz, M. (2011). "The inherent dangers of unidirectional emotional bonds between humans and social robots," in *Robot Ethics: The Ethical and Social Implications of Robotics*, eds P. Lin, K. Abney, and G. A. Bekey (Cambridge, MA: MIT Press), 205.

Schifferstein, H. N., and Zwartkruis-Pelgrim, E. P. (2008). Consumer-product attachment: measurement and design implications. *Int. J. Design* 2, 1–13.

Sharkey, A., and Sharkey, N. (2011). Children, the elderly, and interactive robots. *IEEE Robot. Automat. Mag.* 18, 32–38. doi: 10.1109/MRA.2010.940151

Sharkey, A., and Sharkey, N. (2012). Granny and the robots: ethical issues in robot care for the elderly. *Ethics Informat. Technol.* 14, 27–40. doi: 10.1007/s10676-010-9234-6

Sharkey, N., and Sharkey, A. (2010). The crying shame of robot nannies: an ethical appraisal. *Interact. Stud.* 11, 161–190. doi: 10.1075/is.11.2.01sha

Shim, J., and Arkin, R. C. (2013). "A taxonomy of robot deception and its benefits in HRI," in *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on* (Manchester: IEEE), 2328–2335. doi: 10.1109/SMC.2013.398

Sparrow, R., and Sparrow, L. (2006). In the hands of machines? The future of aged care. *Minds Mach.* 16, 141–161. doi: 10.1007/s11023-006-9030-6

Sullins, J. P. (2012). Robots, love, and sex: the ethics of building a love machine. *IEEE Trans. Affect. Comput.* 3, 398–409. doi: 10.1109/T-AFFC.2012.31

Unies, N. (2015). *World Population Prospects. The 2015 Revision, New York, United Nations*. New York, NY: Population Division.

Van Maris, A., Zook, N., Caleb-Solly, P., Studley, M., Winfield, A., and Dogramadzi, S. (2018). "Ethical considerations of (contextually) affective robot behaviour," in *Hybrid Worlds: Societal and Ethical Challenges - Proceedings of the International Conference on Robot Ethics and Standards (ICRES 2018)* (Troy: CLAWAR Association Ltd.), 13–19.

Wagner, A. R., and Arkin, R. C. (2011). Acting deceptively: providing robots with the capacity for deception. *Int. J. Soc. Robot.* 3, 5–26. doi: 10.1007/s12369-010-0073-8

Weijers, S. (2013). *Exploring Human-Robot Social Relations*. Master's thesis, University of Twente, Twente, The Netherlands.

Winfield, A. (2019). Ethical standards in robotics and AI. *Nat. Electr.* 2, 46–48. doi: 10.1038/s41928-019-0213-6

Winkle, K., and Van Maris, A. (2019). "Social influence and deception in socially assistive robotics," in *Hybrid Worlds: Societal and Ethical Challenges - Proceedings of the International Conference on Robot Ethics and Standards (ICRES 2018)* (London: CLAWAR Association Ltd.), 45–46.

Conflict of Interest: 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.

ORIGINAL RESEARCH
published: 21 February 2020
doi: 10.3389/frobt.2020.00005

Exploring Teens as Robot Operators, Users and Witnesses in the Wild

Elin A. Björling[1*], *Kyle Thomas*[1], *Emma J. Rose*[2] *and Maya Cakmak*[3]

[1] *Momentary Experience Lab, Human Centered Design and Engineering, University of Washington, Seattle, WA, United States,* [2] *School of Interdisciplinary Arts and Sciences, University of Washington Tacoma, Tacoma, WA, United States,* [3] *Human-Centered Robotics Lab, Paul G. Allen School of Computer Science & Engineering, University of Washington, Seattle, WA, United States*

As social robots continue to show promise as assistive technologies, the exploration of appropriate and impactful robot behaviors is key to their eventual success. Teens are a unique population given their vulnerability to stress leading to both mental and physical illness. Much of teen stress stems from school, making the school environment an ideal location for a stress reducing technology. The goal of this mixed-methods study was to understand teens' operation of, and responsiveness to, a robot only capable of movement compared to a robot only capable of speech. Stemming from a human-centered approach, we introduce a Participatory Wizard of Oz (PWoz) interaction method that engaged teens as operators, users, and witnesses in a uniquely transparent interaction. In this paper, we illustrate the use of the PWoz interaction method as well as how it helps identify engaging robot interactions. Using this technique, we present results from a study with 62 teens that includes details of the complexity of teen stress and a significant reduction in negative attitudes toward robots after interactions. We analyzed the teens' interactions with both the verbal and non-verbal robots and identified strong themes of (1) authenticity, (2) empathy, (3) emotional engagement, and (4) imperfection creates connection. Finally, we reflect on the benefits and limitations of the PWoz method and our study to identify next steps toward the design and development of our social robot.

Keywords: social robots, participatory, adolescence, empathy, Wizard of Oz, mental health, human-centered design

OPEN ACCESS

Edited by:
Séverin Lemaignan, Bristol Robotics Laboratory, United Kingdom

Reviewed by:
Patrícia Alves-Oliveira, University Institute of Lisbon (ISCTE), Portugal
Vicky Charisi, European Commission, Joint Research Centre, Belgium

****Correspondence:***
Elin A. Björling
bjorling@uw.edu

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *14 June 2019*
Accepted: *09 January 2020*
Published: *21 February 2020*

Citation:
Björling EA, Thomas K, Rose EJ and Cakmak M (2020) Exploring Teens as Robot Operators, Users and Witnesses in the Wild. Front. Robot. AI 7:5. doi: 10.3389/frobt.2020.00005

1. INTRODUCTION

Teens are now the most stressed age group, with 27% percent of US teens reporting very high levels of daily stress, and 31% reporting feeling overwhelmed as a result of negative stress (American Psychological Association, 2014). Increased stress has been shown to lead to depression (Maughan et al., 2013) and negatively impacts cognitive function that affecting learning (Vogel and Schwabe, 2016). Many schools lack the resources (time and personnel) to implement and maintain school-based mental health programs (Eiraldi et al., 2015).

Social robots have the potential to improve mental health, especially in teens. Given that teens' lives are mediated through a variety of digital technologies, using a digital device to support them may be contextually appropriate. To address the mental health challenges of teens, Project EMAR (Ecological Momentary Assessment Robot) aims to develop a social robot for teens that will be stationed at schools to gather accurate momentary data about teen stress and provide micro-interventions to reduce stress. We used a participatory design approach, involving teens in all research and design activities and decisions.

In the last 3 years, our research team has conducted a number of exploratory high school visits, a social robot design challenge with 7 participating high schools (Rose et al., 2019), and several participatory design and interaction studies that were all in the wild at schools (Rose and Björling, 2017; Björling et al., 2018). Our current investigation involves teens as co-researchers and co-designers to help us explore (1) differences between movement-only and speech-only robot interactions and (2) appropriate robot responsiveness to teen stressors. The goal of conducting this study was to explore both movement and speech behaviors to inform our larger project.

In this paper, we present the results of a study to explore how teens teleoperate and respond to two distinctly different robots. One is a soft-bodied, movement robot with no speech capabilities. The other is an immobile, boxy-robot with speech capability. First, we provide a background on teen stress, social robots, and participatory, human-centered design. Second, we detail the methods of the study and how it was conducted, including the development of a novel method to investigate teen and robot interactions called of participatory Wizard of Oz. Third, we share the findings of the study that explore the complexity of stress, attitudes about robots, comparison of two robot prototypes, and the themes from an analysis of teen engagement with the robot prototypes. Fourth, we discuss the findings and reflect on how the results of the study can inform social robot design for teens. We conclude with a discussion on limitations and next steps.

2. BACKGROUND

2.1. Teen Stress and Mental Health

Eighty-three percent of teens report school as a primary negative stressor (Thapar et al., 2012; American Psychological Association, 2014). Recent evidence shows the cumulative impact of everyday sources of negative stress is highly prevalent and impactful on teens (Hamilton et al., 2016). Although positive stress is experienced by adolescents and appears to benefit their well-being (Branson et al., 2019), chronic negative stress is a known risk factor for both physical and mental health problems (Juster et al., 2010) as often stressful life events precede the onset of adolescent depression (Mazurka et al., 2016). The developing adolescent brain makes adolescents especially vulnerable to the cumulative insults of chronic stress (McEwen and Morrison, 2013). A nationwide survey of high school students in the United States found that 16% of students reported that they were seriously considering suicide, 13% reported creating a plan, and 8% reporting trying to take their own life in the 12 months preceding the survey. Adolescents also exhibit the highest rates of self-harm, including attempted suicide (Ting et al., 2012).

Effective, school-based stress reduction interventions for adolescents exist. The Mindfulness-Based Stress Reduction program for teens has been shown to reduce teen stress and decrease the possibility of mental health problems (Biegel et al., 2014; Edwards et al., 2014). In addition, cognitive behavioral and dialectical behavioral school-based therapy programs have both been successful at reducing stress and incidence of depression (Werner-Seidler et al., 2017). However, most school-based interventions are cost-prohibitive and require significant commitment of staff and student time, which is not possible for many schools, especially those that are under-resourced (Eiraldi et al., 2015). Therefore, designing assistive technologies to support teens in reducing stress can result in increased access to necessary mental health tools.

Technologies designed and aimed to reduce stress and improve mental health do exist. Chatbots such as Woebot (Gabriels, 2019) and Vivibot (Greer et al., 2019) have been shown effective in reducing anxiety and depression in adults. And, although a chatbot has been successful in smoking cessation for adolescents (Simon et al., 2019), a large review of chatbots for mental health concluded more reserach is needed to understand the true effect on mental health and none of the chat agents were focussed specifically on stress reduction (Abd-alrazaq et al., 2019).

2.2. Assistive Social Robots

Social robots provide a variety of benefits and can assist humans by fulfilling unmet needs (Feil-Seifer and Mataric, 2005). Social robots have been suggested as an appropriate tool for mental health applications, providing therapeutic and assessment capabilities in a variety of populations (Breazeal, 2011) including those that are vulnerable (Kim et al., 2013). Several robots have shown promise in terms of therapeutic interventions. For example, the social robot Therabot (Duckworth et al., 2015) is an animated dog designed to support those who have survived trauma and experience feelings of being overwhelmed. Additionally, Paro (Wada et al., 2005) is a plush seal designed for seniors in assisted living environments to reduce stress and stimulate interaction.

Social robots have also been designed specifically for their therapeutic effect for children. Researchers have identified the importance of empathy in social robot interactions with children (Leite et al., 2012; Giannopulu et al., 2018). Social robots have been highly effective for increasing social interactions and communication for children with autism (Fernaeus et al., 2010; Scassellati et al., 2012; Kim et al., 2013). In addition, social robots have been shown to reduce anxiety in children who are hospitalized (Jeong, 2017; Logan et al., 2019). Little work has specifically explored the relationship between teens and social robots.

The design of a social robot to specifically help measure and address teen stress, is a timely expansion of the application of social robots with potential for significant benefits. Project EMAR is an interdisciplinary project using human-centered design to develop a social robot to capture stress and mood data from teens while providing a micro-intervention to relieve stress. EMAR is designed to live in a school setting and collect aggregate, anonymous data and be a tool for teens to better understand and manage their stress, see Björling et al. (2018) for more detail.

2.3. Participatory, Human-Centered Design With Teens

In designing and developing social robots, our project uses human-centered design (HCD), an approach to developing technology that focuses on people and their needs throughout the design process, defined by ISO 9241-210:2010(E). It is a process with a philosophical commitment to upholding human dignity and human rights (Buchanan, 2001; Walton, 2016). Within HCD, this project employs participatory design methods to engage participants (Schuler and Namioka, 1993) which is an appropriate way to engage vulnerable populations such as teens.

While the methods for this study are detailed further in the methods section below, we differentiate our approach from other common approaches to designing social robots. First, the research in this study is conducted in the wild, rather than in a lab. Lab studies can not adequately account for the open-ended encounters that happen between people and robots that are context-dependent (Šabanović et al., 2006). Studies in HRI often privilege the technological capabilities of robots over important factors such as social context and needs of a diverse group of users or stakeholders (Šabanović et al., 2014).

A variety of methods are appropriate for engaging people in design in the wild. Each of these methods have strengths and weaknesses. Contextual inquiry (Beyer and Holtzblatt, 1998) is a method where the researcher engages in detailed observation and interviewing in the context where a product or design will be used, in the wild. This method places the researchers in an apprentice relationship and privileges the expertise and perspective of the target user. However, while this method is well suited to gaining an understanding of how existing processes and procedures, specifically in work settings are completed, it is less appropriate for groups of people, engaging in loosely structured interactions with novel technologies in a social setting. Further, ethnography is a helpful method for understanding a culture or group of people as a way to inform design (Millen, 2000; Olson and Kellogg, 2014). Ethnography is helpful to understand the use and adoption of novel technologies, including social robots, over time (Forlizzi, 2007; Sabelli et al., 2011). However, ethnography is not always an appropriate method choice when designing new or novel technologies that are not fully functional and still require formative feedback and iteration.

Given our participatory approach, we were primarily interested in choosing methods that specifically engaged teens as collaborators in both the research and design process. Other methods that recognize the expertise of users, include the approach of Ladner (2015) to design for user empowerment which calls on the HCI community to build infrastructure and design opportunities to promote the ability for more people to be engaged during the end to end process of design. As he states, "In design for user empowerment, users develop the project, design the requirements and features, develop the prototypes, test the prototypes, and analyze the results of testing to refine the design" (p. 27). Engaging people throughout design in a meaningful and fully engaged way can create more appropriate design solutions.

Other approaches that engage users in the design of social robots that comes closer to Ladner's vision, includes the technique of body storming, where one person role plays being a robot in order to explore design ideas, interactions, and scripts (Oulasvirta et al., 2003). An additional method that includes even more interactivity from this type of role-play is the Wizard of Oz technique, a common approach for simulating the functionality of design (Kelley, 1984) where an operator simulates key features of a technology. The challenge of the Wizard of Oz approach is that it often includes deception and is not transparent about what aspects of the technology are simulated or functional. The Wizard of Oz technique has not been explored extensively in participatory design, with the exception of one study asking participants to create their own gestural interfaces using the Wizard of Oz technique (Akers, 2006).

Participatory Design (PD) has been used in other contexts to develop social robots. Many robotics projects start with a technological stance or intervention, whereas PD starts with an aggregation and analysis of the concerns of a community or group. Members of that group become active in the design process throughout the project. Designs are synthesized from that stand point rather than exclusively gathering feedback on existing designs (Rodil et al., 2018). PD invites participants to use their experiences and bring "their lifeworlds through their design" (DiSalvo et al., 2008; Rodil et al., 2018) and engage in "critical engagements" that can reveal and question existing beliefs about technology (DiSalvo et al., 2008).

3. STUDY DESIGN: EXPLORING DIFFERENT TYPES OF ROBOTS DURING TEEN-ROBOT INTERACTION

The long term goal of our research is to design and develop a social robot that can capture and aggregate data about perceived adolescent stress in schools and offer interventions to help to reduce teen stress. In the past 3 years, we have gathered input from teens to inform the design using a number of different methods. In this paper, we focus on the context of teenagers talking to the robot (in free form) to share details of what stresses them (e.g., an upsetting interaction or an upcoming exam). Our goal was to better understand how the robot should behave such that users feel heard. In particular, we sought to gather input about different embodiments (section 3.2) and understand parameters of robot behaviors (how it should move, what it should say) to give the sense of being heard. In the following section, we describe our method for gathering data from teens in the wild (i.e., at schools), describe the procedure we followed, and enumerate the types of data we gathered.

3.1. Participatory Wizard of Oz Method

While the advantages of observing human-robot interactions in the wild are clear, the researchers' ability to do so is often limited by the availability of social robot platforms that allow rapid prototyping of robust interactions. To address this challenge and take full advantage of conducting studies in the wild with teen participation, we developed a new method that extends the Wizard of Oz technique. Our method, which we call *Participatory Wizard of Oz*, involved removing the typical deception of the WoZ method where a researcher operated the robot without the participant's awareness. Similar to the suggested framework of Druin (2002) where child participants were placed in multiple roles, e.g., user, tester, informant, or

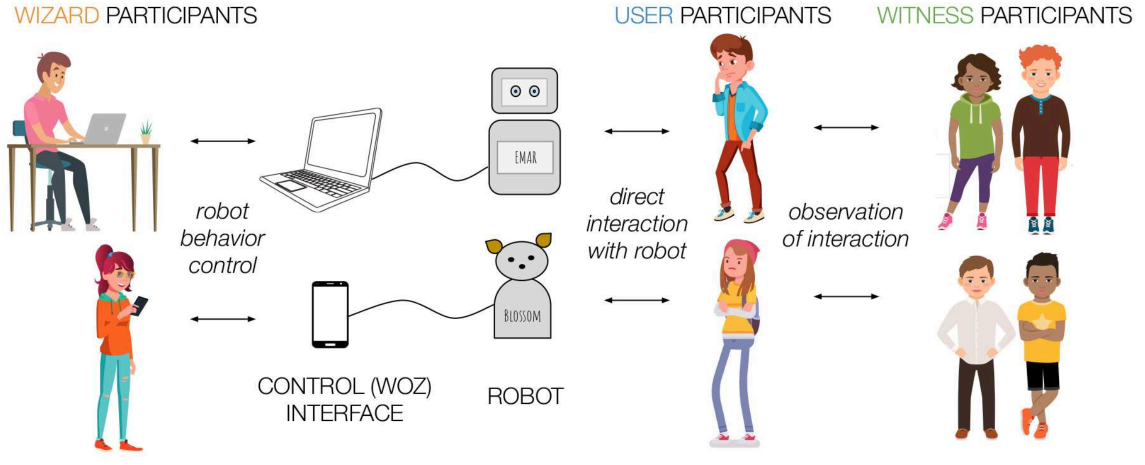


FIGURE 1 | Illustration of the Participatory Wizard of Oz in the Wild Method used in our research. Wizards control robot behavior, users directly interact with the robot, and witnesses observe user-robot interaction.

design partner, our Participatory WoZ (PWoZ) method was fully transparent with the following characteristics:

- Research was conducted in the wild (*in situ*).
- Participants were the creators of interaction content.
- Participants were wizards of the robot.
- Participants were users of the robot.
- Participants were witnesses to the robot interaction.

The characteristics are illustrated in **Figure 1** and described in more detail below.

3.1.1. In the Wild

Conducting research "in the wild" is critical to the goal of developing a social robot that will ultimately be implemented in schools and will require continued engagement from teens in order to have an impact on helping them cope with stress. Having teens interact with our robot prototype in the same context in which the robot will be implemented, enabled them to consider environmental contexts that may not be evident in a lab setting. Further, maintaining ecological validity (Oulasvirta et al., 2003; Carter et al., 2008), greatly strengthened our data stemming from this PWoZ method. Unlike laboratory studies, our study embraces the numerous, uncontrollable variables that exist in the wild. We allow for freedom of choice, the influence of social factors and interactions, and real world distractions. We utilized real-world spaces to conduct our studies in the wild (e.g., flexible spaces and classrooms). In addition, researchers stepped out of the way during interactions and were often not visible, making the interactions even more contextually valid. Finally, studying the interaction in context gave us the opportunity to study how the interaction might be perceived by observers.

3.1.2. Four Participant Roles

In this method, participants played all the key roles in the social interaction, and provided data about their experience of each role including (1) content creators, (2) robot operators (Wizards), (3) robot users, and (4) interaction witnesses. By asking participants to fulfill each of these roles, researchers primarily became facilitators of the method, rather than, wizards or witnesses, therefore allowing teens to be more naturalistic in their group interactions. Researchers set up the study and provided an overview, but faded into the background of the research context as teens design and drove the interactions.

In this method, teens participated in the research in multiple capacities providing design input and feedback from different perspectives. Wizard participants provided input about how they thought the robot should behave by controlling the robot's actions, such as what the robot said, what facial expression it displayed, or how it moved in reaction to the user's story. User participants directly interacted with the robot prototype demonstrating how the interaction unfolded and gaining first hand experience about how the interaction felt. Witness participants observed the interaction from a third person perspective. After experiencing the interaction from three different perspectives, participants in all three roles provided feedback about the robot's behavior in the interaction (chosen by the wizard participant) as well as attributes (e.g., size, material, look) and capabilities (e.g., voice, range of motion) of the robot.

This method combined advantages of many alternative methods discussed in section 2.3: (1) in the wild, (2) low fidelity prototype (like WoZ), (3) significant involvement of teens in multiple roles. One clear distinction of this method from the traditional Wizard of Oz method was that the user participants who interacted with the robot were not under the impression that the robot is operating autonomously. Teens were fully aware that the robot was controlled by their peer, making the experience more teen-centric.

3.2. Robot Platforms

Our prior research with teens resulted in a number of design requirements, but due to the variability we observed in teens' preferences we had not yet fully committed to a particular robot platform. In this work, we explored the use of two different robot platforms, both with high degrees of customizability. Given we wanted to explore the functions of speech and movement, the

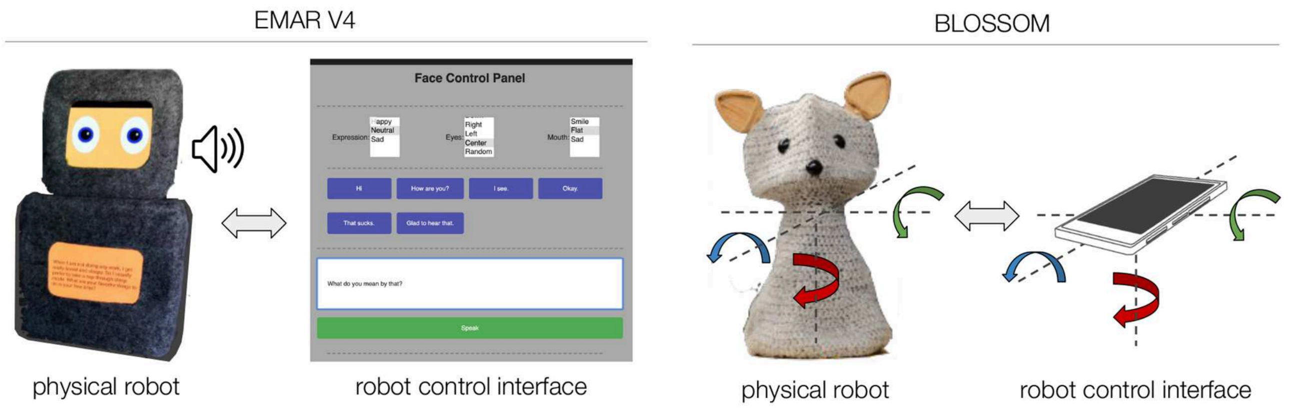


FIGURE 2 | The two robot platforms used in our studies. **(Left)** EMAR V4 and the browser-based control interface for making the robot speak and changing its facial expression. **(Right)** Blossom and the mobile phone interface for controlling the robot's pose and movement.

TABLE 1 | Participant ages and grade levels.

School	*n*	Age (m)	Grade (m)
1	23	17.61	11.96
2	20	16.07	11.96
3	14	16.25	10.58
4	5	17.00	11.00

robots were chosen given their specific functionalities (speech or movement) and limitations (lack of speech or movement). In addition, both of these platforms had simple and intuitive control interfaces, which was key to enabling our research method in which teens take the role of wizard to control the robot. The two platforms and their control interfaces are shown in **Figure 2**.

EMAR V4 is a social robot designed for facial expression and speech communication. It has a box-like structure based upon previous design requirements from teens (Björling et al., 2018). It has two Nexus 7 tablets encased in a soft felt body. One tablet is used as the robot's face, which is a web application running on a browser on the tablet. The face has two eyes that blink and its facial expression can be changed. This tablet is also used to make the robot speak using the browser's text-to-speech capability. The other tablet is located at the robot's belly and is intended as an input/output touchscreen for communication with the user.

The robot's actions are controlled through another browser-based "Wizard of Oz" interface. In this study, the primary way in which the robot responded to the user was through speech. The control interface has a small number of buttons corresponding to simple pre-specified utterances (e.g., "I see," "That sucks") that the wizard can trigger in response to the user's utterances. The interface also includes a free form text box that the wizard can type in what they want the robot to say. In addition to making the robot speak, the wizard had control over what facial expression the robot displayed (neutral, happy, sad) and where the robot looked (center, up, down, left, right, randomized). Components of the robot (two tablets) and the control interface communicated through a real-time database. The robot can be customized in different ways such as changing features of the face, using shells of different size, color, or material, and dressing the robot with additional accessories.

Blossom is a soft-bodied, flexible robot with a crocheted outer shell and a 3 degrees of freedom inner mechanism (Suguitan and Hoffman, 2018). The robot has no facial expression or speech capabilities and therefore represents movement as its only response. The mechanism allows the robot to rotate around the vertical axis (pan) and bend its neck down in any direction (tilt).

Blossom is teleoperated in real time using a smartphone with a gyroscope and magnetoscope. The pan/tilt angles of the smartphone determined by these sensors are directly mapped to the robot's neck pan-tilt angles. This is done in a tight loop that enables continuous motions of the robot to be transformed into continuous robot motions such as nodding or shaking the robot's head.

4. METHODS

4.1. Sample

The study was conducted in four Pacific Northwest urban, public high schools. Teens were recruited from a physics class, a computer science class, an after school STEM club, and a Girls Who Code club. Participants were asked demographic questions including age, grade, and self-reported gender and ethnicity. No identifying information (names or contact information) was gathered. We captured data from 62 teens between the ages of 14 and 18 (M = 16.77) and in grades 9–12 (M = 11.13), see **Table 1**. Twenty-four females, 32 males, and 5 teens who identified as non-binary, participated in our interaction study. Teens were invited to self-identify their ethnicity in an open question. See **Figure 3** for a summary of reported ethnicities.

4.2. Instruments

4.2.1. Negative Stress

It was important to understand teens' stress levels. Therefore, self-reported stress reflecting on the past month was captured using the Perceived Stress Scale (Cohen and Williamson, 1988) as part of our intake questionnaire. The PSS is a 10-item questionnaire that measures the degree to which situations in one's life are appraised as stressful.

4.2.2. Robot Attitudes

In order to capture teens' beliefs about robots, participants completed a slightly modified, 10-item version of the Negative Attitudes Toward Robots Scale (NARS) (Nomura et al., 2006). NARS has been used in many experiments to evaluate participant attitudes toward many kind of robots. It consists of three subscales:

S1: Negative Attitude toward Situations and Interaction with Robots (6 items).
S2: Negative Attitude toward Social Influence of Robots (5 items).
S3: Negative Attitude toward Emotions in Interaction with Robots (3 items).

To make NARS appropriate for teenagers we removed questions that were written from an adult's perspective, such as "I am afraid that robots may negatively influence children's mind" (S2) and "I feel anxiety when I imagine that I may be employed and assigned to a workplace where robots should be used" (S1). We retained all of S3 as we were most interested in teen's attitudes related to emotions in robot interactions. We also added three new items related sharing data with robots and the general role of robots (Q8–Q10 in **Table 2**). We used the standard NARS 5-point Likert scale from Strongly Disagree (1) to Strongly Agree (5) for all items.

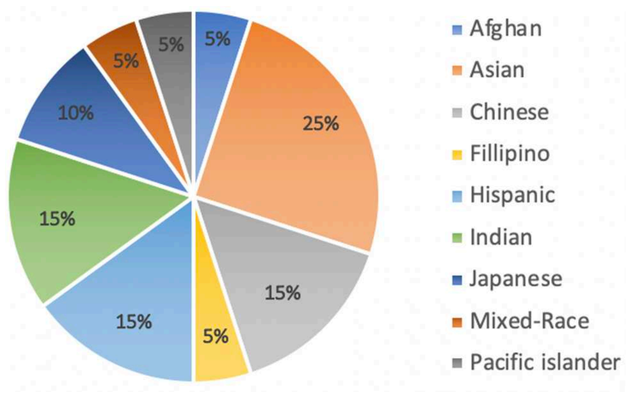


FIGURE 3 | Self-reported ethnicities of our non-white participants ($n = 22$).

4.2.3. Interaction Survey

We created a brief survey to capture data from the PWoz interactions. We asked wizards, users and witnesses to respond to the survey after each teen-robot interaction activity. The PWoz survey consisted of a brief 7-point Likert scale in response ranging from 1 = not at all to 7 = very much to two items about the interaction for the users and the witnesses. For the operator, we asked an open-ended question shown below which led to descriptions regarding how they tried to operate the robot to show it was listening.

- Operator: How did you try to communicate that the robot was listening? (open-ended).
- Users: How much do you think the robot was listening to your stress story? (7-point Likert).
- Witnesses: How much do you think the robot was listening to the speaker? (7-point Likert).

4.2.4. Exit Interview

The exit interview was a customized, single question prompt with probes targeted toward concerns teens may have about the robot. "If a robot were in your school to help with stress, what concerns might you have?"

4.3. Ethics

The research was reviewed and approved by university Internal Review Board and school district research review. Students who were under 18 also obtained parental permission for their participation in our research study. No personal identifiers were captured during the study, only study ID numbers were assigned to identify participants. Photos and videos were taken for research purposes and parents and teens had the option to

TABLE 2 | Baseline measurement of teen attitudes toward robots.

		Question	Mean	*SD*
Q1	NARS	I would feel uneasy if robots really had emotions.	2.69	1.16
Q2*	NARS	I would feel relaxed talking with robots.	3.26	1.08
Q3*	NARS	If robots had emotions, I would be able to make friends with them.	3.23	1.19
Q4*	NARS	I feel comforted being with robots that have emotions.	2.74	0.981
Q5	NARS	**I would feel very nervous just standing in front of a robot.**	**4.02**	1.09
Q6	NARS	I would feel nervous talking with a robot in front of other people.	3.11	1.24
Q7	NARS	**I would feel paranoid talking with a robot.**	**3.54**	1.12
Q8*		I would trust a robot with my data.	2.57	1.10
Q9*		I would feel comfortable sharing my emotional data with a robot.	2.98	1.18
Q10*		**I think robots can help people.**	**4.46**	0.91

**Indicates reverse coded. There were $n = 61$ teen participants who completed the attitude survey before participating in robot design activities. Items where most participants selected "Strongly Disagree" or "Strongly Agree" (reverse coded) are in bold.*

also opt in to give permission for the photos to be used for social media and research publications. Given the importance of maintaining trusting relationships with teens (Björling and Rose, 2019) and their school communities, no deception was used in our study. In fact, teens were told up front about our research and our project, our intention for this particular study, as well as our process of using a participatory, human-centered design approach. Teens were made aware of how the robots were programmed and operated. All teens had the option to opt out of any activity at any time. A few teens refused or forgot to complete a survey, but all teens engaged in the interactions and often seemed disappointed when the activity was over. No personal data (names or contact information) were captured at any time. Teens were assured their video and interaction data would be used for research purposes only. Data were stored in password protected and university approved online database and were accessible only to the research team.

4.4. Study Procedure

The following section describes the study procedures and data collected as part of the study.

4.4.1. Introduction to the Study

All studies were conducted in high school classrooms. We arrived and set up multiple stations including multiple versions of V4 and Blossom with one researcher facilitating the interactions for each station. Before beginning, we presented an overview of the project to the whole group as well as provided some background on the process of human-centered design. Teens were reminded of the consent process and their option to disengage at any time during any of the activities.

4.4.2. Questionnaires, Scripts, and Storyboards

Teens completed intake questionnaires (Demographics, NARS, PSS) and were then divided into pairs to create a stress story (either a script or a storyboard scenario). Teens collaborated together to create these materials for use in the study. We observed that this collaboration was engaging for teens and elicited a great deal of data as teens worked together to illustrate their experiences of stress.

4.4.3. Teen-Robot Interaction Activity

After completing stress stories, teens were assigned to groups of 3-4 and were directed to interaction activity stations. The number of groups was dependent upon the sample at the site which can be referenced in **Table 1**. At larger sites (schools 1 and 2), teen groups were randomly assigned to an interaction station with either Blossom or V4. At smaller sites (schools 3 and 4), teens had time to interact with each of the robots.

At their interaction station, teens received a brief overview of the platform (V4 or Blossom) and how it is controlled. They chose a role (user, wizard, witness) and were assured they could alternate roles if desired. The wizard was shown how to control the platform and decided upon a comfortable position for operation. The wizard was often visible to the user during interaction. The user then chose to share their own stress story, or a script of other stories written by teens. Witnesses were given seats where they could witness the entire interaction. Once the teens were ready for their roles, the researcher started video recording and moved away and observed from a distance to help the teens to feel comfortable. At the end of the interaction, the researcher returned and handed each participant the brief PWoZ questionnaire. Then teens then had the option to change roles to experience another side of the interaction. All groups were able to rotate roles at least once, offering each teen at least two different roles.

4.4.4. Wrap Up and Group Interviews

After the teen-robot interaction activities, the group came back together to complete a final NARS questionnaire. All data were collected and then teens were broken into groups of 4–7 for a group interview. During the group interview, researchers asked questions about their experience and opinions of the robots. Finally, the teens were offered a chance to ask any questions about the study or the robots and given robot stickers as a thank you for their time.

4.5. Analysis

Both the NARS and the PSS data were reverse coded appropriately and then scored. Descriptive analysis was used to explore total scores and individual items. Statistical normality tests were performed. In addition, a repeated measures t-test in SPSS, version 24 was used to detect any differences in the pre and post NARS total or individual items. A one-way ANOVA was used to explore differences in both the NARS and PSS in relation to grade, age, or gender. Likert scale responses from interaction witnesses and users were analyzed using independent samples t-tests to determine any differences between the two robots, EMAR V4 and Blossom.

Exit interviews, storyboards, video interaction data, and open ended responses from wizards were explored qualitatively using a collaborative, applied thematic analysis (Guest et al., 2011). The team of four authors divided data sources by study site and began with review and immersion into the raw data. Using the method of open coding and extraction of salient excerpts, the team made an effort to maintain the context of extracted quotations. The team used the collaborative, online Miro.com (Miro, 2019) site to capture data excerpts including video clips, text and images. The team then met to explore all of the extracted data, the associated context, and preliminary, emergent codes. Through this discussion, substantive themes were collaboratively identified and documented along with their associated evidence. Each author then used a priori coding method on the data again to explore further confirmation of existing themes.

5. FINDINGS

5.1. Complexity of Stress

Given our larger project's aim to design a robot appropriate to both gather stress data, while providing a micro intervention, it is important to capture the general stress level of our teen co-designers. As has been found previously (Björling and Singh, 2017; Björling et al., 2019), the teens in this particular study had high to very high stress levels as indicated by the Perceived Stress Scale (PSS) instrument. As mentioned in the instrumentation section, the PSS is a retrospective self-report referencing the individual's past month. These data allow us to understand the

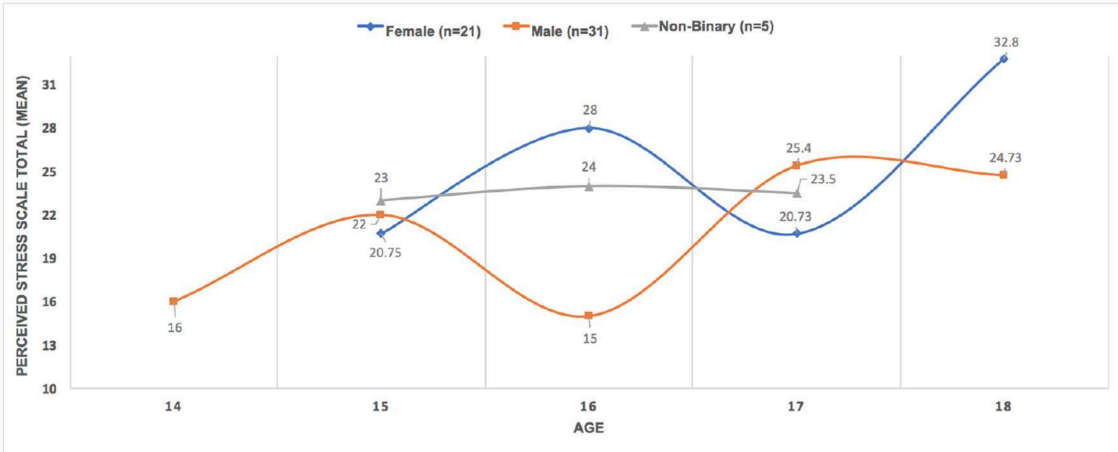


FIGURE 4 | Gender differences in participant mean stress scores.

current context in which our participants are interacting with the robots. The participants' mean stress score of 23.16 (SD 6.67) was much higher than the PSS published norm ($m = 14.2$). Nineteen percent of teens ($n = 12$) scored at the low stress level, 67.7% ($n = 42$) scored at the moderate level, and 9.7% ($n = 6$) scored at the high level. Stress scores did not significantly differ by age or grade, although they were higher for teens in 11th grade. See **Figure 4** for more detail. However, PSS scores were higher for females ($m = 21.09$), compared to males ($m = 17.76$), and significantly higher for participants who identified as non-binary/fluid ($m = 23.25$) [$F = 3.321$ ($df = 2$) $p = 0.043$]. However, it is important to note that that the non-binary/fluid group was very small consisting of only five participants.

Stress was a ubiquitous experience among all the teens in the study. They had no trouble illustrating stress stories or storyboards depicting their recent or common experiences of stress. Teen stories of stress typically illustrated academic stress, commonly related to grades, test scores, or college. Their stress stories illustrated the breadth of their stressors including, relationships, financial worries, and feeling alone. As one teen stress story illustrated, "I just feel like nobody cares about my problems" [Group activity, School 1]. Stress stories included experiences of feeling pressure from teachers, coaches and parents, e.g., "The pressure Ted had been receiving from his father figure was dampening his whole life." [Group storyboard, School 1] Teen stress stories also illustrated the experience of competing priorities, typically described as academic and extracurricular (sports) or paid work. For example, "Sam feels super stressed knowing he has to study for all of his tests while juggling the rest of his life. He has sports practice each day and doesn't know how he'll be able to make everything work" [Group storyboard, School 1].

Finally, some outliers were illustrated in the teen stress stories including the articulation of being stressed about a "sexist teacher" and not knowing how to handle the situation. "There is a very sexist teacher named X at school, but when Hank confronts him, he implies that Hank is stupid… Hank doesn't know what to do" [Group storyboard, School 1].

Teens also shared reasoning for why they are not sharing their stressors with friends and family. One female teen stated, "I know I should be able to talk to people, but I don't want to disappoint anyone" [Group storyboard, School 1]. Another male teen verbally expressed that it is often difficult to talk about stressors with friends and family as often they are part of what is creating stress [Group interview, School 1].

5.2. Attitudes Toward Robots

Total NARS scores ($m = 32.61$) were not significantly different by age. However, similar to the stress scores, the NARS scores did differ significantly by gender with the highest mean score reported by participants who identified as non-binary/fluid ($m = 38.5$) followed by males ($m = 33.3$), and then females ($m = 30.70$) [$F = 3.35$ ($df = 2$), $p = 0.042$]. As far as items on the NARS, most teens strongly disagreed with feeling nervous about standing in front of a robot or talking to a robot. They also felt strongly that robots could help people. See **Table 2** for more detail.

After the teens interacted with the robots in our study as wizards, users, and witnesses, their overall negative attitude scores decreased significantly. Forty-seven participants fully completed all 10 items of the NARS before and after interacting with or operating robots. Significant differences were found including: (1) decreased uneasiness in talking to robots, (2) increased comfort with robots who have emotions, (3) and increased belief that robots can help people. Strongly scored items at intake such as disagreement with the statement about feeling paranoid talking with a robot, and nervousness standing in front of a robot remained stable. See **Table 3** for more detail.

5.3. Comparing EMAR V4 Blossom

A key intention of our study was to explore differences between teen responsiveness to a non-moving, verbal robot (EMAR V4) and a moveable, non-verbal robot (Blossom). However, from our post-interaction surveys with witnesses and users showed no significant differences for EMAR V4 or Blossom. However, on

TABLE 3 | Participant change in NARS item.

Paired differences	*n*	m (diff)	*SD*	*SE*	*t*	*df*	Sig. (2-tailed)
Uneasy	**49**	**−0.469**	**0.915**	**0.131**	**−3.59**	**48**	**0.001**
Relaxed	49	−0.18367	1.01393	0.14485	−1.268	48	0.211
Friends	49	−0.16327	0.74574	0.10653	1.533	48	0.132
Comforted	**49**	**−0.26531**	**0.83605**	**0.11944**	**−2.221**	**48**	**0.031**
Nervous Standing	49	−0.224	0.985	0.141	−1.596	48	0.117
Nervous Talking	49	0.102	1.388	0.198	0.515	48	0.609
Paranoid	48	−0.083	0.986	0.142	−0.586	47	0.561
Trust	49	0.12245	0.9494	0.13563	0.903	48	0.371
Sharing Emotion	48	−0.16667	1.01758	0.14687	−1.135	47	0.262
Robots Help	**49**	**−0.32653**	**0.94401**	**0.13486**	**−2.421**	**48**	**0.019**
NARS Total	**47**	**−1.382979**	**3.892876**	**0.567834**	**−2.436**	**48**	**0.019**

Statistically significant items are in bold.

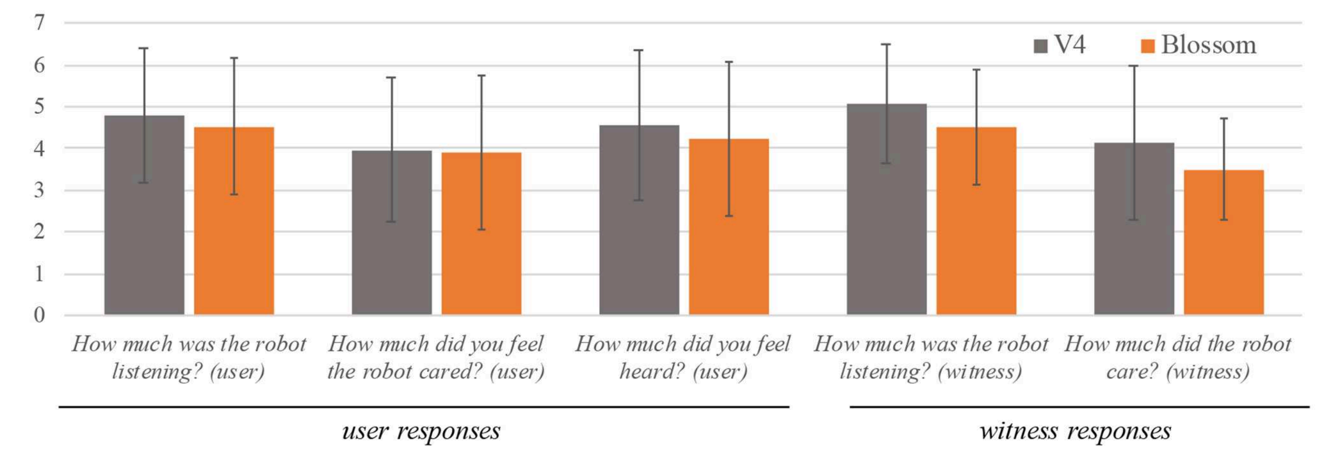


FIGURE 5 | Post-interaction survey responses comparing EMAR V4 and blossom.

TABLE 4 | V4 Operator verbal responses and associated outcomes.

Theme	School	Operator utterance	Outcome	User response
1. Advice	4	"That's too bad, you should try to study a bit more next time."	Disengagement	"Thanks robot, I didn't really need advice."
2. Suggestion	1	"Go listen to some music."	Connection	"Okay. Thanks robot."
3. Empathy: Active	4	"People do care, I care."	Emotional Connection	Touches heart and says, "Thank you. Thanks for hearing me out."
4. Empathy: Passive	4	"That sucks."	Emotional Connection	"Yeah, it does."
5. Humor	1	"I have to deal with miserable kids like you day after day. They should really give me AI so I can help you."	Engagement	Lots of laughter, connection to robot.
6. Inquiry	3	"Why do you think that was so hard for you?"	Engagement	Further discussion / articulation.
7. Reassurance	2	"I am sure everything is going to be okay."	Emotional Connection	Sigh, "Thank you."

all item responses, participants reported EMAR V4 responses as slightly higher than Blossom. See **Figure 5** for more detail.

5.4. Categorizing Wizard Responses and Associated Outcomes

In an overall analysis of all teen-robot interactions including V4 and Blossom, there were several distinguishable categories related to how teens operated the robots. From these interactions, we identified specific responses that led to increased engagement and those that led to disengagement.

5.4.1. V4—Verbal Responses

Teen wizards of V4 had the opportunity to type what V4 said in response to the user's stress. Some of the wizard's responses

Image	Movement	Interpretation	Response
	Slow head tilt	Robot appears to be listening	Engagement, positional mimicry from user
	Looking downward	Robot appears to be sad	Engagement, mimicry from user
	Fast head rotation	Robot appears sympathetic, or in disbelief	Engagement, leaning in from user, connection
	Looking away	Robot appears disengaged	User disengages, looks to other humans
	Looking upwards to speaker	Robot appears ready to listen	Cues user to talk again

FIGURE 6 | Blossom operator behaviors and associated responses.

were very successful at creating engagement or a connection with the user and others were not. For a summary of detailed verbal response themes, see **Table 4**. In almost every example, offering advice to the user led to disengagement, whereas offering an empathic response led to an emotional connection between the teen and the robot. Teens also seemed to greatly appreciate humorous responses and these often led the user to connect to both the robot and the wizard.

5.4.2. Blossom—Movement Responses

Teen wizards of Blossom had the opportunity to use movement as a form of responsiveness to the user's story of stress. **Figure 6** illustrates the six main categories of wizard responses and their associated outcomes. Most teens intuitively used a head tilt, or gentle head turning to convey listening or empathy. Rapid head movement with a downward nod, similar to head shaking, was often used to convey disbelief or understanding in relation to a user's stress. Occasionally, teens turned Blossom's head away from the user (sometimes unintentionally) which often signaled looking at witnesses for confirmation. Surprisingly, even when Blossom's movements were not in line with a typical human response, rarely was there any disengagement with the user or the witnesses.

5.5. Teen-Robot Interactions

Through the data sources: teen stress stories, open-ended wizard surveys, and interaction video data, we observed teens as operators and the effect their operation had on the users' interaction with the robots. As operators, teens attempted to conduct the robot in an appropriate manner and one that would resonate with their peers. In our analysis of the interactions that occurred as a result of teens operating the two different robots, four key themes emerged: (1) Authenticity, (2) Empathy, (3) Emotional Engagement and (4) Imperfection Creates Connection. These themes were supported by multiple pieces of data captured in our interaction study.

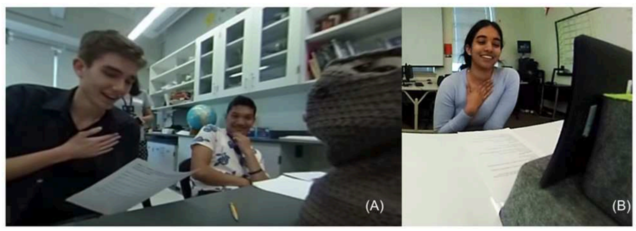


FIGURE 7 | Examples of the *Heartfelt* theme expressed by users during an emotional interaction with the robot. **(A)** Male participant interacting with Blossom. **(B)** Female participant interacting with EMAR.

5.5.1. Authentic Operators

Authenticity appeared numerous times in reviewing the open-ended survey and interview data from teens who had operated EMAR V4 and Blossom robots. We saw several examples of teens articulating their attempts to be "real" or "authentic" in their operation of the robot and even in their responses to the robots. One operator of EMAR V4 said, "I tried to say things that I would say to my friend and say things that seemed genuine" (Group interview, School 3). Authenticity also appeared before we had conducted any robot interactions. At our first study site (School 1), we had crafted scripts for the users to read to read the robot in the event they did not want to share their own stress story. Although based in teen data, these scripts were written by our research team. Two teens immediately commented in the margins on the script about our manufactured scripts. One noted, "This is now just making fun of stress rather than dealing with it." Therefore, we iterated on the method and immediately asked teens to write the actual scripts, verbatim, to use for future interactions. Apparently these teen-written scripts seemed authentic as we never again received negative feedback.

5.5.2. Operating for Empathy

Teen operators were asked to try to make the user feel heard during the interaction. Many teens described attempting to operate the robot in an "empathetic" or "sympathetic" manner. One teen reported, "I was trying to be empathetic, I wish there was more preloaded conversations" [P151, Operator survey, School 4]. As users, teens also expressed feeling empathy from the robot, "The robot didn't say anything, but the movements showed it cared" [P74, Operator survey, School 2]. Overall teens seemed genuinely interested in showing empathy through the robot which often led to an emotional response from the users and witnesses.

5.5.3. Emotional Engagement

The manner in which the teens operated the robots often led to an emotional response from the teen user and sometimes the witnesses.

As users, many of the teens felt the robots cared about them. Throughout participant interactions with the V4 and Blossom robots, our research team observed participants exhibiting engaged facial expressions, body behavior, and verbal responses. We considered the highest level of engagement to be the combination of all three response types. We witnessed on multiple occasions the "heartfelt" gesture which combined a smile (facial expression), hand across the chest (body behavior), and "Aww…" verbal response (**Figure 7**).

Teens often made strong eye contact (**Figure 8A**) during conversation with the robots. At times, they looked for cues of responsiveness from others (especially in relation to Blossom) before continuing their story. Teens also from time to time touched the robot, not typically during interaction, but in-between interactions or when an interaction was completed (**Figure 8B**). Many times a users' response to the robot was laughter and engagement (**Figure 8C**).

Teens also used social referencing (**Figure 8D**) in response to a particularly salient moment in the robot interaction. During social referencing a teen looks for another teen to acknowledge their experience in that moment. Teens often did this when the robot said something funny, surprising, or truly empathetic. Witnesses were important participants in the interaction. During the study, witnesses were often part of the collective response to the interactions between the wizards and the users interacting with the robot. Witnesses often had similar reactions to the person interacting with the robot. For example, when the wizard had the robot say or do something empathetic, the witnesses would also respond in a similar way to the user interacting the robot. For example, when an operator said "That sucks" in a response to a teens stress story, the witnesses laughed along with the user having the interaction. Further, we observed the witnesses glancing and making eye contact with one another during the robot interactions.

5.5.4. Imperfection Creates Connection

When a novice operator commanded a robot, human error became a part of the human-robot interaction. An EMAR V4 operator would press the return key multiple times and a phrase would be repeated by the robot. A Blossom operator would turn the mobile phone controller too far, causing the robot to spin completely around. These unintended actions, resulting from operator "error" caused unexpected outputs, but interestingly evoked strong engagement (often a response of smiling and laughter) in the user and the witnesses.

In one example, a study participant in the operator role submitted a command to have V4 say, "Wow." V4 executed the command and spoke out loud, "Wow, wow, wow, wow." This caused smiling and laughter with all three study participants: the operator, user, and witness. Instead of evoking frustration or irritation from any of the teens involved, they seemed to genuinely enjoy the operator, and thereby robot, imperfections. One might expect that after multiple errors, teens would become frustrated or disengaged, but the opposite seemed to be true in many cases.

Finally, in an exit interview a male participant described Blossom, "I feel like it has a little personality. The way it moves. …even it being a little hard to control makes it seems a little bit real" [Group interview, School 1]. In this example the link between imperfection and realness, suggests that being real or authentic is good and that being fallible is part of that realness.

6. DISCUSSION

The breadth of physiological and cognitive responses to robots is challenging to observe without entering more bias into the system. Responses are dynamic, conscious or unconscious, microscopic or macroscopic, and often differ from expressed attitudes. This is why we structured our study to gather attitudinal and behavioral data from multiple viewpoints.

Teens enjoyed participating in this method and found it engaging. And although we found no significant difference between Blossom and V4 measures post-interaction, we did find measurable reduction on the Negative Attitudes Scale, likely resulting from robot interaction.

FIGURE 8 | Examples of the emotional responses teens had during robot interaction. Eye contact **(A)**, physical contact **(B)**, laughing **(C)**, social referenceing **(D)**.

6.1. Authenticity and Imperfection

We repeatedly saw teens attempt to be genuine and authentic operators as well as strong emotional responses from users in both types of interaction. The importance of authenticity for teens is not a new concept (Ullman, 1987; Chessick, 1996). More recently, authenticity has been found an important component in teen health education (Grabowski and Rasmussen, 2014), the success of older teens in college (Lenz et al., 2016), and has been shown to be an integral component of successful mental health counseling for teens (Holliman and Foster, 2016). So it should be no surprise that teens attempted and appreciated authentic behavior authentic behavior, even through the use of a social robot agent.

Robot imperfection has also been studied previously. Mirnig et al. (2017) purposefully programmed faulty behavior into a robot's interaction in order to understand the impact of faults on likability. They found participants preferred the faulty robot interaction significantly more than the flawless interaction. They also showed that shifts in gaze and laughter are typical human reactions to unexpected and imperfect robot behavior. This is very similar to what we saw in our teen-robot interactions.

The teens desire for authenticity, discussed above, might also explain why teen imperfection during robot operation increased engagement and human to human connection during the interactions. Mistakes are human and therefore, reveal the authentic behaviors of humans.

6.2. Active Listening

"It could move… it felt more like it was actively listening to you" [P154, School 4].

We identified that successful teen-robot interactions, ones that gained a positive response and strengthened engagement, often included components of active listening. Active listening is an empathetic and therapeutic human to human interaction focused on reflection and empathetic expressions described by Carl Rogers a psychotherapist (Rogers and Farson, 1957). Rogers felt that when people are fully listened to, rather than given advice, or asked to think differently, they can hear themselves more clearly, thus bringing about emotional maturity. The powerful human to human engagement of active listening was the impetus for the first chatbot, Eliza (Rzepka and Araki, 2015) and perhaps a reason that the Eliza program became so engaging for users.

Findings from our study point to the importance of non-verbal and verbal signs of active listening in the human-robot relationship. In the case of non-verbal communication, study participants displayed the strongest emotional connection with the robot when smiling, laughing, and making direct eye contact. These non-verbal communications could be reflected or mirrored by a robot as a sign of attentive listening in future versions of EMAR.

6.3. Human to Human Connections

Finally, it is worth noting that in the design and development of our social robot, we have heard concerns from adults (researchers, teachers and colleagues) about the downside of creating a digital agent that teens find truly engaging. The concern raised is often that the robot engagement will mimic that of cell phones and social media which have been suggested as addictive and potentially leading to poor mental health in teens (Twenge, 2019). The main concerns raised by adults about our social robot is that as teens increasingly engage with a digital device, they will further disconnect with the humans in their lives. However, during this interaction study, we see quite the opposite effect. Just as the teen-robot engagement is strong, so are the teen-teen connections during robot interactions. Teens use of social referencing to connect with other teens during interactions and to seek out teen-teen engagement during teen-robot interactions is reassuring that a social robot may be encouraging human-human interaction. This also has been studied in HRI research (Wada et al., 2005; Kim et al., 2013). Finally, teens chose to engage with the robot operator during this study, suggesting that they see through the robot to the operator and can connect with both agents simultaneously. Given that the robot is being designed for a public space and will be interacting with groups of teens, paying attention to the collective responses of the witnesses and the resulting group interactions helps to better understand the communal interaction.

6.4. Reflections on the Participatory Wizard of Oz Method

Due to our desire to collect data from the study that was congruent with a human-centered design and participatory approach, we developed a new method of Participatory Wizard of Oz (PWoz). This method has a variety of benefits for collecting data and also some limitations.

In terms of conducting a study, PWoZ offers benefits for participants. First, it provides a level of ecological validity by having the teens themselves develop the scripts and operate the robots for the study. Their actions and choices provide the primary direction and data for the study. Further, making both the constraints and limitations of the robot makes the technology more transparent. The low fidelity nature of the robots are revealed to the participants in the study providing a more realistic impression of the technological capability of the robots, which is quite modest. It both refuses to over promise the ability for social robots to function completely autonomously and also reveals the limitations of the data collected. Second, it provides a more authentic interaction between teens and robots. As evidenced of the teen scripts and the ability to operate the robots, which is more authentic than the adults taking part in the study and making assuming or presupposing what interactions teens might want or find to be authentic. Third, students were highly engaged, and described their participation as fun and enjoyable. The interaction added more humanness into the prospect of designing a robot. Exposing teens to the design of social robots and concepts related to human centered design are promising to engage more young people in STEM related activities and could potential stoke future interest.

In addition to the enjoyable experience of being in the study, in the study, this method provides rich and layered data to inform the design of robots. We used 360° cameras to capture the activity of all the participants in the study: Wizards, users, and Witnesses. This method gives us as much data about the users and witnesses as it does about the Witnesses who are watching the interaction. Further, we can see this data in real time and simultaneously. The layered aspect to this data allows the team to look at the different view points of the interactions. Finally, this approach allows for input from users earlier in the design process, it does not require a fully functional or autonomous robot to get direction from end users about a whole host of considerations for design.

While there were a variety of significant benefits to using this method, there were also limitations to consider and adapt to in the future. There are clear technical limitations given the inability to control the interaction, possibly resulting in a less systematic exploration of possible robot behaviors. Further, this method eliminates the illusion of interacting with an autonomous robot, thus making it difficult to determine how interactions may change once the robot is autonomous.

Implementing this method also proved challenging. First, the tone and directions need to be clearly communicated to participants. We felt fortunate in this study that teens were engaged, cooperative, and interested. This was in part due to the relationship building and partnerships with schools, teachers, and advisors that we had established over time. We could imagine that another group that had not been primed or was less socially connected might pose challenges in having an effective data collection session. Numerous challenges exist when designing and testing social robots in the wild. As discussed by Šabanović et al. (2014) difficulties measuring interactions are compounded by our naturalistic environment, but did offer access to a situated, real-world user experience. Another challenge to doing PWoz in the wild is that each group and each location can provide different challenges. Given that the study was conducted in 4 different locations, sometimes the constraints in each location lead to changes or adjustments in the set up. Finally, while the scripts were written by teens and contained details that were authentic and based on teens experience of their lives, the process of reading them aloud to channel a specific emotion was somewhat artificial. The participants in the study were willing to do this but often did it in a playful way, it was still a simulation. Overall, we found the method to be a promising way to engage teens in the design of a social robot and plan to continue to implement and refine it in future studies.

7. LIMITATIONS AND NEXT STEPS

We chose to conduct all of our studies in the wild (school settings) in order to maintain ecological validity, however, this also meant that many factors were out of our control. For instance, students' operation of or responsiveness to the robot could have been heavily moderated by the room they were in, or the students in their group, as they were self-selected. In addition, our sample was fairly homogeneous given our geographic location and thus, it is important to consider similar studies in other locations, e.g., rural. There may also be cultural aspects (school culture and ethnic cultures) that have influenced our data or how we perceive our data. therefore, continuing to diversify our participant sample and our research team is important. Finally, the teens were greatly limited by the prototype technology and had we presented them with a more robust device, they may have had very different experiences. Teens also did not have much time to become comfortable with the device and the devices were very novel. Both of these factors likely influenced our data and need to be taken into consideration.

8. CONCLUSION

Using social robots to help teens address stress is a promising application. In this study, we privileged the experiences and voices of teens through human-centered design and participatory design to learn more about their needs and preferences for interactions with a social robot prototype. While there were no significant differences between the two social robots that teens interacted with, the rich data collected through the PWoZ method lead to a variety of insights about teens' desires for robots to be authentic, imperfect, and active listeners.

DATA AVAILABILITY STATEMENT

The datasets for this article are not publicly available due to the identifiable nature of our data and as it incorporates minors. Requests to access the datasets should be directed to the corresponding author.

ETHICS STATEMENT

This study involved human participants and was reviewed and approved by University of Washington Institutional Review Board. Verbal, informed consent to participate in this study as well as the option to consent to the use of a participant's image for research communication purposes was obtained. In addition, written parental/guardian permission for child participation and optional use of child images was obtained in the case of participants who were minors at the time of data collection.

AUTHOR CONTRIBUTIONS

EB and MC conceived of the study idea and administered the study with a team of students. MC developed the technology (EMAR V4) specifically for this study. EB created the study design, and analyzed the quantitative data. EB, KT, ER, and MC reviewed all of the data, both quantitative and qualitative, and collaborated to analyze the qualitative data and determine salient findings, and contributed to the writing and revising of the paper.

FUNDING

This study was funded in part by the National Robotics Initiative, NSF award. This work was supported by the National Science Foundation award SES-1734100 NRI: INT: Design and Development of a Social Robot for Gathering Ecological Momentary Stress Data from Teens.

ACKNOWLEDGMENTS

We acknowledge the many teens who participated in this study and gave us valuable feedback.

REFERENCES

Abd-alrazaq, A. A., Alajlani, M., Alalwan, A. A., Bewick, B. M., Gardner, P., and Househ, M. (2019). An overview of the features of chatbots in mental health: a scoping review. *Int. J. Med. Informat.* 132:103978. doi: 10.1016/j.ijmedinf.2019.103978

Akers, D. (2006). "Wizard of oz for participatory design: inventing a gestural interface for 3d selection of neural pathway estimates," in *CHI'06 Extended Abstracts on Human Factors in Computing Systems* (Montreal, QC: ACM), 454–459.

American Psychological Association (2014). *Stress in America 2013 Highlights: Are Teens Adopting Adults' Stress Habits*. American Psychological Association.

Beyer, H., and Holtzblatt, K. (1998). *Contextual Design: Defining Customer-Centered Systems*. San Francisco: Morgan Kauffman.

Biegel, G. M., Chang, K., Garrett, A., and Edwards, M. (2014). *Mindfulness-Based Treatment Approaches: Clinician's Guide to Evidence Base and Applications, 2nd Edn*. London: Academic Press. 189–212.

Björling, E. A., and Rose, E. (2019). Participatory research principles in human-centered design: engaging teens in the co-design of a social robot. *Multimodal Technol. Interact.* 3:8. doi: 10.3390/mti3010008

Björling, E. A., Rose, E., and Ren, R. (2018). "Teen-robot interaction: a pilot study of engagement with a low-fidelity prototype," in *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI '18)*. (New York, NY: Association for Computing Machinery), 69–70.

Björling, E. A., and Singh, N. (2017). Exploring temporal patterns of stress in adolescent girls with headache. *Stress Health* 33, 69–79. doi: 10.1002/smi.2675

Björling, E. A., Stevens, C., and Singh, N. B. (2019). Participatory pilot of an art-based mindfulness intervention for adolescent girls with headache. *Art Ther.* 36, 86–92. doi: 10.1080/07421656.2019.1609325

Branson, V., Palmer, E., Dry, M. J., and Turnbull, D. (2019). A holistic understanding of the effect of stress on adolescent wellbeing: a conditional process analysis. *Stress Health.* 35, 626–641. doi: 10.1002/smi.2896

Breazeal, C. (2011). "Social robots for health applications," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE* (Boston, MA: IEEE), 5368–5371. doi: 10.1109/IEMBS.2011.6091328

Buchanan, R. (2001). Human dignity and human rights: thoughts on the principles of human-centered design. *Design Issues* 17, 35–39. doi: 10.1162/074793601750357178

Carter, S., Mankoff, J., Klemmer, S. R., and Matthews, T. (2008). Exiting the cleanroom: on ecological validity and ubiquitous computing. *Hum. Comput. Interact.* 23, 47–99. doi: 10.1080/07370020701851086

Chessick, R. D. (1996). Heidegger's "authenticity" in the psychotherapy of adolescents. *Am. J. Psychother.* 50, 208–216. doi: 10.1176/appi.psychotherapy.1996.50.2.208

Cohen, S., and Williamson, G. (1988). "Perceived stress in a probability sample of the United States," in *The social Psychology of Health: Claremont Symposium on Applied Social Psychology*, eds S. Spacapan and S. Oskamp (Newbury Park, CA: Sage), 31–67.

DiSalvo, C., Nourbakhsh, I., Holstius, D., Akin, A., and Louw, M. (2008). "The neighborhood networks project: a case study of critical engagement and creative expression through participatory design," in *Proceedings of the Tenth Anniversary Conference on Participatory Design 2008* (Bloomington, IN), 41–50.

Druin, A. (2002). The role of children in the design of new technology. *Behav. Informat. Technol.* 21, 1–25. doi: 10.1080/01449290110108659

Duckworth, D., Henkel, Z., Wuisan, S., Cogley, B., Collins, C., and Bethel, C. L. (2015). "Therabot: the initial design of a robotic therapy support system," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts* (Portland, OR), 13–14. doi: 10.1145/2701973.2701993

Edwards, M., Adams, E. M., Waldo, M., Hadfield, O., and Biegel, G. M. (2014). Effects of a mindfulness group on latino adolescent students: examining levels of perceived stress, mindfulness, self-compassion, and psychological symptoms. *J. Spec. Group Work* 39, 145–163. doi: 10.1080/01933922.2014.891683

Eiraldi, R., Wolk, C. B., Locke, J., and Beidas, R. (2015). Clearing hurdles: the challenges of implementation of mental health evidence-based practices in under-resourced schools. *Adv. School Ment. Health Promot.* 8, 124–140. doi: 10.1080/1754730X.2015.1037848

Feil-Seifer, D., and Mataric, M. J. (2005). "Defining socially assistive robotics," in *9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005* (Chicago, IL: IEEE), 465–468.

Fernaeus, Y., Håkansson, M., Jacobsson, M., and Ljungblad, S. (2010). "How do you play with a robotic toy animal?: a long-term study of pleo," in *Proceedings of the 9th international Conference on interaction Design and Children* (Barcelona: ACM), 39–48.

Forlizzi, J. (2007). "How robotic products become social products: an ethnographic study of cleaning in the home," in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction* (Arlington, VA: ACM), 129–136.

Gabriels, K. (2019). Response to "uncertainty in emotion recognition". *J. Informat. Commun. Ethics Soc.* 17, 295–298. doi: 10.1108/JICES-07-2019-0072

Giannopulu, I., Terada, K., and Watanabe, T. (2018). Emotional empathy as a mechanism of synchronisation in child-robot interaction. *Front. Psychol.* 9:1852. doi: 10.3389/fpsyg.2018.01852

Grabowski, D., and Rasmussen, K. (2014). Authenticity in health education for adolescents: a qualitative study of four health courses. *Health Educ.* 114, 86–100. doi: 10.1108/HE-06-2013-0020

Greer, S., Ramo, D., Chang, Y.-J., Fu, M., Moskowitz, J., and Haritatos, J. (2019). Use of the chatbot "vivibot" to deliver positive psychology skills and promote well-being among young people after cancer treatment: randomized controlled feasibility trial. *JMIR mHealth uHealth* 7:e15018. doi: 10.2196/15018

Guest, G., MacQueen, K. M., and Namey, E. E. (2011). *Applied Thematic Analysis*. Thousand Oaks, CA: Sage Publications.

Hamilton, J. L., Potter, C. M., Olino, T. M., Abramson, L. Y., Heimberg, R. G., and Alloy, L. B. (2016). The temporal sequence of social anxiety and depressive symptoms following interpersonal stressors during adolescence. *J. Abnorm. Child Psychol.* 44, 495–509. doi: 10.1007/s10802-015-0049-0

Holliman, R. P., and Foster, R. D. (2016). Embodying and communicating authenticity in adolescent counseling. *J. Child Adolesc. Counsel.* 2, 61–76. doi: 10.1080/23727810.2016.1160353

Jeong, S. (2017). *The impact of social robots on young patients' socio-emotional wellbeing in a pediatric inpatient care context*. (Ph.D. Thesis). Massachusetts Institute of Technology, Cambridge, MA, United States.

Juster, R.-P., McEwen, B. S., and Lupien, S. J. (2010). Allostatic load biomarkers of chronic stress and impact on health and cognition. *Neurosci. Biobehav. Rev.* 35, 2–16. doi: 10.1016/j.neubiorev.2009.10.002

Kelley, J. F. (1984). An iterative design methodology for user-friendly natural language office information applications. *ACM Trans. Informat. Syst.* 2, 26–41. doi: 10.1145/357417.357420

Kim, E. S., Berkovits, L. D., Bernier, E. P., Leyzberg, D., Shic, F., Paul, R., et al. (2013). Social robots as embedded reinforcers of social behavior in children with autism. *J. Autism Dev. Disord.* 43, 1038–1049. doi: 10.1007/s10803-012-1645-2

Ladner, R. E. (2015). Design for user empowerment. *Interactions* 22, 24–29. doi: 10.1145/2723869

Leite, I., Castellano, G., Pereira, A., Martinho, C., and Paiva, A. (2012). "Long-term interactions with empathic robots: evaluating perceived support in children," in *International Conference on Social Robotics* (Chengdu: Springer), 298–307.

Lenz, A. S., Holman, R. L., Lancaster, C., and Gotay, S. G. (2016). Effects of relational authenticity on adjustment to college. *J. Coll. Counsel.* 19, 2–16. doi: 10.1002/jocc.12027

Logan, D. E., Breazeal, C., Goodwin, M. S., Jeong, S., O'Connell, B., Smith-Freedman, D., et al. (2019). Social robots for hospitalized children. *Pediatrics* 144:e20181511. doi: 10.1542/peds

Maughan, B., Collishaw, S., and Stringaris, A. (2013). Depression in childhood and adolescence. *J. Can. Acad. Child Adolesc. Psychiatry* 22:35.

Mazurka, R., Wynne-Edwards, K., and Harkness, K. (2016). Stressful life events prior to depression onset and the cortisol response to stress in youth with first onset versus recurrent depression. *J. Abnorm. Child Psychol.* 44, 1173–1184. doi: 10.1007/s10802-015-0103-y

McEwen, B. S., and Morrison, J. H. (2013). The brain on stress: vulnerability and plasticity of the prefrontal cortex over the life course. *Neuron* 79, 16–29. doi: 10.1016/j.neuron.2013.06.028

Millen, D. R. (2000). "Rapid ethnography: time deepening strategies for HCI field research," in *Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (New York City, NY: ACM), 280–286.

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

Miro (2019). *Online Whiteboard for Visual Collaboration*. Available online at: https://miro.com/app/dashboard/ (accessed November 11, 2019).

Nomura, T., Suzuki, T., Kanda, T., and Kato, K. (2006). Measurement of negative attitudes toward robots. *Interact. Stud.* 7, 437–454. doi: 10.1075/is.7.3.14nom

Olson, J. S., and Kellogg, W. A. (2014). *Ways of Knowing in HCI*, Vol. 2. New York City, NY: Springer-Verlag.

Oulasvirta, A., Kurvinen, E., and Kankainen, T. (2003). Understanding contexts by being there: case studies in bodystorming. *Pers. Ubiquit. Comput.* 7, 125–134. doi: 10.1007/s00779-003-0238-7

Rodil, K., Rehm, M., and Krummheuer, A. L. (2018). "Co-designing social robots with cognitively impaired citizens," in *Proceedings of the 10th Nordic Conference on Human-Computer Interaction, NordiCHI '18* (New York, NY: ACM), 686–690.

Rogers, C. R., and Farson, R. E. (1957). *Active Listening*. Chicago, IL: Industrial Relations Center of the University of Chicago.

Rose, E. J., Björling, E., and Cakmak, M. (2019). "Participatory design with teens: a social robot design challenge," in *Proceedings of the 18th ACM International Conference on Interaction Design and Children* (ACM), 604–609.

Rose, E. J., and Björling, E. A. (2017). "Designing for engagement: using participatory design to develop a social robot to measure teen stress," in *Proceedings of the 35th ACM International Conference on the Design of Communication* (Boise, ID: ACM), 7.

Rzepka R., Araki K. (2015) "ELIZA Fifty Years later: an automatic therapist using bottom-up and top-down approaches," in *Machine Medical Ethics*. Intelligent Systems, Control and Automation: Science and Engineering, Vol. 74, eds S. van Rysewyk and M. Pontier (Cham: Springer).

Šabanović, S., Michalowski, M. P., and Simmons, R. (2006). "Robots in the wild: observing human-robot social interaction outside the lab," in *Advanced Motion Control, 2006. 9th IEEE International Workshop on* (Istanbul: IEEE), 596–601.

Šabanović, S., Reeder, S. M., and Kechavarzi, B. (2014). Designing robots in the wild: *in situ* prototype evaluation for a break management robot. *J. Hum. Robot Interact.* 3, 70–88. doi: 10.5898/JHRI.3.1.Sabanovic

Sabelli, A. M., Kanda, T., and Hagita, N. (2011). "A conversational robot in an elderly care center: an ethnographic study," in *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on* (Lausanne: IEEE), 37–44.

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

Schuler, D., and Namioka, A. (1993). *Participatory Design: Principles and Practices*. Boca Raton, FL: CRC Press.

Simon, P., Krishnan-Sarin, S., and Huang, T.-H. (2019). On using chatbots to promote smoking cessation among adolescents of low socioeconomic status. *arXiv: 1910.08814*.

Suguitan, M., and Hoffman, G. (2018). "Blossom: a tensile social robot design with a handcrafted shell," in *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (Chicago, IL: ACM), 383–383.

Thapar, A., Collishaw, S., Pine, D. S., and Thapar, A. K. (2012). Depression in adolescence. *Lancet* 379, 1056–1067. doi: 10.1016/S0140-6736(11)60871-4

Ting, S. A., Sullivan, A. F., Boudreaux, E. D., Miller, I., and Camargo, C. A. Jr. (2012). Trends in us emergency department visits for attempted suicide and self-inflicted injury, 1993–2008. *Gen. Hosp. Psychiatry* 34, 557–565. doi: 10.1016/j.genhosppsych.2012.03.020

Twenge, J. M. (2019). More time on technology, less happiness? Associations between digital-media use and psychological well-being. *Curr. Dir. Psychol. Sci.* 28, 372–379. doi: 10.1177/0963721419838244

Ullman, C. (1987). From sincerity to authenticity: adolescents' views of the "true self". *J. Pers.* 55, 583–595. doi: 10.1111/j.1467-6494.1987.tb00453.x

Vogel, S., and Schwabe, L. (2016). Learning and memory under stress: implications for the classroom. *npj Sci. Learn.* 1:16011. doi: 10.1038/npjscilearn.2016.11

Wada, K., Shibata, T., Saito, T., Sakamoto, K., and Tanie, K. (2005). "Psychological and social effects of one year robot assisted activity on elderly people at a health service facility for the aged," in *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on* (Barcelona: IEEE), 2785–2790.

Walton, R. (2016). Supporting human dignity and human rights: a call to adopt the first principle of human-centered design. *J. Tech. Writ. Commun.* 46, 402–426. doi: 10.1177/0047281616653496

Werner-Seidler, A., Perry, Y., Calear, A. L., Newby, J. M., and Christensen, H. (2017). School-based depression and anxiety prevention programs for young people: a systematic review and meta-analysis. *Clin. Psychol. Rev.* 51, 30–47. doi: 10.1016/j.cpr.2016.10.005

Conflict of Interest: 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.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 15 April 2020
doi: 10.3389/frobt.2020.00043

Toward an Automated Measure of Social Engagement for Children With Autism Spectrum Disorder—A Personalized Computational Modeling Approach

Hifza Javed[1], *WonHyong Lee*[2] *and Chung Hyuk Park*[1]*

[1] *Assistive Robotics and Telemedicine Laboratory, Department of Biomedical Engineering, School of Engineering and Applied Science, The George Washington University, Washington, DC, United States,* [2] *School of Computer Science and Electrical Engineering, Handong Global University, Pohang, South Korea*

OPEN ACCESS

Edited by:
Ginevra Castellano,
Uppsala University, Sweden

Reviewed by:
Catharine Oertel,
Delft University of Technology, Netherlands
Samuel Spaulding,
Massachusetts Institute of Technology, United States

***Correspondence:**
Chung Hyuk Park
chpark@gwu.edu

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *19 July 2019*
Accepted: *12 March 2020*
Published: *15 April 2020*

Citation:
Javed H, Lee W and Park CH (2020) Toward an Automated Measure of Social Engagement for Children With Autism Spectrum Disorder—A Personalized Computational Modeling Approach. Front. Robot. AI 7:43. doi: 10.3389/frobt.2020.00043

Social engagement is a key indicator of an individual's socio-emotional and cognitive states. For a child with Autism Spectrum Disorder (ASD), this serves as an important factor in assessing the quality of the interactions and interventions. So far, qualitative measures of social engagement have been used extensively in research and in practice, but a reliable, objective, and quantitative measure is yet to be widely accepted and utilized. In this paper, we present our work on the development of a framework for the automated measurement of social engagement in children with ASD that can be utilized in real-world settings for the long-term clinical monitoring of a child's social behaviors as well as for the evaluation of the intervention methods being used. We present a computational modeling approach to derive the social engagement metric based on a user study with children between the ages of 4 and 12 years. The study was conducted within a child-robot interaction setting that targets sensory processing skills in children. We collected video, audio and motion-tracking data from the subjects and used them to generate personalized models of social engagement by training a multi-channel and multi-layer convolutional neural network. We then evaluated the performance of this network by comparing it with traditional classifiers and assessed its limitations, followed by discussions on the next steps toward finding a comprehensive and accurate metric for social engagement in ASD.

Keywords: computational model, personalization, social engagement, autism spectrum disorder, convolutional neural network

INTRODUCTION

Social engagement of a child is an indicator of his/her socioemotional and cognitive states. It is the interaction of a child with the environment in a contextually appropriate manner and reflects a complex internal state that signifies the occupation of the child with a person or a task. Much of the research so far has relied on the perceptual evaluation of engagement, utilizing questionnaires and behavioral assessments administered by trained professionals, which typically attempt to identify key behavioral traits that serve as important indicators of social engagement. Automatic quantification of engagement is still limited but can allow not only for an objective interpretation of engagement and the contributing target behaviors, but also help to identify methods to improve engagement in different settings, especially when targeting a specific health condition. Therefore, it serves both as an outcome measure and as an objective measure of the quality of an activity, interaction, or intervention (Kishida and Kemp, 2006).

Social engagement has often been reported to be particularly deficient in children with Autism Spectrum Disorder (ASD). ASD is a neurodevelopmental disorder that causes significant impairment in three broad areas of functioning: communication, social interaction, and restricted and repetitive behaviors (American Psychiatric Association., 2013). This means that children interact with their peers infrequently, thus preventing the formation of lasting and meaningful social relationships and resulting in social withdrawal. These children often feel isolated from or rejected by peers and are more likely to develop behavioral problems (Ollendick et al., 1992) as well as anxiety and depression (Tantam, 2000; Bellini, 2006).

Behavioral and physiological cues can provide insight into the engagement state of a child, with gestures, subtle body language changes, facial expressions, vocal behaviors, and various physiological signals, all carrying significant indications of a child's level of interest and engagement in an interaction. Eye gaze focus, smiling, vocalizations, joint-attention, imitation, self-initiated interactions, and triadic interactions are among the important behavioral cues that can be utilized to assess engagement (Tiegerman and Primavera, 1982, 1984; Wimpory et al., 2000; Nadel, 2002; Ingersoll, 2008; Stanton et al., 2008; Katagiri et al., 2010; Sanefuji and Ohgami, 2011; Tapus et al., 2012; Slaughter and Ong, 2014; Dubey et al., 2015; Contaldo et al., 2016). Heart rate, electrodermal activity, electrocardiography, electromyography, blood pressure etc. are among the key physiological indicators of engagement state (Kushki et al., 2012; Lahiri et al., 2012; Hernandez et al., 2014). A combination of these multi-modal behavioral and physiological features can present a comprehensive feature set for effective engagement evaluation.

A major hurdle in the path toward automated measurement of social engagement is of the identification and classification of these key behaviors. While it may be a simple task for trained professionals to identify these high-level behaviors and infer a fairly accurate engagement state from real-time observations of a child's interactions, it remains a considerable challenge for the state-of-the-art algorithms and machines. Instead, the current technologies are better equipped to extract lower-level behaviors that can be used as a rough estimation of the target behaviors.

This paper presents our first step toward an automated quantifiable measure of social engagement derived from behavioral data collected from two groups of children, one typically developing (TD) and one with ASD. Research from our team thus far has focused on child-robot interaction scenarios that target several ASD symptoms, including sensory processing (Javed et al., 2019), imitation (Bevill et al., 2017), emotion recognition and emotion regulation skills (Javed et al., 2018). In these studies, we collected multi-modal interaction data, including video and audio recordings, as well as motion tracking data. The overall goal of our work is to develop a framework for personalized child-robot interactions for ASD. To this end, our framework aims to (1) sense important features of a child's interaction with a robot, (2) interpret and derive meaningful deductions about a child's engagement in the interaction, (3) identify target behaviors that may be lacking in the detected interaction pattern, (4) reassess the current robot behavior strategy and modulate it to elicit a higher level of engagement from the child. This paper focuses on step 2 of the above approach by processing the multimodal behavioral data collected from this study through a deep learning-based multi-label classification model in order to contribute toward deriving an automated measure of social engagement.

This paper is organized as follows. Section Related Work discusses the previous studies that have designed methods to formulate an automated measure of social engagement. Section Interaction Scenario Design describes the child-robot interaction scenario we used in this study. Sections Multimodal Data Collection and Extracting Ground Truth present the modalities of the data we collected during our experiments and the methods we employed to label these data. Sections Feature Extraction and Network Architecture discuss our feature extraction methods and design of our convolutional neural network for multi-label classification. Sections User Study, Results, and Comparison with Other Machine Learning Classifiersdescribe the user study, its results and a comparison of the proposed network with other classical algorithms. Section Discussion presents a discussion on these findings while Section Conclusion concludes this paper with comments on the future work.

RELATED WORK

Several studies in the past have contributed to this area of research with each method typically varying in terms of the feature set, number of engagement classes and computational model that were used, as well as the demographics of the participants from whom the data were collected. Rajagopalan et al. (2015) showed the feasibility of utilizing low-level behavioral features in the absence of accurate high-level features, and used a two-stage approach to first find hidden structures in the data (using Hidden Conditional Random Fields) and then learn them through a Support Vector Machine (SVM). Only head pose orientation estimates were used to assess engagement and the approach was evaluated by conducting experiments on labeled child interaction data from the Multimodal Dyadic Behavior Dataset (Rehg et al., 2013), obtaining an accuracy of around 70%.

Gupta et al. (2016) designed an engagement prediction system that utilized only the prosodic features of a child's speech as observed during a structured interaction between a child and a psychologist involving several tasks from the Rapid ABC database. Three engagement classes and two levels of prosodic features (local for short-term and global for task-wide patterns) were defined. The system achieved an unweighted average recall of 55.8%, where the best classification results were obtained by using an SVM that utilized both categories of the prosodic features. Another study by Lala et al. (2017) used several verbal and non-verbal behavioral features, including nodding, eye gaze, laughing and verbal backchannels. The authors collected their own dataset comprising audio and video recordings based on conversational scenarios between a human user and a humanoid robot, while human annotators provided labels to establish ground truth. A Bayesian binary classifier was used to classify the user as engaged or not engaged and obtained an AUC (area under the precision-recall curve) score of 0.62.

A study from Castellano et al. (2009) used both behavioral features from the user (gaze focus and smiling) and contextual information from the activity in order to train a Bayesian classifier to detect engagement in users for a child-robot interaction scenario. The labels generated from human coding were based only on the two user behaviors. The authors reported only a slight improvement in the classifier recognition rate when using both behavioral and contextual features (94.79%) vs. when only behavioral features were utilized (93.75%), highlighting the key importance of the behavioral information.

Kim et al. (2016) investigated the use of vocal/acoustic features in determining child engagement in group interaction scenarios. The annotation scheme involves the giving and receiving of attention from other group members. They used a combination of ordinal regression and ranking with SVM to detect engagement in children and found this technique to outperform classification, simple regression and rule-based approaches. Such a system may be acceptable to use with typically-developing children, but since children with ASD may often be non-verbal and/or shy or unwilling to communicate using speech/vocalizations, the exclusive use of acoustic features may not be suited to research involving the ASD population.

Another study from Parekh et al. (2018) developed a video system for measuring engagement in patients with dementia, which uses deep-learning based computer vision algorithms to evaluate their engagement in an activity to provide behavior analytics based on facial expression and gaze analysis. Ground truth was extracted through scoring performed by human annotators by classifying engagement states in terms of attention and attitude. The video system presented in this study was exclusively tested with elderly patients with dementia who were required to participate in a digital interaction while seated directly in front of the camera. Additionally, since only facial expressions and gaze features were utilized, the proximity of the participants to the camera was important, hence, limiting their physical movements.

Oertel and Salvi (2013) studied the relation between group involvement and individual engagement using several features of eye gaze patterns defined as presence, entropy, symmetry and maxgaze. They utilized the Stockholm Werewolf Corpus, which is a video dataset of participants engaging in a game that involved the use of speech and eye gaze. Once again, since only eye gaze patterns were used as features to train a classifier, participants were required to remain seated in front of the cameras.

A study that specifically tested their system on the ASD population was from Anzalone et al. (2015) that used a combination of static (focus of attention, head stability and body posture stability) and dynamic (joint attention, synchrony, and imitation) metrics within two distinct use cases including one where the robot attempts to learn the colors in its environment with the help of a human, and another that elicits joint attention from participating children with ASD. The features were extracted using histogram heatmaps and clustered using the K-means algorithm.

In Rudovic et al. (2018) also targeted the automated measurement of engagement for ASD children with multimodal data collection including features from video (facial expressions, head movements, body movements, poses, and gestures), audio, and physiological (heart rate, electrodermal activity, and heart rate) data. The child-robot interaction setting involved an emotion recognition activity with a humanoid robot that required children to be seated in front of the robot (Rudovic et al., 2017). Participating children belonged to one of two cultures (Eatsern European and Asian) and the cultural differences were also taken into account during engagement estimation. The authors generated ground truth through expert human labelers who marked changes in engagement on a 0–5 Likert scale that is based on the different levels of prompting required from the therapist during the interaction with the robot. In fact, in this work, child engagement is considered to be a function of task-driven behavioral engagement and affective engagement.

Despite the overlap, this approach is significantly different from the one proposed in this paper in several ways. Firstly, we define engagement as a function of several key behavioral indicators that provide an insight into an individual's internal engagement state (Javed et al., 2019), which generates a novel measure to estimate social engagement state i.e., the engagement index. Additionally, our methods do not restrict the movement of the subjects by requiring them to be seated in front of a camera or a robot, and the interaction design allows for free, naturalistic movement in order to closely resemble real-world social settings as opposed to other restrictive experimental approaches. Importantly, this approach toward engagement estimation can be easily generalized to any child, with or without ASD, and to a variety of different, interactive experimental settings that may or may not involve a robot.

The work described in this paper presents a social engagement prediction system for children. It utilizes a combination of features extracted from facial expressions and upper body motion tracking data to train a deep convolutional neural network that can then classify the engagement state of a child. We intentionally designed the experiments to not be strictly structured in order to encourage naturalistic and unguided child-robot interactions during data collection that impose no restrictions on the movement of a child. The nature of the features used in our approach allow for independence of interaction context and can easily be extended to a variety of scenarios within laboratory or home settings. In addition, a unique engagement model is obtained for every individual participant to ensure personalized interaction with the robot, giving it potential to be used as an intervention tool for ASD.

INTERACTION SCENARIO DESIGN

For this work, we used socially assistive robots to design a child-robot interaction that targeted the sensory processing difficulties in ASD, as detailed in our previous work (Javed et al., 2019). In this pedagogical setting, two different mobile robots were used to model socially acceptable responses to potentially overwhelming sensory stimulation that a child is likely to encounter in everyday experiences. The humanoid robot, Robotis Mini (from Robotis) and the iPod-based robot, Romo (from Romotive) both had their unique set of capabilities. While Mini used gestures and speech to communicate, Romo relied mostly on its large set of emotional expressions and some movements.

A maze-like setup consisting of a station for each of the visual, auditory, olfactory, gustatory, tactile and vestibular senses was used, as shown in **Figure 1**. Though one of the goals of the interaction was to leverage the relationship between a robot and a child with ASD, as established by a plethora of previous research (Dautenhahn and Werry, 2004; Scassellati, 2007; Diehl et al., 2012; Cabibihan et al., 2013), the focus of this work (Javed et al., 2019) was to assess the potential of this setup as a tool to socially engage children with ASD and to use the collected data to contribute toward deriving an automated measure of social engagement. Each sensory station simulated an everyday experience, such as encountering bright lights at the *Seeing station*, loud music at the *Hearing station*, scented flowers at the *Smelling station*, different food items at the *Tasting station*, materials with different textures at the *Touching station* and summersaulting to celebrate at the vestibular station (**Figure 2**). These scenarios were chosen to incorporate everyday stimulation that all children experience in uncontrolled environments like malls, playgrounds, cinemas etc. and in the activities of daily living such as eating meals and dressing. This interaction was designed to be highly interactive and engaging, and required the child to participate actively by answering questions from the robots, following their instructions, and "helping" them complete the maze. Details of this study, including the nature of interaction between the children and the robots, can be found in Javed et al. (2019).

FIGURE 1 | Station setup for the sensory maze game (the child's photo rights reserved).

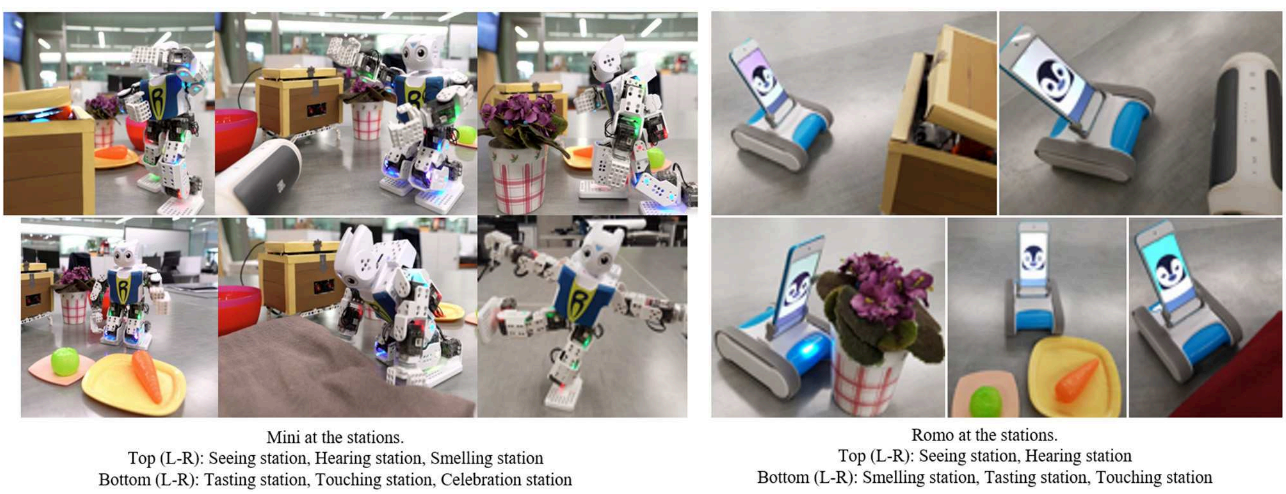


FIGURE 2 | The two robots at each sensory station, adapted from Javed et al. (2019).

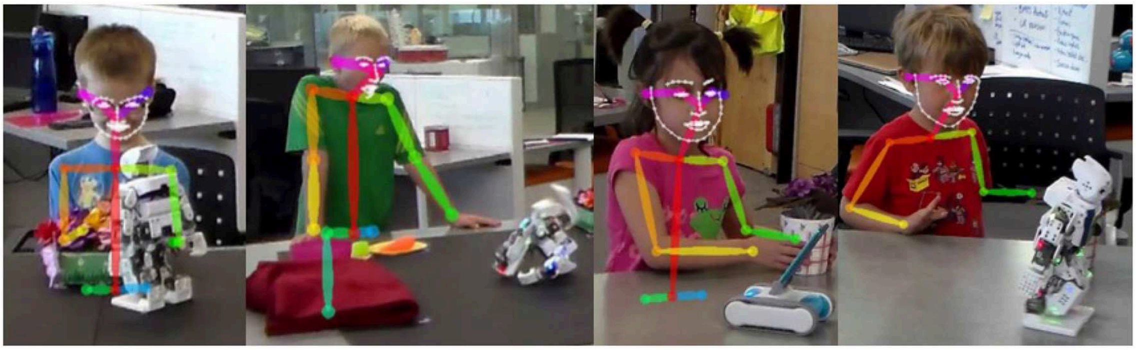


FIGURE 3 | Upper body and facial keypoints generated by OpenPose.

MULTIMODAL DATA COLLECTION

A high-quality measure for social engagement estimation must take into account all behavioral and physiological cues that can serve as quantifiers of social motivation and social interaction. As discussed in Section Introduction, a number of behavioral traits and physiological signals can be used effectively to this end. However, when designing an interaction for autistic children, their unique needs and sensitivities must be taken into account. For this study, this meant that only non-contact sensors could be used in order to limit tactile disturbances to the children and enable free movement to allow for naturalistic interaction. The combination of sensors also needed to provide a wholistic and accurate representation of a child's engagement changes over the length of the interaction.

We collected video recordings of the child-robot interactions with a camcorder placed in one corner of the room, which was repositioned by an instructor as the child moved during the interaction. From these recordings, we were able to extract audio data as well as 2-D motion tracking data with the OpenPose library (Cao et al., 2017). While OpenPose provides full body motion tracking (**Figure 3**), we were only able to utilize upper body data since the chosen experimental setting meant that children were often standing in front of the table that hosted the maze setup, preventing a full-body view from being captured. In addition, OpenPose also allowed for the extraction of facial expression datapoints from the same video data.

EXTRACTING GROUND TRUTH

Unlike some of the previous studies described in Section Related Work, we did not use any existing video datasets to test our methods. Since our goal was to derive an engagement measure specific to the interactions that we designed for children with ASD, we opted to test our methods on the relatively limited data available from our user study. To extract ground truth for a child's engagement in the interaction with the robots, we defined six target behaviors that have been found to be key behavioral indicators of social engagement (Tiegerman and Primavera, 1982, 1984; Wimpory et al., 2000; Nadel, 2002; Ingersoll, 2008; Stanton et al., 2008; Katagiri et al., 2010; Sanefuji and Ohgami, 2011; Tapus et al., 2012; Slaughter and Ong, 2014; Dubey et al., 2015; Contaldo et al., 2016). These included eye gaze focus, vocalizations, smiling, self-initiated interactions, triadic interactions and imitation.

Three raters then coded these videos using the Behavioral Observation Research Interactive Software (BORIS) (Friard and Gamba, 2016) to annotate the start and stop times of each target behavior as it was identified in the video recordings. An inter-coder correlation (ICC) score of 0.8752 ± 0.145 was achieved for the 18 participants, which was used to evaluate the quality of the annotations. Details of the evaluation criteria are reported in Javed et al. (2019).

An eye gaze event was tagged each time the child's gaze moved to the robots or the setup and stopped when the gaze focus was lost. Vocalizations comprised of any verbal expression from the child, including but not limited to a shriek of excitement while interacting with the robots or the utterance of words to communicate sentiments or queries regarding the robots. Smiling recorded all events where a child was observed to visibly express joy in the form of a smile or laugh. Self-initiated interactions involved all interactions with the robots or setup that are initiated by the child. Triadic interactions comprised of an interaction where a child voluntarily involved a third entity in the interaction with the robot, such as sharing their excitement with the parent. Lastly, imitations included all events of voluntary imitation the robot's actions by the child. An in-depth report on the inclusion criteria of the target behaviors, their significance and annotations in video data can be found in Javed et al. (2019).

Based on these annotations, multiple analytics were derived to quantify the social engagement with respect to each robot and target behavior, and across stations to obtain a fine-grained analysis of the child's interaction preferences (Javed et al., 2019). However, for the current work, we have only used the raw time series data of every child's changing engagement state as determined by the chosen target behaviors. These overall engagement changes are shown in **Figure 4**, along with the subplots of each contributing key behavior.

Therefore, each instance of time was mapped to an engagement state. Every behavior contributed a factor of 1/6

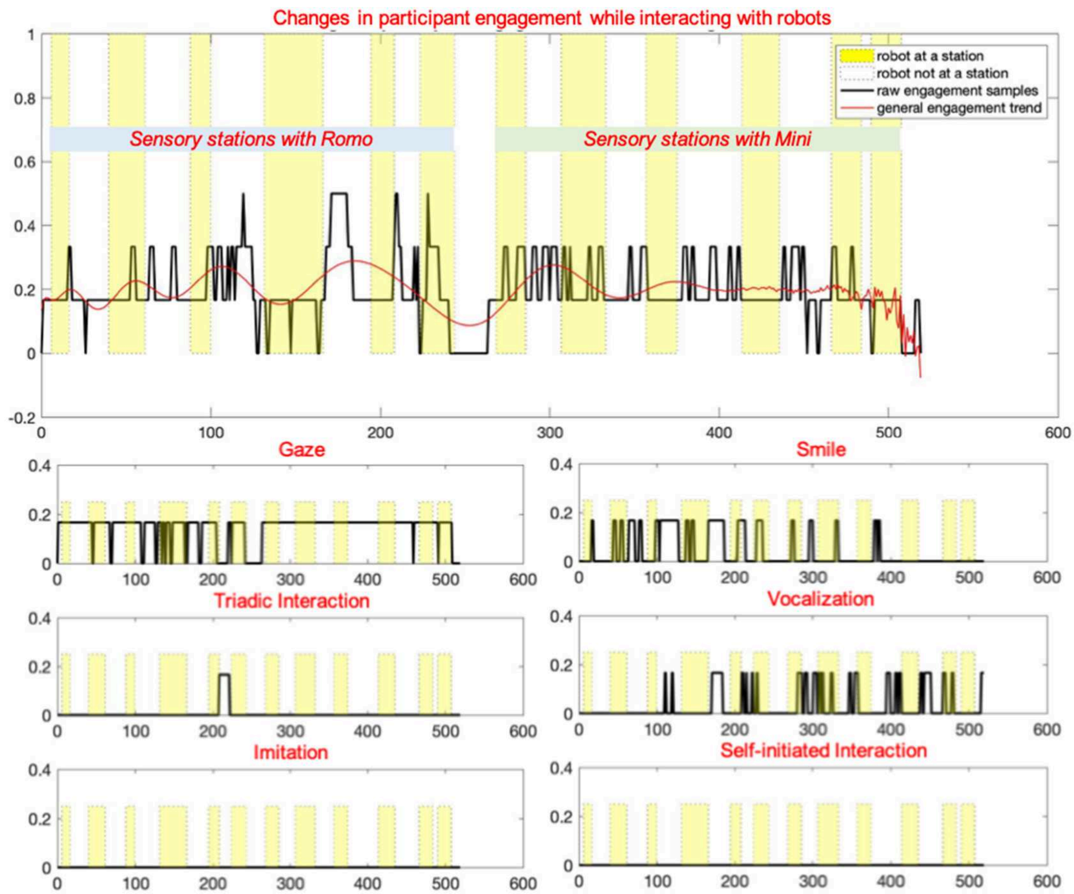


FIGURE 4 | Plots depicting changes in the overall engagement level of a child during an interaction, along with subplots of the target behaviors contributing to this engagement, adapted from Javed et al. (2019).

to the engagement value, thus resulting in a metric with seven distinct values that ranged from 0 (no target behavior observed) to 1 (all target behaviors observed).

FEATURE EXTRACTION

An ideal automated engagement measure in this case would incorporate all of the above behaviors, but also necessitates the automated classification of these behaviors. This is no trivial task, and involves contributions from multiple disciplines including computer vision, speech analysis and machine learning. As a part of a more practical approach that is fitting of a first step toward the derivation of an automated measure of social engagement in ASD, we decided to extract low-level behavioral components from our video data as indicators of engagement in the interactions with the robots. For this purpose, we utilized the 2D body tracking and facial expression data generated by OpenPose (Cao et al., 2017).

Using the body tracking data, we derived three new features based on Laban Movement Analysis (LMA), a method for describing and interpreting all types of human movement (Groff, 1995) used frequently in a variety of fields including dance, acting, music, and physical therapy etc. LMA categorizes all body movements into the categories of body effort, space and shape. Out of the four categories, effort represents the dynamics of human movement and provides an insight into the subtle characteristics of movements with respect to inner intention. This makes it an important feature to use in studies involving the estimation of affect, intention, and engagement states. Effort itself is classified into space, weight and time, which are the three features that we incorporated in our current work. Space represents the area taken up over the course of a movement, weight indicates the power or impact of movement, and time conveys the speed of an action, including a sense of urgency or a lack thereof in a movement. The equations (Masuda et al., 2009; Wakayama et al., 2010) for each of these features are as shown in **Table 1**.

OpenPose generates 50 keypoints for skeletal tracking as described in Cao et al. (2017). In addition to the skeletal data, we also recorded facial keypoints to incorporate the changes in a child's facial expressions in our feature set. **Figure 5** [taken from CMU Perceptual-Computing-Lab (2019)] depicts these datapoints. While a total of 69 facial keypoints is available, we only used the lip and eye keypoints shown on the right. Including the x and y coordinates for each of the 34 facial keypoints and the three Laban features derived from the upper body skeletal keypoints created a total of 71 features in the dataset. A moving window of 1 s, i.e., 30 frames, was used to compute the Laban features in order to incorporate the sequential nature of the movement data. A 1 second interval was chosen to capture meaningful, yet rapidly changing movement patterns in response to the actions of the robot during the child-robot interaction. The number of available datapoints per participant depended on the length of interaction of each participant and ranged between 9,300 and 30,508 datapoints. Further details are listed in **Table 3**.

TABLE 1 | Equations for the derived Laban features adopted from Masuda et al. (2009) and Wakayama et al. (2010).

Feature	Equation	
Space		$space = \left(0.5\lvert\vec{a}\rvert\left\lvert\vec{d}\right\rvert\sin(\theta_1)\right) + \left(0.5\lvert\vec{c}\rvert\left\lvert\vec{b}\right\rvert\sin(\theta_2)\right)$
	where	$\vec{a}$ = *Position vector from left shoulder to left hand*
		$\vec{b}$ = *Position vector from right shoulder to left shoulder*
		$\vec{c}$ = *Position vector from right hand to right shoulder*
		$\vec{d}$ = *Position vector from left hand to right hand*
		θ_1 = *Angle between* $\vec{a}$ & $\vec{d}$
		θ_2 = *Angle between* $\vec{c}$ & $\vec{b}$
Weight		$Weight = \sum_i \tau_i(t)$
	where	$\tau_i = L^2 {\omega_i}^2 \sin(\theta) * mass$
		$\omega_i = \frac{d\theta}{dt}$
		L = *Distance between joints*
		i = *Joint number*
		$\dot{\omega}_i$ = *Angular velocity for joint i*
Time		$Time_i = \sum_i \dot{\omega}_i(t)$
	where	i = *Joint number*
		$\dot{\omega}_i$ = *Angular velocity for joint i*

NETWORK ARCHITECTURE

We used a multi-channel and multi-layer convolutional neural network (CNN) for this temporal multi-label classification problem. The network was composed of two Conv1D layers to identify temporal data patterns (with 5 channels with 64 and 128 filters, respectively, and a kernel size of 3 with 20% dropout) and three dense layers for classification [kernel sizes 256, 256, and 7 (number of output labels: value ranges of engagement level)]. This is illustrated in more detail in **Figure 6**. A 10-fold cross-validation (train/test split of 0.8/0.2) was used for every subject's individual dataset and optimization was performed using the Adam optimizer.

The two Conv1D layers are meant to extract high-level features from the temporal data since the dataset being used has a high input dimension and a relatively small number of datapoints. Since the data have a non-linear structure, the first two dense layers are used to spread the feature dimension, whereas the last one generates the output dimension. The dropout layers are used to avoid overfitting.

USER STUDY

We conducted a user study with a total of 18 children, 13 TD and 5 with ASD between the ages of 4 and 12 years who participated in a one-time interaction with our robots within the setting of a sensory maze game. The average age of the TD group was 7.07 ± 2.56 years and that of the ASD group was 8.2 ± 1.10 years. The TD group consisted of 5 females and 8 males, whereas the ASD group was composed of all male participants. These details are presented in **Table 2**.

The participants were allowed to participate for the entire course of the interaction as designed with the two robots, one after another. The data presented in this study is for one-time interactions between each subject and the robots. The length of the interaction for each participant is listed in **Table 2**. The average TD interaction length was 464.92 s whereas that of the

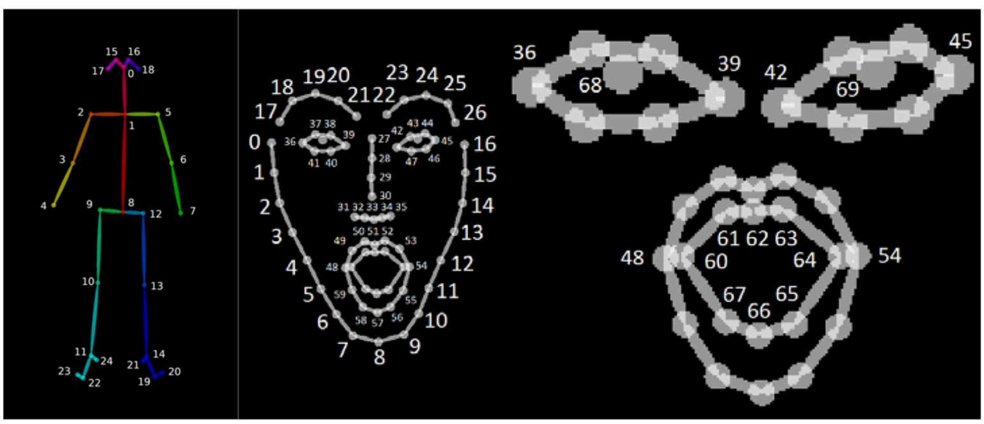


FIGURE 5 | Illustrations of the skeletal and facial keypoints extracted by OpenPose (CMU Perceptual-Computing-Lab, 2019) (permission acquired from the author for using these images with citation).

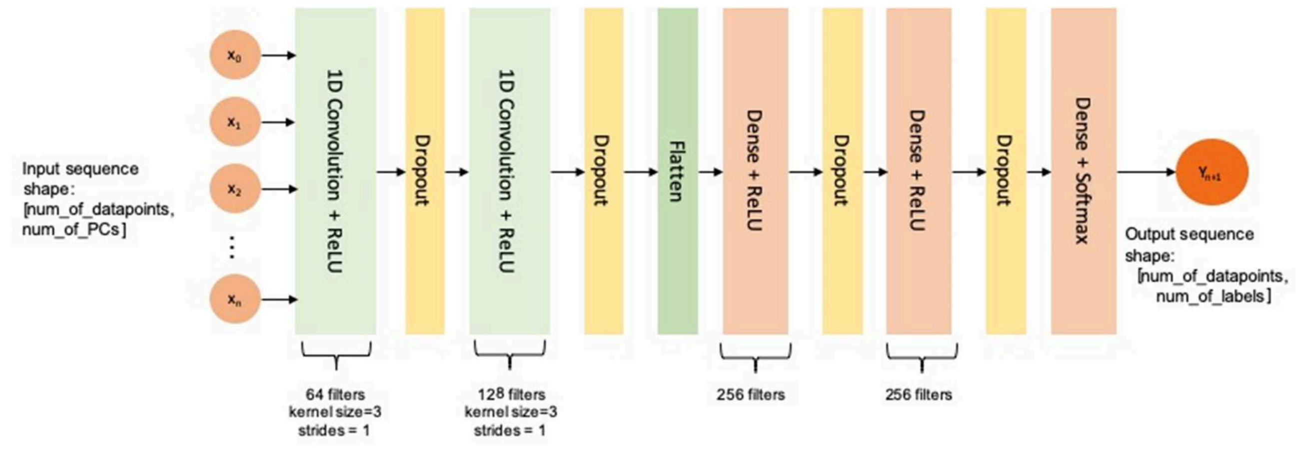


FIGURE 6 | Architecture of the CNN used for multi-label classification.

ASD group was 620 s. Individual engagement prediction models were generated for each participant and their performances were evaluated.

RESULTS

Table 3 presents the detailed results produced by training, validation and testing our network for every subject in the study. The length of interaction is important and provides an insight into the number of video frames, and hence, the datapoints that would be available to the network. The datapoint count is also affected by the processing performed by OpenPose, which can drop some frames where processing could not be completed. This is particularly evident in the case of participant 6 and 12, where the number of available datapoints are far fewer than expected.

Before presenting the results, it must be highlighted that the metrics shown in this work are all weighted metrics, so as to address the impact of the imbalance in engagement level samples within the dataset. The network has an average accuracy of 0.7985 for the TD group and 0.8061 for the ASD group in the training stage. For the test data, the performance remains steady with an average accuracy of 0.7767 for the ASD group and 0.7918 for the TD group. These details are shown in **Table 4**.

Figure 7 depicts the accuracy and loss plots for training and validation data for a participant from each group illustrating the changes in accuracy with respect to the number of epochs. **Figure 8** shows the timeseries plots of the changing engagement states for the participants. The red line shows the true engagement as determined by the annotations (Javed et al., 2019). Predictions made by the network are marked in blue. Since the dataset was randomly partitioned into test and training data, the predictions on the test set appear as a scatter plot.

In addition to the individual models described above, we also trained a group model for each of the two groups by using all the datapoints collected from the participants from each group. The ASD classifier was able to achieve a training accuracy of 0.6389 and a test accuracy of 0.6524, while the TD classifier achieved a slightly higher training accuracy of 0.6733 and a test accuracy of 0.6803. The slightly superior performance of the classifiers on the test data as opposed to the training data can be attributed to the use of regularization techniques used when constructing the classifier structure, in this case, the Dropout layers, which are only applied during the training phase.

We also trained a combined classifier on the data collected from all the participants. This model underperformed slightly compared to the group-specific classifiers, indicating that a group-specific classifier may be better suited for generalization to all participants within the group rather than a single classifier

TABLE 2 | Demographic details of the subjects.

ID	Age	Gender	Group
1	10	M	TD
2	4	F	TD
3	5	F	TD
4	11	F	TD
5	9	M	TD
6	10	F	TD
7	9	M	TD
8	5	M	TD
9	5	F	TD
10	5	M	TD
11	5	M	TD
12	5	M	TD
13	9	M	TD
14	7	M	ASD
15	8	M	ASD
16	10	M	ASD
17	8	M	ASD
18	8	M	ASD

TABLE 3 | Performance metrics for the individual classifiers (TD Group: ID1–ID13, ASD Group: ID14–ID18).

ID	Interaction length (s)	No. of datapoints (frames)	Train		Validation		Test
			Accuracy	Loss	Accuracy	Loss	Accuracy
1	315	9,444	0.8101	0.5028	0.7790	0.6681	0.7946
2	519	15,357	0.6499	0.7278	0.6398	0.7797	0.6393
3	540	16,412	0.6703	0.8723	0.6407	1.0095	0.6526
4	658	10,933	0.8302	0.4189	0.8131	0.4923	0.8240
5	797	22,996	0.9255	0.1903	0.9198	0.2484	0.9159
6	696	9,300	0.9200	0.2850	0.8925	0.3856	0.9124
7	316	9,388	0.7821	0.5423	0.7417	0.7946	0.7338
8	457	13,725	0.7561	0.6065	0.7418	0.6796	0.7483
9	574	10,463	0.6671	0.8486	0.6535	0.9333	0.6364
10	780	16,627	0.9104	0.2253	0.8831	0.3907	0.8698
11	726	12,726	0.8390	0.3843	0.8303	0.4039	0.8283
12	685	9,723	0.8118	0.5162	0.7715	0.6980	0.7720
13	540	12,879	0.8084	0.4296	0.7812	0.5858	0.7702
14	517	15,502	0.8163	0.4417	0.7952	0.5621	0.7907
15	578	14,624	0.9204	0.2276	0.8923	0.3390	0.9108
16	679	15,950	0.6810	0.7582	0.6501	0.9095	0.6398
17	610	16,401	0.8306	0.3946	0.8232	0.4923	0.8366
18	1058	30,508	0.7822	0.5467	0.7759	0.6323	0.7812

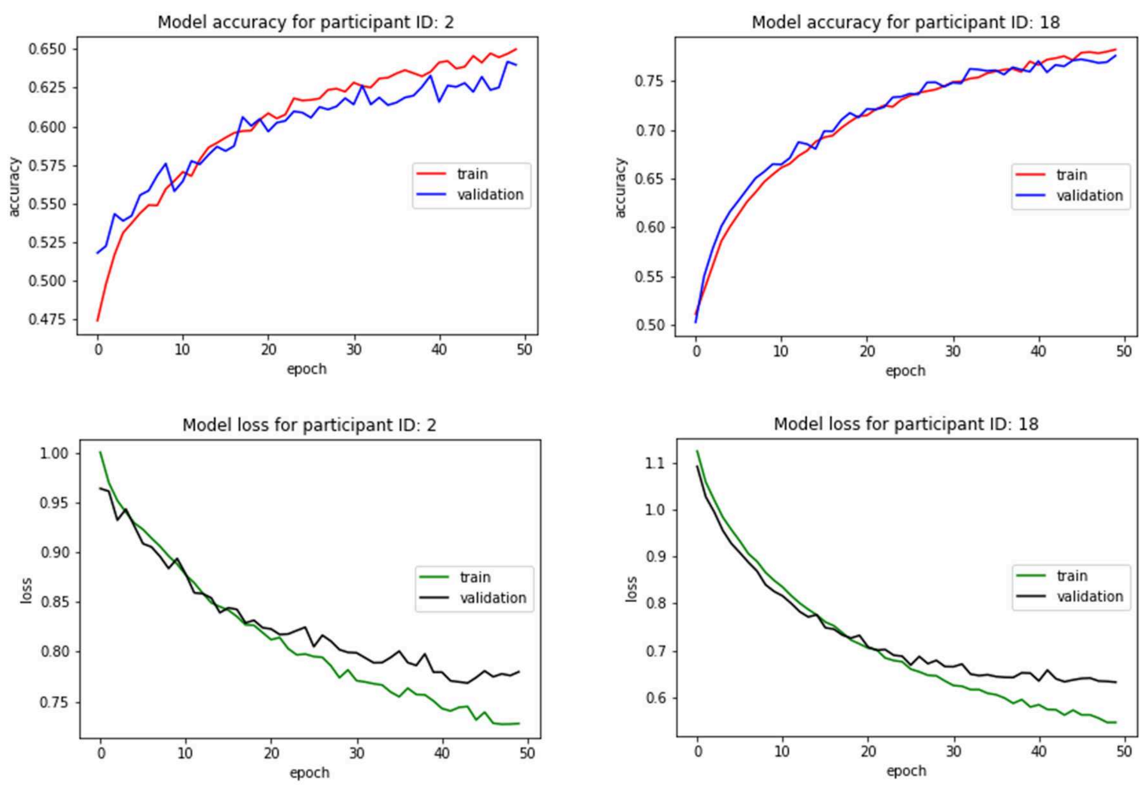


FIGURE 7 | Classifier accuracy and loss with respect to the number of epochs for two different participants.

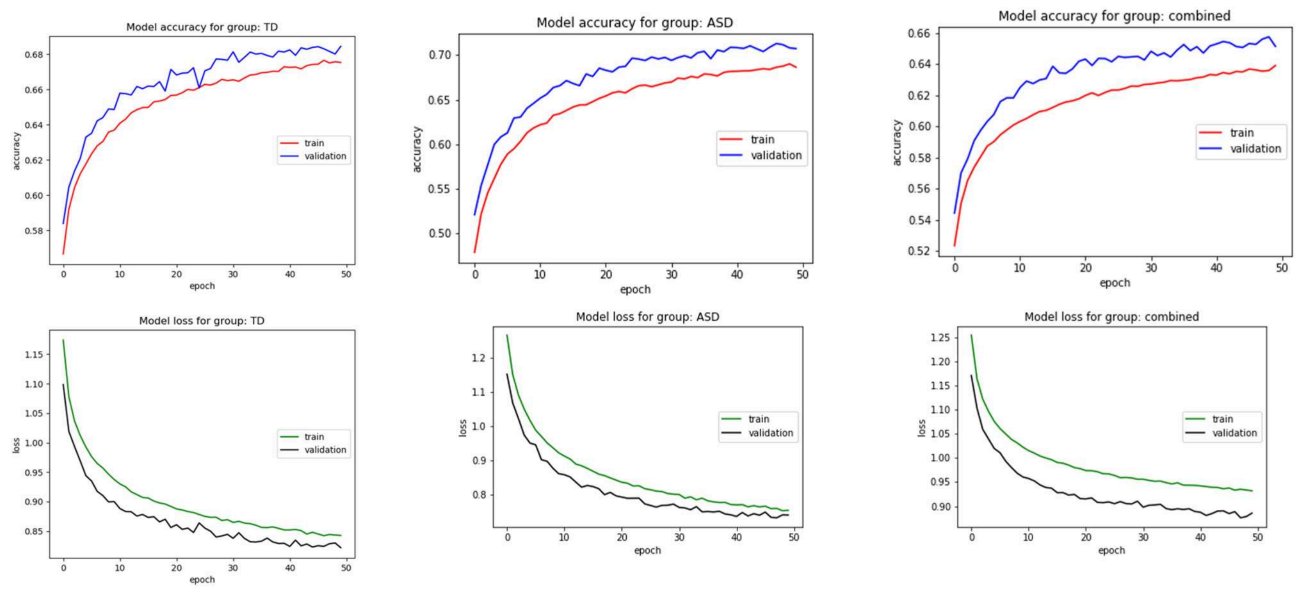


FIGURE 8 | Plots showing the ground truth labels in red and the classifier predictions in blue.

TABLE 4 | Average metrics to compare classifier performance.

ID	Average interaction length (s)	Train		Validation		Test
		Accuracy	Loss	Accuracy	Loss	Accuracy
TD	584.8	0.7985	0.5038	0.7760	0.6207	0.7767
ASD	688.4	0.8061	0.4738	0.7873	0.5870	0.7918

TABLE 5 | Performance metrics for group classifiers.

Classifier	Train		Validation		Test
	Accuracy	Loss	Accuracy	Loss	Accuracy
TD	0.6733	0.8472	0.6800	0.8263	0.6803
ASD	0.6389	0.9320	0.6512	0.8858	0.6524
Combined	0.6733	0.8472	0.6800	0.8263	0.6803

for all participants (**Table 5**). Accuracy and loss plots for the training and validating processes for all three grouped conditions are shown in **Figure 9**.

COMPARISON WITH OTHER MACHINE LEARNING CLASSIFIERS

A number of standard Machine Learning (ML) classifiers were also trained for all the scenarios described above as a way to situate the performance of the CNN, which included Support Vector Classification (SVC), Random Forest (RF), Decision Trees (DT), and K-Nearest Neighbors (KNN). The reported metrics were also averaged across all participants to compare the overall performance of the classifiers. As before, each classifier was trained and tested on entire group datasets to compare performance as a generalized group classifier. These results are shown in **Table 6**.

After averaging over the metrics for all participants, RF is seen to have the best performance followed by KNN and CNN, respectively. A similar trend is seen for grouped classifiers, where RF once again outperforms all other classifiers in terms of both the accuracy and the F1 score, followed again by KNN and CNN, respectively. All classifier performances drop slightly when data from the two groups are combined, suggesting that a single classifier may not be as useful for generalization as a group-specific classifier.

DISCUSSION

In this work, we propose the use of a Deep Learning Convolutional Neural Network to model and predict child social engagement as a part of our larger goal to personalize child-robot interactions. We utilized key social behaviors as indicators of engagement in an interaction, which formed the criterion for the human-generated labels that serves as the ground truth for this engagement classification approach.

We found that the proposed CNN was able to achieve a performance that was comparable to the highest performing classical ML approaches in this work. The RF and KNN classifiers only slightly outperform the CNN in the case of both individual classifiers and grouped classifiers. The individual classifiers serve as personalized engagement prediction networks for the unique behavioral expressions of each individual participant, whereas the grouped classifiers were used to evaluate the potential for a single classifier to generalize the learnt patterns to all the participants within a group.

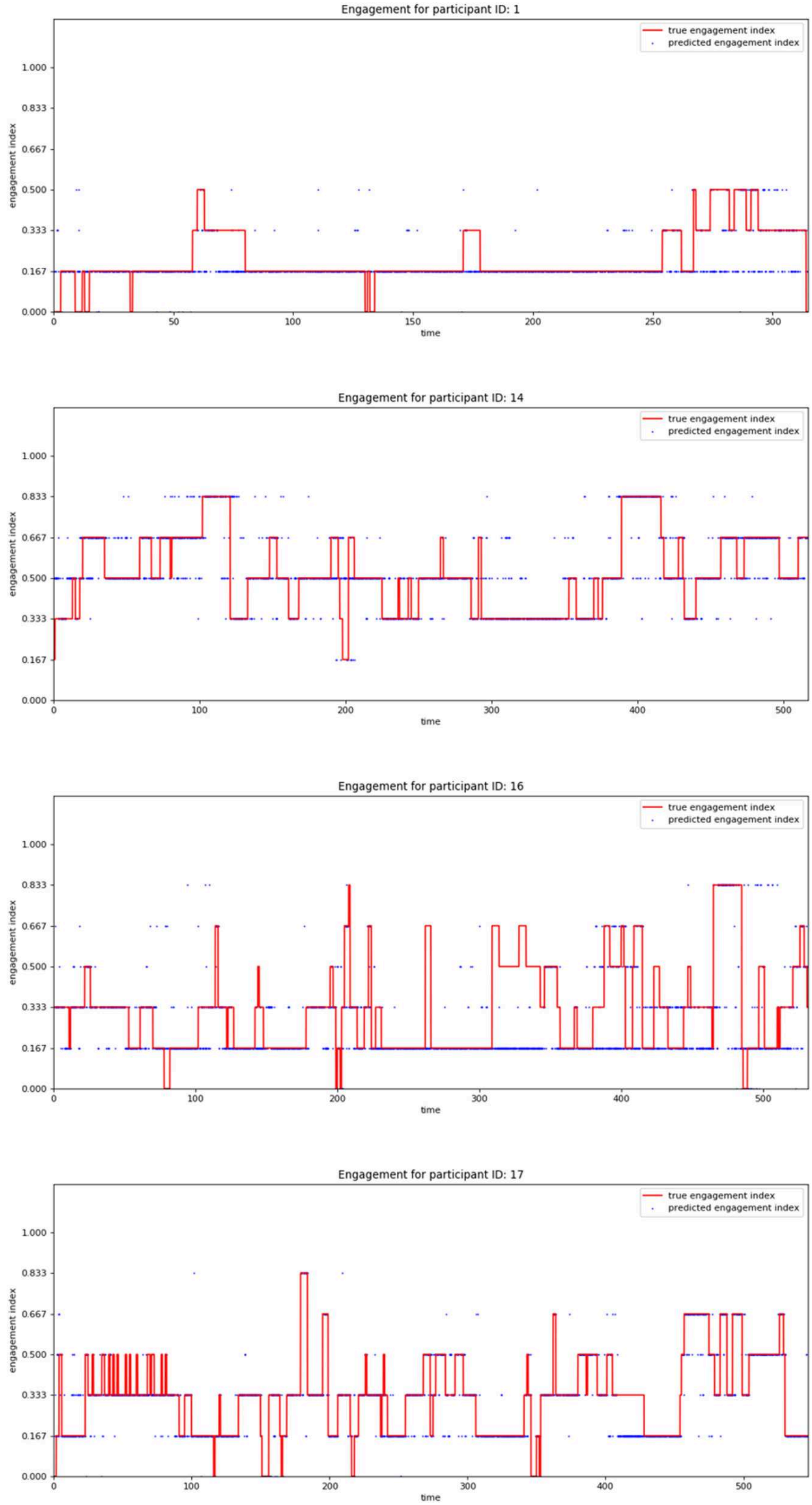


FIGURE 9 | Classifier accuracy and loss for training and test datasets for three grouped conditions.

TABLE 6 | Performance metrics for all classifiers under individual and group conditions.

	Classifier									
	CNN		SVC		RF		DT		KNN	
ID	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1
1	0.79	0.77	0.77	0.72	0.80	0.78	0.77	0.75	0.81	0.79
2	0.64	0.62	0.58	0.55	0.75	0.75	0.65	0.64	0.72	0.71
3	0.65	0.59	0.66	0.55	0.67	0.61	0.65	0.58	0.67	0.61
4	0.82	0.79	0.82	0.76	0.83	0.81	0.82	0.79	0.83	0.81
5	0.92	0.91	0.89	0.87	0.93	0.92	0.90	0.89	0.93	0.93
6	0.91	0.89	0.92	0.90	0.90	0.89	0.91	0.89	0.92	0.90
7	0.73	0.73	0.61	0.59	0.80	0.80	0.72	0.71	0.80	0.80
8	0.75	0.74	0.51	0.47	0.82	0.82	0.66	0.66	0.82	0.81
9	0.64	0.57	0.63	0.56	0.65	0.60	0.63	0.57	0.67	0.61
10	0.87	0.87	0.79	0.77	0.88	0.87	0.82	0.82	0.85	0.85
11	0.77	0.76	0.69	0.65	0.78	0.77	0.72	0.71	0.76	0.74
12	0.83	0.78	0.81	0.74	0.84	0.81	0.82	0.79	0.84	0.80
13	0.77	0.77	0.73	0.69	0.79	0.80	0.77	0.77	0.79	0.80
14	0.79	0.79	0.70	0.69	0.82	0.81	0.73	0.73	0.81	0.81
15	0.91	0.90	0.87	0.83	0.92	0.90	0.90	0.88	0.92	0.91
16	0.64	0.62	0.61	0.57	0.67	0.65	0.62	0.60	0.68	0.66
17	0.84	0.84	0.70	0.69	0.88	0.88	0.76	0.75	0.84	0.84
18	0.78	0.78	0.63	0.60	0.79	0.78	0.61	0.58	0.78	0.78
Average	0.78	0.76	0.72	0.68	0.81	0.79	0.75	0.73	0.80	0.79
TD	0.68	0.65	0.63	0.58	0.74	0.74	0.64	0.61	0.74	0.73
ASD	0.72	0.71	0.60	0.58	0.77	0.76	0.61	0.60	0.76	0.76
Combined	0.65	0.62	0.59	0.54	0.74	0.71	0.60	0.56	0.71	0.71

On the individual level, the CNN was able to attain a best case accuracy of 0.92 (participant 5) and a worst case accuracy of 0.64 (participant 2). On the other hand, the RF classifier reached a highest accuracy of 0.93 (participant 5) and lowest accuracy of 0.65 (participant 9). For the averaged metrics as well as the grouped metrics, the RF accuracy is no more than 2% higher than that of the CNN.

The individual ASD and TD classifiers were generally found to achieve a higher accuracy than the single classifier trained on data from all the participants. This points the possibility of a generalized group classifier that can be used effectively to classify social engagement for all the children in each group while providing a high level of personalization in the interaction.

The CNN is a complex structure with a large number of tunable parameters that generally requires much larger datasets to fully exploit the potential of deep networks. Given the number of input features, the number of output classes and the size of the dataset (generated by single session child-robot interactions only) used in this study, the CNN was able to achieve a performance comparable to simpler ML classifiers but not exceed them. We anticipate that as we continue to collect interaction data from additional participants for a long-term study involving multiple sessions, the proposed deep learning network will likely become a more suitable choice for social engagement classification.

It must also be pointed out that in terms of deployment to a robotic platform, a CNN may also be a more suitable option since the traditional algorithms require expensive resources when deployed to mobile platform in real-world applications, whereas deep learning algorithms can fully take advantage of the scalable computing platforms with GPUs that have low-cost modules (like the NVidia Jetson Nano) while retaining the capacity to handle much larger datasets.

The current work is limited in that it only utilizes single session data for each participant based on which the classifiers are trained. Classifier performance is likely to improve as subsequent sessions are conducted and larger datasets are collected. Another limitation of this work is that the datasets for the two groups are unbalanced, with 13 participants in the TD group and only five in the ASD group generating much larger training dataset for the TD classifier than ASD. Conducting long-term studies with a population such as ASD remains a considerable challenge for all researchers in the field and explains the lack of open multi-modal datasets to benefit the ASD research community.

Since our focus in this work was to evaluate social engagement in a naturalistic interaction setting, the video recordings of the sessions mainly focused on the participant but also included other members of the research team and/or parent in several segments of the videos as the child moved around the room to interact with the robots. OpenPose was chosen to process the movements of the participants particularly because it offers a feature to track multiple persons by assigning each a fixed ID. In practice, however, this ID assignment was found to lack reliability, which we discovered by visualizing the participant's skeletal tracking data. In addition, we also found that the number of frames in the input video and the number of frames generated as output by OpenPose were often inconsistent, contributing to the loss of data.

It would be interesting to see how the classifier performance changes over long-term interactions between the children and robots. Child engagement is likely to vary with continued exposure to the robots and inclusion of additional temporal features in the dataset may become important. We also aim to incorporate additional modalities to our dataset, including physiological signals like heart rate, electrodermal activity, body temperature and blood pressure, as well as audio features. For this complex feature set, we foresee a deep learning network to be a more suitable classifier choice capable of identifying patterns relating to different levels of social engagement in children.

CONCLUSION

In this paper, we presented a multi-label convolutional neural network classifier to formulate an automated measure of social engagement for children. To provide a personalized metric that is the best representation of the unique expression of emotion, interest and intention of each individual, we trained a separate classifier for each subject and then evaluated its performance. We designed the study to ensure the participants were not restricted in their movements at all in order to closely mimic naturalistic interactions in the real world. The use of this setting increases the complexity of data collection and analysis but enables the generalization of the presented analysis techniques to other interaction scenarios and populations, which sets this work apart from other research studies in this domain.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Office of Human Research, The George Washington University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the minor(s)' legal guardian/next of kin for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

CP conceived this project, designed and developed the systems, and mentored the overall project. HJ developed and conducted experiments with CP to collect the data, trained the CNN model with the data with WL, and analyzed the results. WL designed the basic CNN structure for training and mentored on the application to the data. HJ wrote the initial manuscript. WL and CP provided comments and edits to finalize this manuscript.

FUNDING

This research was supported by the National Institutes of Health (NIH) under the grant #R01-HD082914 through the National Robotics Initiative program (NRI) and the National Science Foundation (NSF) under the grant #1846658.

ACKNOWLEDGMENTS

The authors would like to acknowledge the contributions of Fatima Mumtaza Tourk and Baijun Xie for their assistance in the design of the sensory station setup and the behavioral coding.

REFERENCES

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub. doi: 10.1176/appi.books.9780890425596

Anzalone, S. M., Boucenna, S., Ivaldi, S., and Chetouani, M. (2015). Evaluating the engagement with social robots. *Int. J. Soc. Rob.* 7, 465–478. doi: 10.1007/s12369-015-0298-7

Bellini, S. (2006). The development of social anxiety in adolescents with autism spectrum disorders. *Focus Autism Other Dev. Disabl.* 21, 138–145. doi: 10.1177/10883576060210030201

Bevill, R., Cowin, S., and Park, C. H. (2017). *Motion Learning for Emotional Interaction and Imitation Of Children With Autism Spectrum Disorder* (Washington, DC: GW Annual Research Days).

Cabibihan, J. J., Javed, H., Ang, M., and Aljunied, S. M. (2013). Why robots? A survey on the roles and benefits of social robots in the therapy of children with autism. *Int J Soc Robot.* 5, 593–618. doi: 10.1007/s12369-013-0202-2

Cao, Z., Simon, T., Wei, S. E., and Sheikh, Y. (2017). "Realtime multi-person 2d pose estimation using part affinity fields," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7291–7299.

Castellano, G., Pereira, A., Leite, I., Paiva, A., and McOwan, P. W. (2009). *Detecting User Engagement With a Robot Companion Using Task and Social Interaction-Based Features*. Cambridge, Ma. doi: 10.1145/1647314.1647336

CMU Perceptual-Computing-Lab (2019). *OpenPose Demo – Output*. Retrieved from: https://github.com/CMU-Perceptual-Computing-Lab/openpose/blob/master/doc/output.md

Contaldo, A., Colombi, C., Narzisi, A., and Muratori, F. (2016). The social effect of being imitated in children with autism spectrum disorder. *Front. Psychol.* 7:726. doi: 10.3389/fpsyg.2016.00726

Dautenhahn, K., and Werry, I. (2004). Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragm. Cogn.* 12, 1–35. doi: 10.1075/pc.12.1.03dau

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

Dubey, I., Ropar, D., de, C., and Hamilton, A. F. (2015). Measuring the value of social engagement in adults with and without autism. *Mol. Autism* 6:35. doi: 10.1186/s13229-015-0031-2

Friard, O., and Gamba, M. (2016). BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods Ecol. Evol.* 7, 1325–1330. doi: 10.1111/2041-210X.12584

Groff, E. (1995). Laban movement analysis: Charting the ineffable domain of human movement. *J. Phys. Edu. Recreat. Dance* 66, 27–30. doi: 10.1080/07303084.1995.10607038

Gupta, R., Bone, D., Lee, S., and Narayanan, S. (2016). Analysis of engagement behavior in children during dyadic interactions using prosodic cues. *Comput. Speech Lang.* 37, 47–66. doi: 10.1016/j.csl.2015.09.003

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 *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Seattle, WA: ACM), 307–17. doi: 10.1145/2632048.2636065

Ingersoll, B. (2008). The social role of imitation in autism: Implications for the treatment of imitation deficits. *Infants Young Child.* 21, 107–119. doi: 10.1097/01.IYC.0000314482.24087.14

Javed, H., Burns, R., Jeon, M., Howard, A. M., and Park, C. H. (2019). A robotic framework to facilitate sensory experiences for children with autism spectrum disorder: a preliminary study. *ACM Trans. Human-Robot Interact. (THRI)* 9, 1–26. doi: 10.1145/3359613

Javed, H., Jeon, M., Howard, A., and Park, C. H. (2018). 'Robot-assisted socio-emotional intervention framework for children with Autism Spectrum disorder," in *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (ACM), 131–132. doi: 10.1145/3173386.3177082

Katagiri, M., Inada, N., and Kamio, Y. (2010). Mirroring effect in 2-and 3-year-olds with autism spectrum disorder. *Res. Autism Spectr. Disord.* 4, 474–478. doi: 10.1016/j.rasd.2009.11.004

Kim, J., Truong, K. P., and Evers, V. (2016). "Automatic detection of children's engagement using non-verbal features and ordinal learning," in *WOCCI* (San Francisco, CA), 29–34. doi: 10.21437/WOCCI.2016-5

Kishida, Y., and Kemp, C. (2006). Measuring child engagement in inclusive early childhood settings: Implications for practice. *Austral. J. Early Childhood* 31, 14–19. doi: 10.1177/183693910603100204

Kushki, A., Andrews, A. J., Power, S. D., King, G., and Chau, T. (2012). Classification of activity engagement in individuals with severe physical disabilities using signals of the peripheral nervous system. *PLoS ONE* 7:e30373. doi: 10.1371/journal.pone.0030373

Lahiri, U., Bekele, E., Dohrmann, E., Warren, Z., and Sarkar, N. (2012). Design of a virtual reality based adaptive response technology for children with autism. *IEEE Trans. Neural Syst. Rehab. Eng.* 21, 55–64. doi: 10.1109/TNSRE.2012.2218618

Lala, D., Inoue, K., Milhorat, P., and Kawahara, T. (2017). Detection of social signals for recognizing engagement in human-robot interaction. *arxiv [Preprint] arXiv*: 1709.10257.

Masuda, M., Kato, S., and Itoh, H. (2009). "Emotion detection from body motion of human form robot based on laban movement analysis," in *International Conference on Principles and Practice of Multi-Agent Systems* (Heidelberg: Springer), 322–34. doi: 10.1007/978-3-642-11161-7_22

Nadel, J. (2002). "Imitation and imitation recognition: functional use in preverbal infants and nonverbal children with autism," in *The imitative mind: Development, Evolution, and Brain Bases* 6, 42–62. doi: 10.1017/CBO9780511489969.003

Oertel, C., and Salvi, G. (2013). "A gaze-based method for relating group involvement to individual engagement in multimodal multiparty dialogue," in *Proceedings of the 15th ACM on International Conference on Multimodal Interaction* (Sydney, NSW: ACM), 99–106. doi: 10.1145/2522848.2522865

Ollendick, T. H., Weist, M. D., Borden, M. C., and Greene, R. W. (1992). Sociometric status and academic, behavioral, and psychological adjustment: a five-year longitudinal study. *J. Consult. Clin. Psychol.* 60:80. doi: 10.1037/0022-006X.60.1.80

Parekh, V., Foong, P. S., Zhao, S., and Subramanian, R. (2018). "AVEID: automatic video system for measuring engagement in dementia," in *23rd International Conference on Intelligent User Interfaces* (Tokyo: ACM), 409–413. doi: 10.1145/3172944.3173010

Rajagopalan, S. S., Murthy, O. R., Goecke, R., and Rozga, A. (2015). "Play with me—Measuring a child's engagement in a social interaction," in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Vol. 1* (Ljubljana: IEEE), 1–8. doi: 10.1109/FG.2015.7163129

Rehg, J., Abowd, G., Rozga, A., Romero, M., Clements, M., Sclaroff, S., et al. (2013). "Decoding children's social behavior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Portland, OR), 3414–3421. doi: 10.1109/CVPR.2013.438

Rudovic, O., Lee, J., Dai, M., Schuller, B., and Picard, R. W. (2018). Personalized machine learning for robot perception of affect and engagement in autism therapy. *Sci. Robotics* 3:19. doi: 10.1126/scirobotics.aao6760

Rudovic, O., Lee, J., Mascarell-Maricic, L., Schuller, B. W., and Picard, R. W. (2017). Measuring engagement in robot-assisted autism therapy: a cross-cultural study. *Front. Robotics AI* 4:36. doi: 10.3389/frobt.2017.00036

Sanefuji, W., and Ohgami, H. (2011). Imitative behaviors facilitate communicative gaze in children with autism. *Infant Ment. Health J.* 32, 134–142. doi: 10.1002/imhj.20287

Scassellati, B. (2007). "How social robots will help us to diagnose, treat, and understand autism," in *Robotics Research* (Berlin, Heidelberg: Springer), 552–563. doi: 10.1007/978-3-540-48113-3_47

Slaughter, V., and Ong, S. S. (2014). Social behaviors increase more when children with ASD are imitated by their mother vs. an unfamiliar adult. *Autism Res.* 7, 582–589. doi: 10.1002/aur.1392

Stanton, C. M., Kahn, P. H Jr., Severson, R. L., Ruckert, J. H., and Gill, B. T. (2008). "Robotic animals might aid in the social development of children with autism," in *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction* (Amsterdam: ACM), 271–278. doi: 10.1145/1349822.1349858

Tantam, D. (2000). Psychological disorder in adolescents and adults with Asperger syndrome. *Autism* 4, 47–62. doi: 10.1177/1362361300004001004

Tapus, A., Peca, A., Aly, A., Pop, C., Jisa, L., Pintea, S., et al. (2012). Children with autism social engagement in interaction with Nao, an imitative robot: a series of single case experiments. *Interaction Studeies* 13, 15–347. doi: 10.1075/is.13.3.01tap

Tiegerman, E., and Primavera, L. (1982). Object manipulation: an interactional strategy with autistic children. *J. Autism Dev. Disord.* 11, 427–438. doi: 10.1007/BF01531617

Tiegerman, E., and Primavera, L. H. (1984). Imitating the autistic child: facilitating communicative gaze behavior. *J. Autism Dev. Disord.* 14, 27–38. doi: 10.1007/BF02408553

Wakayama, Y., Okajima, S., Takano, S., and Okada, Y. (2010). "IEC-based motion retrieval system using Laban movement analysis," in *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems* (Berlin, Heidelberg: Springer), 251–260. doi: 10.1007/978-3-642-15384-6_27

Wimpory, D. C., Hobson, R. P., Williams, J. M. G., and Nash, S. (2000). Are infants with autism socially engaged? A study of recent retrospective parental reports. *J. Autism Dev. Disord.* 30, 525–536. doi: 10.1023/A:1005683209438

Conflict of Interest: 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.

frontiers
in Robotics and AI

ORIGINAL RESEARCH
published: 30 April 2020
doi: 10.3389/frobt.2020.00058

In the Wild HRI Scenario: Influence of Regulatory Focus Theory

*Roxana Agrigoroaie, Stefan-Dan Ciocirlan and Adriana Tapus**

Autonomous Systems and Robotics Laboratory, U2IS, ENSTA-Paris, Institut Polytechnique de Paris, Palaiseau, France

Research related to regulatory focus theory has shown that the way in which a message is conveyed can increase the effectiveness of the message. While different research fields have used this theory, in human-robot interaction (HRI), no real attention has been given to this theory. In this paper, we investigate it in an in the wild scenario. More specifically, we are interested in how individuals react when a robot suddenly appears at their office doors. Will they interact with it or will they ignore it? We report the results from our experimental study in which the robot approaches 42 individuals. Twenty-nine of them interacted with the robot, while the others either ignored it or avoided any interaction with it. The robot displayed two types of behavior (i.e., promotion or prevention). Our results show that individuals that interacted with a robot that matched their regulatory focus type interacted with it significantly longer than individuals that did not experience regulatory fit. Other qualitative results are also reported, together with some reactions from the participants.

Keywords: HRI, regulatory focus, in the wild, acceptance, social robotics

OPEN ACCESS

Edited by:
*Chung Hyuk Park,
George Washington University,
United States*

Reviewed by:
*Vicky Charisi,
European Commission, Joint
Research Centre, Belgium*
*Myounghoon Jeon,
Virginia Tech, United States*

****Correspondence:***
*Adriana Tapus
adriana.tapus@ensta-paris.fr*

Specialty section:
*This article was submitted to
Human-Robot Interaction,
a section of the journal
Frontiers in Robotics and AI*

Received: *16 May 2019*
Accepted: *06 April 2020*
Published: *30 April 2020*

Citation:
*Agrigoroaie R, Ciocirlan S-D and
Tapus A (2020) In the Wild HRI
Scenario: Influence of Regulatory
Focus Theory. Front. Robot. AI 7:58.
doi: 10.3389/frobt.2020.00058*

1. INTRODUCTION

It is a well-known fact since ancient times that people approach pleasure and avoid pain. Looking at this from a different perspective, we can imagine that people approach or engage in tasks which they find enjoyable, and avoid tasks or situations which brings them pain, or that they do not find enjoyable. While not all tasks that individuals have to perform in their every-day working life can be viewed as only enjoyable or not enjoyable, the question arises at to which strategies will they apply in order to achieve their goals? In the psychology literature, Higgins (1997) introduces a theory stating that individuals adopt one of two possible approaches in achieving a goal.

The theory is the Regulatory Focus Theory (RFT) (Higgins, 1997) and the two approaches are: promotion and prevention. In Crowe and Higgins (1997), the authors characterize promotion type individuals as individuals that guide their actions toward achieving their goals. Whereas, prevention type individuals guide their actions in order to avoid failure.

According to Higgins (2000), regulatory fit is defined as an increased motivational intensity that is experienced when the manner in which an individual engages in an activity sustains his/her current interests. As an example, in order to successfully pass a course, a promotion type individual will be more inclined to read supplementary material in order to maximize their results, while a prevention type individual will be careful to fulfill the minimum course requirements in order to pass.

Furthermore, it was also shown in Higgins (2000, 2005) that people that experience regulatory fit, engage more strongly in their current activity. Therefore, it is expected that individuals who experience regulatory fit will engage for longer in a given task. Moreover, regulatory fit can be used to effectively change attitudes and behaviors, and to improve the quality of life in interpersonal conflicts. For instance, regulatory fit and non-verbal cues can be used to increase persuasiveness (Cesario and Higgins, 2008) (e.g., body gestures, movement speed, or speech rate).

Robots are more and more present in our every-day lives. As a result, more research is being carried in which robots play a social role in human-centric environments. Their roles can be diverse, ranging from a teacher for children (Tazhigaliyeva et al., 2016), to personal companion (Breazeal, 2017). By taking into consideration the regulatory focus theory and the more and more social role of robots, we can imagine that robots have the potential of helping individuals to achieve their goals, to increase their motivation (Nakagawa et al., 2011; Andrist et al., 2015).

The authors of Faur et al. (2015) have designed artificial agents based on RFT. They designed a game scenario and have found that regulatory fit had an effect on the prevention type end-users (i.e., likability of the game). The RFT was also successfully used in designing persuasive technologies (Rezai et al., 2017). On the other hand, a review of the HRI literature has shown that RFT has not received much attention. The first study using RFT in an HRI scenario is presented in Cruz-Maya et al. (2017). The authors have improved the performance of the participants in a Stroop task, by matching the behavior of the robot with the regulatory focus type of the participants. The same authors, have continued to use of RFT in a negotiation type scenario with a humanoid robot (Cruz-Maya and Tapus, 2018). Their results show that RFT and regulatory fit can be successfully used in HRI scenarios.

The study presented in this paper is based on the works of Cruz-Maya et al. (2017), Cruz-Maya and Tapus (2018). Thus, the purpose of this study is to investigate if RFT can be applied in an in the wild HRI scenario. Our main research question is **RQ: "How do individuals (based on their RFT type) react when a robot appears at their doorway to ask them to perform a short questionnaire?"**. By applying different strategies (either promotion or prevention) we wanted to investigate if individuals will be more inclined to perform the task. Of interest for this study is not the answers given to the questionnaire, but if the participants approached the robot to interact with it or they just ignored it based on the robot's behavior and user's regulatory focus type (promotion or prevention type).

The paper is structured as follows. Section 2 is dedicated to the presentation of the interaction scenario, robot navigation, and robot behavior. The results are presented in section 3, while section 4 shows a discussion of these results. Lastly, the paper is concluded in section 5.

2. METHODOLOGY

This study is designed as a 2 (behavior of robot, i.e., promotion or prevention) × 2 (regulatory focus type of the participants) between participants experimental study. In **Table 1** is presented the distribution of the participants into the four conditions.

According to Higgins (2000, 2005), individuals who experience regulatory fit engage more strongly in the activity they are performing. Therefore, we hypothesize that the participants that experience regulatory fit will engage for longer with the robot than the participants that do not experience regulatory fit.

H: Participants that have a matching regulatory focus type with the behavior of the robot (i.e., regulatory fit) will interact with the robot for longer than the participants that do not have a matching regulatory focus type with the behavior of the robot.

TABLE 1 | Distribution of participants based on different factors.

Knowledge about robotics				
1 ("Not at all")	2	3	4	5 ("Very much")
4	4	9	6	6
Regulatory focus results				
Promotion				Prevention
19				10
Conditions				
Robot	Participant			
	Promotion	Prevention		
Promotion	11	6		
Prevention	8	4		

Taking into account the different office layouts and the different times needed by the participants to reach the robot, we consider the measure **time_interaction** as the time needed by the participants to fill in the questionnaire (i.e., between pressing the **START** and **QUIT** buttons) and it represents the interaction time between the participants and the robot.

To test our research hypothesis, we consider the measure time_interaction as the *dependent variable*, and the regulatory fit/no fit as the *independent measure*.

2.1. Scenario

For this study, Tiago, a robot developed by PAL Robotics[1], was used. The robot features a mobile base, a lifting torso, a touch screen mounted on its torso (as shown in **Figure 1**), and a head. The eyes of the robot are equipped with an RGB-D camera and the speakers are located between the head of the robot and the touch-screen.

The study presented in this paper was carried out at the university where the authors are located. The participants are some of the employees from the various departments of the university and they were not informed in any way about this study. The diagram of the experimental scenario is shown in **Figure 2**.

The experiment starts with the robot loading the map (see section 2.2) and the points of interest (POI) corresponding to the doorways of the offices in the university. Each POI had a corresponding flag that indicated if the office was visited before or not. The robot started with the first POI and then continued to visit each office, until all offices were visited.

Once arrived at a POI, by using its laser, the robot checked if the current door was open. However, due to frequent laser malfunctions, this information was given to the robot by the investigators. Even if the robot was navigating autonomously, the

[1] http://tiago.pal-robotics.com/

FIGURE 1 | Tiago robot.

investigators were always in the close proximity of the robot, just to make sure that there were no problems during the interaction. Moreover, the investigators made sure that the participants did not see them.

Next, the robot started checking how many people were in the office. For small interaction distances (i.e., <1.5 m), the robot is able to accurately detect how many people are in an office by using the face detector provided by the Dlib toolkit (King, 2009). However, since there were no two offices with the same layout, and the lightning conditions were very different from one office to another, using an automatic face detector proved not to be very reliable for this scenario. Therefore, as a fail-safe method, we decided to manually determine how many people were in an office, by checking the video-feed provided by the RGB-D camera located in the eyes of the robot.

The interaction was designed for at most two people in an office. If more than two people were detected, the robot would turn around and leave the office. Otherwise, it randomly chose its behavior (i.e., either promotion or prevention) and it would say the message presented in section 2.3. It had a waiting time of 30 s (Timeout, in **Figure 2**). This moment is considered as the first ping (i.e., the first time that the person hears the message from the robot). If during the first waiting time there was no reaction from the person, the robot approached the desk and it would repeat the same message after saying *"Excuse me, can you please listen to me?"*. This moment is considered as the second ping (i.e., the second time that the person hears the message). When the robot approached the desk it would use a moving speed appropriate to the behavior that is currently displaying (see section 2.3). Then, it would wait for 15 s (Timeout2), and if there was no reaction from the person, it would say again the same message (i.e., third ping) and waited for another 30 s (Timeout) for a reply. If the participant still did not want to interact with the robot, it thanked the person and approached the second person in the office (if there was one) or it left the office. When leaving the office the robot set the flag for the office as visited and it approached the next office.

By reaction from the person it is understood that the person would approach the robot and press on the **START** button displayed on the tablet of the robot [see **Figure 3** (left)]. The participants could stop at any time by pressing on the **Quit** button located on the upper right corner of the screen [see **Figure 3** (right)]. The participant could see at all times the number of the current question and the total number of questions.

The task that the participants had to perform, was a 28 questions questionnaire that was displayed on the tablet of the robot. The questions concerning stress at work were selected from the Copenhagen Psychosocial Questionnaire (COPSOQ) (Kristensen et al., 2005). The dimensions selected from the questionnaire were: Cognitive demands (e.g., "Does your work require you to make difficult decisions?"), Work engagement (e.g., "I am enthusiastic about my job"), Stress (e.g., "How often have you had problems relaxing?"), Cognitive stress (e.g., "How often have you had problems concentrating?"), and Self-efficacy (e.g., "I feel confident that I can handle unexpected events"). Some other questions were added that were not part of the questionnaire (e.g., "Do you have enough time in a day to complete your work?").

2.2. Robot Navigation

A total of four maps were created of the entire environment. Each floor of the school contains offices as well as laboratories and small classrooms. We created maps only for the office regions on each floor. For this purpose, the advanced navigation system designed by Pal Robotics was used. The navigation module is based on the ROS (Quigley et al., 2009) 2D navigation stack[2].

The navigation software can be used to perform the mapping as well as to enable the robot to autonomously navigate on the selected map. The mapping system of the robot (i.e., GMapping[3]) uses the readings from the 2D laser scanner which is located on the mobile base to create an Occupancy Grid Map (OGM). Usually when there is somebody in an office, the office door is left open. Therefore, as the mapping was done after the usual working hours, most of the offices were already closed (see **Figure 4** (left) for the original map). Therefore, the map was modified in

[2]http://wiki.ros.org/navigation?distro=indigo
[3]http://wiki.ros.org/gmapping

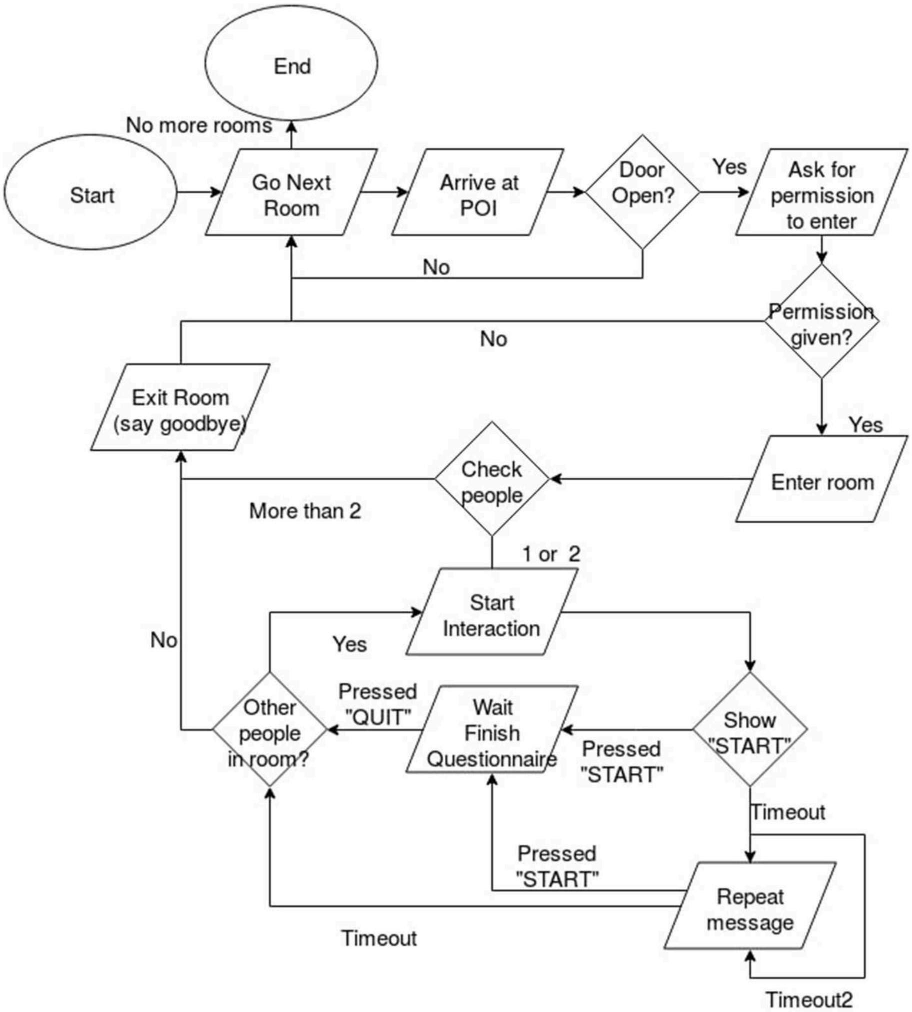


FIGURE 2 | Experimental scenario.

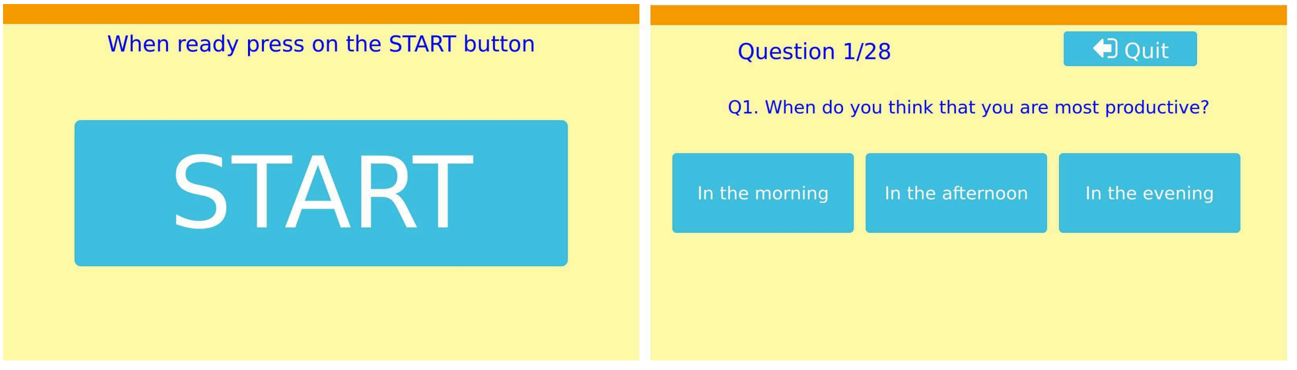


FIGURE 3 | (Left) Start button for starting the questionnaire; **(Right)** Example question.

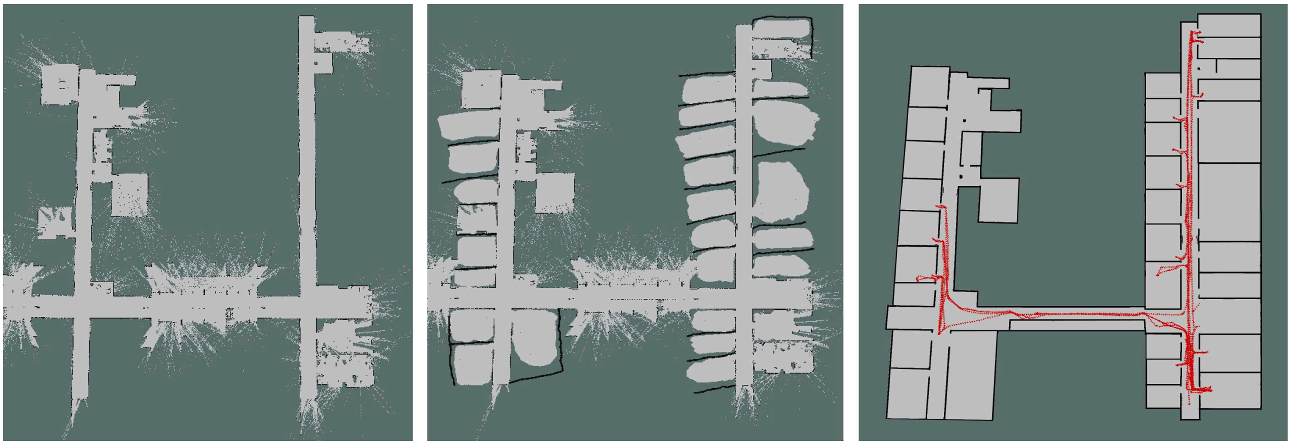


FIGURE 4 | (Left) Map created by the robot; **(Middle)** Map used for navigation; **(Right)** Example of one of the paths of the robot overlayed on one of the modified maps.

GIMP[4] so as to contain the approximate shape of the offices [see **Figure 4** (middle)]. Otherwise, the robot did not know about the existence of the office and the global planner would not allow the robot to enter the office. This modified map was used by the robot to localize itself and to navigate autonomously in the environment. We used the local and the global planners designed by Pal Robotics.

The doorway of each office was designated as a POI. The robot could then easily plan its path from its current location to any of the POIs defined on the map. For each office, we also defined a secondary POI immediately inside the office, in front of the door. This enabled the robot to easily enter each office. At the end of each interaction the robot navigates to the primary POI of the office and then to the POI of the next office door. As the investigators were always in the close proximity of the robot, in case the robot's local planner would not find a possible path and detect that it could not exit the office, they controlled the robot to reach the doorway and then let the robot autonomously move to the next POI. The intervention of the experimenters was required only for a few instances. In the majority of the interactions, the robot successfully found a path for it to exit the office and to reach the next POI.

Next, the behavior of the robot is presented.

2.3. Robot Behavior

As shown in section 1, the way in which a message is conveyed can increase the effectiveness of that message. Therefore, of importance is the way in which the message is framed by the robot in order to persuade the participants to stop whatever they were doing and to start an interaction with it.

The robot could display one of two behaviors (i.e., promotion type or prevention type). As shown in Lee and Aaker (2004), a regulatory fit can be created by using an eager framing for promotion type individuals, and by using a vigilant framing for prevention type individuals. The research presented in Cesario and Higgins (2008) describes which non-verbal cues can be used to define a vigilant and an eager type behavior. Thus, an eager type behavior is characterized by: *fast body movement, fast speech rate, and open hand movements*, among others. On the other hand, a vigilant type behavior is characterized by: *slower body movement, a slower speech rate, and by gestures that show precision*, among others. As our robot does not feature any arms, we could only change the body movement and the speech rate between the two behaviors. Moreover, we also changed the speed of the approach accordingly. No indication was found in the literature for the specific values for the speech rate and the approach speed for the robot. They were empirically set by the experimenters, by taking into consideration also some of the hardware limitations of the robot. Therefore, as the maximum speed of the robot is of 1 m/s, we decided to select an approach speed of 0.6 m/s for the promotion type robot, and an approach speed of 0.2 m/s for the prevention type robot. While preparing the study, the robot did not have a French TTS engine, thus we used a different TTS[5] engine to generate the audio files that contained the speech of the robot. From the online TTS engine, we chose the slow speech rate for the prevention robot (which corresponds to a speech rate of ∼150 words/min) and the fast speech rate for the promotion robot (which corresponds to a speech rate of ∼198 words/min). For each interaction the robot randomly selected between the two types of behavior (i.e., promotion or prevention).

Another important aspect to be considered is how the message is presented. More specifically, the framing of the message can show the recipient the desirable or the undesirable outcomes from successfully or unsuccessfully pursuing a certain goal (Higgins, 2005). Therefore, for the promotion type behavior, the message had to be framed so as to show what was the desirable outcome for the robot if the individual successfully pursuits the task asked by the robot. However, for the prevention type behavior, the message had to be framed in such a manner as for

[4]https://www.gimp.org

[5]http://www.fromtexttospeech.com

the individual to understand which was the undesirable outcome if he/she does not successfully pursuit the task asked by the robot. Based on these considerations, we designed the following messages and non-verbal cues:

2.3.1. Promotion Type Robot

The robot had a moving speed of 0.6 m/s and the speech rate of 198 words/min, with a total speech time of 18 s. The message was the following:

> Hello. My name is Tiago. I am trying to learn more about stress at the workplace. I have 28 questions for you. **If you answer at least 20 of them, I will be able to learn more about what it is like to be active in the workforce.** You can stop at anytime you want by pressing on the QUIT button. When you are ready you can press on the START button.

2.3.2. Prevention Type Robot

The speed of the robot was set to 0.2 m/s and the speech rate of 150 words/min, with a total speech time of 25 s. The message was the following:

> Hello. My name is Tiago. I am trying to learn more about stress at the workplace. I have 28 questions for you. **If you do <u>not</u> answer at least 20 of them, I will <u>not</u> be able to learn more about what it is like to be active in the workforce.** You can stop at anytime you want by pressing on the QUIT button. When you are ready you can press on the START button.

2.4. Participants

For this experiment, a total of 42 participants were approached by the robot. Out of these, 29 (69%) interacted with it, while the others either avoided it completely or simply ignored it. At the end of the experiment, the 29 participants (8 female and 21 male) that interacted with Tiago, signed a consent form that allows us to use their data for research purposes. Moreover, they were also asked to fill in the questionnaires presented at the end of this section, and to answer some demographic questions. The ages of the participants ranged between 23 and 52 years old ($M = 36.42$, $SD = 9.86$). When asked about their background, 14 of them had a computer science background, 7 had a technical background, while for the other participants their backgrounds were diverse, including, linguistics, statistics, human resources, or art history. All the participants were asked to rate their knowledge about robotics, on a scale from 1 ("Not at all") to 5 ("Very much"). The results are shown in **Table 1**. Even if the majority of the participants (17 out of 29) had no serious knowledge about robotics, 25 of them interacted with a robot before.

To determine the regulatory focus type of the participants, the Regulatory Focus Questionnaire—proverb form (Faur et al., 2017) was given to each of them upon the completion of the experiment. Therefore, the experimenters did not know before the interaction the regulatory type of the participants. The questionnaire contains 18 proverbs and it was originally developed in French. The proverbs were translated into English for the 6 participants that were not French native. The distribution of the participants is shown in **Table 1**.

From the BIG5 (Goldberg, 1990) personality questionnaire only the questions related to the Conscientiousness personality trait were selected. As shown in the review paper (Barrick and Mount, 1991), research has shown that an individual with high conscientiousness is dependable, hard-working, persevering. Therefore, we believe that the level of conscientiousness will influence the number of questions answered during the interaction.

The last questionnaire that the participants had to fill was a custom designed post-questionnaire, in which the participants were asked to rate, on a Likert scale from 1 ("Strongly Disagree") to 5 ("Strongly Agree") their thoughts about the robot's behavior (e.g., polite, persuasive, motivating, intimidating).

3. RESULTS

3.1. Hypothesis Results

As previously shown, our research tries to show that participants will interact for longer with a robot that displays a behavior that matches their regulatory focus type than with a robot that displays a behavior that does not match their regulatory focus type.

The two assumptions for the ANOVA analysis are the normal distribution of the data and the variance across groups has to be homogeneous. First, we tested the normal distribution of the data by applying a Shapiro-Wilk normality test. With a p-value > 0.05 ($W = 0.94, p = 0.16$), we can conclude that our data is normally distributed. Next, we apply Levene's test for homogeneity of variance across groups. Based on our results, [$F_{(1,27)} = 1.79, p = 0.19$] we can assume the homogeneity of variances in the two groups.

Therefore, we can apply one-way ANOVA analysis to test our hypothesis. The results of the test, as well as the summary statistics by groups (i.e., count, mean, standard deviation) are presented in **Table 2**. Based on these results, we can conclude that our **research hypothesis is validated**. The participants that interacted with a robot that matched their regulatory focus type interacted with it for longer than the participants that interacted with a robot that did not match their regulatory focus type. This result is also represented graphically as a raincloud plot[6] in **Figure 5**. To further validate our results we have performed a power analysis for the one-way analysis of variance by applying the specific function from the **pwr** R package that implements the power analysis as outlined by Cohen (2013). With the two groups (fit, no fit), a common sample size in each group of 14 participants and a power of 0.8, and a significance level of 0.05, our results show that the effect size for our one-way ANOVA analysis is equal to 0.55. According to Cohen (2013), this result represents a large effect size.

Next, we investigated separately the results for the Promotion type individuals, as well as for the Prevention type individuals. Raincloud plots were created for each group, as shown in **Figure 6** for promotion type individuals, and in **Figure 7** for the prevention type individuals, respectively. The average interaction times for each group are shown in **Table 3**. While it is clear

[6]https://micahallen.org/2018/03/15/introducing-raincloud-plots/

from the table that the participants that experienced regulatory fit interacted longer than participants that did not experience regulatory fit, the differences between the two robot behaviors are not significant [$F(1, 17) = 2.83, p = 0.11$ for promotion type individuals, and $F(1, 17) = 1.78, p = 0.22$ for prevention type individuals, respectively]. Further investigation is needed in order to determine if there are significant differences between the promotion and prevention types individuals.

3.2. Qualitative Results

We believe it is also noteworthy to present a selection of the qualitative results. They provide valuable insight into how individuals react when they are approached by a robot, without being told beforehand. We consider as qualitative results some of the reactions of the individuals that either interacted, avoided, or ignored the robot. First, we present examples of the reactions of the individuals that did not interact with the robot.

A total of 13 individuals (5 females and 8 males) refused to interact with Tiago. One individual completely ignored it, by putting back his headphones. Other two individuals just looked at the robot while it talked to them, but did not display any intention of interaction [see **Figure 8** (left)].

Two individuals first wanted to interact with the robot, however, they had to leave their offices due to work obligations. One individual came to our office and approached us to ask if it is necessary for her to interact with the robot, as she was very busy. We tried to limit the interaction with her as much as possible by telling her that it is totally optional and that it is her choice if she wants to interact with the robot or not.

In one office, the occupants were very angry when they saw the robot in their doorway. They thought that the robot was very invasive and they demanded for the robot to be removed from their office door. As the investigators were seeing and hearing remotely the reaction of these participants, the robot

TABLE 2 | Results for research hypothesis.

	Summary statistics by groups		
Group	**Count**	**Mean**	**SD**
Fit	15	268	88.7
No Fit	14	210	60.2
	Anova results		
	Df	**F**	**Pr (>F)**
Fit	1	4.21	**0.049***
Residuals	27		

**Represents the standard way of representing a significant result for a p-value less than 0.5.*
Bold values indicates that the result is significant.

TABLE 3 | Interaction times for each group.

	Interaction times (s)	
	Promotion type Mean (SD)	**Prevention type Mean (SD)**
Promotion behavior	265.42 (94.05)	225.90 (26.89)
Prevention behavior	197.25 (76.27)	274.24 (84.79)

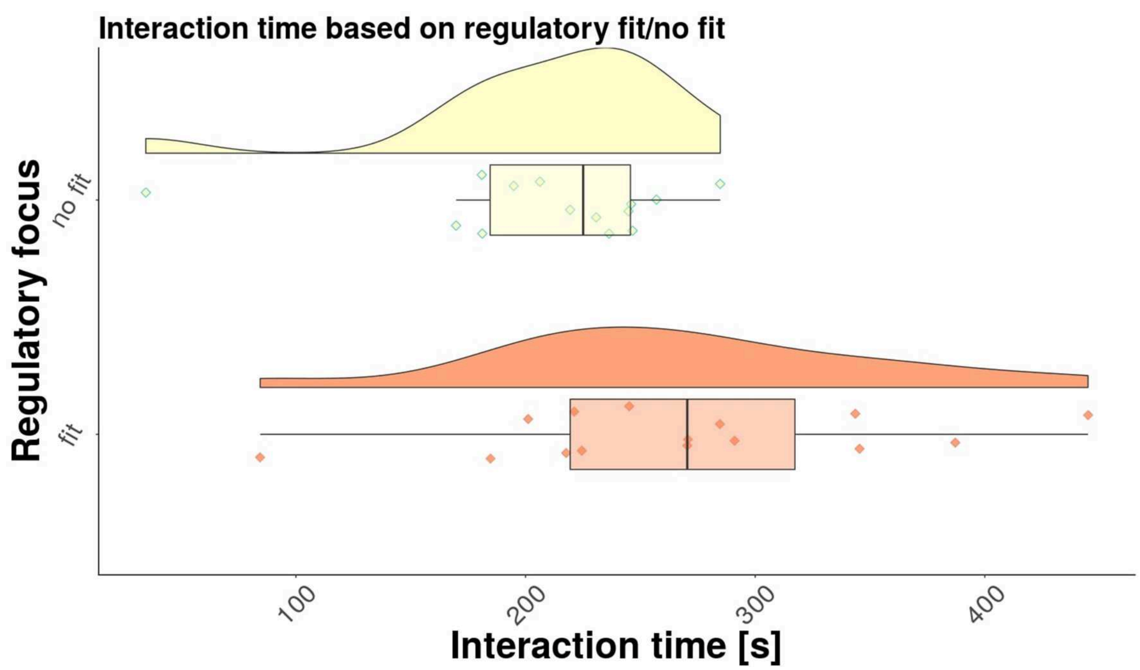


FIGURE 5 | Raincloud plot for interaction time based on regulatory fit/no fit.

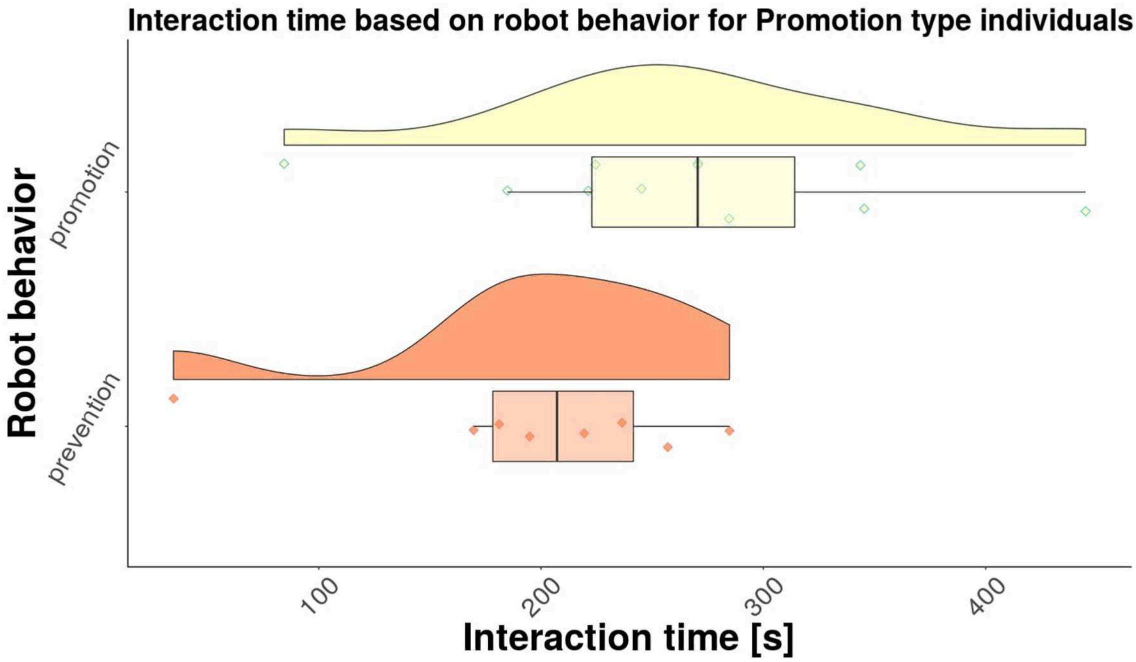


FIGURE 6 | Interaction times for promotion type individuals.

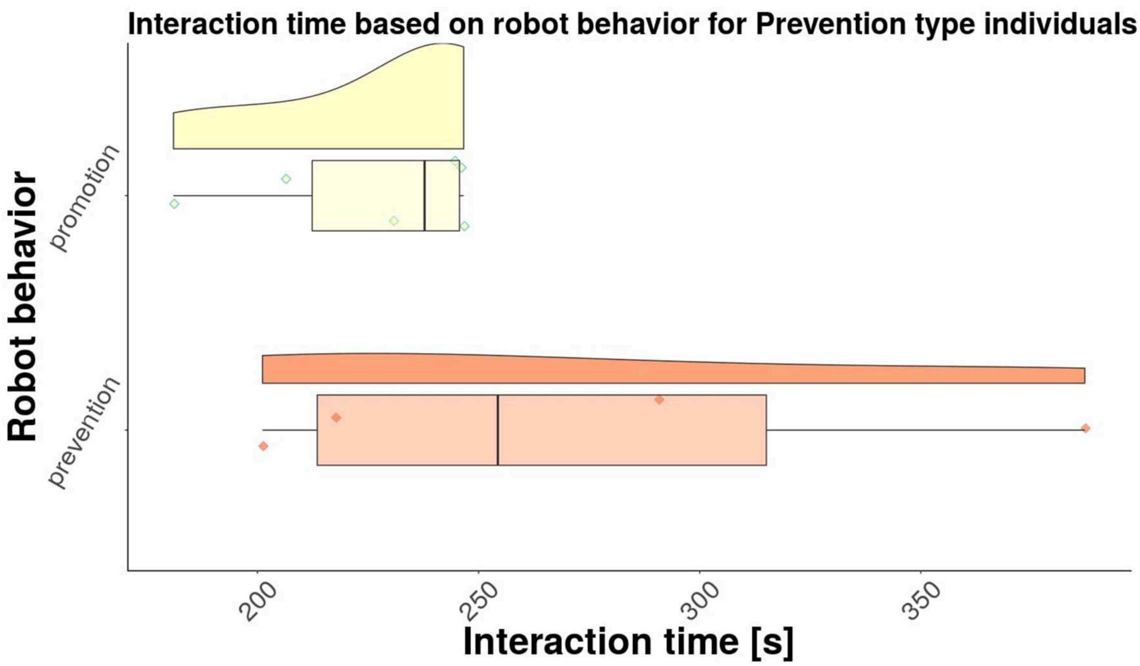


FIGURE 7 | Interaction times for prevention type individuals.

was remotely controlled to say "*Bye Bye! Thank you for the interaction*" and to leave the office.

Another interesting reaction from the individuals that avoided the robot consisted of closing the office when the robot approached. For example, two individuals (from two different offices) showed real interest when they saw the robot in the hallway. However, when the robot tried to approach their doors, they shut the door clearly showing that they had no interest to interact with it.

From the interactions with the 29 participants there were a couple of unexpected reactions. One participant when seeing the

FIGURE 8 | (Left) Individual just looking at the robot without interacting; **(Middle)** Participant filming Tiago; **(Right)** One participant gave Tiago some eye-brows.

robot, put two post-its on the head of the robot representing the eye-brows [see **Figure 8** (right)]. Other participants were very excited and started taking pictures of the robot or even filming it [see **Figure 8** (middle)].

There were participants that were very serious while filling the questionnaire, while others smiled and continuously spoke with Tiago. One female participant saw the robot in the hallway and she started talking to it and saying things like, "*Come and follow me Tiago. I want to interact with you.*" She saw that the name Tiago was written on the back of the robot, so she supposed that the robot is called Tiago.

TABLE 4 | Distribution of participants based on different factors (1 - "Strongly disagree"; 5 - "Strongly agree").

	Regulatory fit					Regulatory no-fit				
	1	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
Polite	–	–	2	10	3	–	–	1	6	7
Friendly	–	1	3	10	1	–	1	3	7	3
Intimidating	5	4	2	3	1	3	4	4	1	1
Motivating	–	5	7	3	–	1	2	6	4	1
Persuasive	1	2	3	7	1	1	2	3	7	1

3.3. Other Results

First, we looked at the number of questions answered by the participants. From the 29 individuals that started the questionnaire, 26 answered all questions, one participants answered two questions, one participant answered 4 questions and one participant answered 11 questions. Taking into consideration these results, we do not have enough participants that did not complete the entire questionnaire in order to investigate if the conscientiousness level of the participants had any influence. Of the 29 participants, 28 had a conscientiousness score ≥ 3 (on a scale from 1 to 5).

Of the total participants, 24 approached the robot after the first ping (i.e., the first time that the robot said the message presented in section 2.3). One participant approached the robot after the second ping (promotion type individual interacting with a promotion type robot), and four participants approached the robot only after the third ping. Again, considering the distribution of the participants based on the number of pings, we cannot investigate further these results.

Next, we were interested in finding out what were the impressions of the participants of the robot. By using a post-questionnaire we assessed on a scale from 1 ("Strongly disagree") to 5 ("Strongly agree") if the participants thought that the robot was: polite, friendly, intimidating, motivating and persuasive. In **Table 4** are shown the results of the post-questionnaire. The majority of the participants, both in the regulatory fit group, as well as in the regulatory no-fit group, agreed or strongly agreed with the statement that the robot was polite, with an average rating of $M = 4.24$ ($SD = 0.63$). A similar result was observed for the statement that the robot is friendly. Eleven participants in the regulatory fit group either agreed or strongly agreed that the robot is friendly, while 10 participants in the regulatory no-fit group either agreed or strongly agreed with the statement. An average rating of $M = 3.79$ ($SD = 0.77$) was obtained. With an average rating of $M = 2.51$ ($SD = 1.32$), the participants in the study did not agree with the statement that the robot is intimidating. Similar average ratings were found in the regulatory fit group ($M = 2.4$) and in the regulatory no-fit group ($M = 2.64$).

After completing the experiment, the path of the robot was overlayed on the extended map of each floor. An example of a final result of a map is shown in **Figure 4** (right). From the figure it can be seen that the robot moved very much in the area around the lower right corner of the map. This is due to the fact that is the location of the laboratory from where the experiment started.

4. DISCUSSION

Even if not all the individuals approached by the robot chose to interact with it, the reaction of all 42 individuals is interesting and needs to be considered. It is quite remarkable that some people would choose to intentionally get up from their desks and to close the door of their offices when the robot tried to interact with them. Even more so, as the participation was completely voluntary. The robot could be ignored, or, as the robot also told them, the interaction could have been stopped at any given time, by simply pressing on the Quit button. Therefore, we consider that this is one of the limitations of this study. It is possible that some of the individuals that chose to deliberately ignore the robot were indeed quite busy and that is why they chose to close the door. Maybe at a different time of the day, or even during a different day they would have been more inclined to interact with the robot. Another aspect to consider is that in some offices there were multiple individuals. This could have influenced if the individuals chose to interact with the robot or not and also the length of the interaction.

When the interaction was over and the investigators approached the participants to give them the questionnaires, some participants stated that when they reached question 20, the number that had to be reached so that the robot can learn more about stress at the work place, they thought about pressing on the Quit button. However, they chose to complete the questionnaire, saying that, as long they started the questionnaire, they might as well finish it.

One results that we found shows that when interacting with a promotion type robot, the participants started approaching the robot faster than when interacting with a prevention type robot. RFT states that increased motivation happens when there is a regulatory fit between the regulatory profile of the person and the behavior of the agent it interacts with (Higgins, 2005). However, in Cesario and Higgins (2008) it was shown that the effectiveness of a message can be increased by using faster rates for conveying a message. This also leads to an increase in the competence and credibility of the source of the message. Therefore, we can conclude, that the participants that interacted with the promotion type robot (i.e., 17 out of 29 interactions), were more eager to approach the robot. This result can be of potential interest for designing HRI scenarios.

Even though, at some moments there was still a need for the investigators to intervene in order to make sure that the interaction was as natural as possible, the robot was autonomous most of the time of the interaction. It has to be taken into consideration that the interaction was carried out in the wild and not in a controlled environment (e.g., in a laboratory). Furthermore, the participants were not aware that they will interact with a robot. They were not previously recruited to take part in the experiment. They were performing their everyday tasks at the workplace and suddenly the robot appeared in their doorway. Thus, the reactions of some individuals are totally understandable (e.g., ignoring the robot, going outside their offices and looking in the hallway to see if they can find the operator of the robot), while others can be considered as surprising (e.g., intentionally going and closing the door).

Further research in the wild is needed in order to better understand how individuals of all ages react toward robots. Of course, these results might have been different if the interaction were to take place with students, with the elderly, or with different groups of individuals. Furthermore, results might have been different in other countries.

One limitation of this study is related to the relatively small number of individuals approached by the robot (i.e., 42 individuals). The RFT was mostly studied in the psychology literature. Therefore, the number of participants in these studies is in the hundreds of participants, while for our study we recruited 42 participants. However, if we consider the related studies in the HRI literature, we can find a similar number of participants as in our study [e.g., in Faur et al. (2015) 20 participants were recruited, while in Cruz-Maya and Tapus (2018) a total of 40 participants took part in the study]. Therefore, we consider that before performing a large scale study, it was important to investigate and to try to understand how individuals might react in such a scenario. Further research is currently planned based on the currently obtained results. We hypothesize that a social robot displaying a behavior in accordance with the regulatory focus theory (i.e., promotion or prevention) can be used in different tasks (e.g., to play cognitive games, to motivate individuals to finish undesirable tasks) and with different populations (e.g., with children, with the elderly).

5. CONCLUSION AND FUTURE WORK

In this paper, we have presented a study carried out with 42 participants in which a humanoid robot approached them in their own offices without being previously informed by the investigators that the interaction will take place (i.e., in the wild type of interaction). Out of the 42 individuals approached by the robot, only 29 interacted with it. The other 13, either avoided the robot or ignored it. In the interaction, the robot displayed one of two types of behaviors: promotion type or prevention type. The behavior was modeled on RFT that exists in the psychology literature. More specifically, in a promotion type of behavior the robot moved faster, spoke with a higher speech rate, and the message communicated was framed in the context of the desirable outcomes that can be obtained from successfully carrying out the task suggested by the robot. On the other hand, a prevention type behavior means lower moving speed, lower speech rate, and a message framed so as to show the undesirable outcomes that result from unsuccessfully pursuing a certain goal.

Our results show that the interaction time with a robot that matches the regulatory focus type of an individual is significantly longer than the interaction time with a robot that does not match the regulatory focus type of the individual. Therefore, we posit that the regulatory focus theory has to be considered when designing interactions between robots and humans.

Our future work will be focused on using a different RGB-D sensor so that the face detector can be more reliable. Furthermore, the approach behavior of the robot will be improved so that no intervention from the human operator is required. And finally, the French TTS will be installed on the robot and a speech recognition module will be used for a more natural dialog between the end users and the robot. Concerning the results of our work, we consider them as a basis for our future work. We plan on doing more in the wild experiments in order to test how the two robot behaviors (i.e., promotion and prevention) can be used in HRI.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by ENSTA Paris. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

RA and S-DC contributed to the design of the experiment, to run the experiment, data analysis, and writing the article. AT contributed to the design of the experiment, to run the experiment, and writing the article.

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 (ACM)* (Seoul), 3603–3612. doi: 10.1145/2702123.2702592

Barrick, M. R., and Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Pers. Psychol.* 44, 1–26. doi: 10.1111/j.1744-6570.1991.tb00688.x

Breazeal, C. (2017). "Social robots: from research to commercialization," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (ACM)* (Vienna), 1. doi: 10.1145/2909824.3020258

Cesario, J., and Higgins, E. T. (2008). Making message recipients "feel right" how nonverbal cues can increase persuasion. *Psychol. Sci.* 19, 415–420. doi: 10.1111/j.1467-9280.2008.02102.x

Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences*. Cambridge, MA: Routledge.

Crowe, E., and Higgins, E. T. (1997). Regulatory focus and strategic inclinations: promotion and prevention in decision-making. *Organ. Behav. Hum. Decis. Process.* 69, 117–132. doi: 10.1006/obhd.1996.2675

Cruz-Maya, A., Agrigoroaie, R., and Tapus, A. (2017). "Improving user's performance by motivation: matching robot interaction strategy with user's regulatory state," in *International Conference on Social Robotics* (Tsukuba), 464–473. doi: 10.1007/978-3-319-70022-9_46

Cruz-Maya, A., and Tapus, A. (2018). "Negotiating with a robot: analysis of regulatory focus behavior," in *2018 IEEE International Conference on Robotics and Automation (ICRA)* (Brisbane). doi: 10.1109/ICRA.2018.8460611

Faur, C., Martin, J.-C., and Clavel, C. (2015). "Matching artificial agents' and users' personalities: designing agents with regulatory-focus and testing the regulatory fit effect," in *Annual Meeting of the Cognitive Science Society (CogSci 2015)* (Pasadena, CA).

Faur, C., Martin, J.-C., and Clavel, C. (2017). Measuring chronic regulatory focus with proverbs: The developmental and psychometric properties of a french scale. *Pers. Indiv. Differ.* 107, 137–145. doi: 10.1016/j.paid.2016.11.037

Goldberg, L. (1990). An alternative 'description of personality': the big-five factor structure. *J. Pers. Soc. Psychol.* 59, 1216–1229. doi: 10.1037/0022-3514.59.6.1216

Higgins, E. T. (1997). Beyond pleasure and pain. *Am. Psychol.* 52, 1280–1300. doi: 10.1037/0003-066X.52.12.1280

Higgins, E. T. (2000). Making a good decision: value from fit. *Am. Psychol.* 55:1217. doi: 10.1037/0003-066X.55.11.1217

Higgins, E. T. (2005). Value from regulatory fit. *Curr. Direct. Psychol. Sci.* 14, 209–213. doi: 10.1111/j.0963-7214.2005.00366.x

King, D. (2009). Dlib-ml: a machine learning toolkit. *J. Mach. Learn. Res.* 10, 1755–1758. doi: 10.1145/1577069.1755843

Kristensen, T., Hannerz, H., Hgh, A., and Borg, V. (2005). The Copenhagen psychosocial questionnaire-a tool for the assessment and improvement of the psychosocial work environment. *Scand. J. Work Environ. Health* 31, 438–449. doi: 10.5271/sjweh.948

Lee, A. Y., and Aaker, J. L. (2004). Bringing the frame into focus: the influence of regulatory fit on processing fluency and persuasion. *J. Pers. Soc. Psychol.* 86:205. doi: 10.1037/0022-3514.86.2.205

Nakagawa, K., Shiomi, M., Shinozawa, K., Matsumura, R., Ishiguro, H., and Hagita, N. (2011). "Effect of robot's active touch on people's motivation," in *Proceedings of the 6th International Conference on Human-Robot Interaction (ACM)* (Lausanne), 465–472. doi: 10.1145/1957656.1957819

Quigley, M., Gerkey, B., Conley, K., Faust, J., Foote, T., Leibs, J., et al. (2009). "ROS: an open-source robot operating system," in *ICRA Workshop on Open Source Software* (Kobe), Vol. 3, 5.

Rezai, L. S., Chin, J., Bassett-Gunter, R., and Burns, C. (2017). "Investigating efficacy of regulatory fit theory in design of persuasive systems that promote physical activity," in *2nd International Workshop on Personalization in Persuasive Technology, PPT 2017* (Amsterdam), 21–25.

Tazhigaliyeva, N., Diyas, Y., Brakk, D., Aimambetov, Y., and Sandygulova, A. (2016). "Learning with or from the robot: exploring robot roles in educational context with children," in *International Conference on Social Robotics* (Kansas City, MO: Springer), 650–659. doi: 10.1007/978-3-319-47437-3_64

Conflict of Interest: 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.

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