Small Group Interactions with Voice-User Interfaces: Exploring Social Embodiment, Rapport, and Engagement

Anastasia K. Ostrowski, Vasiliki Zygouras, Hae Won Park, Cynthia Breazeal
Media Lab, Massachusetts Institute of Technology
Cambridge, MA
akostrow@media.mit.edu,vzygoura@mit.edu,haewon@media.mit.edu,cynthiab@media.mit.edu

Abstract

More and more voice-user interfaces (VUIs), such as smart speakers like Amazon Alexa or social robots like Jibo or Cozmo, are entering multi-user environments including homes. VUIs can utilize multimodal cues such as graphics, expressive sounds, and movement to convey social engagement, affecting how users perceive agents as social others. Reciprocal relationships with VUIs, i.e., relationships with give-and-take between the VUI and user, are of key interest as they are more likely to foster rapport and emotional engagement, and lead to successful collaboration. Through an elicitation study with three commercially available VUIs, we explore small group interactions (n = 33 participants) focused on the behaviors participants display to various VUIs to understand (1) reciprocal interactions between VUIs and participants and among small groups and (2) how participants engage with VUIs as the interface's embodiment becomes more socially capable. The discussion explores (1) theories of sociability applied to the users' behaviors seen with the VUIs, and (2) the group contexts where VUIs that build reciprocal relationships with users can become a powerful persuasive technology and a collaborative companion. We conclude the discussion with recommendations for promoting reciprocity from participants and, therefore, fostering rapport and emotional engagement in VUI interactions.

CCS Concepts

• **Human-centered computing → User studies**; **Natural language interfaces**.

Keywords

voice-user interface; social robot; behavioral analysis; reciprocity; small group

ACM Reference Format:
Anastasia K. Ostrowski, Vasiliki Zygouras, Hae Won Park, Cynthia Breazeal. 2021. Small Group Interactions with Voice-User Interfaces: Exploring Social Embodiment, Rapport, and Engagement. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21), March 8–11, 2021, Boulder, CO, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3434073.3444655

HRI '21, March 8–11, 2021, Boulder, CO, USA

ACM ISBN 978-1-4503-8289-2/21/03.
https://doi.org/10.1145/3434073.3444655

1 Introduction

When technology engages with people in seemingly social ways, humans respond socially to the device [63, 71]. The human-robot interaction (HRI) community has long been interested with how interactive technologies, including robots and voice-user interfaces (VUI), will be used within collaborative and multi-user spaces [12, 41]. VUIs offer a "natural" interface through which users interact with the device through their voice. Consumers are interacting with these devices, including smart speakers and robots, in a wide range of places including their homes, museums, airports, and malls [10, 26, 55, 75], often in small groups of 2–6 people. While engaging with these devices, reciprocity (the "give-and-take" in relationships) can promote positivity, emotional engagement, rapport, and higher attraction [5, 17, 81]. Understanding how users respond to various types of VUIs can inform how people will engage with these devices and how different modalities support these technology-mediated, oftentimes social, interactions.

Our work contributes to the growing research on VUIs in small group settings with an elicitation study that explores small group interactions with three commercially available VUI agents in an adapted "speed dating" format [89]. This format represents a natural situation, such as a retail store, where families are shopping for robots or voice agents while comparing them. We often already see this format with Amazon Alexa and Google Home products displayed next to each other at various stores for customers to experience and make purchasing decisions. Our speed dating approach focuses on dissecting users' reciprocal interactions with VUIs through a behavioral analysis (details of reciprocal interaction definition is provided in Section 3.5). To our knowledge, this is the first example of these methods applied to VUI agents mimicking small group interactions with multiple VUIs. We focus on reciprocal interactions, or the "give-and-take" in interactions [60], as they are critical in establishing rapport and emotional engagement in relationships [5, 17, 81]. By understanding what features enable VUI agents to build reciprocal relationships with users, we can extend voice agents as a persuasive technology that helps users achieve their goals as a collaborative partner through social influence [25]. We hypothesize that (1) people will engage in reciprocal interactions with all VUIs, but more so as a VUI's embodiment becomes more socially capable; (2) participants would exchange more reciprocal interactions with their group as the VUI's embodiment becomes more socially capable; and (3) more reciprocal interactions would correlate to stronger positive emotion, emotional engagement, and rapport with the VUI agent.

The primary contributions of this work to HRI are: (1) an understanding of the reciprocal interactions between participants and VUIs in a small group setting; (2) a discussion of human-VUI and human-human reciprocal interactions with regards to established theories of reciprocity and contexts where reciprocity is a necessary feature for future persuasive and collaborative agent design; and (3) recommendations for promoting user reciprocal interactions that lead to positive emotion, emotional engagement, and rapport during interactions with VUI agents.

2 Background

VUIs are designed with human-like characteristics including language, personality, emotion, and gender [54]. As outlined in the research paradigm, Computers Are Social Actors (CASA), individuals socially interact with computers equipped with anthropomorphic cues, behaving "towards computers as they might towards other human beings, despite knowing intuitively that computers are not animate" [12, 54, 63]. Users also respond with social behaviors to VUIs, as VUIs are in social settings displaying social behaviors and interacting among people [22, 27, 58, 62].

As VUIs are becoming more present in our social settings, it is important that we understand how these technologies are being used in varied contexts. VUIs operating in a small group setting are challenged with the complexity of multi-user interaction[66]. The same challenge also exists in HRI scenarios and only a few works have attempted formulating the problem space and proposing frameworks for multi-user group interactions [41, 57]. As technology can not only promote, but also demote human-human interactions and relationships, there is a great need for understanding what role VUIs play in various small group contexts especially as robot systems increasingly enter social contexts [39, 41, 66, 76]. In order to further understand and design for these multi-user situations with VUIs, we explore how people exhibit reciprocal cues toward agents with different embodiments and other people in the group while interacting with VUI agents.

2.1 Building Relationships with VUIs

Our work explores reciprocal interactions with VUIs and their characterization within human-agent relationships. *Impressions of people are formed within the first minutes of an interaction.* In these minutes, we decide if the person is safe to interact with and has the potential for a future relationship [1, 4, 33, 64]. Similarly, users establish impressions of and attribute human characteristics to social agent technologies such as VUIs in their first encounter, even if they only interact with the technology for a few minutes [20, 31, 64, 68]. Competence, perceived anthropomorphism, and likability of robot systems have been found to be established in the first two minutes of an interaction and maintained across multiple sessions over time [64], highlighting the importance of first impressions of social technology. This also suggests that studying first impressions offers a lens to forecasting users' long-term perception of VUIs. Here, we build upon these findings and explore what reciprocal interactions emerge in a small group setting during a first encounter with various VUIs.

Reciprocal behaviors often take the form of **verbal and non-verbal social cues** [3, 7]. The social cues used by conversational agents in agent-human interaction can be mapped to human-human interaction in terms of kinesics (visually perceivable body movement and gestures), visual, verbal, and auditory modalities [23]. Voice assistants without a physical embodiment such as Apple's Siri lack kinesic cues while robots have very visible movements and gestures. Amazon Echo, a smart speaker, is placed somewhere in between with its tower body offering physical presence but without kinesic cues. It instead uses a light ring to signal its attention toward the direction of the speaker and to indicate various states. VUIs' social cues act to provide feedback to the user of the device's active and attentive states, and they serve to create a transparent interaction between the user and device [26]. Since VUIs leverage these social cues, they are able to elicit social responses such as reciprocal behaviors from users in return. It also suggests that agents with more social modalities may be able to elicit more reciprocal behaviors [23, 63].

VUIs' **embodiment**, often leveraging social cues, can bring a socio-emotional companionship element to interacting with technology that can build rapport and emotional connection. VUI agents designed to foster emotional engagement often convey its social presence through a variety of methods including having a personality, establishing (and maintaining) social relationships, interacting using natural social cues, and behaving in a social manner (e.g., delivering greetings/farewells, being polite, expressing humor, etc.) [21, 26, 53]. Social presence has been found to be greater in physically embodied agents, such as robots, than virtual agents [29, 78, 85]. Physical embodiment also provides more natural and social cues that can be utilized to communicate intentions and internal states [21, 59]. A VUI's embodiment can affect users' engagement, response, and, therefore, reciprocal behaviors, making it a critical feature to consider in VUI agent development.

Stemming from a sociological perspective, **reciprocity** (shown through reciprocal behaviors) is defined as *"the principle of give-and-take"* [16, 60]. It can be a direct "give-and-take" or an unconscious behavioral response. Reciprocity also occurs in human-computer interaction [14, 70]. Specifically applied to robots, Krämer et al. [50] outlines a theoretical framework of sociability, including reciprocity. The micro-level is where reciprocity affects people, creating a relationship between user and robot. The meso-level is dedicated to relationship building. Users build relationships with robots as explained by the media equation theory and CASA paradigm. The last level, the macro-level, is where the role-assignment occurs, highlighting potential roles and personas for VUIs. In our work, we focus on the reciprocal behaviors between VUIs and users on the mirco- and meso-levels, as it is reported that relationships that have reciprocal give-and-take have increased positivity, intimacy, emotional engagement, rapport, and higher attraction [5, 17, 81]; a key area of interest in this research.

2.2 Linking Embodiment, Social Presence, Reciprocity, and Rapport

When interacting with a VUI, factors including embodiment, social presence, reciprocity, and rapport are interwoven together in the users' experience. Deng et al. [21] trace the connections of embodiment, social presence, and rapport based upon results from Segura et al. [77]: *"...for tasks that are relationship-oriented (e.g., a home companion), social engagement is important for maintaining rapport, and physical embodiment is beneficial for increasing social presence, and in turn, engagement and rapport."* Results from Jung & Lee [42] corroborate that physical embodiment creates higher levels of social presence, further supporting the CASA paradigm [63]. Embodiment is also connected to reciprocity. Embodiment is linked to reciprocity, rapport, and social presence. Our work acknowledges these connections and the inter-connections these factors have on one another, openly exploring how people initially engage and start building relationships with VUIs.

2.3 Robot Interactions in Small Groups

As more robots and VUIs become present in small group (2–6 people) settings, it is important to consider how various design features can influence people's behavioral responses to VUIs. Researchers have explored how individuals and small groups of people interact with robots through field studies in airports [37], sensory therapy sessions with older adults [15], hospitals [58], malls [9], cafés [67, 73], homes [27], classrooms [57], workplaces [62], etc. Specific to families, researchers have studied family members' utterances to the robot and to others in the group [79] and how non-family intergenerational groups interact with several types of robots [38]. Our research is unique in studying familial interactions with multiple different types of VUIs in one setting.

In order to design VUIs for small groups, we must understand group members' behavioral responses to VUIs, recognizing that embodiment, social presence, and context influence people's perceptions of them [12, 24, 30, 47, 61]. Sebo et al. [76] highlights that it is critical to understand the role of the robot and how it affects users in a group setting, at an individual and a group level. With this in mind, we used a family group as a unit to understand how one person's interaction with a robot in a group context influences other people in the group. Studying VUIs holistically enables us to explore all "necessary dimensions" that make up a device as one unit [12]. Therefore, we investigated small groups' reciprocal behaviors through an elicitation study with three commercially available VUIs along a spectrum of different levels of social embodiment to holistically understand how current VUIs affect small group engagement.

3 Methods

We designed our study to explore how small groups, in this case families, interact with VUIs and exhibit reciprocal behaviors. We also wanted to understand how varying levels of social embodiment in commercially available agents would affect these behaviors. We refer to *social embodiment* as *an agent (social being) inside a physical form (embodiment) that can leverage social cues, persona, and, at times, their physical form to engage with users and its environment in a social context* (adapted from [19]). In an earlier work, researchers studied how people, from children to older adults, interact with different types of VUI devices (i.e., smart speakers and social robots) over at-home deployment periods up to 4 weeks [80]. For instance, they observed that users tend to engage with and use social robots (Jibo) more than smart speakers (Amazon Echo & Google Home). However, this prior work did not provide insights on the role of VUI social embodiment and user reciprocal interactions in small groups. Our research is guided by the following research question: How do small groups exhibit reciprocity behaviors when interacting with VUIs?

3.1 Hypotheses

Based on our research question, we hypothesize that:

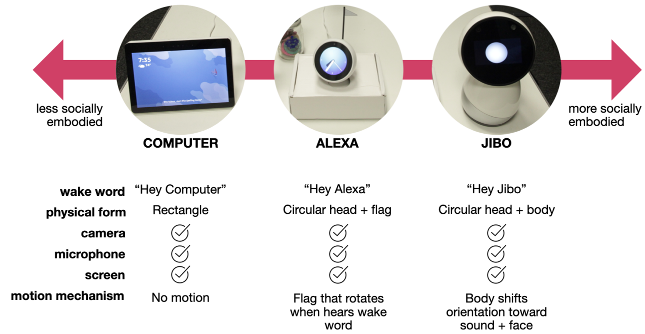

Figure 1: Spectrum of VUIs that represents three commercially available VUIs mapped on to varying levels of embodiment going from less socially embodied to more socially embodied.

- **H1:** People engage in reciprocal interactions with all VUIs, but more so as the VUI's embodiment becomes more socially capable (i.e., more socially embodied).
- **H2**: People exchange more reciprocal interactions with their group as the VUI becomes more socially embodied.
- **H3:** Higher occurrence of reciprocal interactions correlate to stronger positive emotion, emotional engagement, and rapport with the VUI agent.

3.2 VUI Agents

We used three commercial VUI devices shown in Figure 1. All have a touchscreen and a digital assistant persona (Alexa or Jibo). They each represent different levels of social embodiment and interaction, with the square smart display as the least socially embodied, the round smart display (more head-like) in the middle, and the social robot being the most socially embodied. Motion mechanism and wake word (more name-like to more device-like) were configured to also convey this range. Details of these features are discussed per agent below.

Jibo is an 11-inch tall and 6-inch wide table-top VUI with a touchscreen face and three degree-of-freedom expressive body that provides contingent motion during an interaction, such as orienting its face and body toward the user upon being called and when a face is detected in its range of view [35]. It also makes small swiveling movements as it speaks. This VUI's wake word is "Hey Jibo". The Amazon Echo Spot is a 3.8-inch tall and 4.1-inch diameter device with a round screen and the Alexa agent [2]. The Amazon Echo Spot was extended to have a motorized flag that continuously rotates above its screen and increases rotation speed when the participants spoke the activation word, "Hey Alexa", to signal attention through a less socially embodied, mechanical movement. It was designed to appear as an integrated unit with the rotating flag as an additional interaction cue. The Amazon Echo Show is a 7.4-inch tall and 7.4-inch wide device with a rectangular screen and the Alexa agent [2]. We set it's wake word to "Hey Computer", and it has no motion mechanism. All of the devices' screens are a touch-based GUI interface to support alternate ways of interacting beyond voice.

3.3 Participants

For this study, 12 families (33 participants) from a range of sociodemographic backgrounds engaged in interactions with VUI agents and reflective activities, in a study room with a couch and a coffee table. Participants were between 6 and 56 years of age (female=69.26%, age M=24.42, SD=17.70), including 17 children (female=41.18%, age M=9, SD=3.16) and 16 adults (female=37.50%, age M=39.85, SD=10.72). Of the 12 families, 7 families were one parent and one child; 3 were 1 parent and 2 children; 1 was 4 adults and 2 children; and 1 was 2 parents and 2 children. Five of the 33 participants came from lower income brackets and 28 of the 33 participants came from higher income brackets. Families were recruited through emails to the local community and word-of-mouth. All participants volunteered to participate and signed an IRB approved consent form. No incentives were offered.

We purposefully recruited small groups not owning an Amazon Alexa, Google Home, or Jibo. Most participants had not interacted with the mentioned VUIs (one adult had interacted with Google Home before the study). Eight people owned an iPhone with the Siri voice agent, and one person owned a smartphone with the Google Assistant voice agent. Others did not acknowledge having voice agents on their smartphones. Therefore, we can view this study as our participants' first encounter with embodied VUIs, unbiased by previous ownership of a VUI.

3.4 Activity Procedures

The overall study was structured as an elicitation study with the goal of understanding the reciprocal behaviors of participants when interacting with the VUIs. Elicitation studies were originally proposed as a participatory design methodology to understand users' preferences for specific interactive situations, such as gestures or symbolic input [84, 87]. We've adapted this methodology to our study to understand users' reciprocal behavior when interacting with VUIs. Since we cannot control for individual features of the commercialized products, such as appearance, size, degree of freedom, voice, or persona of the agent, we structured this study as an elicitation study to maximize the "guessability" of user interactions with VUIs (i.e., understanding how users will reciprocally interact with VUIs) [87]. Additionally, we chose an elicitation study format comparing the three agents side-by-side as elicitation studies focus on investigating first impressions and exploring people's interactions and perceptions[87, 88]. This study format was selected over users interacting with one agent at a time because we wanted to compare which agent users prefer to interact with in various candidate scenarios and also what information family members exchange while making this decision. Such setting closely mimics a real-world situation such as a family shopping for robots or voice agents in a retail store with multiple options available for purchase. Overall, the elicitation study format enables us to explore interactions with the VUIs holistically as a sum of their features, comparing the complete VUIs side-by-side without strictly comparing one feature at a time.

We designed an agent exploration activity to examine how people behave with each type of VUI agent. In the agent exploration activity, the VUI agents were placed on a coffee table in front of family members sitting on a couch. The order of the agents was randomized for each family. Participants completed an action sheet with 24 directives (see Appendix A. Agent Interaction Sheet). The 24 directives were selected so that each agent could answer. The directives also fit into three categories: information tasks (e.g. news,

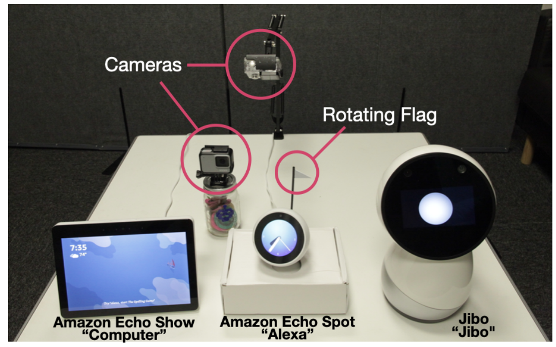

Figure 2: Study setup with commercially available at-home VUIs (left to right: Amazon Echo Show (Computer); Amazon Echo Spot (Alexa) plus a rotating flag to convey mechanical motion and attention; social robot (Jibo). Two cameras were used for recording the interactions and activities.

weather), entertainment tasks (e.g. dance, jokes, music), and interpersonal tasks (questions-and-answers about agent's personality or "thoughts"). As a group, participants completed all of the actions on the sheet. They had the ability to choose which agent to ask the question to or command it to conduct the action. After all the actions were completed, they engaged in free play with the agents, exploring within and beyond the actions that were presented in the action sheet for as long as participants wished.

3.5 Data Collection and Reciprocal Behavior Analysis

The sessions were video recorded with a front-face and table-top view (Figure 2). The front-face view recorded the interactions and responses and was used in the analyses. On average, it took participants 28.74 minutes to completed the action exploration activity and interaction with the VUIs. Overall, 1,711 interaction episodes were coded with an average of 142.58 episodes per family and an average length of 7.33 seconds. The first and second authors each coded 50% of the data individually and then reviewed the other author's coding. Any discrepancies were discussed and resolved upon agreement, following similar coding approaches as [18, 52].

As in Lee et al. [55], our coding scheme was centered on grounding and relational behaviors. Grounding behaviors included waving, acknowledgement, and relevancy (if a person builds on an agent's response). Relational behaviors included politeness, positive behaviors, and negative behaviors. Grounding and relational behaviors were identified from the videos during each interaction episode with an agent (user trigger & agent response) and the period directly proceeding the interaction as these behaviors could be most linked to a reciprocal interaction following the agent's actions. *Reciprocal behaviors* were defined as *the user's verbal and nonverbal reactions during or in-response to a VUI's actions, i.e., the behavior "given back" to the VUI or shared with another family member(s).* These behaviors could occur between a participant and the agent or between the participant and their family members. The coding scheme was developed from an initial viewing of the videos with multiple iterations to develop a finalized list of behaviors to code.

We reorganized the behaviors to be classified as either a type of verbal (e.g., speech, conversation) or nonverbal (e.g., gaze, smile,

Reciprocal Behavior	Behavior Description	Jibo mean±std	Alexa mean±std	Computer mean±std	Friedman test	Post-hoc Wilcoxon with Holm correction
Total	Total number of occurrences of 37 coded behaviors	28.79 ± 18.89	12.22 ± 7.96	11.21 ± 6.52	χ^2(2, N=1711)=19.22 p<1e-04****	Jibo vs Alexa: Z=4.75, p<.01** Jibo vs Comp: Z=4.91, p<.001***
Agent Luring Attention	Agent's behavior promotes group member to direct attention & body language to the agent	1.62 ± 1.19	0.23 ± 0.44	0.0 ± 0.0	χ^2(2, N=24)=19.95 p<1e-04****	Jibo vs Alexa: Z=2.87, p<.01** Jibo vs Comp: Z=3.06, p<.01**
Relevancy	Group member builds on an agent's response	2.1 ± 1.25	0.55 ± 0.99	0.75 ± 0.85	χ^2(2, N=68)=24.99 p<1e-04****	Jibo vs Alexa: Z=3.60, p<.001*** Jibo vs Comp: Z=3.44, p<.001***
Smiling	Group member smiles due to agent's action	4.87 ± 3.53	1.18 ± 1.38	1.29 ± 1.23	χ^2(2, N=241)=38.66 p<1e-04****	Jibo vs Alexa: Z=4.77, p<1e-04**** Jibo vs Comp: Z=4.55, p<1e-04****
Physical Imitation	Group member imitates agent's response using their body	1.25 ± 0.5	0.0 ± 0.0	0.0 ± 0.0	χ^2(2, N=5)=8.00 p<0.05*	N/A due to small number of samples
Looking Away	Group member breaks eye-contact with the agent to look elsewhere	2.74 ± 1.89	1.48 ± 1.55	1.26 ± 1.06	χ^2(2, N=148)=14.02 p<.001***	Jibo vs Alexa: Z=2.76, p<.01** Jibo vs Comp: Z=3.33, p<.001***
Laughing	Group member laughs at the agent's response	3.86 ± 3.00	1.21 ± 2.10	1.0 ± 1.09	χ^2(2, N=170)=28.72 p<1e-04****	Jibo vs Alexa: Z=3.83, p<.001*** Jibo vs Comp: Z=4.25, p<1e-04****
Complimenting	Group member gives a compliment to the agent	1.27 ± 0.79	0.09 ± 0.30	0.18 ± 0.60	χ^2(2, N=17)=11.64 p<.01**	Jibo vs Alexa: Z=2.50, p<.05* Jibo vs Comp: Z=2.04, p<.05*

Table 1: Comparing reciprocal behaviors observed around the three VUI agents that vary in the degree of social embodiment. The overall result across the combined 37 coded behaviors and the seven behaviors that showed statistically significant difference between the agents are highlighted.

lean, etc.) behavior. We then divided the verbal category as acknowledgement & relevancy, verbal with positive characteristics, and verbal with negative characteristics. The nonverbal category was divided based on what physical part of body was involved – torso, head, hand, eye, and face. The three verbal behaviors were related to grounding (including acknowledgement & relevancy) and verbal relational behaviors (both positive and negative characteristics). The five nonverbal categories focused on physical behaviors which is an additional dimension from Lee et al. [55]. In total, 37 different behaviors were coded for in the videos (see Appendix B. Behavior Code Set). Lastly, all behaviors were coded along two additional dimensions: (1) in-sync or mismatched sentiment between the VUI and user; and (2) positive or not positive user sentiment. In-sync and mismatched referred to if the participant responded to the agent as intended by the agent or not. For example, if a robot tells a joke, the expected reciprocal behavior would be a smile or laugh. Positive or not positive sentiment (i.e., neutral or negative) referred to how the participant regarded the interaction with the agent. It was interpreted as positive or not by the participant's behavior. It was revealed that positive sentiment behaviors were observed significantly more frequently than neutral or negative. Context was accounted for in the coding scheme for each behavior. For example, laughing after an agent's incorrect response was coded as not positive and mismatched. We also tracked which agent users' attentions were directed at while themselves or others were interacting with each agent.

4 Results

Our results demonstrate how small groups interact with commercial VUIs displaying reciprocal behaviors individually (H1) and with each other (H2), increasingly as VUIs become more socially embodied (H1 & H2). Results also show that there may be more reciprocal verbal, nonverbal, and in-sync interactions with VUIs that are more socially embodied, establishing positive emotion, emotional engagement, and rapport (H3).

4.1 Interaction Vignettes

Interaction episodes revealed participants engaging in reciprocal behaviors with the agents and with each other. The following vignettes provide qualitative scenario descriptions of reciprocal behaviors occurring when interacting with the VUIs. These five sample vignettes are representative of many observed interactions.

Vignette #1: Looking Away, Glancing at Each Other, Smiling, Laughing, Furrowing Brow, and Shaking Head Two children, ages 7 and 10, and their parent, age 43, are interacting with the three VUIs. The 10 year old looks at Computer and asks, "Computer, what are you thinking?". Computer responds, "I don't recommend dating light bulbs". As Computer is answering the question, both children *look away* from it. Once Computer finishes talking, the children *look at each other* and *smile*. The 7 year old *laughs*, while the 10 year old *furrows their eyebrow* and *shakes their head* at the answer, as if they can't believe the answer.

Vignette #2: Defending a VUI A child, age 7, and their parent, age 41, are sitting on a couch, listening to Jibo give them the news. Jibo ends its news sharing saying, "...and that's what's new in the news." The parent responds, "Really?" as if unsatisfied with Jibo's answer. The child *jumps to Jibo's defense* saying, "Don't listen to my mom…she's really rude to you right now.", laughing at the end of the statement.

Vignette #3: VUI Luring Attention and People Smiling and Pointing Towards VUI Two children, ages 7 and 10, and their parent, age 43, are listening to Alexa give the definition of water. As Alexa is talking, the children become disengaged as the definition is very long, more quickly than during other interactions and any other VUIs' definition of water. The 7 year old stands up attracted by Jibo moving about and *moves closer to it*, lured by its gaze. The 7 year old *smiles* and points toward Jibo, drawing the attention of the 10 year old to Jibo.

Vignette #4: Building on the VUI's Response (Relevancy) A child, age 6, and their parent, age 44, are interacting with the three agents. They ask Jibo to tell them a secret. Jibo responds, "Sure, I'll tell you a secret…I like meatballs.". Jibo ends the "secret" by playing a funny noise like the sound of a trumpet. The parent *responds back to Jibo* saying, "We knew that Jibo, tell us another secret", prompting Jibo to tell another secret to *build on the conversation.*

Vignette #5: VUI Agent Perception Qualitative responses revealed that participants distinguished between all three of the agents' embodiment. Participants reflected on Computer as it "doesn't do things", "doesn't move", "not responsive", and "not interactive enough". Reflections, both positive and negative, about Alexa included "flag annoying", "mechanical noise distracting" and Alexa "could hear better because of the flag". Participants overall saw the flag demonstrating Alexa's attentiveness and making it more "responsive". Jibo was perceived as "more social", "more interactive", and having "more engagement" than other agents.

These five vignettes contextualize the types of interactions, corresponding reciprocal behaviors, and perceptions that we saw in the process of coding the episodes. They also provide a foundation for the next sections regarding the quantitative results for the reciprocal behaviors with VUIs, interactions as small groups, and behavior types.

4.2 Reciprocal Behaviors with VUIs

Thirty-seven reciprocal behaviors were coded for in the analysis to compare small groups' interaction with the three agents. There was no significant difference between children and adults for behavior (Mann-Whitney U test; $U(1)= 30253.00$; $p>.05$), and therefore, we do not report the age group effect in the following results. When comparing across three agents, we used Friedman Chi-Squared test and post-hoc Wilcoxon tests with Holms correction. We used a non-parametric test since most of our data samples were less than 30. In Table 1, we report comparative test results for the total number of reciprocal behaviors for all coded behaviors combined, and for the individual seven behaviors that showed statistically significant difference between the three agents in the amount participants exhibited the behavior.

The behavior distributions between the agents show that participants expressed some of the reciprocal behaviors much more around the social robot Jibo compared to other agents. It's interesting to note that the only agent that was physically imitated by the participants was Jibo and that the significant behaviors include key reciprocal interactions in human communication such as building upon interactions, smiling, laughing, and complimenting. These results support **H1**.

4.3 Interactions as Small Groups

In the analysis, four codes represented group-focused behaviors: glancing at one another, having conversations among themselves that could not be heard by the agent, physically touching another person, or defending the agent against another person. Glancing at one another was when a family member removed attention from the agent and looked at another person, sharing attention with one another. These behaviors were very frequently observed with all agents, but there was a significant difference in the amount of human-human group behaviors occurred around each agent (Friedman $\chi^2(2, N{=}428){=}27.28$, $p{<}1\text{e-}04^{****}$) (Figure 3(a)). A post-hoc Wilcoxon test with Holm correction revealed that the trend was driven by Jibo (7.72 ± 6.34) versus Alexa (2.91 ± 2.12); Z=4.35, $p{<}1\text{e-}04^{****}$, and Jibo versus Computer (2.75 ± 2.91); Z=4.44, $p{<}1\text{e-}04^{****}$). It is also interesting to note that Jibo was the only agent that participants defended from other members of the group. The high-prevalence of human-human group behaviors suggests that VUIs encouraged group social interactions, and the significant difference between the agents suggest that participants engaged in human-human behaviors differently around each agent. Overall, the results

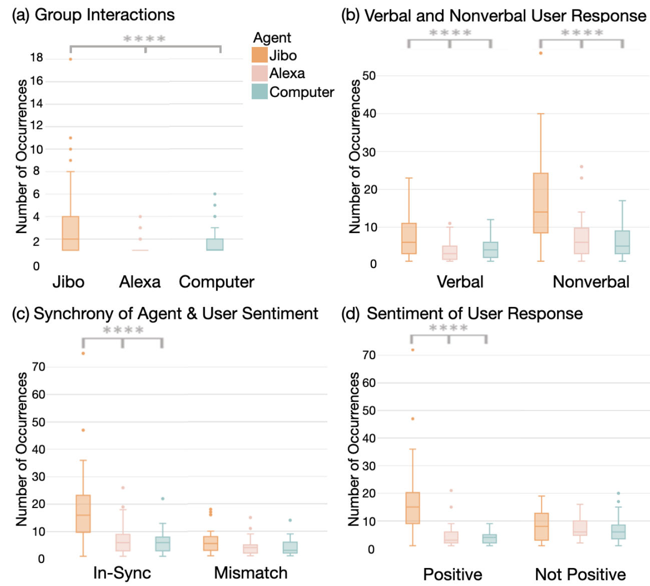

Figure 3: (a) Group Human-Human Interactions. (b)–(d) Interactions split by behavior type were classified within the context of the situation. There is a significant difference between the agents across categories, presence of verbal and nonverbal responses, synchronous sentiment between agent and users, and positive sentiment toward the agent.

of these behaviors demonstrate how interacting with the agents was (1) both an individual participant-agent interaction and a group interaction among members, and (2) small groups behaved more reciprocally with one another with a more socially embodied agent. Our observations support **H2**.

4.4 A High-Level View at the Behavior Types

Behaviors were additionally coded for (1) verbal or nonverbal; (2) in-sync or mismatched VUI-user sentiment; and (3) positive or not positive (i.e. neutral or negative) user sentiment. Across the behavioral categories, verbal (Friedman $\chi^2(2, N{=}469){=}21.76$, $p{<}1\text{e-}04^{****}$), nonverbal (Friedman $\chi^2(2, N{=}994){=}36.66$, $p{<}1\text{e-}04^{****}$), in-sync (Friedman $\chi^2(2, N{=}1048){=}41.15$, $p{<}1\text{e-}04^{****}$), and positive sentiment (Friedman $\chi^2(2, N{=}798){=}42.54$, $p{<}1\text{e-}04^{****}$) behaviors showed statistically significant difference across the agents (Figure 3(b)–(d)).

A post-hoc Wilcoxon test with Holm correction supported this trend. The positive sentiment behavior mean for Jibo (0.33 ± 0.14) was significantly different from the positive sentiment behavior mean for Alexa (0.10 ± 0.08; Z=4.83; $p{<}1\text{e-}04^{****}$) and Computer (0.09 ± 0.12; Z=4.98;$p{<}1\text{e-}04^{****}$), respectively. The verbal behavior mean for Jibo (0.15 ± 0.08) was significantly different from the verbal behavior mean for Alexa (0.08 ± 0.04; Z=4.02; $p{<}1\text{e-}04^{****}$) and Computer (0.08 ± 0.05; Z=4.02; $p{<}1\text{e-}04^{****}$), respectively. The nonverbal behavior mean for for Jibo (0.32 ± 0.12 was significantly different from the nonverbal behavior mean for Alexa (0.13 ± 0.08; Z=4.84; $p{<}1\text{e-}04^{****}$) and Computer (0.13 ± 0.11; Z=4.83; $p{<}1\text{e-}04^{****}$), respectively. The in-sync behavior mean for Jibo (0.37 ± 0.15) was significantly different from the in-sync behavior mean for Alexa

(0.13 ± 0.07; Z=4.88; p<1e-04****) and Computer 0.14 ± 0.11; Z=4.98; p<1e-04****), respectively. It is surprising that Alexa and Computer were not significantly different as Alexa was equipped with a mechanical movement in the form of a rotating flag. The variable speed rotating flag was assumed to communicate attentiveness and attract interest as an additional interaction cue and users perceiving it as more socially embodied than Computer. We discuss this observation in more detail in Section 5.2. In summary, we hypothesized more reciprocal verbal, nonverbal, and in-sync interactions with VUIs that have higher level of social embodiment. Verbal, nonverbal, in-sync, and positive sentiment were proven significantly different through post-hoc analysis, supporting **H3**.

5 Discussion

Our work furthers HRI research in how users, especially in small groups, interact with VUIs in multi-user environments using an elicitation approach that focuses on the VUI as a holistic unit across a spectrum of social embodiment. In our discussion below, we reflect on how our findings of small groups interacting with VUI agents build upon theoretical frameworks of sociability including reciprocity, contexts where reciprocity behaviors are especially necessary with regards to future persuasive and collaborative agent design, and recommendations for promoting reciprocity in agent interactions.

5.1 Frameworks of Sociability & Reciprocity

Relationships that have reciprocity are more likely to be positive, foster intimacy and emotional engagement, and build rapport [5, 17, 81]. Our findings also show that our participants displayed positive sentiment, nonverbal, verbal, and in-sync synchrony interactions with all three VUIs. Between VUIs, the difference was significant between Jibo & Alexa and Jibo & Computer, with Jibo eliciting higher occurrences of reciprocal behaviors from the users. This result suggests that the social robot was more effective at engaging in relationship building with users, which further links to fostering emotional engagement, rapport, and trust [5, 17, 81]. This significant trend was also observed in small group dynamics, in which the group members exchanged more reciprocal interactions between each other in response to Jibo than to other agents. Through the observed reciprocal interaction exchange between VUIs and users and between users, we can confirm that the three VUIs offered varying levels of relationship creation and building, the micro- and meso-levels of Krämer et al.'s sociability framework [50].

5.1.1 The Micro-Level The micro-level is the prerequisite for interaction between a user and a VUI, crucial for creating reciprocity [50]. In the case of the VUIs and small groups, participants explored the agent to understand what the agent "knows" and corresponded to the agent's actions using verbal and nonverbal reciprocal behaviors, such as physical imitation, glancing at group members, and complimenting the agents on their abilities. These behaviors are a demonstration of the Interaction Adaptation Theory [11]. Successful interactions reinforce an understanding of the system and can further promote positive interactions between users and the VUIs. Interactions that are unsuccessful can lead to behaviors expressing confusion or a lower amount of reciprocal behaviors [50]. This may have been the case with Computer as there were less reciprocal behaviors displayed by the users.

5.1.2 The Meso-Level The meso-level is the foundation for relationship building [50] and is demonstrated in how people engage with technologies [63]. People will be more attracted to and engaged in relationship building activities if the other person (or VUI) behaves as if they like interacting with them, both verbally and nonverbally (i.e. reciprocal liking) [6, 45, 46, 51]. In our work, the reciprocal behaviors can translate to reciprocal liking: if the participants perceived that the agent likes interacting with them, then they would be more likely to engage in reciprocal behaviors. The amount of reciprocal behaviors we observed in our study suggests that users can have satisfactory social exchanges with the VUIs, whether it is from gaining information from the agents, feeling joyful and entertained, or fulfilling curiosity toward a new technology. With VUIs, such as the social robot, participants could have felt a greater level of social exchange with it, given the fact that a significant proportion of reciprocal interactions happened around Jibo.

5.2 Mechanisms of Sociability

Embodiment, social presence, and social cues are common factors that are explored to understand VUI engagement and social interaction, and to design for users' desires for relationship building, feelings of trust, companionship, and rapport [23, 74, 82]. These factors are often displayed via nonverbal social cues that can be translated to VUIs and that contribute to social dialogue by expressing emotions, supporting discussion, and communicating attitudes [3, 7]. Reciprocal liking can be conveyed through nonverbal behaviors such as leaning toward the speaker, increased and direct gaze, smiling, nodding, and direct body and facial orientation [3, 72]. These behaviors were all seen in participants interacting with the three VUIs. Two of the VUIs had a combination of movement (Jibo had socially mediated movement and Alexa had a mechanical flag) and attention indicating features (Jibo had gaze and movement and Alexa had a mechanical flag) that may have promoted reciprocal behaviors from the participants. Computer did not have any movement that would have likely influenced participants' decreased amount of interest and reciprocal behaviors toward the agent.

Different social cues work together to create a social presence for VUIs [21, 26, 53], often strengthened by the VUI's embodiment [29, 78]. Increased feelings of social presence can promote more favorable behaviors toward VUIs as social actors [54] and provoke more responsive reciprocal behaviors from the users. Likewise, in our study, a higher amount of reciprocal behavior was shown toward the agent and between the family members when they were interacting with Jibo, supporting theories around social exchange and mutual knowledge.

We hypothesized that Alexa's embodiment through the flag's mechanical movement signaling attentiveness would also promote reciprocal behaviors with individual users and between family members to a lesser extent than Jibo but greater than Computer, but the result showed that Alexa and Computer were not significantly different from one another. This result may suggest that a repetitive, less socially embodied movement is no more beneficial than a light ring indicator that both Alexa and Computer possessed.

5.3 VUI Reciprocity Applied to Social Contexts

Small group, multi-user scenarios have been of great interest to the HRI community [41]. In this study, we focused the small group interaction topic around families. Studying and designing reciprocal agents can open doors to make VUIs a powerful persuasive technology and collaborative partners in helping users achieve their goals [25]. Building rapport and reciprocal relationships with technology, often with social agents, is known to boost users' sustained engagement and adherence to therapy, education, healthy habits, and wellness [8, 28, 32, 34, 48, 86]. Reciprocal relationships will expand the limited capability of the functionalities VUIs can provide today and make them become supportive partners, motivating allies, and collaborative companions [34, 58, 65, 69, 83]. VUIs can enter these roles in social group contexts including, but not limited to, education [43, 49], the work force [13, 40], and therapeutic situations [60, 82]. These application areas rely on collaborative teamwork in potentially multi-user environments where social support matters, further highlighting the importance of emphasizing reciprocal behaviors. Depending on the context and the relationships between group members, reciprocal behaviors may vary. Increased knowledge on contextual reciprocal behaviors in small group spaces can aid in understanding the critical design features of VUIs to promote trust [56], a key feature needed for group engagement [36].

5.4 Promoting Reciprocity in VUI Interactions

The findings presented in this paper suggest that VUIs' social embodiment and socially-expressive nature can positively impact reciprocity behaviors. This goes far beyond simple physical movements in embodiment that can attract attention and signal activity such as the mechanical flag we designed on Echo Spot. Participants demonstrated increased positive sentiment, verbal, nonverbal, and in-sync reciprocal behaviors with a VUI that had richer social cues versus VUIs that have mechanical or no physical cues.

Designers can incorporate expressive social embodiment through socially contingent movements and verbal and nonverbal social cues including orienting gaze and body toward the person, sharing gaze, and conveying emotions to: (1) promote reciprocal behaviors in the human-robot interaction; (2) promote reciprocal behaviors between multiple users in contexts such as families but also collaborative teamwork where support (e.g. coaching, tutoring) and joint action (e.g. building together) are key; (3) create more positive (possibly empathetic) reciprocal behaviors such as smiling and laughing to foster reciprocal likability; and (4) purposefully draw attention to the agent, promoting user behaviors such as luring. Our results demonstrate that these design implications can be included in VUIs for both adults and children. Social cues that invoke reciprocal behaviors in users can be used as a method to strengthen rapport and companionship with VUI devices leveraging the relationship and emotional attraction users develop with the technology. The methodologies in this paper provide designers with a new way of investigating how users perceive and interact with VUI agents in a holistic manner across a spectrum of various design features (as was done with social embodiment in this study) and different contexts (i.e. group vs. individual, large age range).

5.5 Considerations & Future Research Directions

In addition to understanding how reciprocal behaviors with VUIs intersect with theories of sociability and CASA, it is important (1) to understand how people are interacting with these systems in their first minutes of interaction and developing the lens they perceive the VUIs through and (2) to treat VUIs as holistic systems, i.e., as a compound of all features and not just a sum of each, to explore how users interact with them [12]. Lasting impressions around likability and overall feelings toward the VUIs most likely occur in the first few minutes of interacting with them [64]. The tasks in the study were selected from the categories that all three VUIs can fulfill, and it did not include tasks that only some agents could do, such as calendar syncing, recipe retrieval, audio books, etc. If participants found out that one agent was superior in the functionalities it can offer, it could have affected the impression of the agents. Since this was not the scope of the study, we controlled for this factor intentionally, but future studies should explore reciprocal behaviors with VUIs with varying levels of utility. Participants' behaviors influenced and were influenced by others in the group. We did not account for such group effect in our the statistical analysis, but a mixed-effects model could be used in future work. As this study was conducted in the United States, we should not assume that the results from this study would apply to those in other cultures [44]. This study was structured as an in-lab study, so an extended study could be conducted in real-world settings outside of the lab.

Our results indicate that social and expressive cues (verbal and nonverbal) may impact how people respond reciprocally to VUIs. Further studies can investigate how specific social-emotional design features encourage interaction, explore which tasks these design features are best used for VUI interaction, and empirically measure trust and companionship. Another important area to understand is how social embodiment impacts long-term interaction, how they can be leveraged to create socially persuasive technologies, and the impact these could have on our interactions with technology and our social settings. Future work should also replicate and reproduce this study with other social agent embodiments to further understand the impact of social embodiment.

6 Conclusion

In summary, small groups interacted with three commercial VUIs in an elicitation study to explore how people reciprocally behave with VUIs, investigating VUIs as a holistic unit that can be compared on a spectrum of social embodiment. Our findings indicate that participants individually and with other group members reciprocally engage more with VUIs that have a higher level of social embodiment. They also express greater positive emotion, emotional engagement, and rapport with those VUIs. Our discussion centered these results in the context of sociability frameworks and small group settings where building reciprocal relationships with VUIs hold great promise and potential. As a conclusion, we provide recommendations for promoting reciprocal interactions from users through socially embodied technology design. Lastly, we discuss the limitations of our work and suggest future work that investigates how specific design features encourage reciprocal behaviors and their effects on small groups in longer term, expanding from our work.

7 Acknowledgments

We are thankful to Jenny Fu for helping us conduct the study and collecting data.

References

[1] Linda Albright, David A Kenny, and Thomas E Malloy. 1988. Consensus in personality judgments at zero acquaintance. *Journal of personality and social psychology* 55, 3 (1988), 387.

[2] Amazon. 2018. Amazon Echo. https://www.amazon.com

[3] Michael Argyle. 2013. *Bodily communication*. Routledge.

[4] Moshe Bar, Maital Neta, and Heather Linz. 2006. Very first impressions. *Emotion* 6, 2 (2006), 269.

[5] Barbara Becker and Gloria Mark. 1999. Constructing social systems through computer-mediated communication. *Virtual reality* 4, 1 (1999), 60–73.

[6] Ellen Berscheid and Elaine Hatfield. 1969. *Interpersonal attraction*. Vol. 69. Addison-Wesley Reading, MA.

[7] Timothy Bickmore and Justine Cassell. 2005. Social dialongue with embodied conversational agents. In *Advances in natural multimodal dialogue systems*. Springer, 23–54.

[8] Timothy W Bickmore, Laura M Pfeifer, and Brian W Jack. 2009. Taking the time to care: empowering low health literacy hospital patients with virtual nurse agents. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1265–1274.

[9] Dražen Bršćić, Hiroyuki Kidokoro, Yoshitaka Suehiro, and Takayuki Kanda. 2015. Escaping from children's abuse of social robots. In *Proceedings of the tenth annual acm/ieee international conference on human-robot interaction*. 59–66.

[10] Wolfram Burgard, Armin B Cremers, Dieter Fox, Dirk Hähnel, Gerhard Lakemeyer, Dirk Schulz, Walter Steiner, and Sebastian Thrun. 1999. Experiences with an interactive museum tour-guide robot. *Artificial intelligence* 114, 1-2 (1999), 3–55.

[11] Judee K Burgoon, CA White, and JO Greene. 1997. Researching nonverbal message production: A view from interaction adaptation theory. *Message production: Advances in communication theory* (1997), 279–312.

[12] Julia Cambre and Chinmay Kulkarni. 2019. One Voice Fits All? Social Implications and Research Challenges of Designing Voices for Smart Devices. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–19.

[13] Jennifer Casper and Robin R. Murphy. 2003. Human-robot interactions during the robot-assisted urban search and rescue response at the world trade center. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 33, 3 (2003), 367–385.

[14] Karin Knorr Cetina. 1997. Sociality with objects: Social relations in postsocial knowledge societies. *Theory, culture & society* 14, 4 (1997), 1–30.

[15] Wan-Ling Chang, Selma Šabanovic, and Lesa Huber. 2013. Use of seal-like robot PARO in sensory group therapy for older adults with dementia. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 101–102.

[16] Robert B Cialdini. 2009. *Influence: Science and practice*. Vol. 4. Pearson education Boston, MA.

[17] Nancy L Collins and Lynn Carol Miller. 1994. Self-disclosure and liking: a meta-analytic review. *Psychological bulletin* 116, 3 (1994), 457.

[18] Shanna R Daly, Seda Yilmaz, James L Christian, Colleen M Seifert, and Richard Gonzalez. 2012. Design heuristics in engineering concept generation. (2012).

[19] Kerstin Dautenhahn. 1998. Embodiment and interaction in socially intelligent life-like agents. In *International Workshop on Computation for Metaphors, Analogy, and Agents*. Springer, 102–141.

[20] Kerstin Dautenhahn. 2004. Robots we like to live with?!-a developmental perspective on a personalized, life-long robot companion. In *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No. 04TH8759)*. IEEE, 17–22.

[21] Eric Deng, Bilge Mutlu, and Maja J Mataric. 2019. Embodiment in Socially Interactive Robots. *Foundations and Trends® in Robotics* 7, 4 (2019), 251–356. https://doi.org/10.1561/2300000056

[22] Jeanne Dietsch. 2010. People meeting robots in the workplace [industrial activities]. *IEEE Robotics & Automation Magazine* 17, 2 (2010), 15–16.

[23] J Feine, U Gnewuch, S Morana, and A Maedche. 2019. A Taxonomy of Social Cues for Conversational Agents. *International Journal of Human-Computer Studies* (July 2019). https://doi.org/10.1016/j.ijhcs.2019.07.009

[24] Kerstin Fischer, Katrin Lohan, and Kilian Foth. 2012. Levels of embodiment: Linguistic analyses of factors influencing HRI. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 463–470.

[25] Brian J Fogg. 2002. Persuasive technology: using computers to change what we think and do. *Ubiquity* 2002, December (2002), 2.

[26] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. 2003. A survey of socially interactive robots. *Robotics and Autonomous Systems* 42, 3 (2003), 143–166. https://doi.org/10.1016/S0921-8890(02)00372-X

[27] Jodi Forlizzi and Carl DiSalvo. 2006. Service robots in the domestic environment: a study of the roomba vacuum in the home. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. 258–265.

[28] Jonathan Gratch, Ning Wang, Jillian Gerten, Edward Fast, and Robin Duffy. 2007. Creating rapport with virtual agents. In *International workshop on intelligent virtual agents*. Springer, 125–138.

[29] Marcel Heerink, Ben Kröse, Vanessa Evers, and Bob Wielinga. 2010. Assessing acceptance of assistive social agent technology by older adults: the Almere model. *International Journal of Social Robotics* 2 (2010), 361–375. https://doi.org/10.1007/s12369-010-0068-5

[30] Laura Hoffmann, Nikolai Bock, and Astrid M Rosenthal vd Pütten. 2018. The Peculiarities of Robot Embodiment (EmCorp-Scale) Development, Validation and Initial Test of the Embodiment and Corporeality of Artificial Agents Scale. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 370–378.

[31] Joachim R Höflich and Afifa El Bayed. 2015. Perception, acceptance, and the social construction of robots—exploratory studies. In *Social robots from a human perspective*. Springer, 39–51.

[32] Lixing Huang, Louis-Philippe Morency, and Jonathan Gratch. 2011. Virtual Rapport 2.0. In *International workshop on intelligent virtual agents*. Springer, 68–79.

[33] Lauren J Human, Gillian M Sandstrom, Jeremy C Biesanz, and Elizabeth W Dunn. 2013. Accurate first impressions leave a lasting impression: The long-term effects of distinctive self-other agreement on relationship development. *Social Psychological and Personality Science* 4, 4 (2013), 395–402.

[34] Sooyeon Jeong, Deirdre E Logan, Matthew S Goodwin, Suzanne Graca, Brianna O'Connell, Honey Goodenough, Laurel Anderson, Nicole Stenquist, Katie Fitzpatrick, Miriam Zisook, et al. 2015. A social robot to mitigate stress, anxiety, and pain in hospital pediatric care. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*. 103–104.

[35] Jibo. 2017. Jibo.

[36] Gareth R Jones and Jennifer M George. 1998. The experience and evolution of trust: Implications for cooperation and teamwork. *Academy of management review* 23, 3 (1998), 531–546.

[37] Michiel Joosse and Vanessa Evers. 2017. A guide robot at the airport: First impressions. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. 149–150.

[38] Swapna Joshi and Selma Šabanović. 2019. Robots for inter-generational interactions: implications for nonfamilial community settings. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 478–486.

[39] Malte Jung and Pamela Hinds. 2018. Robots in the wild: A time for more robust theories of human-robot interaction.

[40] Malte F Jung, Jin Joo Lee, Nick DePalma, Sigurdur O Adalgeirsson, Pamela J Hinds, and Cynthia Breazeal. 2013. Engaging robots: easing complex human-robot teamwork using backchanneling. In *Proceedings of the 2013 conference on Computer supported cooperative work*. 1555–1566.

[41] Malte F Jung, Selma Šabanović, Friederike Eyssel, and Marlena Fraune. 2017. Robots in groups and teams. In *Companion of the 2017 ACM conference on computer supported cooperative work and social computing*. 401–407.

[42] Younbo Jung and Kwan Min Lee. 2004. Effects of physical embodiment on social presence of social robots. *Proceedings of PRESENCE* (2004), 80–87.

[43] Elena Karpova, Ana-Paula Correia, and Evrim Baran. 2009. Learn to use and use to learn: Technology in virtual collaboration experience. *The Internet and Higher Education* 12, 1 (2009), 45–52.

[44] Yasuhiro Katagiri, Clifford Nass, and Yugo Takeuchi. 2001. Cross-cultural studies of the computers are social actors paradigm: The case of reciprocity. *Usability evaluation and interface design: Cognitive engineering, intelligent agents, and virtual reality* (2001), 1558–1562.

[45] David A Kenny. 2013. Using the social relations model to understand relationships. In *Theoretical frameworks for personal relationships*. Psychology Press, 123–140.

[46] David A Kenny and Lawrence La Voie. 1982. Reciprocity of interpersonal attraction: A confirmed hypothesis. *Social Psychology Quarterly* (1982), 54–58.

[47] Sara Kiesler, Aaron Powers, Susan R Fussell, and Cristen Torrey. 2008. Anthropomorphic interactions with a robot and robot–like agent. *Social Cognition* 26, 2 (2008), 169–181.

[48] Jacqueline M Kory-Westlund and Cynthia Breazeal. 2019. Assessing children's perceptions and acceptance of a social robot. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*. 38–50.

[49] Akihiro Kosugi, Masatomo Kobayashi, and Kentarou Fukuda. 2016. Hands-Free collaboration using telepresence robots for all ages. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*. 313–316.

[50] Nicole C Krämer, Sabrina Eimler, Astrid Von Der Pütten, and Sabine Payr. 2011. Theory of companions: what can theoretical models contribute to applications and understanding of human-robot interaction? *Applied Artificial Intelligence* 25, 6 (2011), 474–502.

[51] Warren N Kubitschek and Maureen T Hallinan. 1998. Tracking and students' friendships. *Social Psychology Quarterly* (1998), 1–15.

[52] Hee Rin Lee, Haodan Tan, and Selma Šabanović. 2016. That robot is not for me: Addressing stereotypes of aging in assistive robot design. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 312–317.

[53] Kwan Min Lee, Younbo Jung, Jaywoo Kim, and Sang Ryong Kim. 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. *International Journal of Human-Computer Studies* 64, 10 (2006), 962 – 973. https://doi.org/10.1016/j.ijhcs.2006.05.002

[54] Kwan Min Lee, Wei Peng, Seung-A Jin, and Chang Yan. 2006. Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human–robot interaction. *Journal of communication* 56, 4 (2006), 754–772.

[55] Min Kyung Lee, Sara Kiesler, and Jodi Forlizzi. 2010. Receptionist or information kiosk: how do people talk with a robot?. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. 31–40.

[56] SeoYoung Lee and Junho Choi. 2017. Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. *International Journal of Human-Computer Studies* 103 (2017), 95–105.

[57] Iola Leite, Marissa McCoy, Monika Lohani, Daniel Ullman, Nicole Salomons, Charlene Stokes, Susan Rivers, and Brian Scassellati. 2015. Emotional storytelling in the classroom: Individual versus group interaction between children and robots. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 75–82.

[58] Sara Ljungblad, Jirina Kotrbova, Mattias Jacobsson, Henriette Cramer, and Karol Niechwiadowicz. 2012. Hospital robot at work: something alien or an intelligent colleague?. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*. 177–186.

[59] KS Lohan, S Gieselmann, A-L Vollmer, K Rohlfing, and B Wrede. 2010. Does embodiment affect tutoring behavior?. In *International Conference on Development and Learning, 2010*.

[60] Tamara Lorenz, Astrid Weiss, and Sandra Hirche. 2016. Synchrony and reciprocity: Key mechanisms for social companion robots in therapy and care. *International Journal of Social Robotics* 8, 1 (2016), 125–143.

[61] Michal Luria, Samantha Reig, Xiang Zhi Tan, Aaron Steinfeld, Jodi Forlizzi, and John Zimmerman. 2019. Re-Embodiment and Co-Embodiment: Exploration of social presence for robots and conversational agents. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 633–644.

[62] Bilge Mutlu and Jodi Forlizzi. 2008. Robots in organizations: the role of workflow, social, and environmental factors in human-robot interaction. In *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 287–294.

[63] Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. 1994. Computers Are Social Actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, Massachusetts, USA) *(CHI '94)*. ACM, New York, NY, USA, 72–78. https://doi.org/10.1145/191666.191703

[64] Maike Paetzel, Giulia Perugia, and Ginevra Castellano. 2020. The Persistence of First Impressions: The Effect of Repeated Interactions on the Perception of a Social Robot. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. 73–82.

[65] Hae Won Park, Rinat Rosenberg-Kima, Maor Rosenberg, Goren Gordon, and Cynthia Breazeal. 2017. Growing growth mindset with a social robot peer. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. 137–145.

[66] Martin Porcheron, Joel E Fischer, Moira McGregor, Barry Brown, Ewa Luger, Heloisa Candello, and Kenton O'Hara. 2017. Talking with conversational agents in collaborative action. In *companion of the 2017 ACM conference on computer supported cooperative work and social computing*. 431–436.

[67] Martin Porcheron, Joel E Fischer, and Sarah Sharples. 2017. " Do Animals Have Accents?" Talking with Agents in Multi-Party Conversation. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 207–219.

[68] Aaron Powers and Sara Kiesler. 2006. The advisor robot: tracing people's mental model from a robot's physical attributes. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. 218–225.

[69] Alisha Pradhan, Leah Findlater, and Amanda Lazar. 2019. " Phantom Friend" or" Just a Box with Information" Personification and Ontological Categorization of Smart Speaker-based Voice Assistants by Older Adults. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–21.

[70] Byron Reeves and Clifford Ivar Nass. 1996. *The media equation: How people treat computers, television, and new media like real people and places.* Cambridge university press.

[71] C.I. Reeves, B., Nass. 1996. *The media equation: How people treat computers, television, and new media like real people and places.* Center for the Study of Language and Information; Cambridge University Press, Chicago, IL, US; New York, NY.

[72] Virginia P Richmond, James C McCroskey, and Mark Hickson. 2008. *Nonverbal behavior in interpersonal relations.* Allyn & Bacon.

[73] Astrid M Rosenthal-von der Pütten, Nicole C Krämer, Christian Becker-Asano, Kohei Ogawa, Shuichi Nishio, and Hiroshi Ishiguro. 2014. The uncanny in the wild. Analysis of unscripted human–android interaction in the field. *International Journal of Social Robotics* 6, 1 (2014), 67–83.

[74] Alan M Rubin and Elizabeth M Perse. 1987. Audience activity and television news gratifications. *Communication research* 14, 1 (1987), 58–84.

[75] Selma Sabanovic, Marek P Michalowski, and Reid Simmons. 2006. Robots in the wild: Observing human-robot social interaction outside the lab. In *9th IEEE International Workshop on Advanced Motion Control, 2006*. IEEE, 596–601.

[76] Sarah Sebo, Brett Stoll, Brian Scassellati, and Malte F Jung. 2020. Robots in Groups and Teams: A Literature Review. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–36.

[77] Elena Márquez Segura, Michael Kriegel, Ruth Aylett, Amol Deshmukh, and Henriette Cramer. 2012. How do you like me in this: User embodiment preferences for companion agents. In *International Conference on Intelligent Virtual Agents*. Springer, 112–125.

[78] Kazuhiko Shinozawa, Byron Reeves Professor, Kevin Wise, Doctoral C, So hye Lim, Doctoral C, Heidy Maldonado, Doctoral Student, and Futoshi Naya. 2002. Robots as New Media: A Cross-Cultural Examination of Social and Cognitive Responses to Robotic and On-Screen Agents. (2002).

[79] Elaine Schaertl Short, Katelyn Swift-Spong, Hyunju Shim, Kristi M Wisniewski, Deanah Kim Zak, Shinyi Wu, Elizabeth Zelinski, and Maja J Matarić. 2017. Understanding social interactions with socially assistive robotics in intergenerational family groups. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 236–241.

[80] Nikhita Singh. 2018. *Talking machines: democratizing the design of voice-based agents for the home.* Master's thesis. Massachusetts Institute of Technology.

[81] Susan Sprecher, Stanislav Treger, Joshua D Wondra, Nicole Hilaire, and Kevin Wallpe. 2013. Taking turns: Reciprocal self-disclosure promotes liking in initial interactions. *Journal of Experimental Social Psychology* 49, 5 (2013), 860–866.

[82] Andrea Tartaro and Justine Cassell. 2007. Using virtual peer technology as an intervention for children with autism. *Towards universal usability: designing computer interfaces for diverse user populations* 231 (2007), 62.

[83] Angelique Taylor, Hee Rin Lee, Alyssa Kubota, and Laurel D Riek. 2019. Coordinating clinical teams: Using robots to empower nurses to stop the line. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–30.

[84] Radu-Daniel Vatavu and Jacob O Wobbrock. 2016. Between-subjects elicitation studies: Formalization and tool support. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 3390–3402.

[85] J. Wainer, D. J. Feil-seifer, D. A. Shell, and M. J. Mataric. 2006. The role of physical embodiment in human-robot interaction. In *ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*. 117–122. https://doi.org/10.1109/ROMAN.2006.314404

[86] Jacqueline M Kory Westlund, Hae Won Park, Randi Williams, and Cynthia Breazeal. 2018. Measuring young children's long-term relationships with social robots. In *Proceedings of the 17th ACM conference on interaction design and children*. 207–218.

[87] Jacob O Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A Myers. 2005. Maximizing the guessability of symbolic input. In *CHI'05 extended abstracts on Human Factors in Computing Systems*. 1869–1872.

[88] Jacob O Wobbrock, Meredith Ringel Morris, and Andrew D Wilson. 2009. User-defined gestures for surface computing. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1083–1092.

[89] John Zimmerman and Jodi Forlizzi. 2017. Speed dating: providing a menu of possible futures. *She Ji: The Journal of Design, Economics, and Innovation* 3, 1 (2017), 30–50.