Assessing Children's Perceptions and Acceptance of a Social Robot

Jacqueline M. Kory-Westlund
Cynthia Breazeal*
jakory@media.mit.edu
cynthiab@media.mit.edu
MIT Media Lab
Cambridge, MA

ABSTRACT

Children's interactions with social robots and other technologies are increasingly longitudinal, especially in areas such as healthcare, therapy, and education. As such, we need to understand how children perceive social robots over time and the kinds of relationships they develop. Relatively few validated assessments exist that measure young children's relationships or their perception and acceptance of social robots. Thus, we present pilot tests of two assessments created for use with children aged 4–7: the Picture Sorting Task and the Social Acceptance Questionnaire. Through a single-session study and also a long-term study, we found that children responded appropriately to the assessments and that the assessments could capture changes in children's perception and relationship over multiple encounters.

CCS CONCEPTS

• **Social and professional topics** → **Children**; • **Applied computing** → **Psychology**; • **Computer systems organization** → **Robotics**; • **Human-centered computing** → *User studies.*

KEYWORDS

assessments; children; long-term interaction; human-robot interaction; relationships; social robots

1 INTRODUCTION

Social robots and related technologies are increasingly used with children in longitudinal contexts, such as education, healthcare, and therapy [6, 23, 40, 43, 49, 59, 60]. Because the goals of child-robot interactions in these areas—namely learning, behavior change, and improving health—necessarily take time, the interactions children have are thus longer-term, with repeated encounters over weeks or months. As a result, we need to understand how children perceive social robots over time and how their relationships with the robots change and develop.

Prior research has shown that children treat robots as social agents that are not quite like people, and are also more than mere machines. Children apply social judgments to robots and respond to their social cues in ways similar to how the children respond to people [10, 34]. Children ascribe psychological and perceptual properties to robots, and also properties of artifacts and toys [3, 29, 32, 35, 36, 42]. Children often think of and speak about robots and other autonomous technologies in social ways, and compare them to both people and toys [13, 14]. They have shown moral objections to the object-like treatment of robots, such as putting robots away in a closet [29], but may also say that, like other objects, people made the robot, people can own robots, and robots can break [36].

Together, this research suggests that children think of robots as betwixt and between the dualistic categories of alive, animate beings and inanimate objects. However, many of these studies examined single encounters that children had with robots. An open question is how much of children's and perception of the robot might be due to novelty or inexperience. Might children revise their opinions after more time interacting with real robots over multiple encounters?

Some recent research has begun examining children's long-term interactions with social robots, especially in education [1, 25, 30, 33, 40, 49, 56, 57]. In these interactions, robots are often situated as *relational agents*, i.e., as agents that actively try to build and maintain long-term, social-emotional relationships [9]. However, most of the work so far has focused on the efficacy of the interaction on children's learning,

engagement, or on other performance metrics, rather than attempting to measure children's relationships with the robot or perception of the robot as a social agent over time. This is partly because there are relatively few validated assessments for measuring young children's relationships or their perception and acceptance of social robots as relational agents. Many of the existing assessments target older children and adults, often assuming reading, comphrension, and attention capabilities above that of most 4–6-year-olds [2, 5, 22, 45]. The one exception is a recent paper presenting four new assessments for measuring the relationships of children aged 5–6 years with social robots [37], which focused on children's closeness to the robot, intimacy via self-disclosure, perception of the robot as a social-relational agent, and how children talk about the robot versus their human best friend.

In this paper, we present pilot tests of two new assessments for use with children aged 4–7 years: the Picture Sorting Task (PST) and the Social Acceptance Questionnaire (SAQ). These two assessments focus on children's acceptance of the robot and their perception of the robot's animacy and human-likeness, which are important qualities that may reflect how likely the child is to treat the robot as a social, relational agent. We tested both assessments in a single-session study and an eight-session field study. We report two primary contributions. First, we discuss the reliability and appropriateness of each assessment, which we use to show that the assessments are valid. Second, we discuss key results of the studies that show the type of information that can be gained from these assessments, which can be used as an point of comparison for later work regarding children's perceptions of robots.

2 BACKGROUND
Children's Perceptions of Robots

Prior research has found that a robot's social capabilities—such as contingent behavior and appropriate gaze, facial expressions, affect and emotion, gesture, and posture—can increase the likelihood that children will treat robots as social agents, similar to how they interact with people [10, 28, 31, 34, 35, 39, 48]. During long-term interactions with robots, children display social behaviors associated with developing friendships and close relationships; they report feeling close to robots and think of the robot as social and relational agents [37].

Children's perceptions of robots appear to be shaped not only by the robot's own behavior, but also by children's experiences in the world and the opinions of those close to them. For example, children's attributions of intelligence to a mouse or a robot was similar to their parents'; they mirrored their parents' mental models [14]. Stories told to children by an adult about a robot's social capabilities affected their social judgments of the robot [36].

In studying children's long-term interactions with robots, we will need to understand not only the relationships children are developing, but also how children construe the robot. Children's perceptions of the robot's socialness, animacy, and human-likeness may be an important factors in how they ultimately relate to the robot and whether they treat it as a relational agent. Thus, the two assessments we present in this paper focus on measuring children's perception of the robot's animacy and human-likeness.

Assessing Perceptions of Robots

Prior work on long-term child-robot interaction has used various assessments to measure children's engagement and the robot's social presence, often using multiple-choice questionnaires or examining children's gaze, affect, and speech patterns, as well as several newer assessments that explicitly examined children's relationship with the robot through picture-based questions, disclosure, and interviews [37, 41]. With regards to children's perception of the robot's animacy and human-likeness, prior work has primarily used questionnaires or interviews [24, 36, 47]; a few studies have also examined children's behavior, e.g., with a robotic dog versus with a stuffed dog [27]. In human-human studies, younger children's relationships have primarily been assessed through interviews or observational methods, such as coding children's behavior for different aspects of relationships such as disclosure, exclusivity, connectedness, conflict, and proximity [18, 20, 38, 52, 61].

With older children or adults, numerous studies have used the Godspeed questionnaire to measure people's perception of a robot's anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety [4, 5]. There are several other similar questionnaires that measure human-likeness, eeriness, or people's attitudes toward robots [21, 46]. However, assessments that work well for older children and adults, such as written questionnaires, may not work as well for younger children who may be pre-reading, have shorter attention spans, or may be unable to complete standard Likert-style questionnaires [12, 37]. In HCI, picture-based scales and activities have been successfully used with young children to measure, e.g., engagement, fun, and self-concepts of competence and acceptance [15, 19, 50, 51], but have not been used to assess children's perceptions of animacy or their acceptance of robots. Thus, our new assessments both use pictures. The first uses a picture-based activity to probe children's construal of robots, while the second uses an questionnaire assisted by pictorial scales.

3 ASSESSMENTS

The instructions and materials for the assessments are available on figshare: https://doi.org/10.6084/m9.figshare.7575911.

Picture Sorting Task. The PST was intended a way of asking children about important features that distinguish different entities, such as aliveness, perception, cognition, human-likeness, and animacy, without using a language-based activity or a questionnaire. Children were asked to arrange a set of pictures of eight entities along a line. The entities included a baby, a frog, a cat, a teddy bear, a computer, a mechanical robot arm, a robot from a movie (e.g., Baymax, WALL-e, or R2D2, depending on which the child was familiar with), and the robot they interacted with in the study. The line was anchored at one end with a picture of an human adult female and at the other with a picture of a table. We wanted to see where children placed the study robot in relation to the other entities. This could help us gain an understanding of how children construe the robot's animate and human-like qualities as compared to these other entities.

Social Acceptance Questionnaire. The SAQ used several questions taken from the Social Acceptance Scale for Kindergarten Children [16, 17], which measures children's acceptance of peers with disabilities (e.g., peers with limitations in how they can interact, such as being unable to walk, hear, or see). We used these questions because robots often have significant limitations in their interaction modalities, e.g., with hearing and understanding because of technical challenges regarding automatic speech recognition and language understanding. We wanted to understand whether children were generally accepting of a robot with these limitations as compared to their acceptance of human peers with similar limitations.

We selected four of the most relevant questions (Table 2) that asked children directly about whether they would like to be good friends with a child with disabilities. The questions can be answered with "yes/no/maybe" responses using a visual scale of 3 smiley faces, which children can point at. We asked these questions both about another child and about the robot in the study.

4 STUDY 1

Methodology

Participants. We recruited 86 children aged 3–8 (*M* = 5.31 years *SD* = 1.43; 44 female, 42 male) from the general Boston area to participate in the study. Of these, 3 were 3-year-olds, 30 were 4-year-olds, 19 were 5-year-olds, 15 were 6-year-olds, and 9 were 7-year-olds, and 10 were 8-year-olds. We recruited a wide age range in order to obtain a sufficient number of participants and also because we were interested in seeing whether older or younger children might respond to the assessments differently. Regarding ethnicity, 48 were White, 7 Asian, 3 Hispanic or Latinx, 20 mixed, and 6 not reported. Forty-nine children spoke English only; 37 children were bilingual. Only some parents reported income level and

Figure 1: A child talks with the robot Tega.

mother's highest level of education: 1 had an income of $10k-$30k, 1 of $30k-$50k, 6 of $50k-$75k, 3 of $75k-$100k, 11 of $100k-$150k, and 23 of over $150k; 41 not reported; 14 were college graduates, 45 had graduate or professional training, and 27 not reported.

Procedure. The first pilot study was a single-session encounter in the lab, approximately 20 minutes in length, during which children interacted one-on-one with a teleoperated social robot, Tega. The robot, pictured in Figure 1, was designed to act as a peer of about the same age as the children, with a friendly and outgoing demeanor. The robot was introduced by name. Children were told that it enjoyed talking about pictures and telling stories. The robot led children in a conversation about a set of pictures (depicting holidays, school activities, a park, and children's movies). It told children a story, and then asked children to retell the story, similar to the activities used in [35, 49]. We administered the PST and SAQ after children completed the robot interaction.

Data. We recorded children's responses to the both assessments in a spreadsheet.

Data Analysis. We coded children's responses to the SAQ on a 3-point scale, with *"no"* as 0, *"maybe"* as 1, and *"yes"* as 2. We labeled children's placement of the entities in the PST, with the anchor on one end (the human) at position 1 and the anchor at the other (the table) at position 10. Thus, a lower rank indicated that children placed the entity closer to the adult woman. We counted positions to determine what rank was held by each picture. We also computed scores for Tega's rank relative to the other entities. For example, we subtracted the human baby's rank from Tega's rank to get Tega's rank relative to the human baby and human adult.

Following the example of [37], we took a two-pronged approach in our data analysis. We first looked at the appropriateness of children's responses, and the reliability and validity of the assessments. Then, we looked at the assessments
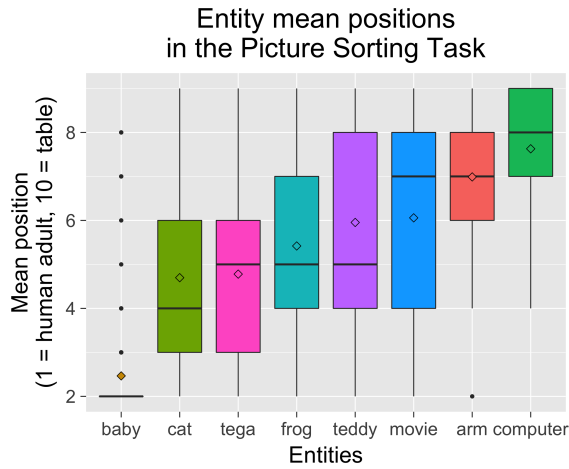
Entity mean positions in the Picture Sorting Task

Figure 2: Children's placement of each entity in the PST in Study 1.

Entity mean distance from Tega Robot in Picture Sorting Task
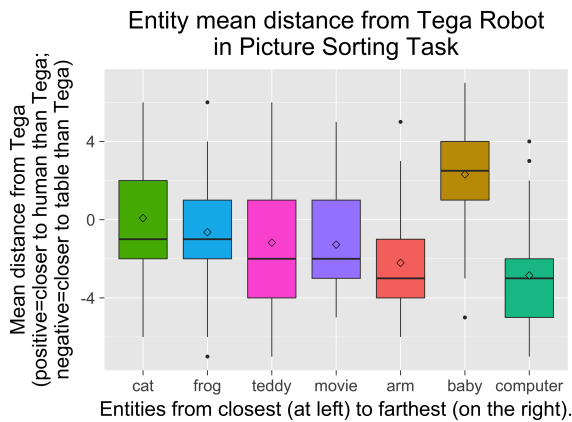
Figure 3: How far each other entity was placed from the Tega robot in the PST in Study 1.

Study 1 SAQ questions about Kids

Figure 4: Girls were slightly more likely to be accepting of other children in Study 1.

Study 1 SAQ questions about Robots
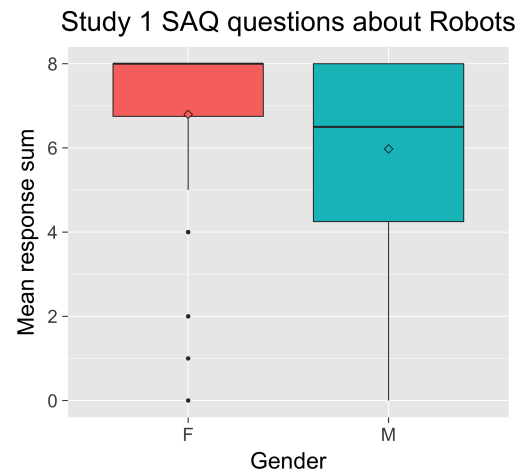
Figure 5: Girls were slightly more likely to be accepting of the Tega robot in Study 1.

with respect to age and gender differences. These analyses, in line with the goals of this paper, served to provide a point of comparision for future work. We were interested in gender differences because prior work has found differences both in how boys and girls develop relationships with their peers as well as with robots [7, 11, 18, 37]. All analyses reported below were planned comparisons.

Results

Picture Sorting Task.

Reliability and appropriateness of response. One-sample t-tests were used to compare the mean of children's responses to chance levels of responding for each PST entity. For chance levels, we used a mean of 5.5, the middle of the scale. If most
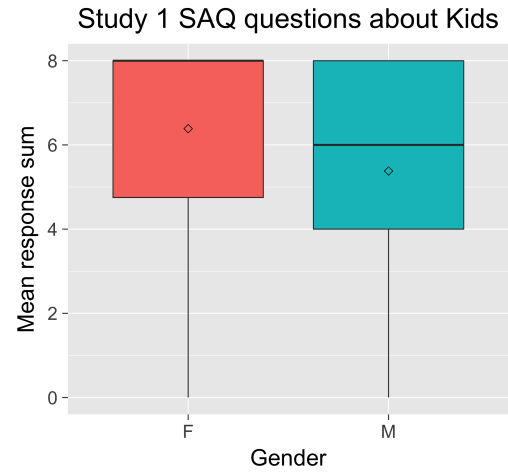
or all entities were not different from chance, this would indicate that children were likely placing the entities randomly. However, if only one or two entities were not different from chance, it would indicate that most entities were being placed consistently toward one end of the scale or the other, which is what we expected to see.

Descriptive statistics and t-test results are shown in Table 1. Children's responses differed from chance as expected: children rated the baby, cat, and Tega as closer to the human adult; the movie robot, robot arm, and computer were closer to the table. The frog and teddy bear were not placed significantly differently from chance, but because these entities could be seen as "in between" the others in terms of

Table 1: Children's overall PST responses. Most differed significantly from chance (mean = 5.5), as shown by one-sample t-tests. Here, "S1" = "Study 1", "S2" = "Study 2", "MAD" = "Median absolute deviation", "df" = "degrees of freedom", "t" = "t-value", and "p" = "p-value".

Entity	Study	Mean (SD)	Median	Range	MAD	df	t	p
Tega	S1	4.78 (1.80)	5	2−9	1.48	85	-3.70	<0.001
	S2 pre	6.17 (2.04)	6	2−9	1.48	45	2.24	0.030
	S2 post	6.02 (1.87)	6	2−9	1.48	45	1.89	0.065
Baby	S1	2.47 (1.31)	2	2−9	0	85	-21.5	<0.001
	S2 pre	2.74 (1.67)	2	2−9	0	45	-11.2	<0.001
	S2 post	2.54 (1.53)	2	2−8	0	45	-13.1	<0.001
Cat	S1	4.70 (1.83)	4	2−9	1.48	85	-4.07	<0.001
	S2 pre	4.48 (1.87)	4	2−9	1.48	45	-3.7	<0.001
	S2 post	4.46 (1.79)	4	2−9	1.48	45	-3.96	<0.001
Frog	S1	5.42 (1.71)	5	2−9	1.48	85	-0.441	0.660
	S2 pre	5.78 (1.93)	6	3−9	2.97	45	0.993	0.326
	S2 post	5.30 (1.91)	5	2−9	1.48	45	-0.696	0.490
Teddy	S1	5.95 (2.19)	5	2−9	2.97	85	1.92	0.058
	S2 pre	6.04 (2.07)	6	2−9	2.97	45	1.78	0.081
	S2 post	6.41 (2.20)	7	3−9	2.97	45	2.82	0.007
Movie	S1	6.06 (2.02)	7	2−9	1.48	85	2.56	0.012
	S2 pre	5.67 (1.97)	6	2−9	1.48	45	0.600	0.552
	S2 post	5.57 (1.89)	6	2−9	2.22	45	0.234	0.816
Arm	S1	6.99 (1.53)	7	2−9	1.48	85	9.02	<0.001
	S2 pre	6.54 (1.79)	7	3−9	1.48	45	3.96	<0.001
	S2 post	6.36 (1.84)	6.5	2−9	2.22	43	3.11	0.003
Computer	S1	7.63 (1.49)	8	4−9	1.48	85	13.3	<0.001
	S2 pre	6.93 (2.34)	8	2−9	1.48	45	4.16	<0.001
	S2 post	7.26 (1.72)	8	3−9	1.48	45	6.95	<0.001

their animacy and human-likeness, this was in line with our expectations.

Differences over time and by gender. A mixed analysis of variance on the entity positions with entity (within: Tega, baby, cat, frog, teddy bear, movie robot, robot arm, computer) and gender (between: male or female) with age as a covariate revealed a significant main effect of Entity, $F(7,588) = 69.9$, $p <0.001$ (Figure 2). Post-hoc tests revealed that the baby was placed closer to the human adult than all other entities. The cat was placed closer to the human adult than all entities except the baby, and its position was not significantly different than that of Tega or the frog. The Tega robot was placed closer to the human adult than the teddy bear, movie robot, robot arm, and computer. The computer and robot arm were placed closer to the table than all other entities; their positions did not differ significantly from each other.

Finally, the positions of the frog, teddy bear, and movie robot did not differ from each other.

We performed a second mixed analysis of variance for the entity positions relative to the Tega robot. We observed a main effect of entity, $F(6,504) = 69.8$, $p = <0.001$ (Figure 3). Post-hoc tests revealed that the baby was placed farther from Tega, and closer to the human adult than Tega was, than all other entities. The computer and robot arm were placed farther from Tega, and closer to the table than Tega, than all other entities. The cat was placed closer to Tega, and closer to the human, than all other entities except the frog; the frog was not placed significantly differently. The frog, movie robot, and teddy bear were not different from each other.

Social Acceptance Questionnaire.

Reliability and appropriateness of response. One-sample t-tests were used to compare the mean of children's responses to chance levels of responding for each question (mean of 1).

Table 2: Summary of children's overall SAQ responses. All differed significantly from chance (mean = 1), as shown by one-sample t-tests. Here, "S1" = "Study 1", "S2" = "Study 2", "df" = "degrees of freedom", "t" = "t-value", and "p" = "p-value". The four questions are Q1: "Would you like to be good friends with a kid/robot who can't see?", Q2: "Would you like to be good friends with a kid/robot who can't hear well?", Q3: "Would you play with a kid/robot even if they had special needs or were disabled?", Q4: "Would you like to be good friends with a kid/robot with special needs?"

Question	Study	Mean (SD)	df	t	p
Q1 Kids	S1	1.47 (0.76)	85	5.66	<0.001
Q2 Kids	S1	1.41 (0.74)	85	5.09	<0.001
	S2 pre	1.43 (0.76)	48	3.93	<0.001
	S2 post	1.42 (0.72)	44	3.92	<0.001
Q3 Kids	S1	1.51 (0.73)	85	6.49	<0.001
Q4 Kids	S1	1.51 (0.70)	85	6.79	<0.001
	S2 pre	1.56 (0.71)	47	5.48	<0.001
	S2 post	1.58 (0.62)	44	6.24	<0.001
Q1 Robot	S1	1.65 (0.63)	85	9.61	<0.001
Q2 Robot	S1	1.50 (0.73)	85	6.34	<0.001
	S2 pre	1.39 (0.81)	48	3.34	0.002
	S2 post	1.49 (0.73)	44	4.51	<0.001
Q3 Robot	S1	1.62 (0.69)	85	8.30	<0.001
Q4 Robot	S1	1.63 (0.63)	85	9.19	<0.001
	S2 pre	1.39 (0.84)	48	3.24	0.002
	S2 post	1.60 (0.69)	44	5.85	<0.001

We expected that more than half the children would respond positively (i.e., be accepting). Descriptive statistics and the t-test results are shown in Table 2. Children's responses differed from chance as expected, with children being more likely to respond positively to the questions.

The reliability of the SAQ was determined by measuring the internal consistency using Cronbach's alpha of the four questions about children and of the four questions about robots separately, since we did not expect that children's opinions about other children and about robots would necessarily be aligned. An alpha coefficient of 0.87 (95% CI: 0.82–0.91) was found for the questions about children. Item reliability was calculated through an item analysis, which revealed that all questions were correlated with the total score, with r values between 0.81–0.88. For the questions about robots, an alpha coefficient of 0.86 (95% CI: 0.82–0.91) was found. Item analysis revealed strong correlations of all questions, with r values between 0.83–0.87 for all items.

Because of the items' reasonably high internal reliability, we computed the sum of the SAQ questions about children and about robots as two composite scores.

Differences over time and by gender. We performed analyses of variance with gender (male vs. female) with age as a covariate on the composite SAQ score. We observed no significant differences, though we did see a trend for a main effect of gender for both the questions about children, $F(1,79) = 3.82$, $p = 0.054$ (Figure 4); and the questions about robots $F(1,79) = 2.98$, $p = 0.09$ (Figure 5). In both cases, girls were more likely than boys to say they would be accepting (children: girls $M = 6.39$, $SD = 2.42$; boys $M = 5.38$, $SD = 2.46$; robots: girls $M = 6.80$, $SD = 2.15$; boys $M = 5.98$, $SD = 2.33$).

5 STUDY 2

Methodology

Participants. We recruited 49 children (23 female, 26 male) aged 4–7 ($M = 5.5$, $SD = 0.93$) from four Boston-area schools to participate, none of whom had participated in Study 1. We recruited from multiple schools because it was not possible to recruit sufficient children from a single school. There were 31 children from School A, 5 children from School B, 6 children from School C, and 7 children from School D. There were 10 four-year-olds, 10 five-year-olds, 24 six-year-olds, and 5 seven-year-olds. Only some parents provided additional demographic information for their children. For ethnicity, 19 children were White, 10 Asian, 1 Black, 1 Hispanic or Latinx, 2 Mixed, and 16 not reported. Twenty-two children were from English-dominant families, 11 were dominant in another language, and 16 were not reported. Three children came from families with an income of $10k-$30k, 4 of $50k-$75k, 4 of $75k-$100k, 6 of $100k-$150k, 12 of over $150k, and 20 not reported. Regarding mother's highest level of education, 4 were high school gradutes or GED, 1 had some college or vocational school, 9 were college graduates, 20 and graduate or professional training, and 16 were not reported.

Procedure. The second pilot study was longitudinal. Children interacted one-on-one with an autonomous social robot, Tega, approximately once per week, for a total of 8 15–20 minute sessions. The robot, shown in Figure 1, was introduced as a peer who enjoyed telling stories; its personality was very similar to that in Study 1. Each session was also similar to that in Study 1: The robot led conversation, told children stories, and asked children to either tell their own or retell the robot's stories.

Due to time limitations, we dropped Q1 from the SAQ due to being less relevant and Q3 due to redundancy with Q4. We administered the modfied SAQ (Q2, Q4) during a pretest prior to children interacting with the robot, and a second time after children's 7th session with the robot. We administered the PST prior to children's 2nd session with the robot and during a posttest after their 8th session with the robot. The differences in when the assessments were

administered were a result of the length of the sessions and other assessments used in the study.

Data. As in Study 1, we recorded children's responses to the both assessments in a spreadsheet.

Data Analysis. Data analysis proceeded the same way as in Study 1. Forty-eight children answered all SAQ questions at the pretest; 1 child did not respond to Q2. Four children were unavailable for the posttest and were excluded from the pertinent analyses. All analyses reported below were planned comparisons.

Results

Picture Sorting Task.

Reliability and appropriateness of response. As in Study 1, one-sample t-tests were used to compare the mean of children's responses to chance levels of responding (mean of 5.5) for each PST entity at the pretest and posttest. Descriptive statistics and the t-test results are shown in Table 1.

As in Study 1, children's responses differed from chance as expected, though in slightly different ways. Children rated the baby and cat as closer to the human adult; the robot arm and computer were closer to the table. The frog and movie robot were not placed significantly differently from chance. At the pretest, Tega was closer to the table; at the posttest, its position was not different from chance. At the pretest, the teddy bear was not different from chance; at the posttest, its position was closer to the table.

Differences over time and by gender. A mixed analysis of variance on the entity positions with entity (within: Tega, baby, cat, frog, teddy bear, movie robot, robot arm, computer), time (within: pre vs. post), and gender (between: male or female) with age as a covariate revealed a significant main effect of Entity, $F(7,656) = 45.9$, $p < 0.001$ (Figure 6). Post-hoc tests revealed that the baby was placed significantly closer to the human adult than all other entities. The cat was placed significantly closer to the human adult than all entities except the baby. The Tega robot was significantly closer to the human adult than the computer, and farther from the adult than the baby and the cat, but was otherwise not placed at a significantly different position from any other entity. The computer was placed significantly closer to the table (and farther from the human adult) than all entities except the robot arm and the teddy bear. Finally, the frog was placed closer to the human adult than the robot arm.

There was a trend toward an interaction of time with gender and entity, $F(7,656) = 1.70$, $p = 0.11$, which suggested, e.g., that girls may have placed the baby closer to the human adult than boys did, that girls may have placed the computer as farther than boys did, and that girls may have moved the teddy bear farther away at the posttest.

We performed a second mixed analysis of variance for the entity positions relative to the Tega robot. We observed a main effect of entity, $F(6,568) = 46.1$, $p < 0.001$. The baby and cat were both placed farther from Tega, and closer to the human adult than Tega was, than all other entities (Figure 7). The computer was farther from Tega than the frog and movie robot were (as well as farther from the adult). There was a trend for an interaction of time with gender and entity, $F(6,568) = 1.74$, $p = 0.11$. These suggested, e.g., that girls may have placed the robot arm closer to the adult than the Tega robot at the posttest than at the S2 test, while boys showed the opposite pattern; girls also tended to place the cat, frog, and computer farther away from Tega than boys did.

Social Acceptance Questionnaire.

Reliability and appropriateness of response. As in Study 1, one-sample t-tests were used to compare the mean of children's responses to chance levels of responding for each question (mean of 1). Descriptive statistics and the t-test results are shown in Table 2. Children's responses differed from chance as expected, with children being more likely to respond positively to the questions at both the pretest and posttest.

As in Study 1, we determined the reliability of the SAQ measuring the internal consistency using Cronbach's alpha and also the Pearson correlation of the two questions about children and of the two questions about robots separately, since we did not except that children's opinions about other children and about robots would necessarily be aligned. At the pretest, an alpha coefficient of 0.63 (95% CI: 0.42–0.83) was found for the questions about children. Item reliability was calculated through an item analysis, which revealed that both questions were correlated with the total score, with r values of 0.87 and 0.84. The items were moderately correlated with each other, $r_p = 0.458$. For the questions about robots, an alpha coefficient of 0.77 (95% CI: 0.65–0.90) was found, with item r values of 0.90 and 0.91. The items were correlated with each other, $r_p = 0.632$.

At the posttest, an alpha coefficient of 0.34 (95% CI: -0.03–0.70) was found for the questions about children, with item r values of 0.81 and 0.74. The items were only somewhat correlated with each other, $r_p = 0.204$. For the questions about robots, an alpha coefficient of 0.42 (95% CI: 0.09–0.74) was found, with item r values of 0.81 and 0.78. The items were slightly correlated with each other, $r_p = 0.264$.

We computed the sum of the SAQ questions about children and about robots as two composite scores and examined test-retest reliability for each composite score. Test-retest reliability was moderate for the questions about children, $r_p = 0.487$, and poor for the questions about robots, $r_p = 0.199$. These results were as expected. We expected children to
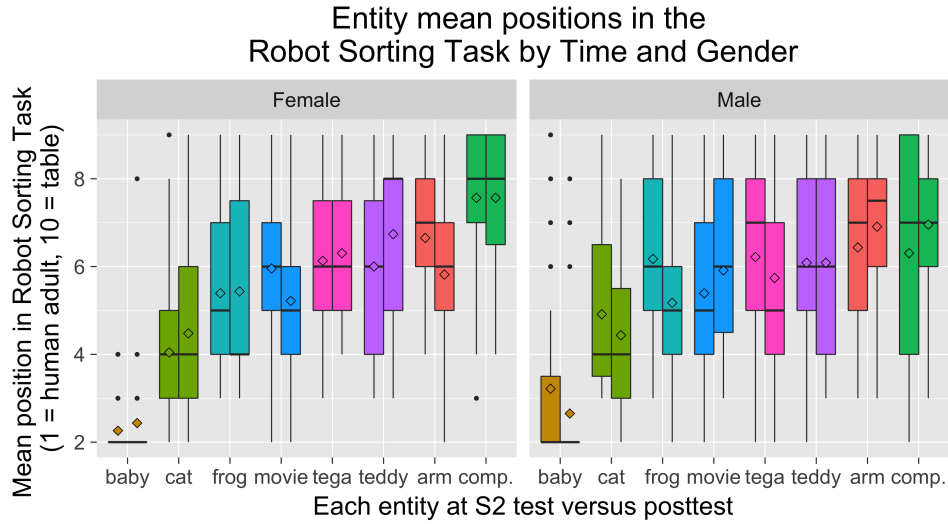
Entity mean positions in the
Robot Sorting Task by Time and Gender

Figure 6: Children's placement of each entity in the PST in Study 2.

Entity mean distance from Tega Robot
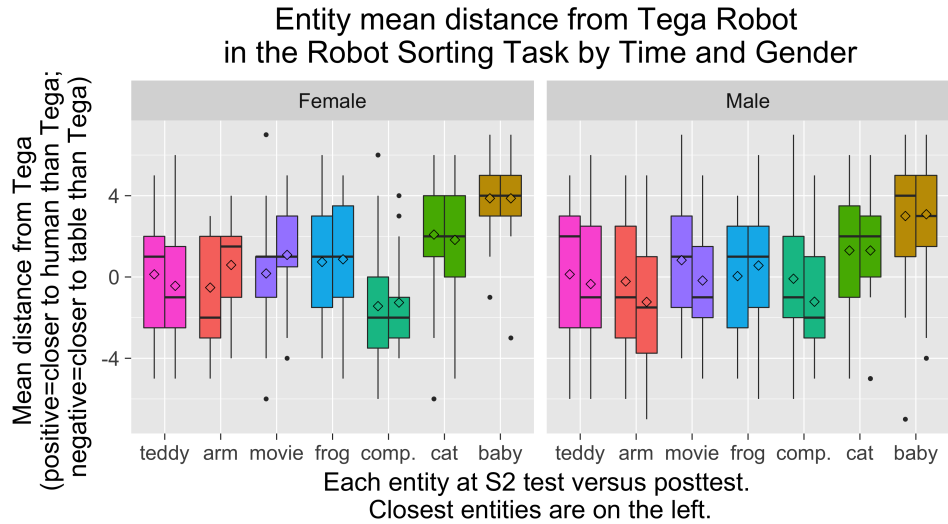## in the Robot Sorting Task by Time and Gender

Figure 7: How far each other entity was placed from the Tega robot in the PST in Study 2.

change their opinions of the robot over time as they grew familiar with it, but change their opinions about other children somewhat less due to existing familiarity and established opinions.

Differences over time and by gender. We performed mixed analyses of variance with time (within: pre vs. post), gender (between: male vs. female) with age as a covariate on the composite SAQ scores. We observed no significant effects, though for the composite children score there was a trend for a main effect of gender, $F(1,40) = 2.04$, $p = 0.16$ (Figure 8). Girls were more likely than boys to say they would be accepting

of other children (girls $M = 3.20$, $SD = 1.13$; boys $M = 2.80$, $SD = 1.09$). For the robot score, there were trends for both main effects of gender, $F(1,40) = 2.05$, $p = 0.16$; and time, $F(1,43) = 1.85$, $p = 0.18$ (Figure 9). Again, girls were more likely than boys to be accepting of the robot (girls $M = 3.14$, $SD = 1.30$; boys $M = 2.70$, $SD = 1.36$). Children were slightly more likely to be accepting of the robot at the posttest (pre $M = 2.73$, $SD = 1.53$; post $M = 3.09$, $SD = 1.12$).

6 DISCUSSION

We presented two new assessments for measuring children's acceptance of a social robot and their perception of a robot's

Study 2 SAQ questions about Kids

Figure 8: Girls were slightly more likely to be accepting of other children in Study 2.

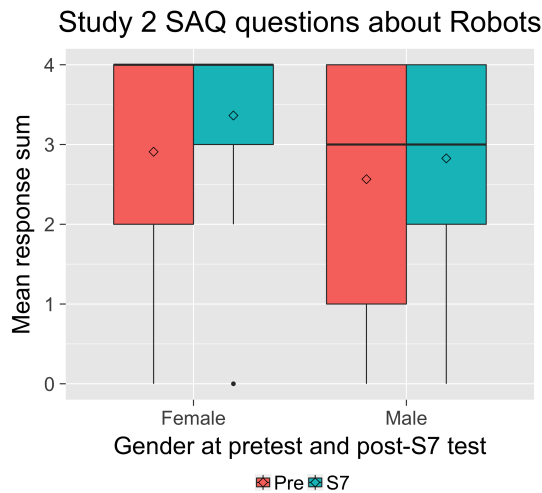## Study 2 SAQ questions about Robots

Figure 9: Children slightly increased their acceptance of the Tega robot from pretest to posttest in Study 2. Girls were slightly more likely to be accepting than boys.

animacy and human-likeness. We tested the assessments in both a single-session study and a long-term study. Below, we discuss the reliabilty and validity results, which demonstrated that children responded appropriately to both assessments. We also discuss the results of using these assessments in the two studies, which show that the assessments can capture differences among children and can be used as a point of comparision for future work.

During the PST, we observed that children's placements of the entities varied, but in general, they tended to place each entity in a reasonable position. For example, children most often placed the baby and cat nearest to the human adult, and placed the robot arm and computer nearest to the table. Children's placements of the different entities showed that children understood several importance features that distinguished the different entities, such as their differing levels of aliveness, animacy, and human-likeness.

Children's placement of the entities in the two studies were fairly similar. Notably, however, in Study 1, the Tega robot was placed closer to the human adult than in Study 2. This is more likely due to the relative autonomy of the robots used in each study rather than differences in activites performed (both studies used similar conversation and storytelling). In Study 1, the robot was teleoperated, which could have lent it a greater degree of responsiveness and human-likeness than the autonomous robot used in Study 2.

The fact that the Tega robot was frequently placed somewhere in the middle, and also frequently close to the frog, movie robot, teddy bear, and robot arm is revealing regarding children's perceptions of the robot. The Tega robot was seen as more human-like than a computer, but definitively less human-like than a baby and a cat. Children seemed to understand that it filled an in-between space: it was not alive in the same way as a human or sophisticated animal; it was also perceiving them and responding to them in a way unlike more static computers and mechanical things. This is in line with prior work suggesting that children may categorize robots as in-between entities, with attributes of both living beings and mechanical artifacts [3, 26, 29, 36, 42, 58]. Children did not confuse the robot with a human—they knew it was different. This may be a result of numerous factors. The robot followed a script; it could not interact in an open-ended way or respond to all a child's questions. Its morphology may have influenced children's perceptions as well, since the Tega robot is fluffy and colorful, and moves more like an animated character than like a humanoid being.

One extension of the PST could be to ask children to speak aloud during sorting about why they are placing each entity in a particular position. Anecdotally, we did attempt a version of this with several children. When they had finished sorting, we asked why they had placed each entity where they did, but these particular children were unable to articulate what features they were using to sort the entities, instead leaving us with shrugs or "I don't know"s. In future work, we may be able to develop an appropriate protocol for asking children to explain more about what they think or feel is different about the entities, as a way of gaining a deeper understanding of how they are construing the animacy and human-likeness of the entities.

Regarding the appropriateness of the SAQ, we saw children use all answer options. They gave a range of responses, though children predominantly responded positively to question about their acceptance of both other children and of the

robot. The SAQ had reasonable internal reliability, more so for the 4-question version used in Study 1 than the 2-question version used in Study 2. Study 2 had used a 2-question version primarily because of time limitations, so we recommend using the full 4-question version if there is time to do so. We found the composite score to be a useful summary score for this assessment. That said, depending on one's research questions, examining children's responses to individual questions may be appropriate and informative (e.g., an individual question about acceptance of a child or robot with a particular limitation if that limitation is being studied in particular).

Children's SAQ responses about children were relatively stable in Study 2 from the pretest to posttest with moderate test-retest reliability. Their SAQ responses about the robot were not so stable, which was expected, since we had hypothesized that children might change their opinions of the robot over time.

One complication with the SAQ was that some children did not know what disabilities or special needs were. The protocol did include a brief conversation about disabilities prior to asking the questions, but did not include a standard definition that all experimenters were required to use if children were unsure. We recommend adding a standard definition to the questionnaire prior to asking the questions along the lines of, "Sometimes kids with disabilities or special needs can't see, or can't hear, or maybe they use a wheelchair. Sometimes they need extra help with things." Any definition used should be tested with children prior to being added to the SAQ.

We saw no significant differences by age for either assessment, suggesting that both were appropriate for our target age group (i.e., 4–7 years). There were a few 3-year-olds in Study 1 as well as several 8-year-olds, so it may be that these assessments will also work for some children outside our target range, though further longitudinal testing with these younger and older children will be needed.

Regarding differences over time, in Study 2, we saw several trends suggesting that children's placements of the entities in the PST changed from the pretest to posttest. While not statistically significant, these trends suggest that children may have changed their construal of the robot's properties relative to these other entities after interacting with it over the course of the study. The lack of statistical significance could be for several reasons, foremost being that the time children spent with the robot may not have been long enough for measurable changes in their construal of the robot to occur. Relationships can take time to develop. If children's construal of the robot as an animate, human-like being is related to their relationship or friendship with the robot, it could take more time for their construal to change, since frequency of interaction has been shown to influence children's perceptions of their friendships with peers [44, 53].

We also saw in Study 2 that children's acceptance of the robot increased over time. Again, this was not statistically significant, perhaps for the same reasons regarding the total length of the interaction. The children in the study also started out with fairly accepting attitudes, so perhaps there was less space for significant change to occur.

We observed several trends for differences between genders for both assessments, such as girls being more accepting of other children and of the robot than boys. These differences jibe with prior research on differences in girls' versus boys' relationships, which has shown, e.g., that girls are often more egalitarian than boys and often rate intimacy and alliance higher than boys do [7, 8, 11, 18]; girls have also rated a robot as more social and relational than boys [37], and have expressed preferences for playing with a robot with a matching gender [54, 55]. The differences we observed, though not statistically significant, suggest that the assessments can capture some individual differences between children.

7 LIMITATIONS AND FUTURE WORK

There are several limitations of this work. First, the two studies included only a small number of participants, with unequal numbers of children of each age group and each demographic group. This is partly a result of the nature of child-robot interaction studies, particularly long-term studies, in which it can be difficult to recruit sufficiently large numbers of children. Future work should endeavor to test the PST and SAQ with a larger, more diverse, and more balanced population of children.

Second, multiple experimenters were involved in the administration of the studies. As such, they may have administered the assessments in slightly different ways, despite being given the same set of instructions regarding administration. This could have affected some children's responses.

Third, the assessments only measured children's acceptance with respect to the robot's abilities, and their construal of the robot's animacy and human-likeness in comparison to a set of other entities. Future work should focus on developing assessments to measure other aspects of children's acceptance and construal of robots, as well as other aspects of children's overall relationships with robots. In addition, because there are no existing measures of these constructs that work well with young children, we could not perform a comparison of these new assessments to existing ones.

Despite these limitations, this work is an important contribution to human-robot interaction. We present these two assessments for others to use, modify, and improve as a way of enabling child-robot interaction research. We hope these assessments will be useful tools for others who want to understand how children think about and relate to social robots—and, potentially, to other technological agents—during both single-session studies and long-term interactions.

ACKNOWLEDGMENTS

8 SELECTION AND PARTICIPATION OF CHILDREN

For Study 1, we recruited 86 children aged 3–8 years to participate from the general Boston area. For Study 2, we recruited 49 children aged 4–7 to participate from four Boston-area schools. We recruited from multiple schools because it was not possible to recruit enough children from a single school. We invited all children in the classrooms whose teachers were interested in the study to participate. In both studies, children's parents gave written informed consent prior to the start of the study, and all children assented to participate. All children had the opportunity to converse and tell stories with a social robot. The protocol was approved by the MIT Committee On the Use of Humans as Experimental Subjects.

REFERENCES

[1] P. Alves-Oliveira, P. Sequeira, and A. Paiva. 2016. The Role That an Educational Robot Plays. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 817–822. https://doi.org/10.1109/ROMAN.2016.7745213

[2] Arthur Aron, Elaine N. Aron, and Danny Smollan. 1992. Inclusion of Other in the Self Scale and the Structure of Interpersonal Closeness. *Journal of Personality and Social Psychology* 63, 4 (1992), 596–612. https://doi.org/10.1037/0022-3514.63.4.596

[3] B. Bartlett, V. Estivill-Castro, and S. Seymon. 2004. Dogs or Robots: Why Do Children See Them As Robotic Pets Rather Than Canine Machines?. In *Proceedings of the Fifth Conference on Australasian User Interface - Volume 28 (AUIC '04)*. Australian Computer Society, Inc., Darlinghurst, Australia, Australia, 7–14.

[4] Christoph Bartneck, Takayuki Kanda, Omar Mubin, and Abdullah Al Mahmud. 2009. Does the Design of a Robot Influence Its Animacy and Perceived Intelligence? *International Journal of Social Robotics* 1, 2 (Feb. 2009), 195–204. https://doi.org/10.1007/s12369-009-0013-7

[5] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics* 1, 1 (Jan. 2009), 71–81. https://doi.org/10.1007/s12369-008-0001-3

[6] Tony Belpaeme, James Kennedy, Aditi Ramachandran, Brian Scassellati, and Fumihide Tanaka. 2018. Social Robots for Education: A Review. *Science Robotics* 3, 21 (Aug. 2018), eaat5954. https://doi.org/10.1126/scirobotics.aat5954

[7] Joyce F. Benenson. 2014. *Warriors and Worriers: The Survival of the Sexes.* Oxford University Press.

[8] Joyce F. Benenson, Ariel Durosky, Jennifer Nguyen, Alexandra Crawford, Evelyne Gauthier, and Éloise Dubé. 2018. Gender Differences in Egalitarian Behavior and Attitudes in Early Childhood. *Developmental Science* 0, 0 (2018), e12750. https://doi.org/10.1111/desc.12750

[9] Timothy W. Bickmore and Rosalind W. Picard. 2005. Establishing and Maintaining Long-Term Human-Computer Relationships. *ACM Trans. Comput.-Hum. Interact.* 12, 2 (June 2005), 293–327. https://doi.org/10.1145/1067860.1067867

[10] Cynthia Breazeal, Paul L. Harris, David DeSteno, Jacqueline M. Kory Westlund, Leah Dickens, and Sooyeon Jeong. 2016. Young Children Treat Robots as Informants. *Topics in Cognitive Science* (March 2016), 1–11. https://doi.org/10.1111/tops.12192

[11] Duane Buhrmester and Wyndol Furman. 1987. The Development of Companionship and Intimacy. *Child Development* 58, 4 (1987), 1101–1113. https://doi.org/10.2307/1130550

[12] Christine T. Chambers and Charlotte Johnston. 2002. Developmental Differences in Children's Use of Rating Scales. *Journal of Pediatric Psychology* 27, 1 (Jan. 2002), 27–36. https://doi.org/10.1093/jpepsy/27.1.27

[13] Stefania Druga, Randi Williams, Cynthia Breazeal, and Mitchel Resnick. 2017. "Hey Google Is It OK If I Eat You?": Initial Explorations in Child-Agent Interaction. In *Proceedings of the 2017 Conference on Interaction Design and Children (IDC '17)*. ACM, New York, NY, USA, 595–600. https://doi.org/10.1145/3078072.3084330

[14] Stefania Druga, Randi Williams, Hae Won Park, and Cynthia Breazeal. 2018. How Smart Are the Smart Toys?: Children and Parents' Agent Interaction and Intelligence Attribution. In *Proceedings of the 17th ACM Conference on Interaction Design and Children (IDC '18)*. ACM, New York, NY, USA, 231–240. https://doi.org/10.1145/3202185.3202741

[15] Allison Druin (Ed.). 1998. *The Design of Children's Technology.* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[16] Paddy C. Favazza and Samuel L. Odom. 1996. Use of the Acceptance Scale to Measure Attitudes of Kindergarten-Age Children. *Journal of Early Intervention* 20, 3 (July 1996), 232–248. https://doi.org/10.1177/105381519602000307

[17] Paddy C. Favazza, 1954, Leslie Phillipsen, and Poonam Kumar. 2000. Measuring and Promoting Acceptance of Young Children with Disabilities. *Exceptional Children* 66, 4 (2000), 491–508.

[18] Tracy R. Gleason and Lisa M. Hohmann. 2006. Concepts of Real and Imaginary Friendships in Early Childhood. *Social Development* 15, 1 (Feb. 2006), 128.

[19] Susan Harter and Robin Pike. 1984. The Pictorial Scale of Perceived Competence and Social Acceptance for Young Children. *Child Development* 55, 6 (1984), 1969–1982. https://doi.org/10.2307/1129772

[20] Willard W. Hartup, Brett Laursen, Mark I. Stewart, and Amy Eastenson. 1988. Conflict and the Friendship Relations of Young Children. *Child Development* 59, 6 (1988), 1590–1600. https://doi.org/10.2307/1130673

[21] Chin-Chang Ho and Karl F. MacDorman. 2010. Revisiting the Uncanny Valley Theory: Developing and Validating an Alternative to the Godspeed Indices. *Computers in Human Behavior* 26, 6 (Nov. 2010), 1508–1518. https://doi.org/10.1016/j.chb.2010.05.015

[22] Adam O. |Greenberg Horvath. 1989/00/00. Development and Validation of the Working Alliance Inventory. *Journal of Counseling Psychology* 36, 2 (1989/00/00), 223–33.

[23] Sooyeon Jeong, Cynthia Breazeal, Deirdre Logan, and Peter Weinstock. 2018. Huggable: The Impact of Embodiment on Promoting Socio-Emotional Interactions for Young Pediatric Inpatients. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 495:1–495:13. https://doi.org/10.1145/3173574.3174069

[24] Jennifer L. Jipson and Susan A. Gelman. 2007. Robots and Rodents: Children's Inferences about Living and Nonliving Kinds. *Child development* 78, 6 (2007), 1675–1688.

[25] Aidan Jones and Ginevra Castellano. 2018. Adaptive Robotic Tutors That Support Self-Regulated Learning: A Longer-Term Investigation with Primary School Children. *International Journal of Social Robotics* 10, 3 (June 2018), 357–370. https://doi.org/10.1007/s12369-017-0458-z

[26] Peter H. Kahn, Batya Friedman, and Jennifer Hagman. 2002. I Care about Him as a Pal: Conceptions of Robotic Pets in Online Aibo Discussion Forums. In *CHI'02 Extended Abstracts on Human Factors in*

Computing Systems. ACM, 632–633.

[27] Peter H. Kahn, Batya Friedman, Deanne R. Perez-Granados, and Nathan G. Freier. 2006. Robotic Pets in the Lives of Preschool Children. *Interaction Studies* 7, 3 (2006), 405–436.

[28] Peter H. Kahn, Heather E. Gary, and Solace Shen. 2013. Children's Social Relationships With Current and Near & Future Robots. *Child Development Perspectives* 7, 1 (2013), 32–37.

[29] Peter H. Kahn, Takayuki Kanda, Hiroshi Ishiguro, Nathan G. Freier, Rachel L. Severson, Brian T. Gill, Jolina H. Ruckert, and Solace Shen. 2012. "Robovie, You'll Have to Go into the Closet Now": Children's Social and Moral Relationships with a Humanoid Robot. *Developmental psychology* 48, 2 (2012), 303.

[30] T. Kanda, R. Sato, N. Saiwaki, and H. Ishiguro. 2007. A Two-Month Field Trial in an Elementary School for Long-Term Human #x2013;Robot Interaction. *IEEE Transactions on Robotics* 23, 5 (Oct. 2007), 962–971. https://doi.org/10.1109/TRO.2007.904904

[31] James Kennedy, Paul Baxter, and Tony Belpaeme. 2017. Nonverbal Immediacy as a Characterisation of Social Behaviour for Human–Robot Interaction. *International Journal of Social Robotics* 9, 1 (Jan. 2017), 109–128. https://doi.org/10.1007/s12369-016-0378-3

[32] W. Bradley Knox, Samuel Spaulding, and Cynthia Breazeal. 2016. Learning from the Wizard: Programming Social Interaction Through Teleoperated Demonstrations (Extended Abstract). In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems (AAMAS '16)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1309–1310.

[33] Jacqueline M. Kory Westlund and Cynthia Breazeal. 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)*. ACM, New York, NY, USA, 65–66. https://doi.org/10.1145/2701973.2701989

[34] Jacqueline M. Kory Westlund, Leah Dickens, Sooyeon Jeong, Paul L. Harris, David DeSteno, and Cynthia L. Breazeal. 2017. Children Use Non-Verbal Cues to Learn New Words from Robots as Well as People. *International Journal of Child-Computer Interaction* (2017). https://doi.org/10.1016/j.ijcci.2017.04.001

[35] Jacqueline M. Kory Westlund, Sooyeon Jeong, Hae Won Park, Samuel Ronfard, Aradhana Adhikari, Paul Lansley Harris, David DeSteno, and Cynthia Breazeal. 2017. Flat versus Expressive Storytelling: Young Children's Learning and Retention of a Social Robot's Narrative. *Frontiers in Human Neuroscience* 11 (2017). https://doi.org/10.3389/fnhum.2017.00295

[36] Jacqueline M. Kory Westlund, Marayna Martinez, Maryam Archie, Madhurima Das, and Cynthia Breazeal. 2016. Effects of Framing a Robot as a Social Agent or as a Machine on Children's Social Behavior. In *The 25rd IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE Press, 688–693.

[37] 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*. ACM, 207–218.

[38] Gary W. Ladd, Becky J. Kochenderfer, and Cynthia C. Coleman. 1996. Friendship Quality as a Predictor of Young Children's Early School Adjustment. *Child Development* 67, 3 (June 1996), 1103–1118. https://doi.org/10.1111/j.1467-8624.1996.tb01785.x

[39] Iolanda Leite, Ginevra Castellano, André Pereira, Carlos Martinho, and Ana Paiva. 2014. Empathic Robots for Long-Term Interaction. *International Journal of Social Robotics* 6, 3 (Aug. 2014), 329–341. https://doi.org/10.1007/s12369-014-0227-1

[40] Iolanda Leite, Carlos Martinho, and Ana Paiva. 2013. Social Robots for Long-Term Interaction: A Survey. *International Journal of Social Robotics* 5, 2 (April 2013), 291–308. https://doi.org/10.1007/s12369-013-0178-y

[41] Séverin Lemaignan, Fernando Garcia, Alexis Jacq, and Pierre Dillenbourg. 2016. From Real-Time Attention Assessment to "With-Me-Ness" in Human-Robot Interaction. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction (HRI '16)*. IEEE Press, Piscataway, NJ, USA, 157–164.

[42] Gail F. Melson, Jr. Kahn, Peter H., Alan Beck, and Batya Friedman. 2009. Robotic Pets in Human Lives: Implications for the Human–Animal Bond and for Human Relationships with Personified Technologies. *Journal of Social Issues* 65, 3 (Sept. 2009), 545–567. https://doi.org/10.1111/j.1540-4560.2009.01613.x

[43] Omar Mubin, Catherine J. Stevens, Suleman Shahid, Abdullah Al Mahmud, and Jian-Jie Dong. 2013. A Review of the Applicability of Robots in Education. *Journal of Technology in Education and Learning* 1 (2013).

[44] Andrew F. Newcomb and Catherine L. Bagwell. 1995. Children's Friendship Relations: A Meta-Analytic Review. *Psychological Bulletin* 117, 2 (1995), 306–347. https://doi.org/10.1037/0033-2909.117.2.306

[45] Tatsuya Nomura, Tomohiro Suzuki, Takayuki Kanda, and Kensuke Kato. 2006. Altered Attitudes of People toward Robots: Investigation through the Negative Attitudes toward Robots Scale. In *Proc. AAAI-06 Workshop on Human Implications of Human-Robot Interaction*. 29–35.

[46] Tatsuya Nomura, Tomohiro Suzuki, Takayuki Kanda, and Kensuke Kato. 2006. Measurement of Negative Attitudes toward Robots. *Interaction Studies* 7, 3 (Jan. 2006), 437–454. https://doi.org/10.1075/is.7.3.14nom

[47] Sandra Y. Okita and Daniel L. Schwartz. 2006. Young Children's Understanding of Animacy and Entertainment Robots. *International Journal of Humanoid Robotics* 03, 03 (Sept. 2006), 393–412. https://doi.org/10.1142/S0219843606000795

[48] Hae Won Park, Mirko Gelsomini, Jin Joo Lee, and Cynthia Breazeal. 2017. 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)*. ACM, New York, NY, USA, 100–108. https://doi.org/10.1145/2909824.3020245

[49] Hae Won Park, Ishaan Grover, Samuel Spaulding, Louis Gomez, and Cynthia Breazeal. 2019. A Model-Free Affective Reinforcement Learning Approach to Personalization of an Autonomous Social Robot Companion for Early Literacy Education. In *AAAI*.

[50] Janet C. Read. 2008. Validating the Fun Toolkit: An Instrument for Measuring Children's Opinions of Technology. *Cognition, Technology & Work* 10, 2 (April 2008), 119–128. https://doi.org/10.1007/s10111-007-0069-9

[51] Janet C. Read and Stuart MacFarlane. 2006. Using the Fun Toolkit and Other Survey Methods to Gather Opinions in Child Computer Interaction. In *Proceedings of the 2006 Conference on Interaction Design and Children (IDC '06)*. ACM, New York, NY, USA, 81–88. https://doi.org/10.1145/1139073.1139096

[52] Ken J. Rotenberg. 1995. Development of Children's Restrictive Disclosure to Friends. *The Journal of Genetic Psychology* 156, 3 (Sept. 1995), 279–292. https://doi.org/10.1080/00221325.1995.9914823

[53] Kenneth H. Rubin, William M. Bukowski, and Jeffrey G. Parker. 1998. Peer Interactions, Relationships, and Groups. In *Handbook of Child Psychology*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470147658.chpsy0310

[54] A. Sandygulova and G. M. P. O'Hare. 2016. Investigating the Impact of Gender Segregation within Observational Pretend Play Interaction. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 399–406. https://doi.org/10.1109/HRI.2016.7451778

[55] Anara Sandygulova and Gregory M. P. O'Hare. 2018. Age- and Gender-Based Differences in Children's Interactions with a Gender-Matching Robot. *International Journal of Social Robotics* 10, 5 (Nov. 2018), 687–700. https://doi.org/10.1007/s12369-018-0472-9

[56] Sofia Serholt. 2018. Breakdowns in Children's Interactions with a Robotic Tutor: A Longitudinal Study. *Computers in Human Behavior* 81 (April 2018), 250–264. https://doi.org/10.1016/j.chb.2017.12.030

[57] Sofia Serholt and Wolmet Barendregt. 2016. Robots Tutoring Children: Longitudinal Evaluation of Social Engagement in Child-Robot Interaction. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*. ACM, New York, NY, USA, 64:1–64:10. https://doi.org/10.1145/2971485.2971536

[58] Rachel L. Severson and Stephanie M. Carlson. 2010. Behaving as or Behaving as If? Children's Conceptions of Personified Robots and the Emergence of a New Ontological Category. *Neural Networks* 23, 8 (2010), 1099–1103.

[59] Lai Poh Emily Toh, Albert Causo, Pei-Wen Tzuo, I.-Ming Chen, and Song Huat Yeo. 2016. A Review on the Use of Robots in Education and Young Children. (2016).

[60] Rianne van den Berghe, Josje Verhagen, Ora Oudgenoeg-Paz, Sanne van der Ven, and Paul Leseman. 2018. Social Robots for Language Learning: A Review. *Review of Educational Research* (Dec. 2018), 0034654318821286. https://doi.org/10.3102/0034654318821286

[61] Lise M. Youngblade, Kathryn A. Park, and Jay Belsky. 1993. Measurement of Young Children's Close Friendship: A Comparison of Two Independent Assessment Systems and Their Associations with Attachment Security. *International Journal of Behavioral Development* 16, 4 (Dec. 1993), 563–587. https://doi.org/10.1177/016502549301600403