Check for
updates

Exploring the Effects of a Social Robot's Speech Entrainment and Backstory on Young Children's Emotion, Rapport, Relationship, and Learning

Jacqueline M. Kory-Westlund* and Cynthia Breazeal

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

In positive human-human relationships, people frequently mirror or mimic each other's behavior. This mimicry, also called entrainment, is associated with rapport and smoother social interaction. Because rapport in learning scenarios has been shown to lead to improved learning outcomes, we examined whether enabling a social robotic learning companion to perform rapport-building behaviors could improve children's learning and engagement during a storytelling activity. We enabled the social robot to perform two specific rapport and relationship-building behaviors: speech entrainment and self-disclosure (shared personal information in the form of a backstory about the robot's poor speech and hearing abilities). We recruited 86 children aged 3–8 years to interact with the robot in a 2 × 2 between-subjects experimental study testing the effects of robot entrainment *Entrainment* vs. *No entrainment* and backstory about abilities *Backstory* vs. *No Backstory*. The robot engaged the children one-on-one in conversation, told a story embedded with key vocabulary words, and asked children to retell the story. We measured children's recall of the key words and their emotions during the interaction, examined their story retellings, and asked children questions about their relationship with the robot. We found that the robot's entrainment led children to show more positive emotions and fewer negative emotions. Children who heard the robot's backstory were more likely to accept the robot's poor hearing abilities. Entrainment paired with backstory led children to use more of the key words and match more of the robot's phrases in their story retells. Furthermore, these children were more likely to consider the robot more human-like and were more likely to comply with one of the robot's requests. These results suggest that the robot's speech entrainment and backstory increased children's engagement and enjoyment in the interaction, improved their perception of the relationship, and contributed to children's success at retelling the story.

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

1. INTRODUCTION

Social robots have been designed as peers, tutors, and teachers to help children learn a variety of subjects (Belpaeme et al., 2018), including math (Clabaugh et al., 2015; Kennedy et al., 2015), language (Movellan et al., 2009; Kory and Breazeal, 2014; Gordon et al., 2016; Kory Westlund et al., 2017a,b; Vogt et al., 2017; Rintjema et al., 2018), reading (Gordon and Breazeal, 2015), handwriting (Hood et al., 2015), social skills (Robins et al., 2005; Scassellati et al., 2018), curiosity (Gordon et al., 2015), and a growth mindset (Park et al., 2017b). Prior work has explored how social robots can best engage children in learning activities and improve learning outcomes, using, e.g., personalization of behavior or curriculum (Gordon and Breazeal, 2015; Hood et al., 2015; Gordon et al., 2016; Baxter et al., 2017; Scassellati et al., 2018), appealing appearance and personality (Kory and Breazeal, 2014), and appropriate nonverbal behaviors (Kennedy et al., 2015; Kory Westlund et al., 2017a,b). One aspect of human-human interpersonal interaction that has been linked to improved learning outcomes in peer tutoring situations is rapport and positive relationships (Sinha and Cassell, 2015a,b). Because of this link, we hypothesize that improving a social robot's capabilities for building rapport and positive relationships with children may similarly lead to improved learning outcomes.

Some prior work with adults provides evidence in support of this hypothesis (Kidd and Breazeal, 2008; Lubold et al., 2016, 2018; Lubold, 2017); however, there is little work yet exploring a social robot's rapport and relationship with young children. Thus, in this paper, we explored whether enabling a social robot to perform rapport-building behaviors, including speech and behavior entrainment, and giving the robot an appropriate backstory regarding its abilities, could help establish rapport and generate positive interactions with children, which we hypothesized could improve children's learning and engagement.

2. BACKGROUND

2.1. Relationships, Rapport, and Learning

We have strong evidence that children's peer relationships provide bountiful opportunities for learning via observing peers, being in conflict with peers, and cooperating with peers (Piaget, 1932; Bandura and Walters, 1963; Bandura, 1971; Vygotsky, 1978; Tudge and Rogoff, 1989; Rubin et al., 1998; De Lisi and Golbeck, 1999; Whitebread et al., 2007). The research so far on children's peer learning discusses how children might learn from other, but does not yet thoroughly address what precisely modulates peer learning. That is: Are all peers approximately equivalent as sources to promote learning, or is there something about some peers that makes them "better inputs" than others? In the context of social robots, what is it about a social robot that could lead children to learn more, or less?

Two possible modulating factors are rapport and a positive relationship. Some recent work has linked rapport to improved learning outcomes in older children's human-human peer tutoring situations (Sinha and Cassell, 2015a,b). In addition, the social bonds between children and teachers can predict learner performance (Wentzel, 1997). Other research has shown that children may learn math concepts from media characters more effectively when they have stronger parasocial relationships with those characters (Gola et al., 2013; Richards and Calvert, 2017).

Many different social and relational factors can increase rapport, trust, and engagement with virtual agents and robots. For example, using appropriate social cues (Desteno et al., 2012; Lee et al., 2013; Breazeal et al., 2016b), contingent backchanneling (Park et al., 2017a), nonverbal mirroring (Bailenson et al., 2005; Burleson and Picard, 2007; Lubold et al., 2018), responsiveness and proactivity (Kim et al., 2006), increased social presence (Lester et al., 1997), and matching ethnic communication styles (Cassell et al., 2009) all have had positive effects.

We chose to implement two rapport- and relationship-building behaviors in a social robot to explore their effects on young children's engagement and learning: speech entrainment and self-disclosure (shared personal information).

2.2. Speech Entrainment

In positive human-human interpersonal interactions, people frequently mimic each other's behavior—such as posture, affect, speech patterns, gestures, facial expressions, and more—unconsciously, without awareness or intent (Davis, 1982; Grammer et al., 1998; Philippot et al., 1999; Provine, 2001; Lakin et al., 2003; Semin and Cacioppo, 2008; Reitter et al., 2011; Borrie and Liss, 2014). This mimicry, also called entrainment, is considered a signal of rapport and has been observed in a variety of human relationships (Tickle-Degnen and Rosenthal, 1990; Dijksterhuis and Bargh, 2001; Rotenberg et al., 2003; Dijksterhuis, 2005; Chartrand and van Baaren, 2009; Wiltermuth and Heath, 2009; Lubold, 2017), as well as with robots and virtual agents (Breazeal, 2002; Bell et al., 2003; Suzuki and Katagiri, 2007; Levitan et al., 2016). While there is less work exploring mimicry and rapport in children, there is some showing that infants and children mimic emotions with humans (Haviland and Lelwica, 1987; Chisholm and Strayer, 1995; Rotenberg et al., 2003) and with robots (Gordon et al., 2016). Thus, enabling a robot to perform entrainment could significantly increase children's rapport with it. We chose speech entrainment because language learning is often a dialogue-heavy activity, and thus, would perhaps be more noticeable and relevant than entraining other behaviors. In addition, given the morphology and technical limitations of the robot platform we had available for this study (the Tega robot, described below), speech entrainment was one of the most feasible behaviors to study, though other behaviors could also be examined in the future (such as posture or affect).

Speech entrainment involves matching the vocal features such as speaking rate, intensity, pitch, volume, and prosody of one's interlocutor. This mimicry tends to happen unconsciously, and more often when rapport has been established—i.e., when one feels closer to or more positively about one's interlocutor (Porzel et al., 2006; Reitter et al., 2011; Borrie and Liss, 2014). Some recent work has explored increasing prosodic synchrony in a speech-controlled child-robot game in order to promote cooperation and improve enjoyment (Chaspari and Lehman, 2016; Sadoughi et al., 2017). In addition, Lubold and colleagues developed several social voice-adaptive robots that adjust the

pitch of the robot's text-to-speech voice to match that of its human interlocutor (Lubold et al., 2015, 2016, 2018; Lubold, 2017). This vocal entrainment contributed to increased learning with undergraduate students as well as middle school students during math tasks, but did not increase self-reported rapport. However, our work differs in several ways. We are investigating the impact of entrainment with younger children in a more social task—language learning—that may be more affected by social relationships. Second, these prior studies compared a robot with a text-to-speech voice to one that had a more expressive (albeit contingently adapted) voice. They did not control for the expressivity of the voice. Other recent work found that a robot with a more expressive voice was more effective as a learning companion, leading to greater engagement and learning, than a robot that used a flat voice, similar to a classic text-to-speech voice (Kory Westlund et al., 2017b). This work raises the question of whether the effects seen in Lubold et al.'s studies are strictly a result of the entrainment or a result of the robot's voice being more expressive. In the work presented here, we control for the robot's expressivity.

2.3. Backstory (Personal Self-Disclosure)

Backstory is the story told by or about an agent, including personal story (e.g., origin, family, hobbies), capabilities, limitations, and any other personal information that might be disclosed. With young children in particular, we expect that sharing information about an agent in a story context could make it easier for children to understand.

Prior work has shown that the story told about a robot prior to interaction can change how people perceive the robot and interact with it. Telling participants that a robot is a machine vs. a human-like, animate agent (Stenzel et al., 2012; Klapper et al., 2014; Kory Westlund et al., 2016b) or giving the robot a name and a story involving greater agency and experience (Darling et al., 2015) can manipulate people's perceptions of the robot as an animate, social agent as well as their empathy for the agent. These studies build on extensive work in social cognition and social psychology literature regarding the idea that framing or priming can influence subsequent behavior and perception (Dijksterhuis and Bargh, 2001; Biernat, 2004). However, it is not only stories told before an interaction, but also the content *of* an interaction that affects people's perceptions of their interlocutor. For example, one aspect of children's friendships and positive relationships is self-disclosure. Children disclose more information, and more personal information, in closer relationships (Rotenberg and Mann, 1986; Rotenberg, 1995). The amount of disclosure during conversation reflects how close two children feel to one another. A robot that discloses personal information may impact not only relationship formation and perception, but the story it tells could also impact how a child perceives how social an agent the robot is.

Backstory can also increase engagement with an agent. For example, in one study, giving a robot receptionist a scripted backstory during a long-term deployment increased engagement, since the story added interesting variation and history to the interactions people had with it (Gockley et al., 2005). However,

no research as yet has examined the impact a backstory can have on young children's learning.

Part of our goal in giving the robot a backstory was to promote a more positive relationship. Thus, we examined specific interventions regarding the acceptance of peers and how these interventions might play into the story told about the robot. Favazza and colleagues explored how to promote the acceptance of peers with disabilities in children's kindergarten classrooms, as well as how to measure that acceptance (Favazza and Odom, 1996; Favazza et al., 2000). One component of the intervention they used involved telling stories with guided discussion about children with disabilities; a second component involved structured play with the peers who had disabilities. We combined the idea of telling a story about one of the robot's relevant difficulties that could be perceived as a disability—namely, its hearing and listening abilities—with the idea of self-disclosure as a component of children's friendships; and followed this story/disclosure with several structured activities with the robot.

There are ethical concerns regarding deception when giving robots stories that may elicit empathy, trust, or acceptance. In this study, the backstory we chose to use was fairly reflective of the actual limitations and capabilities of social robots. It pertained to the robot's difficulties with hearing and listening and was thus fairly realistic and not particularly deceptive, given general difficulties in social robotics with automatic speech recognition and natural language understanding. The remainder of the backstory discussed the robot's interest in storytelling and conversation, which was deceptive in that robots do not really have interests, but served to present the robot as a character with interests in these subjects in order to promote engagement in learning activities.

3. METHODOLOGY

3.1. Research Questions

We wanted to explore whether a social robot that entrained its speech and behavior to individual children and provided an appropriate backstory about its abilities could increase children's rapport, positive relationship, acceptance, engagement, and learning with the robot during a single session.

3.2. Design

The experiment included two between-subjects conditions: Robot entrainment (*Entrainment* vs. *No entrainment*) and Backstory about abilities (*Backstory* vs. *No Backstory*). We abbreviate the four conditions as *E-B, E-NB, NE-B,* and *NE-NB*. In the *Entrainment (E)* condition, the robot's speech was entrained based on each child's speaking rate, pitch, and volume, and exuberance. In the *Backstory (B)* condition, the experimenter explained that the robot was not so good at hearing and needed practice; this backstory was reinforced by the robot later.

3.3. Participants

We recruited 95 children aged 3–8 years (47 female, 48 male) from the general Boston area to participate in the study. We recruited a wide age range in order to recruit a sufficient number of participants and also because we were interested in seeing

Condition	Mean age (SD)	Girls	Boys	Monolingual	Bilingual
E-B	5.40 (1.54)	11	9	12	8
E-NB	5.21 (1.34)	7	9	9	7
NE-B	5.44 (1.67)	13	15	18	10
NE-NB	5.27 (1.35)	13	9	11	11

whether older children (e.g., 6–8 years) or younger children (e.g., 3–5 years) might relate differently to the robot's relational behavior, since children may develop relationships differently as they grow older (Hartup et al., 1988; Rubin et al., 1998).

Nine children were removed from analysis because they did not complete the study[1]. The children in the final sample included 86 children aged 3–8 (44 female, 42 male), with a mean age of 5.31 years (*SD* = 1.43). 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. Forty-nine children spoke English only; 37 children were bilingual.

We used random counterbalanced assignment to assign children to conditions. There were 20 in the *E-B* condition, 16 in the *E-NB* condition; 28 children in the *NE-B* condition; and 22 in the *NE-NB* condition. The imbalance was a result of the children who did not complete the study. **Table 1** lists age, gender, and bilingual status by condition. Age did not significantly differ by condition. We asked parents to rate their children's social behavior on a variety of dimensions; these ratings also did not significantly differ by condition.

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

3.4. Hypotheses

We expected that the robot's entrainment and backstory might affect both children's rapport and social behavior, as well as learning and retention, during a single session with the robot. Accordingly, we used a variety of measures to explore the effects of the robot's entrainment and backstory. We tentatively expected the following results:

Learning

- **H1:** In all conditions, children would learn the target vocabulary words presented in the robot's story. In prior studies, we have seen children learn new words from stories told by robots (Kory, 2014; Kory Westlund et al., 2017b; Park et al., 2019). However, we expected that children would learn more as a result of the robot's entrainment or

from an increased relationship, i.e., the most in the *E-B* condition, followed by the *E-NB* and *NE-B* conditions, and the least in the *NE-NB* condition.
- **H2:** Children who learned the target vocabulary words would also use them in their story retells. We have previously seen children mirror a robot's vocabulary words in their own stories (Brennan, 1996; Iio et al., 2015; Kory Westlund et al., 2017b).
- **H3:** Because of the expected connection between the robot's entrainment and backstory to children's rapport and relationship, as well as prior work showing that the story told about a computer's limitations influenced participants' lexical entrainment (Pearson et al., 2006), we expected the entrainment and backstory would lead to differences in children's mirroring of the robot's story in their retells. Children in the *E-B* condition would produce more vocabulary, longer stories, and phrase mirroring because of more rapport and a closer relationship.

Rapport, Relationship, and Social Behavior

- **H4:** A robot with an appropriate backstory about its abilities (*E-B* and *NE-B* conditions) would lead to greater acceptance by children of the robot and more helping behaviors.
- **H5:** Both entrainment and backstory would lead children to treat the robot as a greater social other, such as laughing and smiling more (Provine, 2001; Smidl, 2006), and affording the robot courtesies such as saying goodbye or considering its preferences (Reeves and Nass, 1996). We expected to see this more in the *E-B* than the other conditions; and least in the *NE-NB* condition.
- **H6:** Children would show greater rapport, entrainment, mirroring, and helping behaviors with a robot that entrained to them (*E-B* and *E-NB* conditions). We also expected that a robot with both an appropriate backstory and entrainment (*E-B*) would promote a stronger relationship, and as a result, greater attention, engagement, rapport, and mirroring than in the *E-NB* condition. Furthermore, children's attention, engagement, and positive emotions would increase—or at least decrease less—over the course of the session than in the other conditions.
- **H7:** Children who reported a closer relationship to the robot would also show more mirroring behaviors, more helping behaviors, greater rapport, greater engagement, and more learning. We expected a connection between children's relationship and their learning because of prior work showing that rapport can facilitate learning in peer tutoring scenarios (Sinha and Cassell, 2015a,b).

3.5. Procedure

Five different experimenters (three female adults and two male adults) ran the study in pairs in a quiet room in the lab. The study setup is shown in **Figure 1**. One experimenter interacted with the child. The second experimenter was present in the room, but sat back behind a laptop and did not interact directly with

---

[1]The children who failed to complete the study were primarily younger children (one 3-year-old, five 4-year-olds, one 5-year-old, and two six-year-olds). Most were very distracted during the session and did not want to play with the robot for the full duration of the session. One 4-year-old and the 3-year-old appeared scared of the robot and did not want to interact at all, even with parental prompting. One of the 6-year-olds had accidentally signed up for the study twice, and this was not noticed until after we began the session.
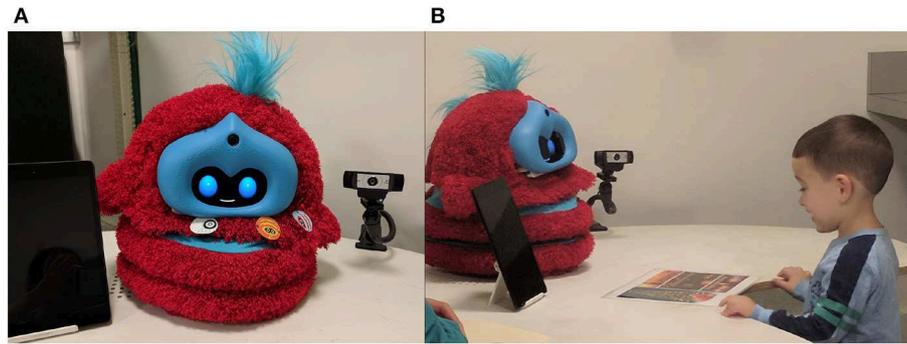
FIGURE 1 | (A) The robot was placed on a table. The tablet was set upright to the left (when facing the robot), and the camera behind the robot and to the right. (B) A child discusses holidays with the robot in the picture conversation task. Written informed consent was obtained to use this image.

the child; their role was to teleoperate the robot and manage the other equipment. Some children wished their parents to stay with them (e.g., if they were particularly shy); in these cases children's parents were instructed to watch only and let their children do as much as possible by themselves.

For each child, the interaction with the robot lasted about 20 min, followed by 5–10 min for the posttests. The interaction script, full interaction procedure, and other study materials are available for download from figshare at: https://doi.org/10.6084/m9.figshare.7175273; they are available for download as **Supplementary Materials**.

The experimenter introduced the sleeping robot, Tega, to the child and explained that it liked looking at pictures and telling stories. If the child was in the Backstory condition, the experimenter also explained that Tega sometimes had trouble hearing: "Do you see Tega's ears? Tega's ears are hiding under all the fur, so sometimes Tega's ears don't work very well. Tega sometimes has a lot of trouble hearing. You should talk to Tega in a loud and clear voice so Tega can hear you. Try to be understanding if Tega needs to hear something again." Then, in all conditions, the experimenter invited the child to help wake up the robot.

The robot interaction had four main sections: A brief introductory conversation (providing context for sharing the backstory, 2–3 min), a conversation about pictures (providing opportunities for speech entrainment and a helping/compliance request, 5–6 min), a sticker task (a sharing/compliance request, 1 min), a storytelling activity (providing opportunities to learn words and mirror the robot's speech, 10–12 min), and a brief closing conversation (1–2 min).

In the introductory conversation, the robot introduced itself, shared personal information about its favorite color and an activity it liked doing, and prompted the child for disclosure in return. Then, in the Backstory condition, the robot reinforced the backstory provided by the experimenter earlier, telling the child, "Sometimes I have trouble hearing and I can't always understand what people tell me. I try really hard, but sometimes I just don't hear things right. I need help and practice to get better!"

The picture conversation took approximately 5 min and was designed to provide many conversation turns for the child, and thus provide the robot with opportunities to entrain its speech to

the child's. The experimenter placed photos one at a time in front of the robot and child (e.g., a collage of holidays or pictures from children's movies). For each picture, the robot introduced the picture content, expressed something it liked about the picture, asked the child a question, responded with generic listening responses (e.g., "Can you tell me more?," "Oh, cool!," "Keep going!"), shared another fact relevant to the picture, and asked another question. At two points during this activity, there were scripted moments where the robot had difficulty hearing (saying, e.g., "I didn't hear that, can you say it again?"), to reinforce its backstory. The experimenter explained that the robot and child had to do at least three pictures, but they could do one more if they wanted—this set up a later compliance/helping task after the third picture, in which the robot asked if the child would do a fourth picture with it to help it practice extra. If the child declined the fourth picture, the experimenter moved on.

The sticker task was used to see how likely the child was to agree to a request by the robot to share a favorite object. The child was allowed to pick out a sticker from a small selection. The robot stated that it wanted the child's sticker and asked for it. The child could spontaneously speak or give their sticker to the robot, or decline. If the child gave their sticker, the experimenter would conveniently find a duplicate sticker in their pocket to replace it, so that the child would not have to forgo their favorite sticker.

The storytelling activity was modeled after the story retelling task used in Kory Westlund et al. (2017b). The robot told a story consisting of a 22-page subset of the wordless picture book "Frog, Where Are you?" by Mercer Mayer. The pages of the book were shown one at a time on the tablet screen. On each page, the robot said 1–2 sentences of the story. Every few pages, the robot asked a dialogic reading comprehension question about the events in the story, e.g., "Where is the deer taking the boy?," 'and "How do you think the boy feels now?" (3 questions total, decreased from the 11 questions in the prior study to decrease the length of the story activity). As in the prior study, the robot responded to children's answers with encouraging, non-committal phrases such as "Mmhm," "Good thought," and "You may be right."

We embedded six target vocabulary words (all nouns) into the story. As in the prior study, we did not test children on their knowledge of these words prior to the storytelling activity because we did not want to prime children to pay attention to

these words, since that could bias our results regarding whether or not children would learn or use the words after hearing them in the context of the robot's story. We used the six key nouns identified in the original story in Kory Westlund et al. (2017b), which were replaced with the target words "gopher"(original word: animal), "crag" (rock),"lilypad" (log), "hollow" (hole), "antlers" (deer), and "cliff" (hill).

After the robot told the story, the robot prompted children to retell the story. Children could use the tablet while retelling the story to go through the story pages, so they could see the pictures to help them remember the story. Twice during the retell, the robot had difficulty hearing ("What? Can you say that again?"), which reinforced the backstory. Children's retellings were used as a measure of their story recall, mirroring of the robot's speech, and expressive use of the vocabulary words.

As part of the closing conversation, we included a goodbye gift task. The experimenter brought out a tray with several objects on it: a small toy frog (because the frog was present in the robot's story), a small book (because the robot expressed great interest in stories), a sticker of the robot's favorite color (blue), and an orange sticker. The child could pick an object to give to the robot, and the experimenter followed up by asking why the child had picked that gift.

After the robot interaction, the experimenter administered a receptive vocabulary test of the six target words in the story. For each word, four pictures taken from the story's illustrations were shown to the child. The child was asked to point to the picture matching the target word. We examined both children's receptive knowledge of the words as well as children's expressive or productive abilities during the story retelling, since children who can recognize a word may or may not be able to produce it themselves.

This was followed by the Inclusion of Other in Self task, adapted for children as described in Kory-Westlund et al. (2018). In this task, children are shown seven pairs of circles that proceed from not overlapping at all to overlapping almost entirely. They are asked to point to the circles showing how close they feel to five different entities: their best friend, their parent, a bad guy they saw in a movie, their pet (or if they have no pet, their favorite toy), and the robot. These five entities were included because we were curious how children might rate the robot compared to other people and things they might feel close to.

Then the experimenter asked several questions taken from the Social Acceptance Scale for Kindergarten Children (Favazza and Odom, 1996; Favazza et al., 2000) regarding how accepting children might be of the robot and its hearing difficulties, as well as of other children who might have hearing difficulties, as described in Kory-Westlund and Breazeal (2019). Finally, children performed a Picture Sorting Task (Kory-Westlund and Breazeal, 2019), in which they were asked to arrange a set 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 Tega. The line was anchored at one end with a picture of an adult human female and at the other with a picture of a table. We wanted to see where children placed the robot in relation to the other entities.

3.6. Materials

We used the Tega robot, a colorful, fluffy squash and stretch robot designed for interactions with young children (Kory Westlund et al., 2016a) (see **Figure 1**). The robot is covered in red fur with blue stripes and uses an Android phone to display an animated face and run control software. The face has blue oval eyes and a white mouth, both of which can change shape to display different facial expressions and mouth movements (visemes) during speech. The robot can move up and down, tilt sideways, rotate from side to side, and lean forward and backward. The experimenters referred to the robot by name (not with pronouns) in a non-gendered way throughout the study.

Speech was recorded by a human adult female and shifted to a higher pitch to sound more child-like. All robot speech was sent through the automated audio entrainment module and streamed to the robot. For the *Entrainment* conditions, all speech was entrained; for the *No Entrainment* conditions, processing still occurred, but the speech simply passed through and was not changed. The reason for this was to incur the same delay (generally a latency of less than 1–2 s) that results from entraining and streaming speech in both conditions. More details regarding entrainment are provided below.

We used a Google Nexus 9 8.9-inch tablet to display the story. Touchscreen tablets have effectively engaged children and social robots in shared tasks (Park et al., 2014), including storytelling activities (Kory and Breazeal, 2014; Kory Westlund et al., 2017b). We used the same custom software on the tablet to display the story pages as in Kory Westlund et al. (2017b), which allowed the teleoperator to turn the pages at appropriate times. This software is open-source and available online under the MIT License at https://github.com/mitmedialab/SAR-opal-base/.

3.7. Teleoperation

As in the prior study (Kory Westlund et al., 2017b), we used custom teleoperation software to control the robot and digital storybook. The teleoperation software is open-source and available online under the MIT License at https://github.com/mitmedialab/tega_teleop/. The experimenters were all trained to control the robot by an expert teleoperator.

Using teleoperation allowed the robot to appear autonomous while removing technical barriers, primarily natural language understanding, since the teleoperator could be in the loop to parse language. The teleoperator triggered when the robot began each sequence of actions (speech, physical motions, and gaze), and when the storybook should turn the page. Thus, the teleoperator had to attend to timing in order to trigger action sequences at the right times. The timing of actions within sequences was automatic and thus consistent across children. There were also several occasions when the teleoperator had to listen to children's speech and choose the most appropriate of a small set of different action sequence options to trigger, namely during the picture conversation task.

The teleoperator performed one of two actions if the child asked an unexpected question or said something unusual. During the conversation portion of the interaction, the teleoperator could trigger one of the generic responses (e.g., "Mmhm!," "Hm, I don't know!") in reply. During the remainder of the

interaction, the teleoperator had to continue in accordance with the interaction script, which essentially ignored unexpected behaviors. While this is not ideal from an interaction standpoint, it was necessary to ensure reasonably consistent behavior on the part of the robot across children.

3.8. Entrainment

In the *Entrainment* condition, the speaking rate and pitch of the robot's voice were automatically adjusted to be more similar to the child. In addition, the robot's volume and exuberance were manually adapted by the teleoperator.

For speaking rate and pitch entrainment, the child's speech was automatically collected via the robot's microphone when it was the child's turn to speak in the conversation. Using automatic software scripts with Praat (audio analysis software), various features of the children's speech were extracted and used to modify the robot's recorded speech files. These modified audio files were then streamed to the robot for playback.

For speaking rate, the robot's speech was sped up or slowed down to match the child's speaking rate. Thus, if a child spoke slowly, the robot slowed down its speech as well. We included ceiling and floor values such that the robot's speech would only ever be sped up or slowed down by a maximum amount, ensuring that the speech stayed within a reasonable set of speeds. We used the Praat script for speaking rate detection from de Jong and Wempe (2009). The code for our entrainment module is open-source and available online under a GNU General Public License v3.0 at https://github.com/mitmedialab/rr_audio_entrainer/.

The mean pitch of the robot's speech was shifted up or down. In doing this, the robot matches two features: (1) the child's age, (2) the child's current mean pitch. In general, people speak at a particular fundamental frequency, but there is variation within an individual (pitch sigma). Thus, we provided a table of mean fundamental frequencies for different age children based on the values computed in prior work (Weinberg and Zlatin, 1970; Bennett, 1983; Sorenson, 1989; Hacki and Heitmüller, 1999; Baker et al., 2008; Gelfer and Denor, 2014). For a given child, all of the robot's speech was first shifted to have the mean pitch for children of that age. Then, since an individual may vary their pitch in each utterance, the pitch of each utterance was also shifted up or down slightly based on whether the child's most recent utterance was higher or lower. Unlike Lubold and colleagues (Lubold et al., 2016, 2018), we did not adapt the pitch contour of the robot's speech. Because the base sounds for the robot's speech were recorded by a human (not flat text-to-speech as in Lubold et al.'s work), the sounds had their own pitch contours. Pilot tests showed that morphing or replacing this contour led to speech that sounded unnatural (e.g., placing emphasis on the wrong syllables).

We also manually adapted the robot's volume and exuberance. During the introduction and first picture in the picture task, the teleoperator observed the child's behavior and personality: were they shy, passive, reserved, or quiet (less exuberant/quiet children)? Or were they loud, extroverted, active, smiley, or expressive (more exuberant/loud children)? Based on this binary division, the teleoperator adjusted the robot's audio playback volume twice, at two specific points during the interaction, to either be slightly quieter (for less exuberant/quiet children) or slightly louder (for more exuberant/louder children). Furthermore, the teleoperator triggered different animations to be played on the robot at six different points during the interaction—more excited and bigger animations for more exuberant/louder children; quieter, slower, animations for less exuberant/quieter children.

3.9. Data

We recorded audio and video of each interaction session using a camera set up on a tripod behind the robot, facing the child. All audio was transcribed by human transcriptionists for later language analyses. Children's responses to the posttest assessments were recorded on paper and later transferred to a spreadsheet.

3.10. Data Analysis

For the analysis of children's story retellings, we excluded the three 3-year-olds because one did not retell the story, and the other two needed extra prompting by the experimenter and were very brief in their responses. Of the remaining 83 children, one child's transcript could not be obtained due to missing audio data. Fifteen children did not retell the story (the number from each condition who did not retell the story was not significantly different). Thus, in total, we obtained story retell transcripts for 67 children (15 *E-B*; 9 *E-NB*; 22 *NE-B*; 21 *NE-NB*).

We analyzed children's transcribed story retells in terms of story length (word count), overall word usage, usage of target vocabulary words, and similarity of each child's story to the robot's original story. We created an automatic tool to obtain similarity scores for each child's story as compared to the robot's story, using a phrase and word matching algorithm. The algorithm proceeded as follows: First, take both stories (the original story and the child's story) and remove stopwords (i.e., words with no significant information such as "the," "uh," and "an"). Second, stem words—i.e., convert words to their original form. For example, "jumping" would be converted to "jump." Third, find all N-grams in each story, where an N-gram is a continuous sequence of N words from both texts. Fourth, remove duplicate N-grams from one text. Fifth, count how many N-grams are the same in both texts. The number of matches is the similarity score. This algorithm produces a score reflecting the number of exact matching phrases in both stories—i.e., words used in the same order by both the child and robot. It also produces a higher match score for texts that have both more matching phrases and longer matching phrases. We also implemented an algorithm for counting similar matches that are close to each other, but not exactly the same. This algorithm was the same as the above, where the fifth step (counting matching N-grams) used a fuzzy string matching algorithm to determine if the N-grams matched.

When running the algorithm to match stories, we used $N = 3$ for computing exact match scores because a smaller N may not retain enough information to be considered actual phrase matching, while a larger N may encompass more information than would constitute a single phrase. For determining similar match scores, we used $N = 4$, so that when phrases differed by

one word, or used a different word in the middle of a similar phrase, they might still match, as would be expected for similar phrases. We combined the exact and similar match scores to get a single overall similarity score for each child's story that reflected the child's overall use of exact and similar matching phrases.

For example, the robot's story included the sentences, "The baby frog liked the boy and wanted to be his new pet. The boy and the dog were happy to have a new pet frog to take home." After stopword removal and stemming, this was converted to: "baby frog like boy want be new pet boy dog happy new pet frog take home." One child's story included the similar section, "Then he hopped on his hand and he wanted to be his pet. And then the dog and the boy was happy to have a new pet," which was converted to: "hop hand want be pet dog boy happy new pet." There were several exactly matching phrases, e.g., *"happy new pet."* There were also several similar matching phrases, e.g., (robot) *"be pet boy dog"*/(child) *"be pet dog boy."*

We obtained children's facial expressions from the recorded videos using Affdex, emotion measurement software from Affectiva, Inc., Boston, MA, USA (McDuff et al., 2016). Affdex can detect 15 facial expressions, which are used to detect whether the face is displaying nine different affective states. Affdex only recognizes outward expressions of affect (i.e., facial configuration patterns), which does not imply detecting any underlying feelings or inferring deep internal states (though they are believed to be correlated). For each frame of a video, Affdex attempts to detect a face. If a face is detected, Affdex scores each affective state as well as the presence of each expression in the range 0 (no expression/affective state detected) to 100 (expression or state fully present); middle values represent an expression or state that is partially present. However, these values are relative and Affdex does not specify what the exact difference between scores means. For more detail on the algorithms used for facial affect classification, see Senechal et al. (2015). We analyzed affect data for 74 children (16 *E-B*; 11 *E-NB*; 26 *NE-B*; 21 *NE-NB*). For the remaining 12 children, little or no affect data were collected as a result of system failures, such as children's faces not being recognized by Affdex.

We focused our analysis on the following affective states and facial expressions: joy, fear, sadness, surprise, concentration, disappointment, relaxation, engagement, valence, attention, laughter, and smiles. We included valence in addition to specific emotions such as joy because Affdex uses different sets of facial expressions to detect the likelihood that a face is showing each affective state. Thus, valence is not detected from, e.g., the emotions joy or sadness; instead, it is calculated from a set of facial expressions that is somewhat different than, though overlapping with, the set of expressions used to calculate other emotions. The expression "concentration" was called "contempt" by Affectiva. Affectiva has no label for concentration or thinking expressions. Affectiva uses brow furrows and smirks to classify contempt; prior work has found that brow furrowing and various lip movements present in smirks such as mouth dimpling and lip tightens are also associated with concentration (Oster, 1978; Rozin and Cohen, 2003; Littlewort et al., 2011). Furthermore, contempt is generally defined as "the feeling that a person or thing is worthless or beneath

consideration," which, as in Kory Westlund et al. (2017b), did not make sense in this context; children's expressions were more indicative of concentration.

We coded children's responses to the Social Acceptance Scale questions 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 Picture Sorting Task, 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. Because Tega's position among the entities was dependent on where children placed the other entities in the task, we examined where children placed all the different entities.

We coded whether children agreed to do the fourth picture and whether they gave the robot their sticker with *"no"* as 0 and *"yes"* as 1. We coded children's selections in the goodbye gift task as follows: *frog* as 4, *book* as 3, *blue sticker* as 2, and *orange sticker* as 1. We also coded the comments children made regarding why they selected a particular gift with the following rubric: 2 if they referenced the robot or the robot's feelings (e.g., "Tega would like it because frog jumped out in story," "Tega likes books," "Because he wanted a sticker"); 1 for a somewhat relevant comment, mentioning the interaction (e.g., "It was in the story"); 0 for no explanation, reference to themselves, or an irrelevant comment (e.g., "It is swamp week at camp," "I don't know").
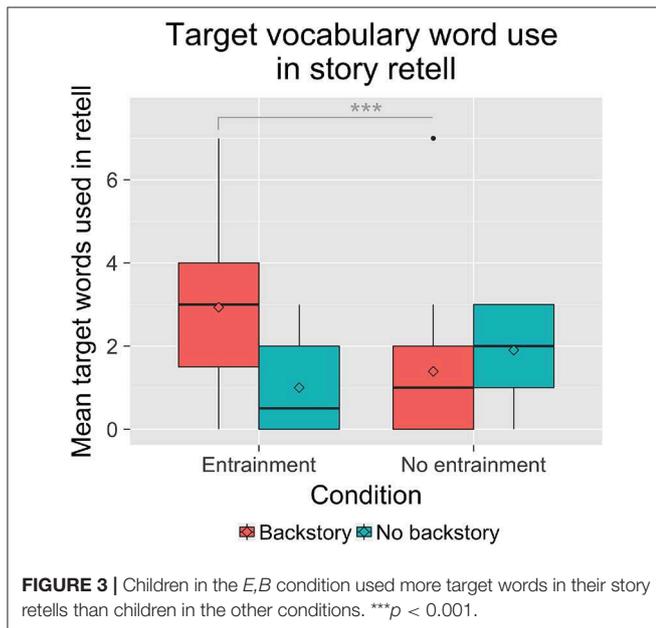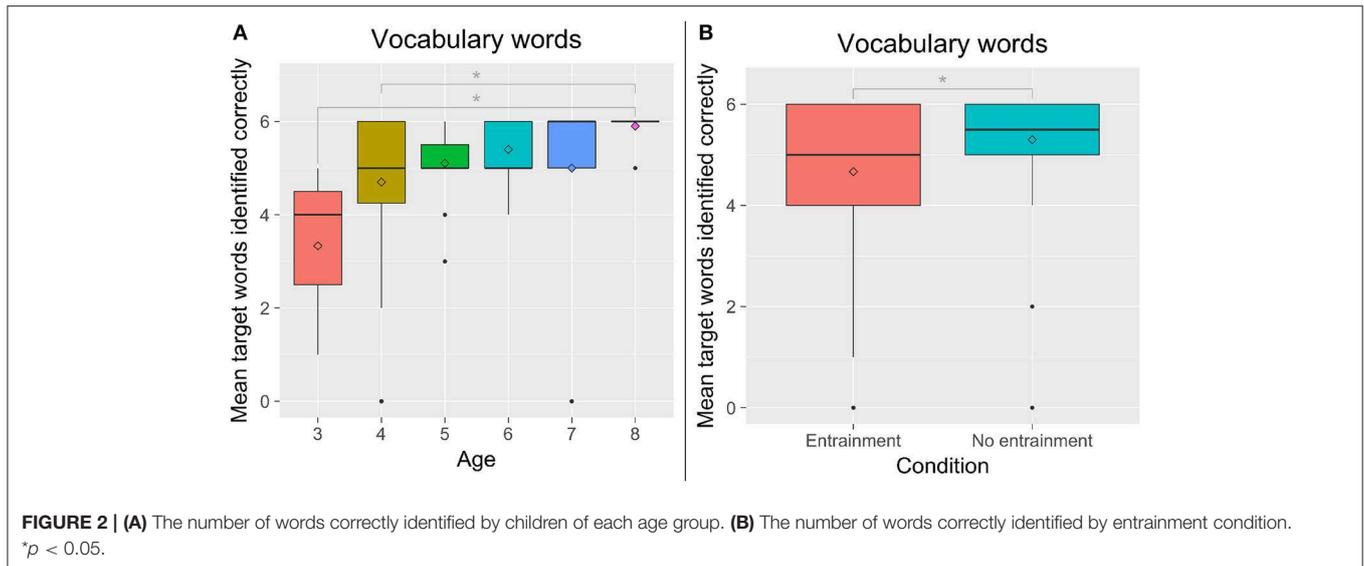
4. RESULTS

Our results are divided below into two parts, each reflecting one of our hypothesis areas: (1) *Learning*: We asked whether the robot's entrainment and backstory would increase children's learning with the robot and emulation of the robot's story; and (2) *Rapport, relationship, and social behavior:* We asked whether children would show greater rapport, acceptance, positive emotion, engagement, and closeness to the robot as a result of its entrainment and backstory.

4.1. Learning (H1, H2, H3)
For all learning-related analyses of variance, we included Age as a covariate because we expected that children's age would be related to their language ability and thus to their vocabulary scores and the complexity and/or length of their stories.

4.1.1. Target Vocabulary Word Identification (H1)
We performed 2×2 between-subjects analyses of variance with Entrainment (*E* vs. *NE*) and Backstory (*B* vs. *NB*) with Age as a covariate. We found a significant effect of Age on the total vocabulary words identified correctly, $F_{(5, 77)} = 2.76$, $p = 0.024$, $\eta_p^2 = 0.15$. Eight-year-olds correctly identified the most words, while 3-year-olds correctly identified the least (**Figure 2A**). We also found a significant effect of Entrainment on children's identification of the target words, $F_{(1, 77)} = 5.47$, $p = 0.022$, η_p^2

FIGURE 2 | (A) The number of words correctly identified by children of each age group. **(B)** The number of words correctly identified by entrainment condition. *$p < 0.05$.

FIGURE 3 | Children in the *E,B* condition used more target words in their story retells than children in the other conditions. ***$p < 0.001$.

$= 0.07$. Contrary to our hypotheses, children in the *NE* condition correctly identified more words than children in the *E* condition; however, in both conditions, there appeared to be a ceiling effect (**Figure 2B**). Older children were more likely to correctly identify words than younger children, $r_{s(85)} = 0.367$, $p < 0.001$.

4.1.2. Target Vocabulary Word Use (H2, H3)

A 2×2 between-subjects analyses of variance with Entrainment (*E* vs. *NE*) and Backstory (B vs. NB) with Age as a covariate revealed a significant interaction between Entrainment and Backstory regarding children's use of the target vocabulary words in the story, $F_{(1,59)} = 9.45$, $p = 0.003$, $\eta_p^2 = 0.14$. Children in the *E,B* condition used significantly more of the target words than children in all three other conditions (**Figure 3**).
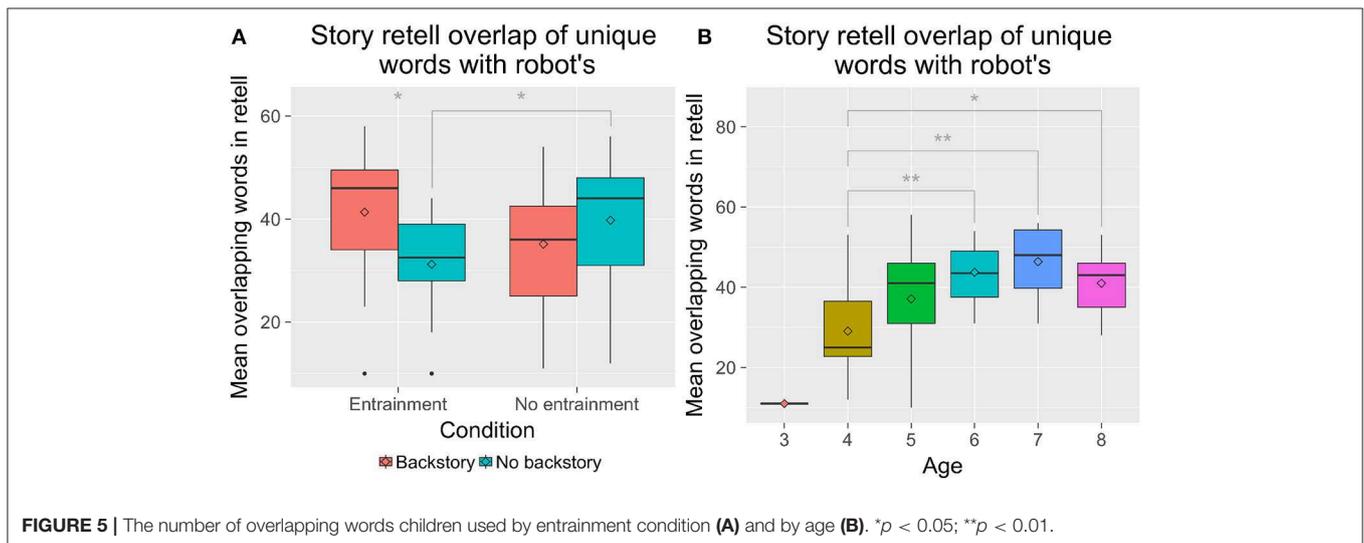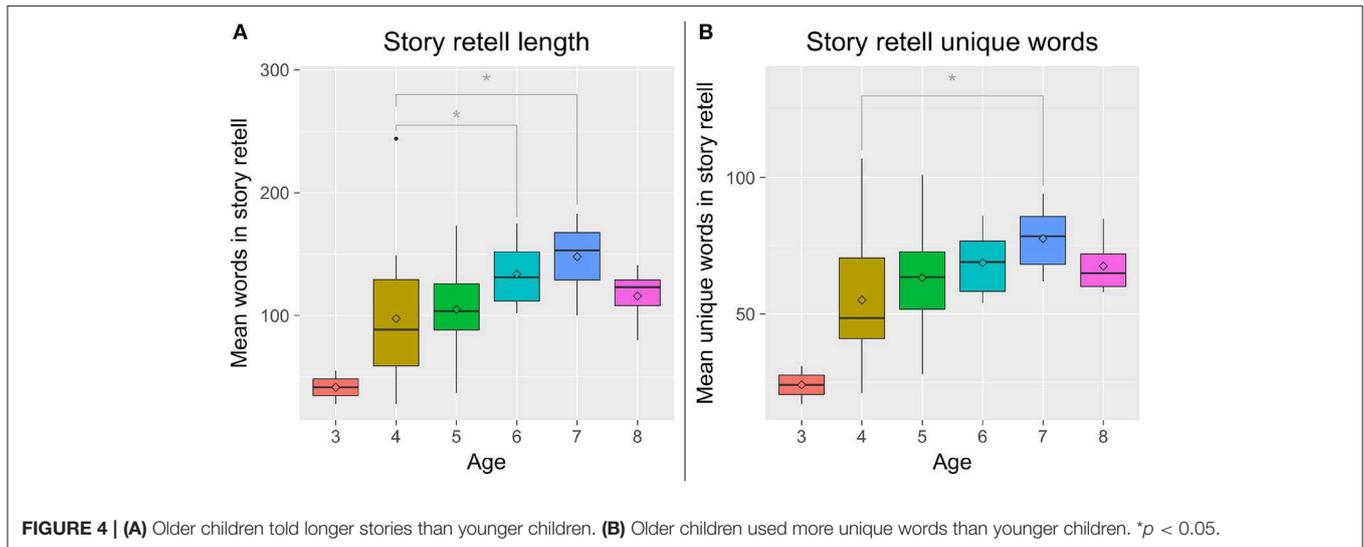
Overall, we saw no correlation between children's recognition of words on the vocabulary test and their subsequent use of those words in their retells, $r_{s(67)} = 0.047$. However, there were trends showing that this did vary by condition, though none of the correlations were significant. If the robot entrained, children were more likely to use the words themselves if they had identified the words correct on the test, *E-B* $r_{s(15)} = 0.253$; *E-NB* $r_{s(10)} = 0.254$; children who did not receive entrainment were less likely to do so, *NE-B* $r_{s(23)} = -0.077$; *NE-NB* $r_{s(21)} = 0.024$.

In summary, given that children's scores on the vocabulary identification test were not significantly different by condition, these results suggest that the robot's entrainment and backstory did not impact children's initial encoding of the words, but did affect children's expressive use of the words in their retelling.

4.1.3. Story Length (H3)

The robot's story was 435 words long, including the dialogic questions. The mean length of children's retells was 304 words ($SD = 110.9$). After stopword removal, the robot's story was 185 words, of which 99 were unique, non-overlapping words. The mean length of children's stories after stopword removal was 113 ($SD = 41.7$), with a mean of 63.1 unique words ($SD = 19.0$).

We performed 2×2 between-subjects analyses of variance with Entrainment (*E* vs. *NE*) and Backstory (*B* vs. *NB*) with Age as a covariate, which revealed a significant effect of Age on the length of children's stories after stopword removal, $F_{(4, 59)} = 3.77$, $p = 0.008$, $\eta_p^2 = 0.20$, and on the number of unique words children used, $F_{(4, 59)} = 3.19$, $p = 0.019$, $\eta_p^2 = 0.17$. *Post-hoc* tests revealed that 6- and 7-year-old children told longer stories than 4-year-old children, and 7-year-old children used more unique words than 4-year-old children (**Figures 4A,B**). The length of children's stories before stopword removal followed the same pattern, but was not statistically significant. This suggests that the primary difference between older (6–7 years) and younger (4–5 years) children's stories was their use of significant content words vs. stopwords.

FIGURE 4 | (A) Older children told longer stories than younger children. (B) Older children used more unique words than younger children. *$p < 0.05$.

FIGURE 5 | The number of overlapping words children used by entrainment condition (A) and by age (B). *$p < 0.05$; **$p < 0.01$.

4.1.4. Mirroring the Robot's Story (H2, H3)

Children used a mean of 37.7 unique words ($SD = 12.3$) in their retells of the 99 unique words that the robot had used in its story. A 2×2 between-subjects analyses of variance with Entrainment (E vs. NE) and Backstory (B vs. NB) with Age as a covariate revealed that the number of overlapping unique words used was significantly different by Age, $F_{(4, 60)} = 6.12$, $p < 0.001$, $\eta_p^2 = 0.29$. We also observed a significant interaction of Entrainment with Backstory, $F_{(1, 60)} = 6.42$, $p = 0.013$, $\eta_p^2 = 0.10$. Post-hoc tests showed that older children overlapped more than younger children (**Figure 5A**). Children in the E-NB condition ($M = 31.2$, $SD = 10.9$) overlapped less than children in the E-B and NE-NB conditions (E-B: $M = 41.3$, $SD = 13.2$; NE-B: $M = 36.2$, $SD = 10.6$; NE-NB: $M = 39.8$, $SD = 13.3$) (**Figure 5B**).

Children's stories received mean scores of 41.3 ($SD = 36.2$) for their use of exact and similar phrases that mirrored the robot's phrases. However, we observed no significant differences

between conditions in children's use of exact and similar matching phrases.

4.2. Rapport, Relationship, and Social Behavior (H4, H5, H6, H7)

4.2.1. Acceptance of the Robot (H4)

We performed 2×2 between-subjects analyses of variance with Entrainment (E vs. NE) and Backstory (B vs. NB) for the questions asked about children's social acceptance of the robot and of other children. We found a significant main effect of Backstory of children's responses to the question "Would you like to be good friends with a robot who can't hear well," $F_{(1, 82)} = 7.55$, $p = 0.007$, $\eta_p^2 = 0.08$. Children who heard the robot's backstory were more likely to respond positively than children who did not hear the robot's backstory. Children who heard the backstory were also somewhat more likely to respond positively to the question, "Would you like to be good friends with a
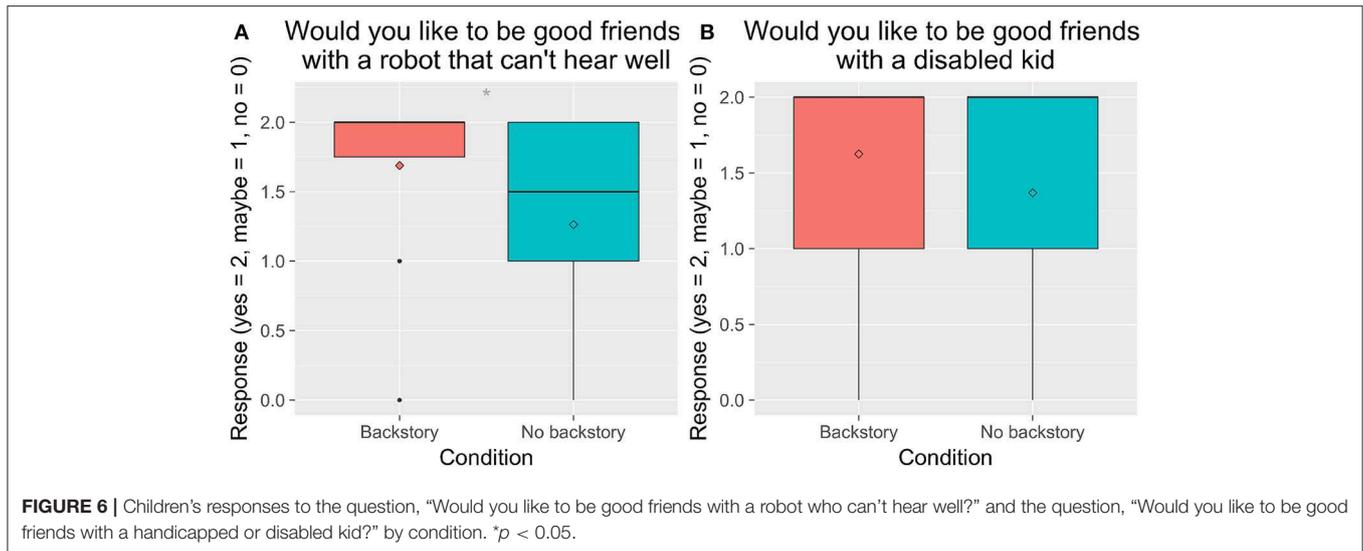
FIGURE 6 | Children's responses to the question, "Would you like to be good friends with a robot who can't hear well?" and the question, "Would you like to be good friends with a handicapped or disabled kid?" by condition. *p < 0.05.

TABLE 2 | Analysis of facial expressions during the interaction by condition.

Expression	Overall	E-B	E-NB	NE-B	NE-NB
Engagement	30.8 (11.7)	33.3 (13.3)	30.5 (12.0)	29.6 (11.2)	30.5 (11.4)
Attention	68.9 (13.4)	62.2 (21.1)	67.8 (15.2)	71.9 (5.56)	72.0 (9.51)
Valence	−0.738 (9.11)	3.51 (8.81)	5.75 (13.72)	−4.13 (5.20)	−2.72 (8.47)
Joy	7.13 (8.04)	9.13 (8.81)	12.1 (12.5)	5.48 (5.02)	5.61 (7.26)
Smiles	8.98 (8.82)	10.9 (9.35)	14.6 (13.4)	7.16 (5.65)	7.52 (8.31)
Laughter	0.13 (0.22)	0.23 (0.31)	0.28 (0.36)	0.08 (0.09)	0.07 (0.11)
Relaxation	3.53 (5.31)	4.13 (5.38)	6.63 (9.61)	2.49 (2.42)	3.06 (5.03)
Surprise	7.21 (6.96)	8.47 (9.22)	4.53 (4.63)	7.40 (5.32)	7.43 (7.84)
Disappointment	4.98 (3.98)	2.58 (2.01)	3.58 (3.03)	6.58 (4.37)	5.72 (4.05)
Fear	1.48 (2.06)	1.00 (1.40)	0.38 (0.66)	1.87 (2.04)	1.93 (2.72)
Concentration	2.92 (2.48)	2.02 (1.79)	2.11 (1.87)	3.20 (2.45)	3.72 (3.03)
Sadness	0.27 (0.46)	0.22 (0.34)	0.49 (0.54)	0.32 (0.59)	0.17 (0.24)

Values can range from 0 (no expression present) to 100 (expression fully present), except Valence, which can range from −100 to 100. Each column lists mean and standard deviation.

handicapped or disabled kid," though it was not statistically significant (**Figure 6**).

4.2.2. Children's Expressivity and Positive Emotion (H5, H6)

Overall, children were highly attentive and engaged, and displayed surprise and other emotions during the story (see **Table 2**). To evaluate whether children showed greater engagement or positive emotion with the robot that entrained, we performed 2×2 between-subjects analyses of variance with Entrainment (E vs. NE) and Backstory (B vs. NB).

We found a significant main effect of Entrainment on children's expressions of joy, $F_{(1, 69)} = 6.25, p = 0.015, \eta_p^2 = 0.070$; fear, $F_{(1, 69)} = 5.31, p = 0.024, \eta_p^2 = 0.074$; concentration, $F_{(1, 69)} = 5.09, p = 0.027, \eta_p^2 = 0.074$; disappointment, $F_{(1, 69)} = 12.7, p < 0.001, \eta_p^2 = 0.17$; attention, $F_{(1, 69)} = 5.66, p = 0.02, \eta_p^2 = 0.091$; laughter, $F_{(1, 69)} = 12.02, p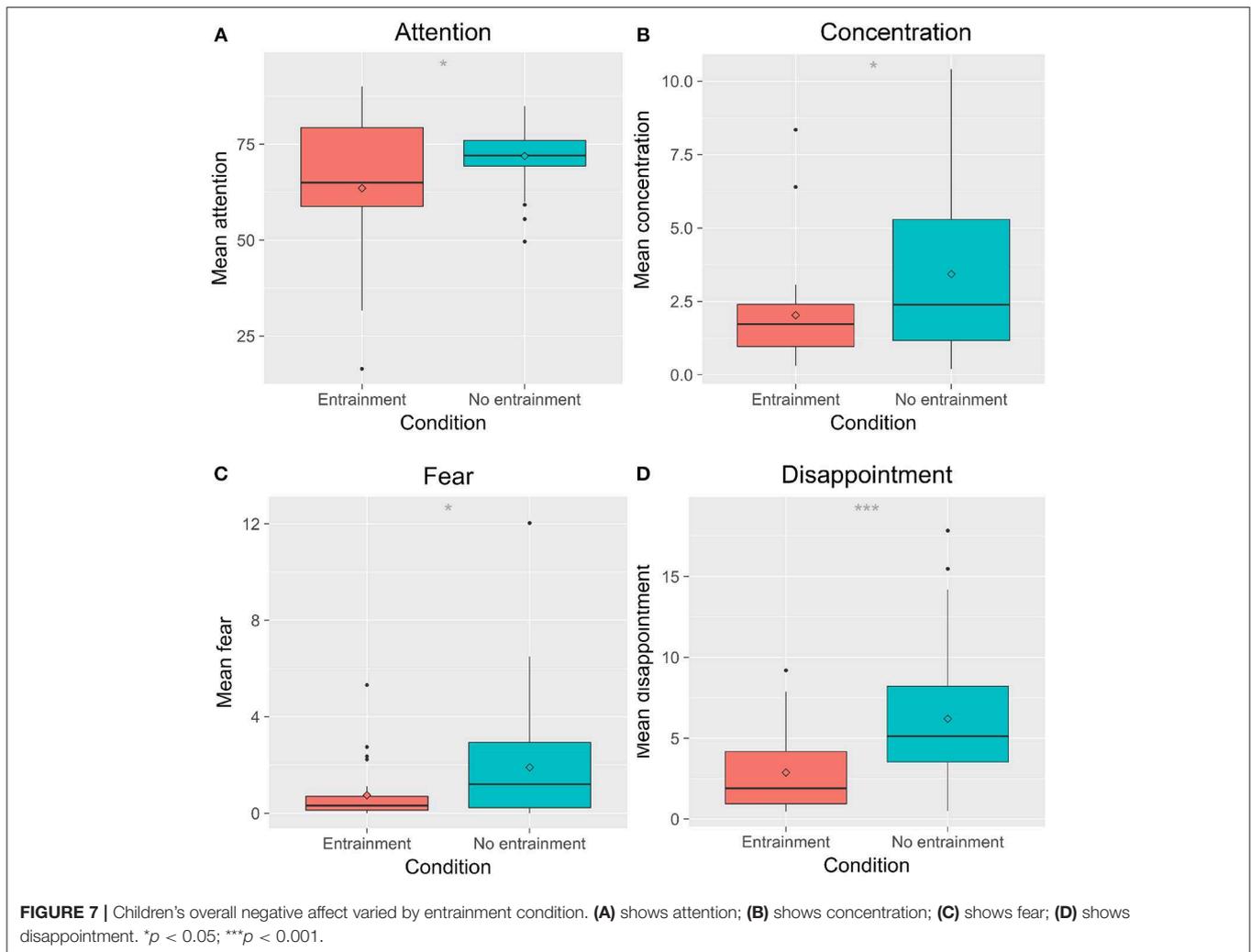 < 0.001, \eta_p^2 = 0.13$; smiles, $F_{(1, 69)} =$ 5.82, $p = 0.019, \eta_p^2 = 0.064$; and valence, $F_{(1, 69)} = 14.7, p = <$ 0.001, $\eta_p^2 = 0.16$. *Post-hoc* tests showed that children expressed less fear, concentration, disappointment, and attention in the E condition than in the NE condition (**Figure 7**). Children showed higher mean joy, laughter, valence (i.e., showed more affect with a positive valence), and more smiles in the E condition than in the NE condition (**Figure 8**). There were no significant differences in sadness, surprise, relaxation, or engagement; however, there was a trend for children in the E condition to show more relaxation than in the NE condition, which could have contributed to the higher valence seen in the E condition.

Next, we asked whether children's affect changed during the session. We split the affect data into the first half of the session and the second half of the session, using the data timestamps to determine the halfway point. We ran a 2×2×2 mixed ANOVA with time (within: first half vs. second half) × Entrainment (between: E vs. NE) × Backstory (between: B vs. NB). Although we hypothesized several changes in children's affect over time as a result of condition, we corrected for multiple comparisons here and only considered results significant when $p < 0.004$.

Like before, we found a significant main effect of Entrainment on disappointment, $F_{(1, 70)} = 14.7, p < 0.001$; laughter, $F_{(1, 70)} =$ 8.94, $p = 0.004$; and valence, $F_{(1, 70)} = 14.6, p < 0.001$. There were trends for a main effect of Entrainment on joy, $F_{(1, 70)} = 4.25, p =$ 0.043; fear, $F_{(1, 70)} = 5.88, p = 0.018$; attention, $F_{(1, 70)} = 4.37, p$ = 0.040; and smiles, $F_{(1, 70)} = 3.99, p = 0.0497$. Children showed fewer expressions of fear and disappointment in the E than in the NE condition (**Figure 9**). Children showed more joy, more smiles, and higher valence in the E than the NE condition.

We found a significant main effect of time on joy, $F_{(1, 67)} =$ 34.6, $p < 0.001$; valence, $F_{(1, 67)} = 17.7, p < 0.001$; engagement, $F_{(1, 67)} = 10.3, p = 0.002$; smiles, $F_{(1, 67)} = 40.5, p < 0.001$; relaxation, $F_{(1, 67)} = 27.2, p < 0.001$; laughter, $F_{(1, 67)} = 11.9, p$ = 0.001. All of these decreased from the first half to the second half of the session.

We saw trends for interactions of Entrainment with time: concentration, $F_{(1, 67)} = 6.79, p = 0.011$; attention, $F_{(1, 67)} =$

FIGURE 7 | Children's overall negative affect varied by entrainment condition. **(A)** shows attention; **(B)** shows concentration; **(C)** shows fear; **(D)** shows disappointment. *$p < 0.05$; ***$p < 0.001$.

5.47, $p = 0.022$; and laughter, $F_{(1, 67)} = 7.82$, $p = 0.007$. Children showed more concentration during the first half in the *NE* than in the *E* condition. Children showed more attention during the first half for *NE* vs. *E*, but they did not differ during the second half. Children laughed more in the first half in the *E* condition than in the *NE* condition, and decreased to the second half, while in the *NE* condition the amount of laughter did not change over time.

We also saw trends for interactions of time with Backstory for fear, $F_{(1, 67)} = 8.55$, $p = 0.005$; sadness, $F_{(1, 67)} = 7.01$, $p = 0.010$; disappointment, $F_{(1, 67)} = 7.70$, $p = 0.007$; attention, $F_{(1, 67)} = 4.88$, $p = 0.031$; and valence, $F_{(1, 67)} = 8.12$, $p = 0.006$ (**Figure 10**). Children expressed less fear in the second half of the session when they did not hear the backstory, but expressed somewhat more fear in the second half if they had heard the backstory. They expressed less sadness in the second half in *NB* condition, but did not change in *B* condition. Children's expressions of disappointment increased slightly in the *B* condition from first to second half, but not for the *NB* condition. Children's attention was higher initially in the *NB* condition and decreased slightly, while children's attention started lower in the *B* condition and
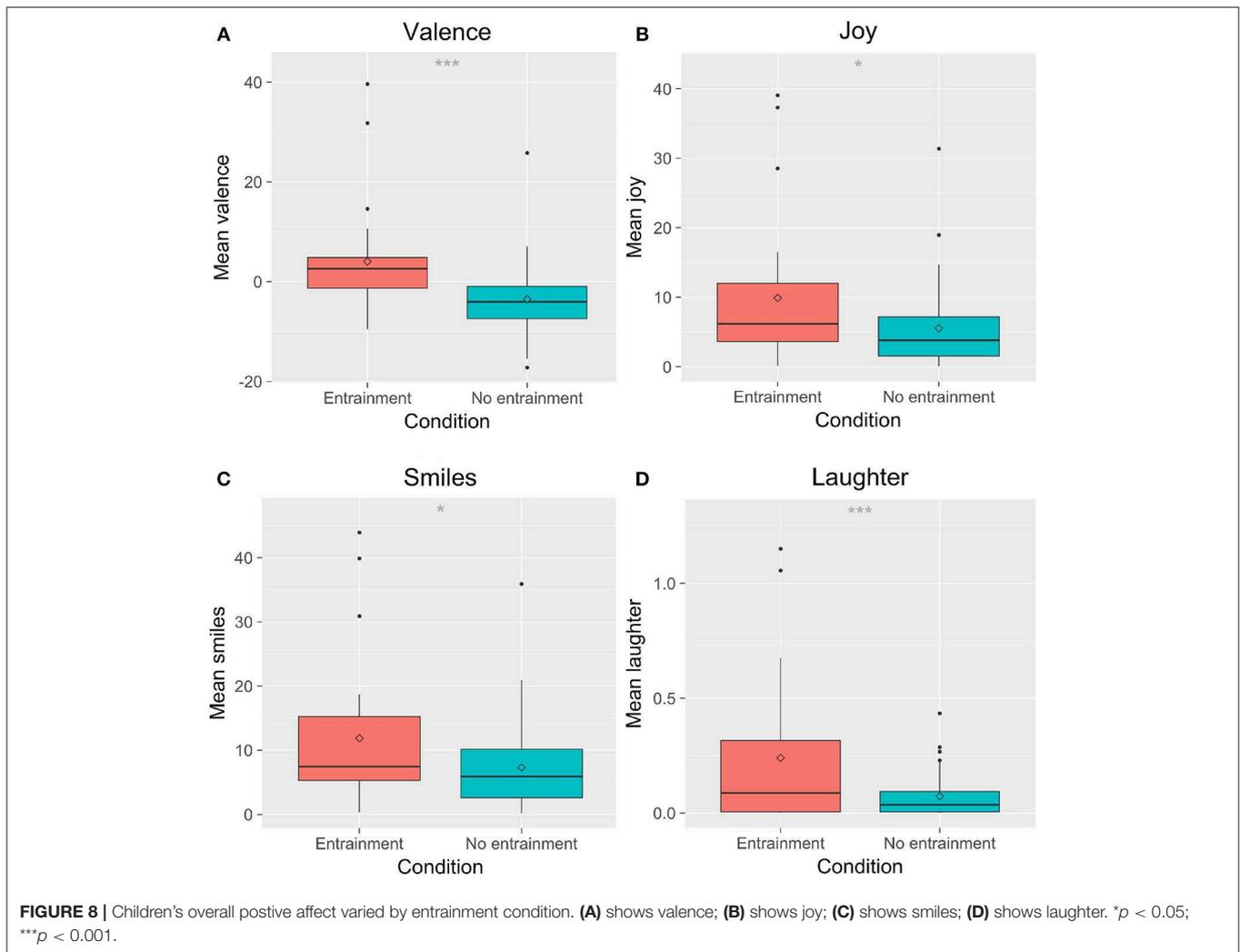
increased slightly. Children showed decreased valence in the *B* condition from first half to second half, but not in the *NB* condition.

4.2.3. Closeness to the Robot (H5, H6)

We performed a 2×2×5 mixed ANOVA with Entrainment (*E* vs. *NE*) × Backstory (*B* vs. *NB*) × IOS agent (within: Friend, Parent, Tega, Pet/Toy, Bad guy). We found a significant effect of agent, $F_{(4, 302)} = 61.9$, $p < 0.001$. *Post-hoc* Tukey's HSD tests showed that the bad guy was rated significantly lower than all other agents. In addition, the robot was rated significantly lower than the friend, but was not significantly different from the parent or pet/toy (**Figure 11A**). Older children were more likely to rate Tega as closer, $r_{s(86)} = 0.410$, $p < 0.001$ (**Figure 13A**).

Regarding the Picture Sorting Task, overall, Tega was placed at a mean position of 4.78 ($SD = 1.80$) (**Figure 11B**). **Figure 12A** shows results by condition for Tega's distance to the human, and **Figure 12B** shows the relative distance of each entity from the Tega robot by condition.

We performed a mixed ANOVA with Entrainment (between: *E* vs. *NE*) × Backstory (between: *B* vs. *NB*) × Entity (within:

FIGURE 8 | Children's overall positive affect varied by entrainment condition. **(A)** shows valence; **(B)** shows joy; **(C)** shows smiles; **(D)** shows laughter. *$p < 0.05$; ***$p < 0.001$.

Tega robot, baby, cat, frog, teddy bear, movie robot, robot arm, computer) for the entity positions, as well as for the entity positions relative to the Tega robot. For entity positions, we observed a significant main effect of Entity, $F_{(7, 574)} = 71.7$, $p < 0.001$. We also observed a significant interaction of Entity with Entrainment, $F_{(7, 574)} = 2.15$, $p = 0.037$; and a significant interaction of Entity with Backstory, $F_{(7, 574)} = 2.35$, $p = 0.022$.

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 for the Tega robot in the *E* condition, and closer to the human than all entities except Tega and the frog in the *NB* condition. In both the *NE* and *B* conditions, the cat was not placed significantly differently from Tega, the frog, movie robot, or teddy bear.

In the *E* condition, the Tega robot was significantly closer to the human adult than the robot arm, computer, movie robot, and teddy bear. It was farther from the human adult than the baby and was not placed in a significantly different position from the cat or frog. In the *NE* condition, Tega was only placed significantly closer to the human adult than the robot arm and computer; it

was not placed significantly differently from the cat, frog, movie robot, or teddy bear. Tega was not placed in a significantly different position from the movie robot in the *B* condition, but was placed significantly farther from it (closer to the human) in the *NB* condition.

The frog was placed significantly closer to the human adult than the robot arm and computer, and significantly farther from the human adult than the baby, but otherwise its position did not differ significantly from any other entities, except in the *NB* condition, where it was placed closer than the movie robot.

In the *NE* condition, the robot arm was placed closer to the table than the frog and movie robot, but in the *E* condition, the robot arm was not placed significantly differently from the frog or movie robot. By Backstory, children in the *B* condition placed the robot arm closer to the table than all other entities except the computer and teddy bear, while in the *NB* condition the robot arm's position was also not signficantly different from the movie robot's. Finally, in the *NE* and *B* conditions, the computer was placed closer to the table than all entities except the robot arm, while in the *E* and *NB* conditions, the computer was also not significantly different from the movie robot.
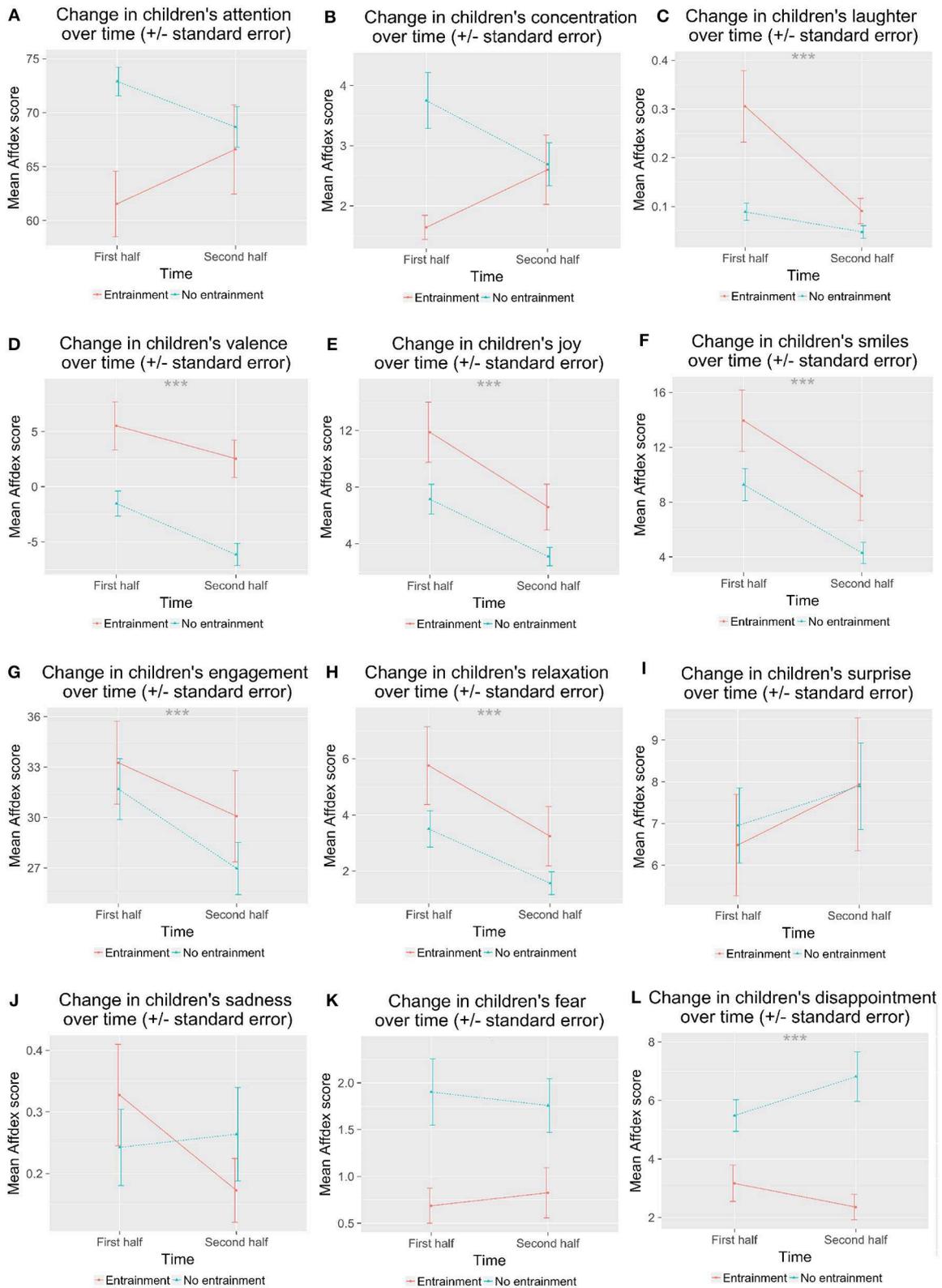
FIGURE 9 | Children's affect during the first half and the second half of the interaction varied by entrainment condition. **(A)** shows attention; **(B)** shows concentration; **(C)** shows laughter; **(D)** shows valence; **(E)** shows joy; **(F)** shows smiles; **(G)** shows engagement; **(H)** shows relaxation; **(I)** shows surprise; **(J)** shows sadness; **(K)** shows fear; **(L)** shows disappointment. ***$p < 0.001$.
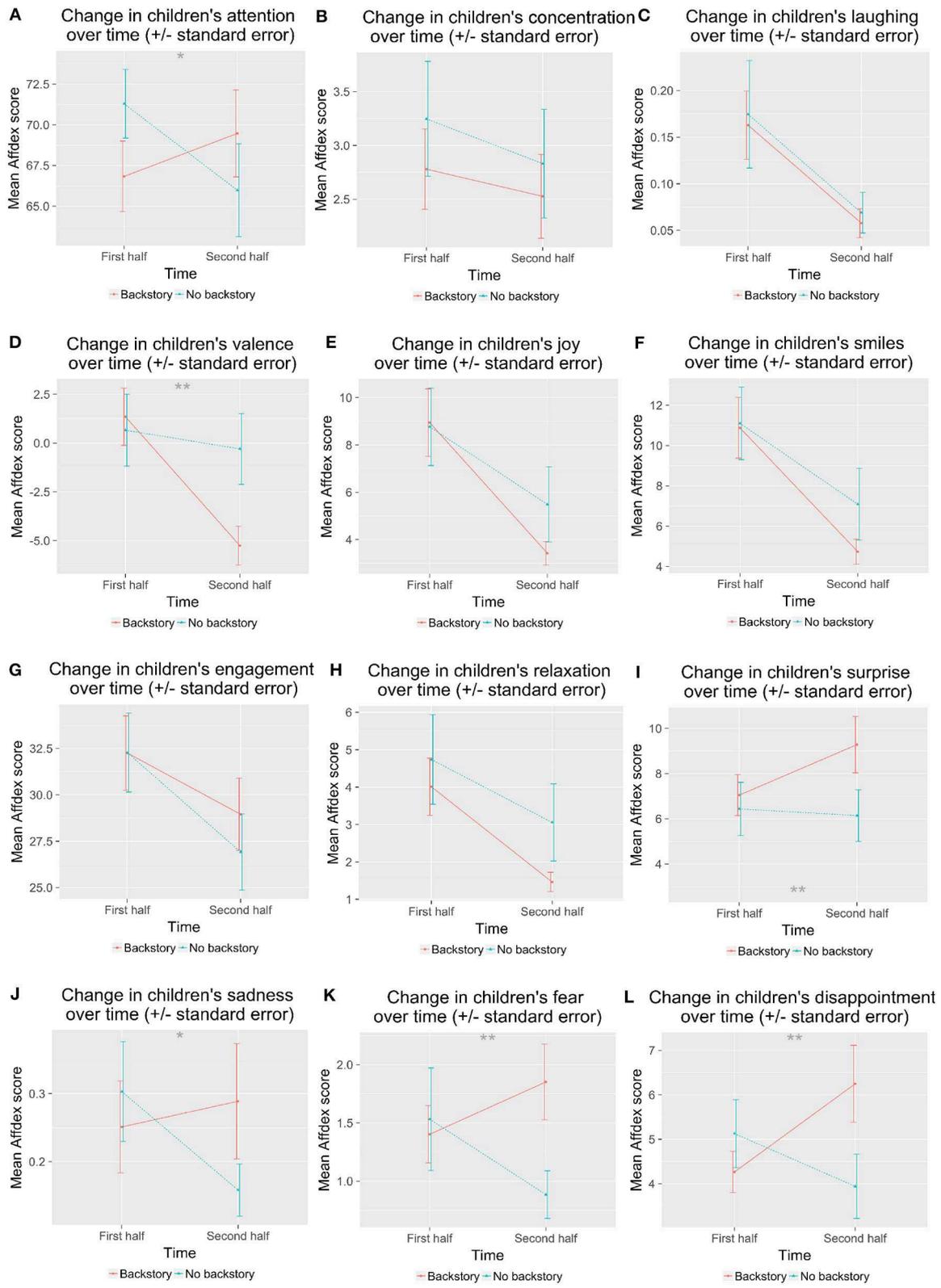
FIGURE 10 | Children's affect during the first half and the second half of the interaction varied by backstory. **(A)** shows attention; **(B)** shows concentration; **(C)** shows laughter; **(D)** shows valence; **(E)** shows joy; **(F)** shows smiles; **(G)** shows engagement; **(H)** shows relaxation; **(I)** shows surprise; **(J)** shows sadness; **(K)** shows fear; **(L)** shows disappointment. *$p < 0.05$; **$p < 0.01$.
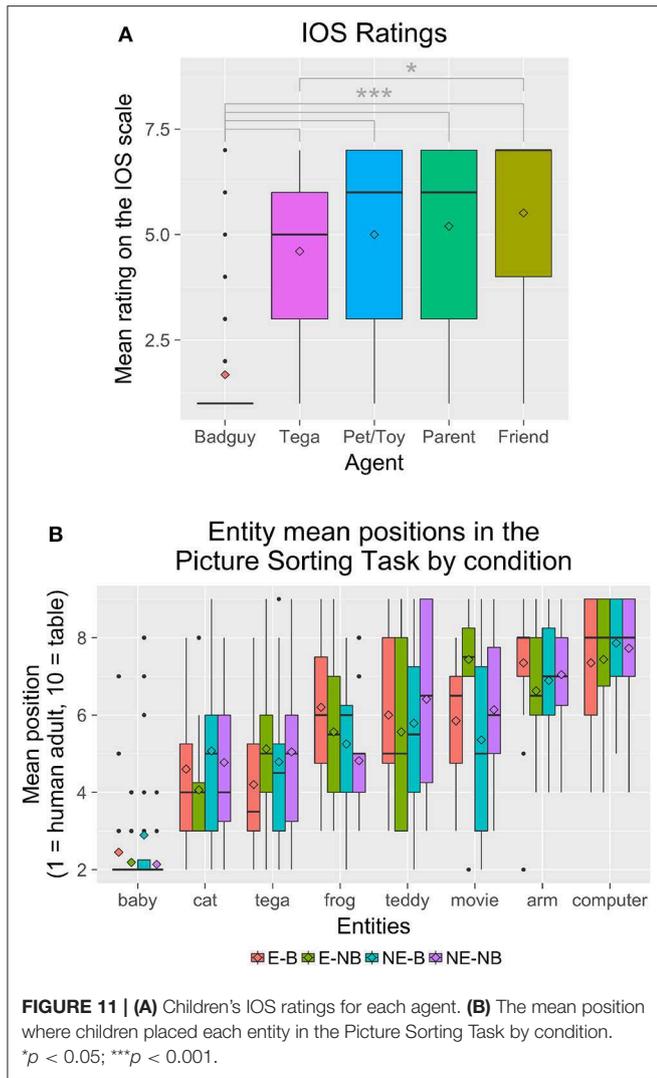
FIGURE 11 | (A) Children's IOS ratings for each agent. **(B)** The mean position where children placed each entity in the Picture Sorting Task by condition. *$p < 0.05$; ***$p < 0.001$.

FIGURE 12 | (A) Tega's mean distance from the human adult in the Picture Sorting Task by condition. **(B)** The distance of each entity from the Tega robot in the Picture Sorting Task by condition. There were trends for the Tega robot to be placed closer to the baby in the *B* condition than in the *NB* condition, closer to the movie robot in the *E* condition than in the *NE* condition, and closer to the frog in the *E-B* condition than in the other conditions.

Regarding the distance of each entity relative to the Tega robot, we observed a significant main effect of Entity, $F_{(6, 492)} = 71.8$, $p < 0.001$. We also observed a significant interaction of Entity with Entrainment, $F_{(6, 492)} = 2.13$, $p = 0.049$; and a trend toward an interaction of Entity with Backstory, $F_{(6, 492)} = 2.11$, $p = 0.051$. *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. There was a trend for children to place the Tega robot closer to the baby (and the baby closer to the human adult than Tega) in the *B* condition (mean difference = 1.83, *SD* = 2.55) than in the *NB* condition (*M* = 2.92, *SD* = 2.01).

The cat was placed closer to Tega than most other entities. It was not placed significantly differently than the teddy bear in the *E* condition; from the frog, movie robot, or teddy bear in the *NE* and *B* conditions; and from the frog in the *NB* condition.

The computer was placed farther from Tega than all entities except the robot arm and, in the *E* and *NB* conditions, the movie robot. The robot arm, in turn, was placed farther from Tega than all entities except the computer and teddy bear. In the *NB* and *NE* conditions, the robot arm was also not different than the movie
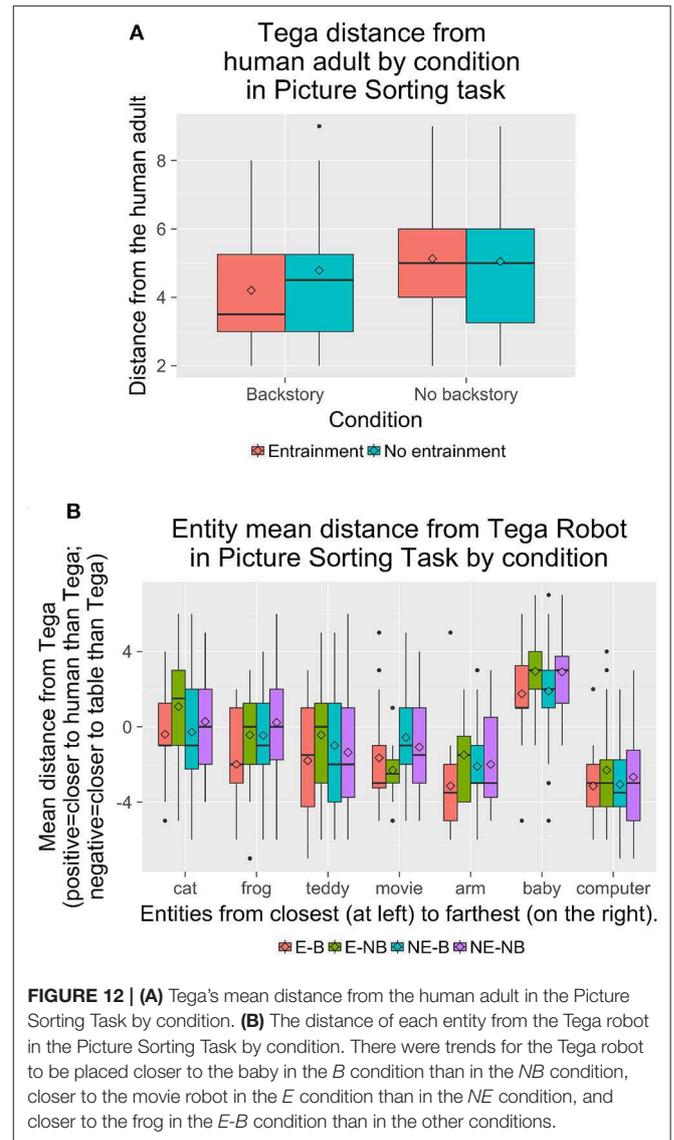
robot; and in the *E* condition, the robot arm was also not different from the movie robot or frog. There was a trend for children to place Tega farther from the movie robot, and closer to the human than the movie robot was, in the *E* condition (*M* = −1.94, *SD* = 2.40) than in the *NE* condition (*M* = −0.80, *SD* = 2.69).

Finally, we also observed trends for Tega to be placed farther from the frog, and also closer to the human adult than the frog was, in the *E* (*E*: *M* = −1.31, *SD* = 2.77, *NE*: *M* = −0.16, *SD* = 2.62) and *B* conditions (*B*: *M* = −1.11, *SD* = 2.76, *NB*: *M* = −0.05, *SD* = 2.60).

We observed no significant differences between conditions regarding whether children were more likely to agree to do the fourth picture with the robot, give the robot their sticker in the sticker task, or give the robot a bigger goodbye gift (in terms of how meaningful the robot might think it to be). About half the children in each condition chose to do the fourth picture; we did not see any effects of the number of picture conversations (i.e., the three required vs. the optional fourth one) on the results. If

we looked at children's likelihood to perform all three activities (adding up the fourth picture, the sticker, and the goodbye gift, rather than any one individually), we saw a trend for children in the *E-B* condition to be slightly more likely to do all three activities, though this was not statistically significant.

4.2.4. Children's Mirroring, Learning, and Relationship (H7)

We found that children who gave Tega a closer score on the IOS task were also more likely to use the target words in their stories, $r_{s(67)} = 0.359$, $p = 0.003$ (**Figure 13C**). They were also more likely to emulate the robot's stories as reflected by the number of exact and similar phrases used in their retells, $r_{s(67)} = 0.273$, $p = 0.025$ (**Figure 13B**). Given that age also correlated with children's ratings of Tega on the IOS task, we might suspect that age is more relevant than how close children felt to the robot. However, age did not correlate with children's use of exact and similar phrases, which suggests a deeper story.

In addition, children who placed Tega closer to the human in the Picture Sorting Task were also more likely to use phrases similar to the robot's, $r_{s(67)} = -0.299$, $p = 0.014$ (**Figure 13D**). There was a trend for children who placed Tega closer to the human to also rate Tega more closely on the IOS task, $r_{s(86)} = -0.197$, $p = 0.069$.

We did not observe any significant correlations of children's vocabulary scores with their phrase mirroring or any of the relationship assessments.

5. DISCUSSION

We asked whether a social robot that entrained its speech and behavior to individual children and provided an appropriate backstory about its abilities could increase children's rapport, positive relationship, acceptance, engagement, and learning with the robot. Below, we discuss the main findings and then discuss the implications of these findings.
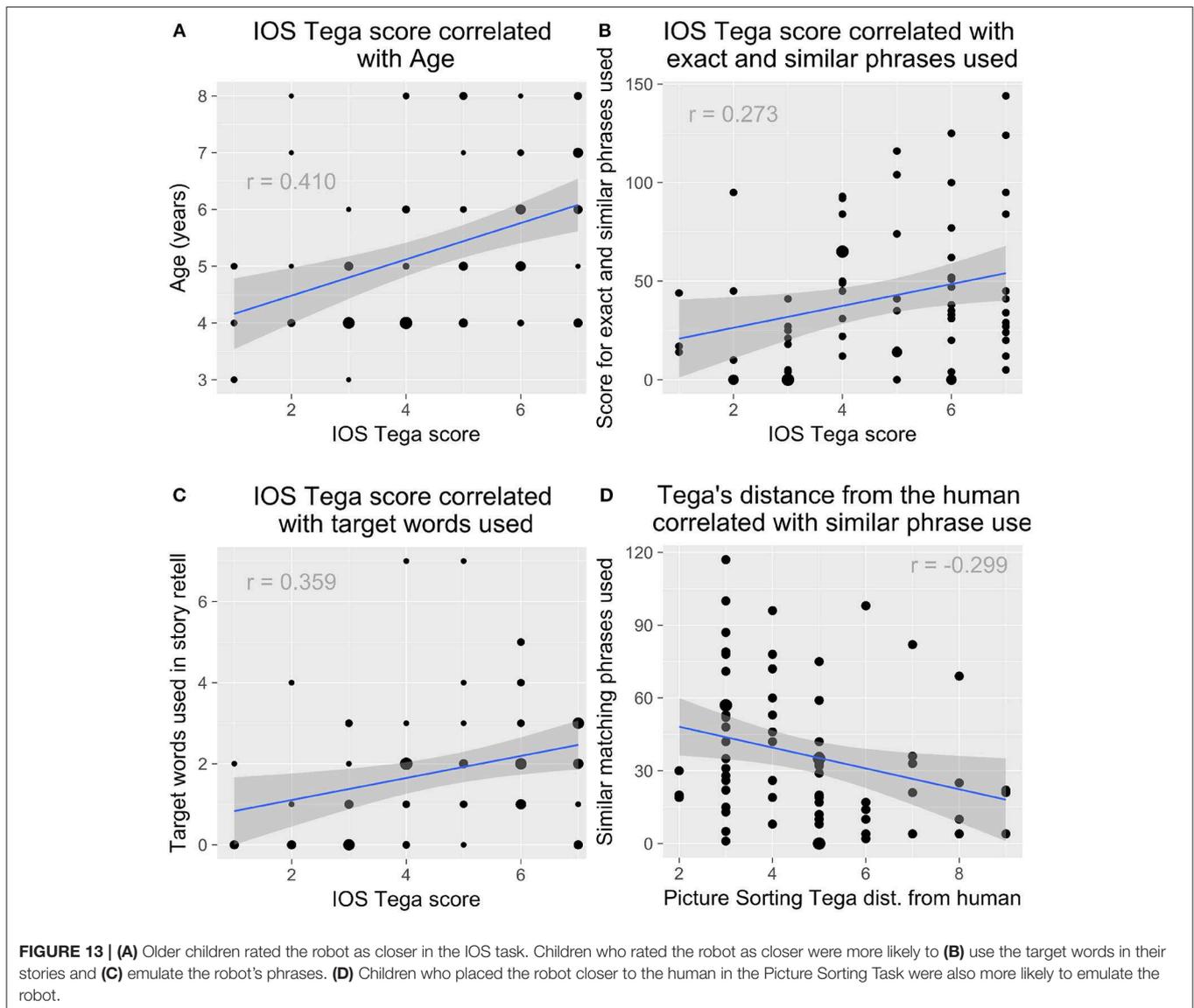
5.1. Learning

Children learned the target vocabulary words in the robot's story and were generally attentive and engaged with the robot regardless of the experimental condition. They showed a variety of emotional expressions throughout the interaction. Children remembered the robot's story as evidenced by their ability to retell the story and their identification of target words on the vocabulary test. These results are in line with the prior study using this story activity (Kory Westlund et al., 2017b), which found significant learning gains.

We did see differences in children's learning by condition. Contrary to our hypotheses (H1), children in the *No Entrainment* condition correctly identified more target words than children in the *Entrainment* condition (**Figure 2B**). This could be for several reasons. A prior study found that a robot tutor that employed social adaptive behaviors led to lower learning gains than a robot that did not act as socially (Kennedy et al., 2015). Thus, perhaps the entraining robot was perceived more socially, which was detrimental in learning. This is contrary to our hypotheses regarding the importance of social behavior, rapport, and relationship in language learning with peers. However, in the

prior study, children performed a math task with the robot tutor. The authors hypothesized that perhaps children were paying attention to the robot's social behavior as opposed to the lessons it was providing, or, alternatively, that the social behavior placed greater cognitive load on children thus inhibiting their ability to perform in the math task. Performance on a math task in a tutoring format may indeed benefit less from a robot's social behaviors than performance in a language-based story activity in a peer-learning format.

A second explanation pertains to the learning results we observed. There was a ceiling effect and little variance in children's responses, with 43% of children correctly identifying all six target words, and 41% correctly identifying 5 of the target words. If a significant number of children were already familiar with the target words, then the vocabulary tests would not reflect their learning during the task with the robot; the difference between conditions may not reflect children's learning in the task. Furthermore, given that children's receptive language abilities may precede their expressive abilities (Bloom, 1974; Ingram, 1974; Sénéchal, 1997), we would expect that children who correctly identified more words to also use more of them in their stories (H2), reflecting greater understanding and deeper encoding of the words (this was also seen in the prior study, Kory Westlund et al., 2017b). However, we did not see this correlation: children's use of the target words was not significantly correlated with correct identification of the words. In fact, children's use of the target words was significantly greater in the *E-B* condition than all others, in line with our hypotheses (H3) (**Figure 3**). Additionally, while the patterns were not significant, children were moderately more likely to use the words if they had identified them correctly in the *Entrainment* condition than in the *No Entrainment* condition. These results suggest that the robot's rapport- and relationship-building behaviors affected either or both of (a) children's learning and deeper understanding of the words such that they were more able to expressively use the words, or (b) children's mirroring of the robot's speech such that they used more of these target words, both of which would be in line with prior work linking rapport to learning (Sinha and Cassell, 2015a,b). This was also a short-term encounter. Given the positive aspects we see here regarding word use and mirroring, we expect that over multiple sessions, we would see greater differences in word learning.

When we examined children's mirroring of the robot's speech, we saw that children did mirror the robot (H2, **Figures 3**, **5**), in line with past work suggesting that children may mirror adults' syntax and speech (Huttenlocher et al., 2004) and earlier work in human-computer interaction showing that adults will entrain to computers and robots (e.g., Pearson et al., 2006; Lubold et al., 2018). However, we saw no significant differences in children's emulation of the robot's phrases, and in fact, less overlap in the number of unique words used by children that mirrored the words the robot used in the *E-NB* condition, and little difference among the other conditions (contrary to H3). This suggests that perhaps entrainment did not affect children's mirroring of the words the robot used so much as their expressive ability to use the key words present in the story. Prior work has shown that social robots can be successful at prompting children to demonstrate expressive vocabulary skills in both

FIGURE 13 | (A) Older children rated the robot as closer in the IOS task. Children who rated the robot as closer were more likely to (B) use the target words in their stories and (C) emulate the robot's phrases. (D) Children who placed the robot closer to the human in the Picture Sorting Task were also more likely to emulate the robot.

vocabulary test and storytelling contexts (e.g., Kory and Breazeal, 2014; Kory Westlund et al., 2017b; Wallbridge et al., 2018). The present study suggests that the robot's entrainment may influence expressive ability.

The lack of difference in phrase mirroring was counter to our hypotheses (H3). Perhaps children did not feel sufficiently more rapport with the entraining robot for this to affect their storytelling. Indeed, in all conditions, the robot was a friendly, expressive character, which children generally said they felt close to—as close as to pet or parent, though less close than to a best friend. The entrainment only affected the robot's speech and some animations (which were played primarily in accompaniment with speech). In particular, if a child was very shy and rarely spoke, then the robot had fewer opportunities to adapt and entrain to that child. Perhaps greater difference would be seen if the robot also entrained other behaviors, such as posture, gesture, or word use. Another explanation is that perhaps language mirroring is not as closely linked to rapport as

we expected; there is limited research so far suggesting this link, and more is needed.

5.2. Rapport, Relationship, and Social Behavior

The robot's entrainment and backstory also affected children's displays of positive emotions during the interaction. All children were engaged, but children in the *E-B* condition showed more positive emotions (e.g., joy, laughter, smiles, and positive valence), as well as fewer negative emotions (e.g., disappointment, fear) (supporting H5 and H6; see **Figures 7–10**). Laughter and smiling are social behaviors (Provine, 2001; Smidl, 2006; Manson et al., 2013). We also saw trends for children to be more helpful and accommodating in the *E-B* condition, as one might expect with a more social agent (Reeves and Nass, 1996), as evidenced by their behavior with fourth picture, the sticker task, and the goodbye gift. This is evidence that the robot's entrainment and backstory improved children's enjoyment of the

interaction and may have perceived it as more of a social agent, perhaps a result of increased rapport (supporting H5 and H6).

Children in the *E-B* condition also showed fewer attentive expressions, though only during the first half of the interaction (they did not differ later on). This could mean that these children were in fact less attentive initially, or it could mean that they were showing more positive attentive expressions that were coded by the affect recognition software as engagement and joy. If they were less attentive, we might expect this to be reflected in their vocabulary scores and story retellings—perhaps this is why these children did not identify as many words correctly. However, children in the *E-B* condition showed just as many expressions of engagement as children in the other conditions, were just as likely to retell the story, and as noted earlier, there were few significant differences by condition in children's story retellings beyond *more* use of the target words by children in the *E-B* condition. An alternative explanation is that perhaps children's attentive looks were related to how much cognitive effort was involved in performing the task. The robot's entrainment and backstory could have improved rapport and made the interaction more fluent, easier, and smoother, thus requiring less intense attention by children. This would be especially apparent earlier in the interaction, immediately following the robot's backstory disclosure and during the picture conversation task, when the robot was entraining more frequently due to the increased number of conversational turns during that task.

Related to this, we saw that children's attention increased over time in the *B* condition, but decreased in the *NB* condition, while multiple negative emotions (fear, disappointment, sadness) were displayed more frequently over time in the *B* condition than in the *NB* condition. For all other affective states measured, the change over time was not significant, though there were patterns for decreases in positive affect (e.g., joy, smiles, etc.) over time for all children. If children's attentive expressions were related to cognitive effort, this could indicate that in the *B* condition, children felt that over time, they had to attend more carefully to the robot (putting in more effort) in order to help it and deal with its hearing limitations. This could, perhaps, have led to increased feelings of difficulty interacting with the robot over time, which could have led to the increased displays of negative emotions that we observed in the *B* condition.

Regarding the decrease in attention in the *NB* condition, it may be that these children became less attentive because they were growing bored or were not as invested in the interaction. Indeed, while not statistically significant, children's engagement did decrease slightly more over time in the *NB* condition than in the *B* condition. There were also no affective states for which children in the *NB* condition increased their expression over time, suggesting that they became less expressive overall, which may be indicative of boredom or less emotional investment in the interaction.

We observed that children showed greater acceptance of the robot when they had heard the robot's backstory, as we expected (H4; **Figure 6**). Children's increased negative affect seen in the *B* condition may also reflect increased sympathy for the robot. Regardless, it seems that the robot's story influenced children's perceptions of it, in line with prior work showing

that a robot's story does influence how people understand and react to it (Stenzel et al., 2012; Klapper et al., 2014; Darling et al., 2015; Kory Westlund et al., 2016b). Interestingly, this effect seemed to carry over to children's ideas about being friends with other children. While only a trend, it suggests room for future interventions using robots to help children understand and accept others different from themselves.

As noted above, children generally felt as close to the robot as they did to a pet, favorite toy, or parent, though not quite so close as to their best friend (**Figure 11A**). They generally placed Tega closer to the human adult than the table in the Picture Sorting Task, and frequently close to the human baby and to the cat (**Figures 11B**, **12**). These results present an intriguing picture regarding children's perceptions of the robot as a peer- or friend-like, non-human, animate entity. Children did not confuse the robot with a human; they knew it was different. Children seemed to clearly find companionship in the robot and to place it in a category between friend, pet, and authority figure. It was not merely a machine or computer; it was seen as more animate and alive—but not in the same category as a human. This jibes with prior work suggesting that children may categorize robots as in-between entities, with attributes of both living beings and mechanical artifacts (Kahn et al., 2002, 2012; Severson and Carlson, 2010). Perhaps children observed that some of the things that are messy about human relationships, such as the kinds of conflict that arise and the emotions that others display, are not the same in robot relationships—perhaps they are more like pet relationships. In this case, the robot did not get overly upset when it did not receive the sticker it wanted in the sticker task; it was generally cheerful throughout the interaction, which perhaps would not have been the case with another child. It is also likely that the robot's morphology influenced children's perceptions, since the robot we used was fluffy, colorful, and moved more like an animated character or sidekick than a humanoid being.

In support of our hypotheses regarding the connection between children's feelings of closeness, rapport, and relationship with learning and mirroring the robot (H7), we observed that children who rated the robot as closer to themselves also used the target words more often and emulated the robot's story more (**Figure 13**). This is in line with earlier work linking rapport to learning (Sinha and Cassell, 2015a,b). However, we also saw that age correlated with children's ratings of Tega on the IOS task. Older children rated the robot as closer; younger children as less closer. Perhaps younger children were less sure of the robot and needed more time to become comfortable with it. Given these correlations, we might suspect that age was more relevant to children's use of the target words and emulation of the robot's story than children's closeness ratings. However, children's age did not correlate with children's emulation of the robot's phrases at all, which suggests that this emulation was in fact related to children's feelings of closeness.

Finally, we also observed a few age differences. The length of children's story retellings differed with respect to their age, but did not vary by condition (**Figure 4**). Notably, the stories told by 6- and 7-year-old children were longest. The stories of 8-year-old children were not quite so long, which may have been because they were less interested in the story, rather than less capable. The

story and activity were designed with 4–7-year-olds in mind. The story may have been a little on the difficult side for the younger children, and on the easy side (and thus perhaps a little boring) for the oldest. However, even the children outside the target age range for the activity were receptive to the social robot, showing engagement, learning, and emulation.

Taken together, these results show that the robot's rapport and relationship-building behaviors do matter in interactions with young children. A robot that deliberately emulates a child's speech in a way similar to how people mirror each other can elicit more positive emotion and greater emulation of key words in a language learning activity. Children's feelings of closeness are related to their emulation of the robot's words in their stories.

5.3. Relation to Related Work

Our results also mirror, to an extent, the results in the prior study that explored a robot's use of expressive vs. flat speech (Kory Westlund et al., 2017b). In both studies, the robot's entrainment, backstory, and expressivity reflected the sensitivity the robot showed to the interaction. This sensitivity influenced children's engagement and learning. This is in line with work examining nonverbal behaviors in human-human learning interactions, in particular, nonverbal immediacy. *Nonverbal immediacy* refers to the *perceptual availability* of one's interaction partner, i.e., the use of nonverbal behaviors including gaze, gesture, posture, facial expressions, and vocal qualities such as prosody to signal general responsiveness and attentiveness. In human-human learning interactions, nonverbal immediacy has been linked to increased learning gains (Mehrabian, 1968; Christophel, 1990; Witt et al., 2004). When we examine prior child-robot interaction studies, we see that they have found a similar pattern of results to these human-human studies: The use of nonverbal immediacy behaviors including socially contingent behavior, appropriate gaze and posture, and vocal expressivity increased children's learning, engagement, and trust in a learning companion (Breazeal et al., 2016a; Kennedy et al., 2017; Kory Westlund et al., 2017a,b). Thus, it may be that the entrainment behaviors used by the robot increased its perceived immediacy and perceived sensitivity to the interaction.

However, in other work on language learning with social robots, the robot's social interactive capabilities have been found to influence children's relationships and social acceptance of the robot, but not their learning (e.g., Kanda et al., 2004, 2007, 2012). Indeed, some work has shown no significant differences in children's word learning from a social robot (with numerous embodied social capabilities) than from a tablet (e.g., Kory Westlund et al., 2015; Vogt et al., 2019). Arguably, these studies suggest a contrary story in which the robot's social capabilities may not affect children's learning that much.

These studies, however, have generally included learning tasks that did not require a robot or much social behavior for learning to proceed. For example, the second language learning activities used by Vogt et al. (2019) involved educational games presented on a tablet, for which the robot provided instructions, feedback, and support, but in which—as the authors acknowledge—the robot appeared to be non-critical for the learning interaction. The robot's social behavior may matter more for conversation and storytelling-based activities than for tablet

games or simpler word learning tasks. Thus, we suspect that the robot's social capabilities (such as nonverbal immediacy) can influence children's learning—as we have seen here and in multiple other studies discussed earlier—but that the influence of social behavior is moderated by other factors, such as the extent to which the robot's sociality is necessary for the learning activity to proceed smoothly (as in the case of conversation and storytelling-based activities), and the extent to which the robot's social behavior helps build rapport.

This hypothesis is supported by Lubold and colleagues' recent work with middle school children and adults, in which a social robot with vocal entrainment contributed to increased learning on math tasks, though not increases in self-reported rapport (Lubold et al., 2016, 2018; Lubold, 2017). Because the vocal entrainment served not only to match pitch and other vocal features, but also made the robot's text-to-speech voice much more expressive, these studies could not disentangle the effects of expressivity from entrainment—however, both expressivity and entrain increase the robot's social capabilities. Our results here are similar to Lubold et al.'s, in that we also found that the robot's vocal entrainment was related to learning, but unlike Lubold's work, we also found connections between the robot's entrainment and aspects of children's relationship and rapport, including increased positive emotion and language emulation. This difference could be for numerous reasons, including the different age groups studied, the different learning matter (math vs. language), and the additional social and expressive capabilities of our robot.

Our results also extend prior work showing that children learn through storytelling with peer-like robot companions in ways that are significantly different from how children learn and engage with other technologies. We are seeing a peer learning dynamic similar to that seen in child-child interactions. Children socially model and emulate the behavior of the robots, like they do with other children. For example, children are more emotionally expressive when the robot is more expressive (Spaulding et al., 2016), show more curiosity in response to a robot's increased curiosity (Gordon et al., 2015), teach new tasks to robot peers (Park and Howard, 2015), and emulate linguistic phrases and vocabulary (Kory Westlund et al., 2017b). This study extends these previous works to explore not only *whether* children will learn with and emulate a robot peer, but the *mechanisms* by which robots can influence peer learning. Rapport and relationship appear to be two such mechanisms.

5.4. Limitations

This study had several limitations. First, we did not control for children's individual differences, particularly with regards to learning ability, language ability, or socio-economic status, all of which may affect individual children's social interactions and learning with the robot. Furthermore, we did not obtain an equal number of children at each age group to participate in the study. Future work should examine a more homogeneous sample as well as explore the stability of results across individual differences and across ages as children grow older.

We also lacked complete story retelling data and affect data for all children. Some children did not retell the story and in

a few cases, we had issues regarding the audio quality of the recorded stories. Some children's faces were not recognized by the Affdex software, and a few videos were missing or insufficiently captured a full frontal view of the children's faces, which was necessary for affect recognition. As a result, the analyses reported are underpowered. Future work should take greater effort to obtain quality audio and video recordings for all children during the study.

As mentioned in Kory Westlund et al. (2017b), the target vocabulary words were uncommon, but some children still may have known them. In particular, older children may have been familiar with some of the words, given the correlation we observed between children's age and the number of words identified correctly. The words' uncommonness may have cued children to pay attention to them; as such, future work should consider using nonce words or include a vocabulary pretest. Including a vocabulary pretest would also help ensure that children's language abilites did not differ by condition.

The robot's automated entrainment was limited to its speaking rate and pitch, so if a child was very quiet or spoke rarely, the robot would not have been able to entrain to that child. Because volume and exuberance were teleoperated, these occurred for all children. Future work could explore ways of encouraging shy children to speak up, or explore other modalities for entrainment, such as posture, gesture, facial expressions, and word use.

It is also unclear how generalizable the results are to robots with different embodiments or morphologies. The Tega robot that we used appears much like a fluffy stuffed animal, and thus is morphology could be seen as more familiar to children than a robot such as the Aldebaran NAO, which is humanoid. Children may feel a different level of comfort or uncanniness with a humanoid robot than with the Tega robot.

Finally, this study explored only a single one-on-one interaction with the robot. As such, any overall effects could be related to the novelty of the robot. However, children had the same amount of exposure to the robot in all conditions, so novelty cannot explain the differences we observed between conditions regarding the effects of entrainment and backstory.

Because learning tends to happen over time, as does the development of relationships, future work should explore longitudinal interactions to help us better understand the relationship between learning and rapport. Furthermore, children are frequently accompanied by friends and siblings in educational contexts. We do not know how multiple encounters with the robot or how interacting in groups might affect children's development of a relationship and rapport with the robot. Exploring group interactions that include multiple children, or children in concert with parents and teachers, could help us learn how to integrate robots into broader educational contexts and connect learning with peers to learning in school and at home.

REFERENCES

Bailenson, J., Swinth, K., Hoyt, C., Persky, S., Dimov, A., and Blascovich, J. (2005). The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence* 14, 379–393. doi: 10.1162/105474605774785235

Baker, S., Weinrich, B., Bevington, M., Schroth, K., and Schroeder, E. (2008). The effect of task type on fundamental frequency in children. *Int. J. Pediatr. Otorhinolaryngol.* 72, 885–889. doi: 10.1016/j.ijporl.2008.02.019

6. CONCLUSION

In this work, we explored the impact of a robot's entrainment and backstory on children's engagement, rapport, relationship, and learning during a conversation and story activity. We found that the robot's rapport- and relationship-building behaviors affected children's emulation of the robot's words in their own stories, their displays of positive emotion, and their acceptance of the robot, and their perception of the robot as a social agent. This study adds to a growing body of work suggesting that the robot's social design impacts children's behavior and learning. The robot's story, use of relationship behaviors, nonverbal immediacy and rapport behaviors, social contingency, and expressivity are all important factors in a robot's social design.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the MIT Committee on the Use of Humans as Experimental Subjects with written informed consent from all child subjects' parents and verbal assent from all child subjects. All child subjects' parents gave written informed consent and all child subjects gave verbal assent in accordance with the Declaration of Helsinki. The protocol was approved by the MIT Committee on the Use of Humans as Experimental Subjects.

AUTHOR CONTRIBUTIONS

The study was conceived and designed by JK-W and CB. Data analysis was performed by JK-W. The paper was drafted, written, revised, and approved by JK-W and CB.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frobt.2019.00054/full#supplementary-material

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

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

Baxter, P., Ashurst, E., Read, R., Kennedy, J., and Belpaeme, T. (2017). Robot education peers in a situated primary school study: personalisation promotes child learning. *PLoS ONE* 12:e0178126. doi: 10.1371/journal.pone.0178126

Bell, L., Gustafson, J., and Heldner, M. (2003). "Prosodic adaptation in human-computer interaction," in *Proceedings of ICPHS*, Vol. 3 (Barcelona), 833–836.

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

Bennett, S. (1983). A 3-year longitudinal study of school-aged children's fundamental frequencies. *J. Speech Lang. Hear. Res.* 26, 137–141. doi: 10.1044/jshr.2601.137

Biernat, M. (2004). *Standards and Expectancies: Contrast and Assimilation in Judgments of Self and Others*. London: Psychology Press.

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

Borrie, S. A., and Liss, J. M. (2014). Rhythm as a Coordinating Device: entrainment With Disordered Speech. *J. Speech Lang. Hear. Res.* 57, 815–824. doi: 10.1044/2014_JSLHR-S-13-0149

Breazeal, C. (2002). Regulation and entrainment in human-robot interaction. *Int. J. Robot. Res.* 21, 883–902. doi: 10.1177/0278364902021010096

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

Breazeal, C., Harris, P. L., DeSteno, D., Kory Westlund, J. M., Dickens, L., and Jeong, S. (2016b). Young children treat robots as informants. *Topics Cogn. Sci.* 8, 481–491. doi: 10.1111/tops.12192

Brennan, S. E. (1996). Lexical entrainment in spontaneous dialog. *Proc. ISSD* 96, 41–44.

Burleson, W., and Picard, R. W. (2007). Gender-specific approaches to developing emotionally intelligent learning companions. *Intell. Syst. IEEE* 22, 62–69. doi: 10.1109/MIS.2007.69

Cassell, J., Geraghty, K., Gonzalez, B., and Borland, J. (2009). "Modeling culturally authentic style shifting with virtual peers," in *Proceedings of the 2009 International Conference on Multimodal Interfaces, ICMI-MLMI '09* (New York, NY: ACM), 135–142.

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

Chaspari, T., and Lehman, J. F. (2016). "An acoustic analysis of child-child and child-robot interactions for understanding engagement during speech-controlled computer games," in *INTERSPEECH* (San Francisco, CA), 595–599.

Chisholm, K., and Strayer, J. (1995). Verbal and facial measures of children's emotion and empathy. *J. Exp. Child Psychol.* 59, 299–316. doi: 10.1006/jecp.1995.1013

Christophel, D. M. (1990). The relationships among teacher immediacy behaviors, student motivation, and learning. *Commun. Educ.* 39, 323–340. doi: 10.1080/03634529009378813

Clabaugh, C., Sha, F., Ragusa, G., and Mataric, M. (2015). "Towards a personalized model of number concepts learning in preschool children," in *Proceedings of the ICRA Workshop on Machine Learning for Social Robotics* (Seattle, WA), 26–30.

Darling, K., Nandy, P., and Breazeal, C. (2015). "Empathic concern and the effect of stories in human-robot interaction," in *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (Kobe), 770–775.

Davis, M. (ed.). (1982). *Interaction Rhythms: Periodicity in Communicative Behavior*. New York, NY: Human Sciences Press.

de Jong, N. H., and Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behav. Res. Methods* 41, 385–390. doi: 10.3758/BRM.41.2.385

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

Desteno, D., Breazeal, C., Frank, R. H., Pizarro, D., Baumann, J., Dickens, L., et al. (2012). Detecting the trustworthiness of novel partners in economic exchange. *Psychol. Sci.* 23, 1549–1556. doi: 10.1177/0956797612448793

Dijksterhuis, A. (2005). Why we are social animals: the high road to imitation as social glue. *Perspect. Imit.* 2, 207–220.

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

Favazza, P. C., and Odom, S. L. (1996). Use of the acceptance scale to measure attitudes of kindergarten-age children. *J. Early Intervent.* 20, 232–248. doi: 10.1177/105381519602000307

Favazza, P. C., Phillipsen, L., and Kumar, P. (2000). Measuring and promoting acceptance of young children with disabilities. *Except. Child.* 66, 491–508. doi: 10.1177/001440290006600404

Gelfer, M. P., and Denor, S. L. (2014). Speaking fundamental frequency and individual variability in caucasian and African American school-age children. *Am. J. Speech Lang. Pathol.* 23, 395–406. doi: 10.1044/2014_AJSLP-13-0016

Gockley, R., Bruce, A., Forlizzi, J., Michalowski, M., Mundell, A., Rosenthal, S., et al. (2005). "Designing robots for long-term social interaction," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2005 (IROS 2005)* (Edmonton, AB), 1338–1343.

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

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

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

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

Grammer, K., Kruck, K. B., and Magnusson, M. S. (1998). The courtship dance: patterns of nonverbal synchronization in opposite-sex encounters. *J. Nonverb. Behav.* 22, 3–29. doi: 10.1023/A:1022986608835

Hacki, T., and Heitmüller, S. (1999). Development of the child's voice: premutation, mutation. *Int. J. Pediatr. Otorhinolaryngol.* 49(Suppl. 1), S141–S144. doi: 10.1016/S0165-5876(99)00150-0

Hartup, W. W., Laursen, B., Stewart, M. I., and Eastenson, A. (1988). Conflict and the friendship relations of young children. *Child Dev.* 59, 1590–1600. doi: 10.2307/1130673

Haviland, J. M., and Lelwica, M. (1987). The induced affect response: 10-week-old infants' responses to three emotion expressions. *Dev. Psychol.* 23, 97–104. doi: 10.1037/0012-1649.23.1.97

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

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

Iio, T., Shiomi, M., Shinozawa, K., Shimohara, K., Miki, M., and Hagita, N. (2015). Lexical entrainment in human robot interaction. *Int. J. Soc. Robot.* 7, 253–263. doi: 10.1007/s12369-014-0255-x

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

Kahn, P. H., Friedman, B., and Hagman, J. (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* (Minneapolis, MN: ACM), 632–633.

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

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

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

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

Kennedy, J., Baxter, P., and Belpaeme, T. (2015). "The robot who tried too hard: social behaviour of a robot tutor can negatively affect child learning," in *Proceedings of HRI*, Vol. 15 (Portland, OR).

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

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

Kim, Y., Baylor, A. L., and PALS Group (2006). Pedagogical agents as learning companions: the role of agent competency and type of interaction. *Educ. Technol. Res. Dev.* 54, 223–243. doi: 10.1007/s11423-006-8805-z

Klapper, A., Ramsey, R., Wigboldus, D., and Cross, E. S. (2014). The control of automatic imitation based on bottom–up and top–down cues to animacy: insights from brain and behavior. *J. Cogn. Neurosci.* 26, 2503–2513. doi: 10.1162/jocn_a_00651

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

Kory Westlund, J. M., Dickens, L., Sooyeon, J., Paul, H., DeSteno, D., and Breazeal, C. (2015). "A comparison of children learning new words from robots, tablets, and people," in *New Friends: The 1st International Conference on Social Robots in Therapy and Education* (Almere).

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

Kory Westlund, J. M., Lee, J. J., Plummer, L., Faridi, F., Gray, J., Berlin, M., et al. (2016a). "Tega: a social robot," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Christchurch).

Kory Westlund, J. M., Martinez, M., Archie, M., Das, M., and Breazeal, C. (2016b). "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)* (New York, NY: IEEE Press), 688–693.

Kory, J. (2014). *Storytelling with robots: effects of robot language level on children's language learning.* (Master's Thesis). Massachusetts Institute of Technology, Cambridge, MA.

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

Kory-Westlund, J. M., and Breazeal, C. (2019). "Assessing children's perceptions and acceptance of a social robot," in *Proceedings of the 18th ACM International Conference on Interaction Design and Children, IDC '19* (New York, NY: ACM), 38–50.

Kory-Westlund, J. M., Park, H. W., Williams, R., and Breazeal, C. (2018). "Measuring young children's long-term relationships with social robots," in *Proceedings of the 17th ACM Conference on Interaction Design and Children* (Trondheim: ACM), 207–218.

Lakin, J. L., Jefferis, V. E., Cheng, C. M., and Chartrand, T. L. (2003). The chameleon effect as social glue: evidence for the evolutionary significance of nonconscious mimicry. *J. Nonverb. Behav.* 27, 145–162. doi: 10.1023/A:1025389814290

Lee, J. J., Knox, W. B., Wormwood, J. B., Breazeal, C., and DeSteno, D. (2013). Computationally modeling interpersonal trust. *Front. Psychol.* 4:893. doi: 10.3389/fpsyg.2013.00893

Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., and Bhogal, R. S. (1997). "The persona effect: affective impact of animated pedagogical agents," in *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, CHI '97* (New York, NY: ACM), 359–366.

Levitan, R., Benus, S., Gálvez, R. H., Gravano, A., Savoretti, F., Trnka, M., et al. (2016). "Implementing acoustic-prosodic entrainment in a conversational avatar," in *INTERSPEECH* (San Francisco, CA), 1166–1170.

Littlewort, G. C., Bartlett, M. S., Salamanca, L. P., and Reilly, J. (2011). "Automated measurement of children's facial expressions during problem solving tasks," in *Face and Gesture 2011* (Santa Barbara, CA), 30–35.

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

Lubold, N., Pon-Barry, H., and Walker, E. (2015). "Naturalness and rapport in a pitch adaptive learning companion," in *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)* (Scottsdale), 103–110.

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

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

Manson, J. H., Bryant, G. A., Gervais, M. M., and Kline, M. A. (2013). Convergence of speech rate in conversation predicts cooperation. *Evol. Hum. Behav.* 34, 419–426. doi: 10.1016/j.evolhumbehav.2013.08.001

McDuff, D., Mahmoud, A., Mavadati, M., Amr, M., Turcot, J., and el Kaliouby, R. (2016). "AFFDEX SDK: a cross-platform real-time multi-face expression recognition toolkit," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '16* (New York, NY: ACM), 3723–3726.

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

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

Oster, H. (1978). "Facial expression and affect development," in *The Development of Affect*, Genesis of Behavior, eds M. Lewis and L. A. Rosenblum (Boston, MA: Springer US), 43–75.

Park, H. W., Coogle, R. A., and Howard, A. (2014). "Using a shared tablet workspace for interactive demonstrations during human-robot learning scenarios," in *2014 IEEE International Conference on Robotics and Automation (ICRA)* (Hong Kong), 2713–2719.

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

Park, H. W., Grover, I., Spaulding, S., Gomez, L., and Breazeal, C. (2019). "A model-free affective reinforcement learning approach to personalization of an autonomous social robot companion for early literacy education," in *AAAI* (Honolulu).

Park, H. W., and Howard, A. M. (2015). "Retrieving experience: interactive instance-based learning methods for building robot companions," in *2015 IEEE International Conference on Robotics and Automation (ICRA)* (Seattle, WA), 6140–6145.

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

Pearson, J., Hu, J., Branigan, H. P., Pickering, M. J., and Nass, C. I. (2006). "Adaptive language behavior in HCI: how expectations and beliefs about a system affect users' word choice," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '06* (New York, NY: ACM), 1177–1180.

Philippot, P., Feldman, R. S., and Coats, E. J. (1999). *The Social Context of Nonverbal Behavior*. Cambridge, UK: Cambridge University Press.

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

Porzel, R., Scheffler, A., and Malaka, R. (2006). "How entrainment increases dialogical effectiveness," in *Proceedings of the IUI*, Vol. 6 (Sydney, NSW: Citeseer), 35–42.

Provine, R. R. (2001). *Laughter: A Scientific Investigation*. London: Penguin.

Reeves, B., and Nass, C. (1996). *How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge, UK: CSLI Publications and Cambridge University Press.

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

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

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

Robins, B., Dautenhahn, K., Boekhorst, R. T., and Billard, A. (2005). Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills? *Univ. Access Inform. Soc.* 4, 105–120. doi: 10.1007/s10209-005-0116-3

Rotenberg, K. J. (1995). Development of children's restrictive disclosure to friends. *J. Genet. Psychol.* 156, 279–292. doi: 10.1080/00221325.1995.9914823

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

Rotenberg, K. J., and Mann, L. (1986). The development of the norm of the reciprocity of self-disclosure and its function in children's attraction to peers. *Child Dev.* 57, 1349–1357. doi: 10.2307/1130414

Rozin, P., and Cohen, A. B. (2003). High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of naturally occurring facial expressions of Americans. *Emotion* 3, 68–75. doi: 10.1037/1528-3542.3.1.68

Rubin, K. H., Bukowski, W., and Parker, J. G. (1998). "Peer interactions, relationships, and groups," in *Handbook of Child Psychology: Social, Emotional, and Personality Development*, eds W. Damon and N. Eisenberg (Hoboken, NJ: John Wiley & Sons Inc), 619–700.

Sadoughi, N., Pereira, A., Jain, R., Leite, I., and Lehman, J. F. (2017). "Creating prosodic synchrony for a robot co-player in a speech-controlled game for children," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (Vienna: ACM), 91–99.

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

Semin, G. R., and Cacioppo, J. T. (2008). "Grounding social cognition: synchronization, coordination, and co-regulation," in *Embodied Grounding: Social, Cognitive, Affective, and Neuroscientific Approaches*, eds G. R. Semin and E. R. Smith (New York, NY: Cambridge University Press), 119–147. doi: 10.1017/CBO9780511805837.006

Sénéchal, M. (1997). The differential effect of storybook reading on preschoolers' acquisition of expressive and receptive vocabulary. *J. Child Lang.* 24, 123–138. doi: 10.1017/S0305000996003005

Senechal, T., McDuff, D., and el Kaliouby, R. (2015). "Facial action unit detection using active learning and an efficient non-linear kernel approximation," in *Proceedings of the 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), ICCVW '15* (Washington, DC: IEEE Computer Society), 10–18.

Severson, R. L., and Carlson, S. M. (2010). Behaving as or behaving as if? Children's conceptions of personified robots and the emergence of a new ontological category. *Neural Netw.* 23, 1099–1103. doi: 10.1016/j.neunet.2010.08.014

Sinha, T., and Cassell, J. (2015a). "Fine-grained analyses of interpersonal processes and their effect on learning," in *Artificial Intelligence in Education* (Cham: Springer), 781–785.

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

Smidl, S. L. (2006). *Portraits of laughter in "kid" ergarten children: the giggles and guffaws that support teaching, learning, and relationships* (Doctoral dissertation). Virginia Tech.

Sorenson, D. N. (1989). A fundamental frequency investigation of children ages 6–10 years old. *J. Commun. Disord.* 22, 115–123. doi: 10.1016/0021-9924(89)90028-2

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

Stenzel, A., Chinellato, E., Bou, M. A. T., del Pobil, A. P., Lappe, M., and Liepelt, R. (2012). When humanoid robots become human-like interaction partners: corepresentation of robotic actions. *J. Exp. Psychol.* 38:1073. doi: 10.1037/a0029493

Suzuki, N., and Katagiri, Y. (2007). Prosodic alignment in human-computer interaction. *Conn. Sci.* 19, 131–141. doi: 10.1080/09540090701369125

Tickle-Degnen, L., and Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychol. Inq.* 1, 285–293. doi: 10.1207/s15327965pli0104_1

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

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

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

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

Wallbridge, C. D., van den Berghe, R., Hernández Garcia, D., Kanero, J., Lemaignan, S., Edmunds, C., et al. (2018). "Using a robot peer to encourage the production of spatial concepts in a second language," in *Proceedings of the 6th International Conference on Human-Agent Interaction, HAI '18* (New York, NY: ACM), 54–60.

Weinberg, B., and Zlatin, M. (1970). Speaking fundamental frequency characteristics of five- and six-year-old children with mongolism. *J. Speech Lang. Hear. Res.* 13, 418–425. doi: 10.1044/jshr.1302.418

Wentzel, K. R. (1997). Student motivation in middle school: the role of perceived pedagogical caring. *J. Educ. Psychol.* 89, 411–419. doi: 10.1037/0022-0663.89.3.411

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

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

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