RELATIONAL AI:
Creating long-term interpersonal interaction, rapport, and relationships with social robots

by

JACQUELINE M. KORY-WESTLUND

M.S. in Media Arts and Sciences, Massachusetts Institute of Technology, 2014

B.A. in Cognitive Science, Vassar College, 2011

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning, in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Media Arts and Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2019

Signature redacted

Signature of Author ....

Jacqueline M. Kory-Westlund
Program in Media Arts and Sciences
March 4, 2019

Signature redacted

Certified by ....

Dr. Cynthia Breazeal
Associate Professor of Media Arts and Sciences
Thesis Supervisor

Signature redacted

Accepted by ....

Prof. Tod Machover
Academic Head
Program in Media Arts and Sciences

RELATIONAL AI:
Creating long-term interpersonal interaction, rapport, and relationships with social robots

JACQUELINE M. KORY-WESTLUND

Submitted to the Program in Media Arts and Sciences on March 4, 2019 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Media Arts and Sciences

ABSTRACT

Children are now growing up with AI-enabled, socially interactive technology. As such, we need to deeply understand how children perceive, interact, and relate to this kind of technology, especially given the many ethical concerns that arise in the context of human-machine interactions, most of which are most contentious with children. To this end, I explore questions about young children's interactions and relationships with one such technology—social robots—during language learning activities. Language learning is a ripe area for exploring these questions because of the social, interactive, interpersonal nature of the activity. In addition, literacy, language, and interpersonal skills are some of the most important skills any child will learn, as they can greatly impact children's later educational and life success.

Through a series of 9 empirical child-robot interaction studies with 347 children and using both teleoperated and autonomous robots, I establish the role of social robots as relational technology—that is, technology that can build long-term, social-emotional relationships with users. I hypothesize that a key aspect of why social robots can benefit children's learning is their social and relational nature. To that end, I demonstrate the capabilities of social robots as learning companions for young children that afford opportunities for social engagement and reciprocal interaction, particularly peer-to-peer mirroring. I discuss how we can understand children's conceptualizations of social robots as relational agents and measure children's relationships over time.

I introduce the term *relational AI* to refer to autonomous relational technologies. I develop a computational relational AI system to examine how using relational AI in a social robot can impact child-robot learning interactions. Through testing the autonomous system in a longitudinal study with 49 children, I explore connections between children's relationship and rapport with the robot and their engagement and learning. I discuss the ethical use and design implications of relational AI. I show that relational AI is a new, powerful educational tool, unlike any other existing technology, that we can leverage to support children's early education and development.

Thesis Supervisor: Dr. Cynthia Breazeal
Title: Associate Professor of Media Arts and Sciences

RELATIONAL AI:
Creating long-term interpersonal interaction, rapport, and relationships with social robots

JACQUELINE M. KORY-WESTLUND

The following people served as readers for this thesis:

Thesis Reader .... Signature redacted ....
Dr. Paul Harris
Professor of Education
Harvard Graduate School of Education

Thesis Reader .... Signature redacted ....
Dr. Rosalind Picard
Professor of Media Arts and Sciences
MIT Media Lab

ACKNOWLEDGMENTS

Probably the most important person to thank is my husband, Randy, for moving to Boston with me in the first place, and being there the whole time I worked on this degree. Notable actions he took that significantly contributed to the completion of this PhD include, but are not limited to: talking over all aspects of my research; advising on software architecture; helping smack down tricky bugs in my code; caring for little Elian while I ran study participants, wrangled data, and put words into sentences; and baking awesome fresh bread (for morale and all that). Oh, and that day he replaced the windshield wipers on the car so I could see through the downpour at 8:30am on my way to schools with the robot, that was great, too. His help and support have been invaluable and there are no words good enough to sufficiently acknowledge that.

Following that, I should thank little Elian for being an energetic bundle of constant delight. He tried so hard to help me write this document, to the chagrin of all who wanted coherent sentences and not a toddler's keyboard-smash as the output (don't worry, I fixed it later). As it turns out, he developed an early knack for climbing onto chairs and desks. He also helped by decorating my laptop with motivational dinosaur stickers.

Thank you to my advisor, Cynthia, for having such a strong and powerful vision for our robot-laced futures. Her vision is really what brought me to MIT in the first place. Her understanding of the kinds of relationships we form with others (human and non-human alike), her ideas for the innovative ways social robots can help humans flourish, her deep feedback on ideas, and so much more made this thesis what it is. Also helpful was the fact that I'm pretty sure she knows absolutely everybody, and could easily help connect me to the right people to answer my questions and provide advice.

Thank you to my committee, Roz and Paul, for not just helping with this dissertation, but for advising me throughout my whole research career at MIT. Paul's careful probing, detailed feedback, deep understanding of how children think and act, novel ideas for assessing children, and expertise in experimental design shaped many of the studies reported here. I particularly appreciate Roz's enthusiasm for understanding people and their emotions in relation to technology, her inquisitive and thoughtful questions, and her optimism that I would get results (along with the challenge to "get the best results ever," paraphrased of course). In the same breath, I should thank David DeSteno, who, besides helpfully posing as a five-year-old when we were first testing out some DragonBot interactions, provided great insight and feedback during several of the early studies. Edith Ackermann deserves thanks as well for her amazing insight into children's minds and their playful relations with technology.

The students, postdocs, and staff of Personal Robots probably each deserve their own paragraph, but I'm sticking you all in here together because I was in the lab for a while and there were a lot of you: Hae Won Park, Sooyeon Jeong, Polly Guggenheim, Nick de Palma, Jin Joo Lee, Siggi Örn, Goren Gordon, Michal Gordon, Sam Spaulding, Brad Knox, Luke Plummer, Kris dos Santos, Palash Nandy, David Nuñez, Randi Williams, Meng Xi, David Robert, Kenton Williams, Nikhita Singh, Ishaan Grover, Pedro Reynolds-Cuéllar, Huili Chen, Safinah Ali, Anastasia Ostrowski... That said. I'm giving special thanks to Sooyeon for being the best officemate ever (back before the construction when we had offices). Our conversations about research, kids, life, robots, research, life, and everything else were so much fun, frequently insightful, a wonderful venue for hashing out new ideas, and whatever other nice adjectives describe an overall good conversation. Nick gets special mention too, for all the times we hung out drinking our hot beverages of choice while chatting about research and random ideas instead of actually doing research. Good times. Polly, of course, gets her own special shoutout for being the robot den-mother who keeps the lab running. Hae Won and Goren deserve additional thanks for their work, effort, and feedback on several of the studies I report on here. Kris, Luke, Siggi, and Meng all receive super bonus thanks for their skills and support in keeping robots running. Robot research doesn't happen without people like you!

I've taken a number of undergraduates under my wing over the years (UROPs in MIT-speak), and they have helped me prepare study materials, run studies, and analyze data. Some of them even took on the time-consuming and tedious task of coding kids' gaze from videos—which I know is time-consuming because I've done it, too! I appreciate their enthusiasm to contribute to the work of PRG: Eileen Rivera, Stacy Ho, Madhurima Das, Lila Jansen, Maryam Archie, Marayna Martinez, Michal Shlapentokh-Rothman, Ryoya Ogishima, Kika Arias, Adam Gumbardo, Farida Virani, and Branden Morioka. Gold stars all around.

Special thanks to our research support staff and Key Kid Wranglers who helped me bring kids and robots together during the final two studies: Olga Sans, Alicia Zamora, Anastasia Ostrowski, and David Cruz. Olga gets extra points because she traveled tirelessly to several schools, multiple days a week, to help out. Which was amazing. It meant I had time to eat lunch and get the next week's robot sessions ready, because surprise to some of you, not absolutely everything was finished for that study months in advance! (My time management skills are stellar, but I'm still human.)

I'd like to acknowledge the many students, postdocs, and professors who were involved in numerous grants and other projects that I was also involved in, most prominently, the infamous SAR grant (for those not in-the-know, it was a five-year NSF Expeditions in Computing grant on Socially Assistive Robotics, spanning numerous institutions, and wow, was that an amazing group of people all combined). Hearing about everyone else's research, seeing the team work together at the yearly grant reviews, learning from everyone else's mistakes and successes—all fantastic and inspiring. Shoutouts as well to a few students from other groups or institutions who were instrumental in performing several of the studies discussed here and overall great to work with: Rebecca Kleinberger, Leah Dickens, and Samuel Ronfard.

But wait, there's more! Thanks to the group of MIT Media Lab Learning Innovation Fellows for being a great bunch to eat sandwiches with, talk about learning research, and bounce ideas. Thanks to Philipp and Katherine for organizing and running the fellowship! Our reading group, organized talks and speakers, and other events brought me several new and highly relevant theories and ideas. Special shoutout to Kristy here; she was also a role model in juggling kids and work.

I also want to acknowledge my academic legacy. The cognitive science program at Vassar College shaped my intellectual future, in particular, Ken Livingston, Gwen Broude, John Long, and Jan Andrews. They showed me the importance of intellectual curiosity and continuing to read, probe, and learn; the importance of context, embodiment, and interdependence; and the importance of language to, well, everything. Thank you to Jenny Walter for introducing me to computer science—Vassar's Computer Science department was an inclusive, welcoming environment that never made me feel like women couldn't do great work with computers. After Vassar, I joined Sidney D'Mello's lab for a year, and I should thank him for being a role model in productivity, focus, and attention to detail.

Last but not least, thanks to all the family and friends who didn't mind hearing about my work endlessly for the past several years (maybe because robots are cool and who doesn't like hearing about robots). Finally, if I completely forgot to name you here and you think that glaring error is egregious, let me know, maybe we can work out a deal involving compensatory cupcakes.

Research presented in this dissertation was variously supported by a MIT Media Lab Learning Innovation Fellowship, and by the National Science Foundation (NSF) under Grants CCF-1138986, IIS-1122886, IIS-1122845, IIS-1123085, IIS-1523118, and Graduate Research Fellowship Grant No. 1122374. Any opinions, findings and conclusions, or recommendations expressed are those of the author and do not represent the views of the NSF.

CONTENTS

1 INTRODUCTION

1.1 AI EVERYWHERE

People now regularly live and interact with a wide range of smart Internet of Things devices (e.g., smart speakers such as Amazon Echo[1], Google Home[2]), AI-enabled digital assistants (e.g., Google Now[3], Siri[4], Cortana[5]); personal home robots for assistance, education, health, and home security (e.g., Jibo[6], Buddy[7], Mabu[8], Dash and Dot[9], Cosmo[10], and Angee[11]); and more. Intelligent machines designed to work with and alongside people are becoming more common in manufacturing, warehouse, and retail environments (e.g., Baxter[12], Kiva[13], and Pepper[14]).

With so many AI-enabled, socially interactive, and collaborative technologies entering everyday life, we need to deeply understand how these technologies affect us. How do people interact with and respond to them? How do people conceptualize and understand them? What kinds of relationships are people forming with their social and AI-enabled technology—because they *are* forming relationships of some kind? What are the long-term consequences of having such technology in our lives—whether benefits or detriments? How can these technologies be used to promote human flourishing? How do we mitigate ethical concerns—and there are many—about the use of social technology and AI in our lives?

In this thesis, I explore some of these questions through the lens of children's interactions and relationships with AI-enabled social technology, specifically, with social robots that act as language learning companions. Children today are growing up with AI, and AI-enabled technology is increasingly being used in both formal and informal educational contexts—in classrooms, in homes, as part of casual question-asking in conversation. Studying children in an educational context can bring insight into how different social and interactive capabilities of an agent can affect children's behavior, engagement, rapport, relationships, and learning over time. We can learn about how children conceptualize robots, whether they understand that robots and other social technology are different than humans, and what their longitudinal relationships might be like. In addition, many of the ethical concerns about social technology and AI are most contentious with children, such as concerns about emotional attachment, decep-

1 Amazon Echo, https://www.amazon.com/Amazon-Echo-And-Alexa-Devices/b?ie=UTF8&node=9818047011, retrieved Jan 16, 2019.
2 Google Home, https://home.google.com/, retrieved Jan 16, 2019.
3 Google, https://www.google.com/search/about/, retrieved Jan 16, 2019.
4 Apple Inc., http://www.apple.com/ios/siri/, retrieved Jan 16, 2019.
5 Microsoft, https://www.microsoft.com/en-us/mobile/experiences/cortana/, retrieved Jan 16, 2019.
6 Jibo, Inc., https://www.jibo.com/, retrieved Jan 16, 2019.
7 Blue Frog Robotics, http://www.bluefrogrobotics.com, retrieved Jan 16, 2019.
8 Catalia Health, http://www.cataliahealth.com/, retrieved Jan 16, 2019.
9 Wonder Workshop, https://www.makewonder.com/, retrieved Jan 16, 2019.
10 Anki, https://www.anki.com/en-us/cozmo, retrieved Jan 16, 2019
11 Angee, http://meetangee.com/, retrieved Jan 16, 2019.
12 Rethink Robotics, http://www.rethinkrobotics.com/baxter/, retrieved Jan 16, 2019.
13 Amazon Robotics, https://www.amazonrobotics.com/, retrieved Jan 16, 2019.
14 SoftBank Robotics, https://www.ald.softbankrobotics.com/en/cool-robots/pepper, retrieved Jan 16, 2019.

tion, social manipulation, privacy, transparency, and security. Studying children can also provide much needed data to ethical discussions as well as our understanding of how to design new social, relational technology in beneficial ways.

1.2 EXPLORING RELATIONSHIPS AND AI THROUGH CHILDREN'S LANGUAGE LEARNING

Language and interpersonal skills are arguably the most important skills any child will learn. Language is the foundation of most human knowledge transfer. Through language, we share our stories and understand the stories of others. We use language with imagination to create new stories and new knowledge. But words alone are not enough: The stories we share have purpose. We communicate using words; words are for creating and sharing meaning with others. Thus, we cannot talk about language without also talking about social, interpersonal skills. Both are equally important for children's development. They are both critical for nearly all subsequent learning and can greatly impact children's later educational and life success.

1.2.1 *Language*

Numerous studies in the United States have shown that children who grow up with an impoverished exposure to English—such as fewer total words heard, fewer novel words heard, and a lack of rich vocabulary-building curricula or cognitively challenging tasks in preschools—show language deficits that may negatively affect the child's entire academic career (Fish and Pinkerman, 2003; Hart and Risley, 1995; Huttenlocher et al., 1991, 2002, 2010; Paez, Tabors, and Lopez, 2007; Perkins, Finegood, and Swain, 2013; Schwab and Lew-Williams, 2016; Snow et al., 2007). For example, Hart and Risley (1995) studied children from 42 families as they learned to speak, observing the children for an hour every month from when they were 1–2 years old until they were 3–4 years old. They used the number of words addressed to children as a measure of children's cumulative language experience. They found significant effects of the families' socio-economic status (SES) on the estimated number of words each child heard: children raised in high-SES families heard an estimated 30 million words more than children from low-SES families. Furthermore, this was linked to children's vocabularies, with low-SES children having significantly smaller vocabularies. Vocabulary use at age 3 was also predictive of children's language skills at ages 9–10 years, as well as being associated with scores on later vocabulary and reading comprehension tests. In addition, there were differences in the type of language children's parents used. For example, parents in high-SES families used more encouraging feedback, with a 6:1 ratio of affirmatives to prohibitions, while low-SES parents used a ratio of 12:7, and welfare families had a ratio of 5:11.

Fish and Pinkerman (2003) performed a longitudinal study following children from infancy through kindergarten in rural Appalachia to examine the impact of maternal behavior, the caregiving environment, their socio-economic status, and child temperament on the children's language skills. The children studied were all from low-SES families. They found that the aforementioned negative low-SES effects were not evident until children were 2–3 years old, but thereafter, the children were comparable to low-SES children from other regions. Their language skills at 15-months predicted their language skills at four years and at the start of kindergarten. Maternal interaction behavior was also significantly related to their language skills: children whose mothers were more facilitative and gave more contingent feedback performed better.

These two studies highlight importance of contingent, encouraging feedback and quality language input for children's language learning. As both Harris, Golinkoff, and Hirsh-Pasek (2011) and Neuman, Pinkham, and Kaefer (2013) have discussed, vocabulary learning is not about memorizing words in isolation, but is about learning the concepts for which words stand. Children learn words that they hear often, that involve concepts and events that interest them or are relevant to them, and that are discovered in interactive, meaningful contexts. Receiving repeated exposure to a great variety of novel words, connecting words to other related words, and hearing words integrated in context can all improve learning. For example, Snow et al. (2007) found that children whose preschool instructors used unfamiliar words and engaged them in cognitively challenging tasks entered kindergarten with higher language abilities. Hearing varied, complex language provides greater opportunity for uncovering grammar patterns.

1.2.2 *Social Language*

I have listed several important behaviors related to children's successful language learning, including quality language input, facilitative interactions, and encouraging feedback. These point to particular beneficial language learning environment: the dialogic context. Children learn language best when they are active participants, engaged as speakers and listeners for the purpose of communicating with words (Bloom, 2000; Duranti and Goodwin, 1992; Teale and Sulzby, 1986; Vygotsky, 1978). A dialogic context involves children directly and actively—they are not passive receivers of input, but social agents, learning to communicate using words, emotions, and other nonverbal signals, with the other social agents around them. Recent work from John Gabrieli and colleagues has found that independent of children's socio-economic status, IQ, and total amount of speech, a greater number of conversational turns in adult-child conversations led to greater left inferior frontal activation (Broca's area, a brain region associated with language and verbal skill) (Romeo et al., 2018a,b). Prior research has demonstrated the importance of social interaction for language development (e.g., Hoff, 2006; Romeo et al., 2018a,b), especially social cues that facilitate joint attention and rapport—such as eye-gaze, motor, and affective mimicry and synchrony (Chartrand and Baaren, 2009; Valdesolo and DeSteno, 2011; Wiltermuth and Heath, 2009). These social cues can impact both children's language learning (Bloom, 2000; Meltzoff et al., 2009; Sage and Baldwin, 2010) andtheir willingness to engage with and learn from instructors (Corriveau et al., 2009; Harris, 2007, 2012). Furthermore, (Kuhl, 2007, 2011) has argued that a lack of social interaction with a partner may even impair language learning, showing that infants learn to differentiate new phonemes that are presented by a live person, but not from a mere audio stream or from a video of a person. In a similar vein, Naigles and Mayeux (2001) suggest that although young children may be able to learn new words from educational television, they may not not learn grammar or complex sentence structures—learning these features of language may require social interaction with a live agent.

Because social interaction is so critical for language development, children's early interpersonal and social-emotional skills are counted among the most important skills children learn (Elias, 2006; Hoffman, 2009; Liew, 2012). Vice versa, language is counted as a key component of social competence, given its communicative, social nature (Coolahan et al., 2000; Gallagher, 1993).

1.2.3 *Support for Language Learning*

Early language learning is linked to later literacy rates (e.g. Griffin et al., 2004; Hart and Risley, 1995). Recent UNESCO reports state that there are approximately 758 million illiterate adults and 122 million illiterate youth around the world[15]. Even in the United States, which is arguably a country better equipped to battle illiteracy than many other countries around the world, the Program for the International Assessment of Adult Competencies (PIAAC) reported in 2014 that 18% of US adults age 16 to 65 fall at or below level 1 proficiency in literacy[16]. These figures are in line with earlier studies that suggested that nearly a quarter of the US population was functionally illiterate (Kaestle et al., 2001; Kutner, Greenberg, and Baer, 2006). These statistics highlight the gap discussed earlier: Despite the importance of early language development, not all children currently receive enough support, instruction, or practice in learning early language, literacy, and interpersonal skills. Some schools have insufficient resources or are understaffed. Some parents cannot spend as much time on literacy-related activities with their children as they might like, or may not be proficient English-speakers themselves. Some children have developmental disabilities or delays and need additional time or therapy. It is imperative for us to address this need.

Many interventions have been developed in recent years, including teacher education programs, parent resources, preschool readiness programs, and educational apps and games. Of these, technology-based interventions in particular have great potential for three key reasons, as listed in (Kory and Breazeal, 2014): (1) accessibility, i.e., real-world deployment at scale as technology becomes cheaper and more accessible; (2) ease-of-use, including rapid customization and the addition of new content; and (3) versatility, in that technology could be used alone, with peers, or with caregivers. Ideally, based on existing evidence, these technologies will target young children aged 3–5 years. Fixing the language gap early will have greater impact, since disparities in children's language exposure at age 3 magnify over time, resulting in children who have little chance of catching up to their peers (Hart and Risley, 1995). Developmentally, some work suggests that language learning may proceed more rapidly in the early years (Johnson and Newport, 1989). However, recent work with a much larger dataset found that grammar-learning ability is preserved until the late teenage years before declining, setting a definite critical period for language learning but one that is later than earlier work suggested (Hartshorne, Tenenbaum, and Pinker, 2018). Thus, while preschool may not be critical in terms of being the critical period for language learning, starting early can dramatically improve later outcomes. Finally, preschool children are also actively developing the social and emotional skills that facilitate language learning, promote their readiness to learn, moderate positive peer interactions, and contribute to their academic and life success (Coolahan et al., 2000; Gallagher, 1993; McClelland and Morrison, 2003; Pahl and Barrett, 2007). Learning language alongside these skills is natural; all these skills interact and support each other. Preschool is a critical age.

---

15 UNESCO, http://www.unesco.org/new/en/education/themes/education-building-blocks/literacy/resources/statistics, retrieved April 8, 2017.

16 Program for the International Assessment of Adult Competencies, https://nces.ed.gov/surveys/piaac/index.asp, retrieved April 8, 2017.

However, there is currently a dearth of quality technological innovation for younger children. Most of the innovation is in apps and games. As Nuñez (2015) reported, in 2012, there were over 20,000 apps labeled "educational" in the Apple app store, but only 5% related to literacy (Chiong and Shuler, 2010; Shuler, Levine, and Ree, 2012). Three year later, there were around 80,000 "educational" apps—largely untested, with unknown actual educational potential (Hirsh-Pasek et al., 2015). The literacy-focused apps tended to target children who were learning to read or already knew how to read, leaving out children who may first need to work on oral language skills. When apps do target oral language, the focus is often exclusively on vocabulary building, not on language as a communicative medium. Furthermore, very few of the literacy-related apps were adequate in terms of pedagogy or engagement, and 75% of the "top" apps were lesson-based, engaging children passively as watchers, listeners, and consumers, rather than being creative, active, and open-ended (Goodwin and Highfield, 2012; Guernsey et al., 2012). But as we have discussed, early language skills are rooted in interpersonal, social interaction. There are many opportunities in this space for quality innovation. Chapter 2 will discuss these technologies further.

When creating technology for children's education, it is important to acknowledge the concern that these technologies will replace teachers, parents, or other caregivers. That is not the goal. Instead, as (Kory, 2014) outlined, the goal is to supplement what caregivers already do, support them in their efforts, and scaffold or model beneficial behaviors that they may not know to use or may not be able to use. For example, one beneficial behavior during book reading is asking dialogic questions, i.e., open-ended and function/attribute questions that prompt the child to think about the story, predict what is happening, and engage more deeply with the material. Numerous studies have shown that dialogic reading can boost children's vocabulary (Hargrave and Sénéchal, 2000; Valdez-Menchaca and Whitehurst, 1992; Whitehurst, G. J. et al., 1988). Several digital storybooks have been developed to support dialogic reading, for example, adding a virtual character that modeled dialogic question asking for parents (e.g., Boteanu et al., 2016; Chang et al., 2012; Nuñez, 2015). Technology can also facilitate child-caregiver interactions. For example, Freed (2012) developed a simple game with a robot and a tablet for learning vocabulary that children and their parents played together. The robot's presence encouraged communication and discussion: without prompting, parents guided and reinforced children's behavior in a way that aligned with the language learning goals of the game. In general, the goal of many new technologies is to augment—not replace—the existing relationships between children and their families, friends, and caregivers.

1.3 THESIS GOALS AND ROADMAP

As stated above, children are now growing up with AI-enabled devices and a variety of social technology. We need to deeply understand how children perceive, interact, and relate to these technologies, especially during longitudinal interactions, and especially given the many ethical concerns that arise in the context of human-machine interactions.

To this end, I explore questions about children's interactions and relationships with social technology during language learning activities. Language learning is a ripe area for exploring these questions because of the social, interactive, interpersonal nature of the activity. I expect that new technology developed to support young children's early oral language development will afford the greatest benefits when it embodies the social and interpersonal, targets younger children, and engages them in rich dialogic contexts as active participants. That is, it will be relational technology: technology that is, inherently, interpersonal and social, reciprocal and relational, and personalized through time—exactly the kind of technology that raises the most ethical questions. Relational technology is not merely social; it includes social behaviors, such as nonverbal cues, contingency, and other behaviors that make an agent *someone to interact with*, but also additional behaviors that contribute more directly to building and maintaining an ongoing relationship. Relational technology could deeply engage children and provide either great benefits, or great harm.

One such technology is social robots. Through a series of empirical child-robot interaction studies, I establish the role of social robots as relational technologies that can successfully support young children in learning new language skills (Chapters 2 and 3). These studies begin my explorations of how different social and interactive capabilities of an agent can affect children's behavior, engagement, and learning.

Next, because learning tends to occur over time, I discuss what we have learned from multiple longitudinal studies, in particular, that social robots have strong capabilities as learning companions that afford opportunities for social engagement and reciprocal interaction, particularly peer-to-peer modeling (Chapter 4). I discuss what we have discovered about the role social robots can play as peers, how children relate to them and learn with them over time, and how this led me to dive more deeply into children's relationships with social robots and the connection children's relationships may have to learning (Chapter 5). In Chapter 6, I introduce the term relational AI to refer to a subset of relational technology that is autonomous and changes through time—the kind of technology that these social robots could be.

In my final two studies, I explore how relational AI affects children's relationships, engagement, and learning. I develop numerous new ways of assessing children's relationships (Chapter 7). I hypothesize that children's learning is influenced by their relationships and their rapport with others with whom they are interacting. I explore mechanisms that could modulate children's peer learning with social robots—specifically, the use of relationship and rapport-building behaviors. In the first of these two studies, I explore the impact of the robot's use of behavioral mimicry and disclosure on children's engagement, rapport, relationship, and learning during a single interaction session (Chapter 8). For the final study, I develop a computational relational AI system with a social robot, which I test in a longitudinal child-robot learning interaction (Chapter 9). This study allows me to deeply explore connections between children's relationship and rapport and their engagement and learning—as well as unexpected connections between children's gender and their relationship formation (Chapter 10).

I discuss what we now know about how social and relational behaviors affect engagement and learning, as well as how we can understand children's conceptualizations of social robots as relational agents, different from the other agents children encounter, and the kinds of relationships children currently form with them (Chapter 11). I discuss the implications my work has for the design of new relational technology (Chapter 12), as well as how we can address a variety of ethical concerns that relational technology raises (Chapter 13). Through all of this work, I argue that relational AI is a new, powerful educational tool, unlike any other existing technology, that we can leverage to support children's early education and development.

1.4 THESIS CONTRIBUTIONS

In this thesis, I make the following technical, empirical, and theoretical contributions (detailed further in Section 14.1):

Technical: I built a new relational AI system and implemented it in an autonomous social robot. During this process, I produced digital materials for reuse, including open-source software and assessment instructions.

Empirical: I developed new assessments for measuring children's relationships with social robots. Through 9 studies with 347 children, I collected new data that provides insight into children's behavior, engagement, rapport, and relationships with relational technology. These data inform our understanding children's conceptualizations of social robots and the kinds of relationships children form with social robots over time. I also present a model showing how social behavior, relational behavior, engagement, rapport, relationships, and learning are linked.

Theoretical: The data I collected can inform our theories about children's conceptualizations of relational technology as well as our understanding of children's language learning and learning with peers. I discuss how our understanding of children's view of technology has changed over time. I present design recommendations informed by this new data and discuss how to address several important ethical concerns about children's interactions with relational technology.

2 TECHNOLOGIES FOR LITERACY AND LANGUAGE LEARNING

Many technologies that support young children's early literacy and language learning have been developed already and are actively in use in homes and schools. While some early research suggested that media and computer-based teaching may not provide specific learning benefits e.g., Clark, 1983, as technology evolved, its potential as a tool for enabling learning of all kinds has become clear. Current technologies fall into three general categories: (a) apps and games that run on computers, phones, tablets, or iPads; (b) virtual agents, including animated pedagogical agents, embodied conversational agents, and virtual peers; and (c) social robots that act as tutors or peers. I review each category below.

2.1 APPS AND GAMES

Tablets and iPads are easily used by young children. Children quickly pick up the touch-based interfaces for interaction and are highly engaged by the interactive content. Research has found that app and game use can be used successfully to promote children's early literacy skills, such as letter and sound recognition, letter shapes, phonics, and vocabulary learning (Chang and Breazeal, 2011; Chiong and Shuler, 2010; Flewitt, Messer, and Kucirkova, 2015; Hwang et al., 2014; Judge, Floyd, and Jeffs, 2015; McClure et al., 2018; Neumann, 2018; Vaala, Ly, and Levine, 2015). While many of the studies have small sample sizes and test specific apps, these studies show the potential these devices and games may have.

When looking at scale, the best and biggest example of a successful global tablet-based digital learning initiative is Curious Learning (Breazeal et al., 2016a; Gottwald et al., 2017; Wolf et al., 2014; Wolf et al., 2017). Tablets filled with literacy-related content were given to children in rural and underserved areas around the world—including many children who might otherwise not have been exposed to literacy or attended school at all. This project found that children, on their own without adult supervision or guidance, learned many early literacy skills, with 30% teaching themselves how

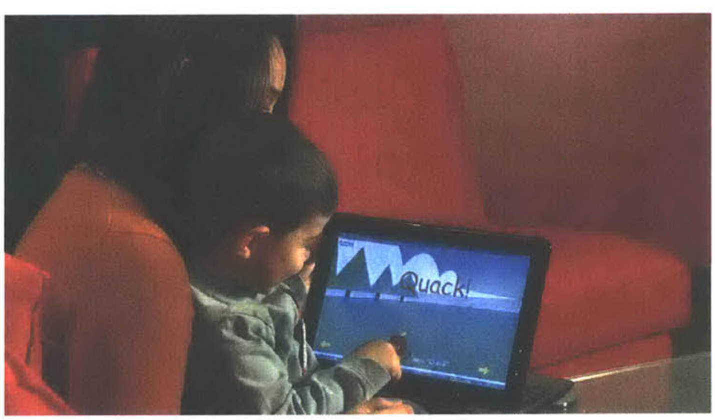


Figure 1: A mother and child play the TinkRBook storybook app together.

to read. The potential digital devices have for reaching underserved populations and increasing literacy around the world is phenomenal. Given that the digital divide continues to be a concern (Selwyn, 2004; van Dijk, 2006), research showing that access to the digital is increasing and can provide substantial benefits is highly encouraging.

However, there is concern, some of it valid, that mobile media can distract and entertain young children to their detriment (Anderson, Economos, and Must, 2008; Hirsh-Pasek et al., 2015; Radesky and Christakis, 2016; Radesky, Schumacher, and Zuckerman, 2015; Sisson et al., 2010). One concern is that increased screen time decreases children's time for active playand may be related to increasing rates of childhood obesity (Anderson, Economos, and Must, 2008; Radesky and Christakis, 2016; Sisson et al., 2010). Active play ishighly important for children's learning and development, and time for free play is decreasing (Golinkoff, Hirsh-Pasek, and Singer, 2006; Gray, 2013; Singer, Golinkoff, and Hirsh-Pasek, 2006; Singer et al., 2009). However, not all of the decrease is due to increased used of screens (e.g., video games, television, mobile apps); some is related to increased amounts of time in structured school and after-school activities.

A second key issue is that very few apps have been properly tested to determine whether they actually provide useful educational content. Many of the available apps do not add value or promote deep learning, despite being highly engaging. In 2012, there were over 20,000 apps labeled "educational" in Apple's App Store (Shuler, Levine, and Ree, 2012). In 2015, the number had grown to 80,000 (Hirsh-Pasek et al., 2015). Now, it is even higher. This plethora of apps shows the promise apps have, the amount of attention and engagement they can garner, and the ease of developing new content. To this end, Hirsh-Pasek and colleagues (Hirsh-Pasek et al., 2015) have argued for a "second wave" of educational apps that build on the science of learning—active learning, engagement in the learning process, meaningful learning, and social interaction—that can build on the successes of the app market while bringing "education" back into the apps labeled "educational". They argue that not all screen time is bad—instead, screen time can benefit children when used the right way in line with research on how learning actually happens. The digital can supplement and support the physical: e.g., providing exposure to new ideas and new words, facilitating collaboration with teachers, parents, and friends, and enabling social interaction via video chat.

Prior work has shown that involving parents in literacy learning activities with technology can be beneficial (Boteanu et al., 2016; Cairney and Munsie, 1995; Chang et al., 2012; Chiong, C. et al., 2012; Connell, Lauricella, and Wartella, 2015; Freed, 2012; Nuñez, 2015; Takeuchi, Stevens, and others, 2011). Co-use of technology andjoint media engagement increase the social interaction and discussion around media use, and help children learn to use technology in appropriate and collaborative ways. For example, (Chang and Breazeal, 2011; Chang et al., 2012) created TinkRBooks, apps that enable children and their parents to actively tinker with and explore concepts in text as they read (Figure 1). They found that including tinkerability in a storybook app encouraged more dialogue, dialogic questions, and print referencing between parents and their children, all of which are beneficial behaviors that promote literacy.

2.2 VIRTUAL AGENTS

Virtual agents have been in use as parts of educational systems for more than thirty years (e.g., André, 2011; Azevedo et al., 2012; Bers et al., 1998; Cassell et al., 2007;

Johnson and Lester, 2016, 2018; Lester et al., 1997; Taub et al., 2016; Veletsianos and Russell, 2014). Known variously as animated pedagogical agents,virtual peers, and embodied conversational agents, they are found in the context of games and apps, accompanying written text, as virtual students, as peer-like projected companions, and more. Virtual agents in education are frequently animated as humans. They take on the roles of teachers, mentors, learning companions, and actors in role plays.

Some of the most prominent work with virtual agents in literacy and language learning was conducted by Justine Cassell and colleagues (Bers et al., 1998; Cassell, 2004; Cassell and Ryokai, 2001; Cassell et al., 2007; Ryokai, Vaucelle, and Cassell, 2003). Much of this work involved creating story-listening systems, which used socially-situated virtual peers to engage children in storytelling activities, such that children could produce as well as consume stories in a playful, more natural context. Storytelling can help children learn emergent literacy; new vocabulary, metalinguistic knowledge about language patterns, structure, and function; and decontextualized language (Cassell, 2004; Curenton, Craig, and Flanigan, 2008; Engel, 1995). These systems were often tested in single sessions with children rather than longitudinally; many were teleoperated though the latest work has shifted toward autonomy. There was frequently a focus on nonverbal behaviors for generating rapport, such as appropriate gaze and attention, and using different conversational strategies and styles of speech (Bickmore and Cassell, 2001; Cassell and Bickmore, 2003; Cassell et al., 2009; Iacobelli and Cassell, 2007; Sinha and Cassell, 2015a,b; Zhao et al., 2016). For example, the Sam the CastleMate was a virtual child projected on a screen behind a toy castle (Ryokai, Vaucelle, and Cassell, 2003). Children told stories about a figurine in the toy castle, then could set the figurine in a special room in the castle, which triggered the agent to tell stories of its own. The agent modeled more advanced narrative language during its stories. During children's stories, it gave verbal and nonverbal feedback to encourage children to continue talking.

Barbara Wise and colleagues created the COLit system, a literacy program that uses a virtual tutor to teach reading and literacy skills (Cole et al., 2007; Wise et al., 2005). The virtual tutor delivered appropriately leveled content based on children's performance in various reading, phonological awareness, and spelling exercises; provided personalized feedback with encouragement, hints, and explanations of incorrect responses; and aid during tasks such as pronouncing words and reading content aloud. The interaction was primarily text-based, but used speech recognition to check the pronunciation of words; it also included face tracking and gaze tracking, which were used primarily to determine which words children were trying to read.

Numerous other systems have been developed since that help users learn vocabulary, pronunciation, and other language skills (Anderson et al., 2008; Bergmann and Macedonia, 2013; Park, 2018; Pütten, Straßmann, and Krämer, 2016; Wik and Hjalmarsson, 2009). These systems are often reasonably effective, and found that additional social cues such as the use of iconic gestures can improve performance.

Many additional educational agents have been developed for STEM learning, such as learning computer literacy (Baylor and Kim, 2004; Baylor, Ryu, and Shen, 2003; Kim and Baylor, 2006) and learning psychological research methods (D'mello and Graesser, 2012). Other agents teach social skills, e.g., by generating interactive, text-based stories (Ong et al., 2017); or aid in puzzle-solving tasks and mirror affective states (Burleson and Picard, 2004; Burleson and Picard, 2007). There are also many virtual agents developed for healthcare and therapy applications (Bickmore et al., 2016; Bickmore, Gruber, and Picard, 2005; Bickmore, Schulman, and Yin, 2010; Bickmore

(a) A boy plays a storytelling game with the DragonBot robot. (b) A girl learns new Spanish words with the Tega robot.

Figure 2: Children learning with social robots.

et al., 2018; Vardoulakis et al., 2012). These projects often tackle similar problems as in education applications, since both domains tend to focus on long-term endeavors that require engagement and rapport over time to achieve the best results. For example, Bickmore and Picard (2005) created a virtual relational agent designed to aid users in an exercise adoption program for a month. Compared to a task-oriented agent, when the agent used more social-emotional and relationship-building behaviors, it led users to trust, respect, and like the agent more.

2.3 SOCIAL ROBOTS

As a research field, social robotics spans approximately two decades (Breazeal, 2004; Breazeal, Dautenhahn, and Kanda, 2016; Breazeal, Takanishi, and Kobayashi, 2008; Feil-Seifer and Mataric, 2011). Interest in using social robots in education started early, with several projectsinvestigating how a technology that combines traditional computers and machines with the embodied, situated world could engage and benefit children (e.g., Kanda et al., 2004); it has only grown since, with numerous workshops and conferences specifically dedicated to the topic of learning with social robots (e.g., New Friends: The 1st/2nd/3rd International Conference on Social Robots in Therapy and Education; (Johal et al., 2017, 2018)). With physical bodies, social robots share physical spaces with humans, and leverage human behaviors—such as speech, movement, and nonverbal signals—to communicate with us in more natural ways. This means they can leverage human means of learning and teaching to engage and educate.

A recent survey of 101 studies of robots in education found that 86% of studies positioned robots as teachers or tutors, while 9% placed the robot in a peer or novice role, 4% gave the robot a mixed tutor/teacher role, and 1% gave the robot another role (Belpaeme et al., 2018). Two-thirds of studies looked at affective learning outcomes, while the remaining third looked at cognitive outcomes. Two-thirds of studies also looked at single learners with the robot; the rest involved groups. Nearly 60% of the studies surveyed involved children.

A recent, albeit brief, review on social robots for literacy and language learning suggests that children generally find robots engaging and feel positively about participating in language learning studies (Kanero et al., 2018). Research frequently focused on vocabulary learning, in particular English language learning (often as a second language). Language production is somewhat less-studied but growing in popularity as technologies for increasing the robot's interactive technologies and recognizing children's speech improve. This review did not mention other literacy-related activities, though there are several that have also been studied, such as reading and handwriting; it also did not discuss the roles robots play in the learning scenarios (e.g., teacher, peer, novice). Thus, I discuss these various categories further here.

With regards to robot role, social robots for literacy and language learning are most frequently positioned as tutors or teachers (e.g., Alemi, Meghdari, and Ghazisaedy, 2014; Chang et al., 2010; Deshmukh et al., 2015; Kennedy et al., 2016b; Lee et al., 2011; Robins et al., 2005; Serholt et al., 2014; You et al., 2006).For example, Alemi, Meghdari, and Ghazisaedy (2015) programmed a robot as an "assistant teacher"that participated in dialogue with the primary teacher and acted out new words to teach new foreign language vocabulary to students. Lee et al. (2011) created an autonomous robot that administered oralEnglish vocabulary lessons to Korean elementary school students. The robot, which had speech recognition, emotional expression, and RFID-enabled person identification capabilities, successfully engaged children, increased their interest in learning English, and helped them improve speaking skills.

Several robots have been situated as peers or slightly advanced peers (e.g., Freed, 2012; Gordon et al., 2016; Kanda et al., 2004; Kory-Westlund et al., 2017b; Kory and Breazeal, 2014) (Figure 2). For example, in my prior work, which was partly inspired by Justine Cassell's earlier work with story listening systems (Cassell, 2004; Ryokai, Vaucelle, and Cassell, 2003), the robot invited children to play a storytelling game (Kory and Breazeal, 2014). It told stories that were matched to the child's general language level, such that the robot's stories were slightly more advanced than what the child might tell, drawing the child into the zone of proximal development (Vygotsky, 1978).

Kanda et al. (2004) created one of the first autonomous robots for English language learning,which interacted with Japanese children aged 6–7 and 11–12 years. The robot was situated as a playmate who spoke a different language. It interacted with children in groups during school, using RFID nametags to identify individual children. It played simple games such as rock-paper-scissors and performed interactive behaviors such as hugging, shaking hands, singing, and pointing. The robot could play over 300 sentences and had speech recognition capabilities for around 50 words, enabling children to practice the new English words they were learning. The idea of robot peer as foreign playmate appeared again in Natalie Freed's work, where a French-speaking robot played a food-vocabulary with young children (Freed, 2012).

Some robots are positioned as younger peers or novices (e.g., Gordon and Breazeal, 2015; Hood, Lemaignan, and Dillenbourg, 2015; Matsuzoe and Tanaka, 2012; Tanaka and Kimura, 2009; Tanaka et al., 2015; Tanaka and Matsuzoe, 2012). For example, Matsuzoe and Tanaka (2012) and Tanaka and Matsuzoe (2012) set up their robot as a "care-receiving" less competent peer whom the children had to help teach. The overall goal was to see whether the robot could help young Japanese children, aged 3–6 years, learn new English verbs. The robot was remote-operated and played a verb-learning game with each child. In the game, the experimenter asked the child or robot to act out new verbs. On the child's turns, if the child was uncertain, the experimenter demonstrated the action for the child. On the robot's turns, initially, the experimenter "taught" the robot the actions; later on, the experimenter asked the child to teach the robot the action. Children were engaged by the activity and did learn new verbs from this format. In a similar setup, Hood, Lemaignan, and Dillenbourg (2015) introduced 7–8-year-old children to a teachable robot that wanted to learn handwriting skills.

Children had to show the robot how to write letters and correct the robot's mistakes, thereby practicing handwriting skills themselves.

One recent study compared a robot that adapted its level of expertise, switching between novice and expert roles in order to maximize children's engagement and learning, to a purely novice robot and a purely expert robot (Chen, 2018). Children aged 5–7 years played a word-learning game with one of the three robots. Playing with the adaptive role switching role led to increased vocabulary learning compared to playing either the expert or novice robots. This was the first study that dynamically changed the role of the robot according to what might lead the child to learn best.

A range of literacy and language activities have been used with social robots. These activities have included vocabulary learning; oral language, conversation, and storytelling; reading; and handwriting. As mentioned earlier, vocabulary learning may be the most common activity (Breazeal et al., 2016b; Chang et al., 2010; Freed, 2012; Gordon et al., 2016; Kanda et al., 2004; Kennedy et al., 2016b; Kory-Westlund et al., 2015a,b; Lee et al., 2011; Movellan et al., 2009; Tanaka and Matsuzoe, 2012), though the mechanisms of vocabulary teaching have varied—e.g., embedding vocabulary words in stories, games, or interaction (Freed, 2012; Gordon et al., 2016; Kanda et al., 2004; Kory and Breazeal, 2014; Movellan et al., 2009); asking children to act out new words (Matsuzoe and Tanaka, 2012; Tanaka and Matsuzoe, 2012); and discussing or demonstrating new words for children (e.g., Alemi, Meghdari, and Ghazisaedy, 2015; Alemi, Meghdari, and Ghazisaedy, 2014; Lee et al., 2011). One pattern that emerges is the difference in the roles the robots play in relation to the kind of vocabulary activity. Robots situated as teachers or tutors were more likely to explicitly teach, demonstrate, or discuss words; robots situated as peers or novices were more likely to teach words through social interaction. Most of the activities involving vocabulary via social interaction also involved language production, storytelling, or conversation.

With regards to written words, the studies so far situated the robot as a teachable novice. Gordon and Breazeal (2015) created a personalized reading robot, portrayed as a younger peer who wishes to learn to read, which selected child-specific assessment-based words for the child to practice reading. In two studies (Chandra et al., 2018; Hood, Lemaignan, and Dillenbourg, 2015), as mentioned above, children showed the robot how to write letters and gave the robot feedback, correcting its mistakes. It appeared that teaching a robot that appeared to improve (i.e., a "learning" robot) increased children's own learning and performance (Chandra et al., 2018).

Given the wide range of literacy-related activities that could exist, there is room for new robots that perform new activities. For example, social robots could interact with younger children to support precursors to literacy such as joint pointing, the performative power of words, and pretense reading (Ackermann, 2002).Because of the importance of social and emotional skills in social interaction—and thus in dialogue and language learning—we could explore new work supporting children's social and emotional development. So far, the majority of the research on social-emotional skills targets children with Autism Spectrum Disorder, using social robots to encourage social interaction (Kim et al., 2013; Ricks and Colton, 2010; Scassellati, Admoni, and Mataric, 2012). Fewer studies have been performed with typically developing children e.g., Leite et al., 2015.

2.4 MIXED SYSTEMS AND VIRTUAL REALITY

Multiple virtual reality systems have been built with the goal of creating an immersive language learning experience (Bailenson et al., 2008; Schwienhorst, 2002a,b; Vázquez et al., 2018; Zheng and Newgarden, 2011; Zheng et al., 2009). Sometimes the user is alone in the virtual reality; sometimes virtual characters are present as well; and sometimes the virtual setting is used to connect multiple human users to each other. For example, Vázquez et al. (2018) created a virtual reality system for a single user that incorporated kinesthetic learning into the learning of word-action pairs, finding that using movement to reinforce word learning increased word retention rates.

Few systems so far have used multiple agents or combinations of systems, such as both a robot and a virtual character, in language learning activities. However, one new study has shown the promise of combining a robot with a virtual human in supporting deaf infants' exposure to sign language (Scassellati et al., 2018b). Because screen-based media are unlikely to support infants' learning by itself (Kuhl, Tsao, and Liu, 2003), they created a socially contingent dialogue between a robot and the virtual human that successfully engaged the infants. This shows the promise that multi-agent systems may have in engaging different populations of children.

2.5 WHY ROBOTS?

From the research presented above, we can see that different technologies can benefit and support children's language and literacy learning in different ways—screen time is not always bad! Games, apps, virtual agents, and social robots have different affordances. We are still learning how to use each technology to its fullest potential, but we are learning. When designing technology for language learning, it appears that technologies that enable social interaction— either with the technology itself (e.g., social robots) or with human others (e.g., video chat, apps that enable collaboration with friends)—may provide more benefit than technologies that ignore the critical social components of language learning.

We have found during interviews with preschool teachers that they are looking for new ways of engaging kids in literacy activities (Kory-Westlund et al., 2016b). As discussed earlier (Chapter 1), using technology can extend the reach of humans, and supplement what parents and caregivers already do in engaging children in language and literacy activities. Many schools are already introducing iPads and tablets with new activities. Should we introduce social robots as another? Some of the concerns around why we might choose apps or virtual agents involve cost, east of use, ease of access, and deployment at scale. Virtual agents, apps, and games could be used on low-cost smartphones or tablets, such as those that already exist in classrooms and homes, and require no new purchases, just installation of new software. That said, there are now commercial robots available that hit the same price points as iPads and tablets. However, if it was found that virtual agents did not lead to as great of learning gains as robots, it could be that they still lead to more gains than having no intervention or no agent at all, and so their lower cost or ease of use might justify using them regardless. Given all this, is there a compelling reason to choose robots for language and literacy activities when virtual agents and apps may arguably be cheaper to deploy?

I recently surveyed 71 peer-reviewed publications comprising 80 unique comparative studies in human-robot interaction that compared virtual agents to co-present and

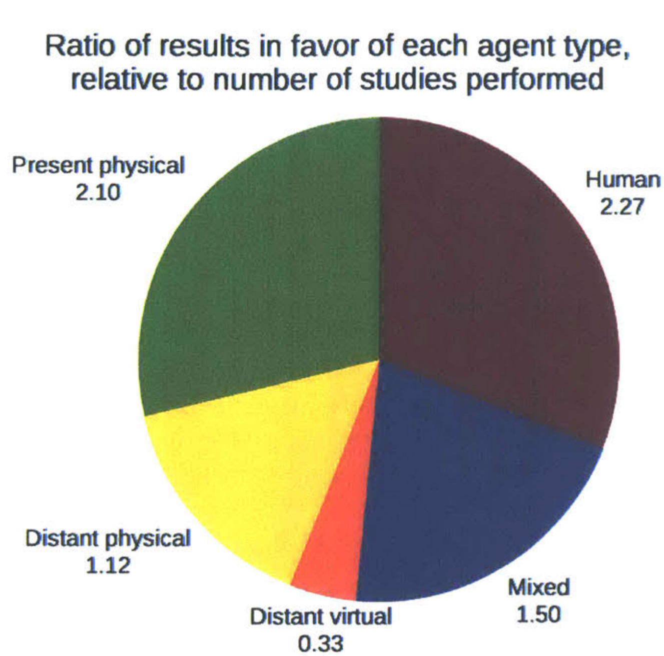


Figure 3: Some agents were studied more often than others. This graph shows the number of results favoring each of the main agent types across all the studies surveyed, adjusted for the number of studies that actually included each agent type.

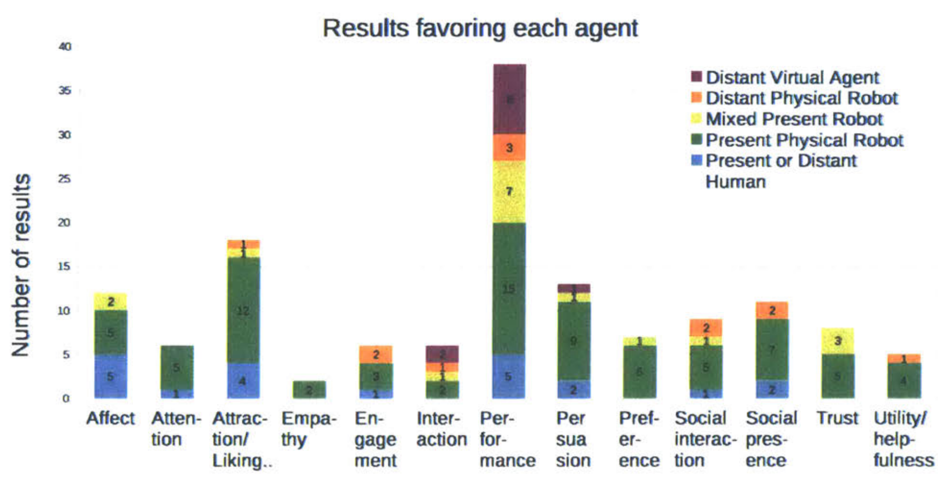
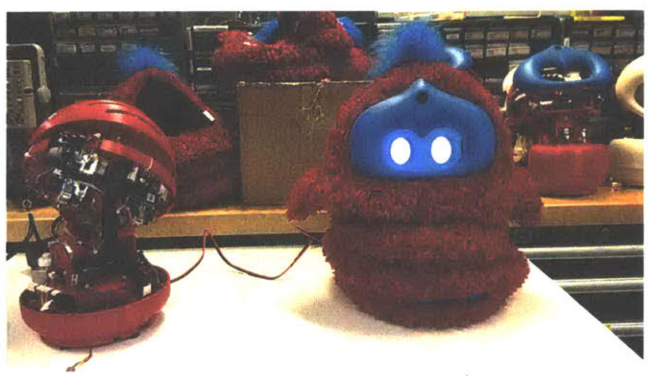


Figure 4: The number of results favoring each main agent type for the dependent variables measured in at least three different studies (affect, attention, attraction/liking, empathy, engagement, interaction, performance, persuasion, preference, social interaction, social presence, trust, utility/helpfulness).

(a) Green and yellow DragonBots.

(b) Blue DragonBot.

(c) Tega.



(d) Tega, earlier version of the face.

Figure 5: The two robots I used.

telepresent robots (Kory-Westlund, Breazeal, and Ostrowski, in review). I categorized the agents studied based on their embodiment (physical, virtual, or a mixed agent with some physical and some digital components) and their presence (present or distant). In addition, multiple studies included other kinds of agents, such as humans, tablets or laptops without an agent, or a voice-only agent with no embodiment at all. More than half the studies compare two agents; about a third compared three; and a handful compared four or more. I also categorized whether the types of tasks performed with the agent included physical components (e.g., a Towers of Hanoi or block-stacking physical puzzle), social components (e.g., conversation, storytelling, judging emotions), or digital components (e.g., tasks shown on a screen, digital puzzles).

In looking at the results favoring each agent type, I found that physically embodied co-present robots affect humans more deeply and strongly than virtual agents and telepresent robots, though this was modulated somewhat by the type of tasks performed with the agent. Figure 3 shows the ratio of results favoring each agent type relative to the number of studies performed that included that agent type. Humans and physically present robots were favored most often, and most often during socially interactive tasks and physical tasks. During digital tasks, the results generally did not favor one agent over another. In addition, humans and present physical robots led to more increases in important social metrics including attention, attraction/liking, empathy, persuasion, and trust (Figure 4).

The results of this survey suggest that the robot's presence and embodiment—but presence more so than embodiment—seem especially important for interpersonal, social tasks. Because language learning is rooted in social interaction (Section 1.2.2), this suggests that a present physical robot should be preferred over a distant robot or virtual agent, as it will likely lead to greater trust, rapport, and performance during language learning tasks. While cost tradeoffs may still need to be considered when deploying at scale, in the work that follows, I use present, mixed robots (physical bodies with digital faces), as I expect they provide greater benefits than distant or virtual agents. Furthermore, I argue that present social robots afford new and different opportunities than other technologies, including increased social engagement, reciprocal and relational interaction, and peer-to-peer modeling.

2.6 ROBOTS USED

In the studies reported in the following chapters, I use two different present, mixed robots: DragonBot (Freed, 2012; Kory, Jeong, and Breazeal, 2013; Setapen, 2012) and Tega (Kory-Westlund et al., 2016c). The robots are shown in Figure 5. Both are squash-and-stretch robots designed for interactions with young children. Their expressive body movement is based on principles of animation (Lasseter, 1987). They can move up and down, tilt to the side, rotate from side to side, and lean forward and backward.

Each robot has an Android phone that displays an animated face and runs control software. The phone's sensors can be used to capture audio and video, which can be used as input for various behavior modules, e.g., affect recognition and affect mirroring, or for use in teleoperation of the robot. Tega uses a Samsung Galaxy S7 and has an additional camera mounted in its forehead, which can be used to augment the phone's more limited camera view.

The robots' animated faces can change to display various facial expressions and mouth movements (visemes) during speech. DragonBot's face has round blue eyes and a blue mouth. In Study 5 (Section 4.4), Tega's face had white oval eyes and no mouth; in the later studies, Tega's face has blue oval eyes and a white mouth. For both robots, some animations that play on the robot use only the face, while others use both the body and face together. The phone's microphone can be used to capture audio.

Each robot is covered in a skin of colorful, fluffy fur. DragonBot has skins of several colors that were used in different studies—blue, green and turquoise, yellow and purple. Tega has a skin with red fur and blue stripes.

For both robots, I recorded all speech used, with two exceptions. In Study 3 (Section 3.5), a second experimenter and I each teleoperated a DragonBot, and we live-streamed speech to the robots, rather than using recorded speech. In Study 5 (Section 4.4), a female undergraduate student recorded the robot's speech. In all cases, all utterances was shifted to a higher pitch to make it sound more childlike.

Finally, in all the studies, the experimenters referred to each robot by name (not with pronouns) in a non-gendered way throughout the study. The names used varied by study and robot skin color. DragonBot was named either Yellow, Green, or Blue; Tega was called either Tega or Red.

3 SOCIAL ROBOTS AS LANGUAGE LEARNING COMPANIONS

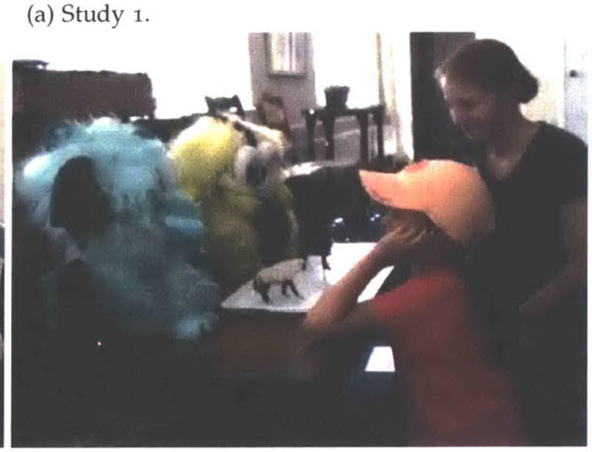
3.1 THE IMPORTANCE OF PEERS

The dynamics of children's interactions with peers are very different from their interactions with teachers, parents, and other adults. Relationships with adults can be construed as asymmetrical, with the adult having higher dominance, power, and expertise (Rubin, Bukowski, and Parker, 1998). Peer relationships, on the other hand, are more balanced, with greater opportunity for openness, exploration, and discovery. A great deal of research in the past several decades has shown that children's peers can greatly impact development and learning. Peer interaction can, e.g., enhance children's overall preschool competency and language growth, especially in the company of more advanced peers (DeLay et al., 2016; Fuchs et al., 1997; Justice et al., 2011; Lin et al., 2016; Mashburn et al., 2009; Mathes et al., 1998; Schechter and Bye, 2007; Topping, 2005). For example, (Mashburn et al., 2009) measured preschool children's receptive and expressive language skills at the start of the school year and again at the end of the school year. They found that children's peers' expressive language ability was positively related to children's own language growth during the year. (Justice et al., 2011) replicated this result. Children benefited most from having higher ability peers around them. Furthermore, children with lower skills were much more affected by the skills of their peers, while children who already had higher language skills were more "resilient" to the skill level around them.

There are several theories regarding how peer learning occurs. Each focuses on learning through a different mechanism: e.g., observation of peers, conflict with peers, or cooperation with peers. Piaget, for example, thought that dialogue and discussion among peers could lead to cognitive development (De Lisi and Golbeck, 1999; Piaget, 1932; Rubin, Bukowski, and Parker, 1998; Tudge and Rogoff, 1989). Social influences were not central to Piaget's theories on children's development; rather, the emphasis was on the child as an individual. Peers could provoke higher thinking or cognitive challenge through discourse, negotiation, and argument—i.e., through conflict. Vygotsky, on the other hand, promoted a social-cultural theory in which the child's social context was highly important and influential for development, from infancy onwards (Rubin, Bukowski, and Parker, 1998; Tudge and Rogoff, 1989; Vygotsky, 1978). For example, more advanced or experienced peers could support a child in the zone of proximal development, helping them practice and acquire skills beyond their current skill level—i.e., through cooperation. Peers co-construct knowledge with each other. Children can also take on explicit instructional roles (Neuman and Roskos, 1992; Stone and Christie, 1996). Finally, Bandura and Walters' social learning theory suggests that children learn through observing and imitating others (Bandura, 1971; Bandura and Walters, 1963; Rubin, Bukowski, and Parker, 1998). Children implicitly reinforce and change each others' behaviors in such a way as to reward social competence and socially appropriate behaviors and to punish non-normative or socially incompetent behaviors. They learn both through this peer feedback as well as through observation of the consequences of their peers' actions.

(a) Study 1.

(b) Study 2.



(c) Study 3.

Figure 6: Children interact with DragonBots during the first three studies.

3.2 ROBOTS AS PEERS

Because peers can significantly and positively affect children's language learning, I postulated that a peer-like robot could lead to similar benefits. Peer interaction has the potential to be highly dynamic and highly fluid. The robot could provide affordances for learning not just through instruction, as in the case of tutor robots, but also via observation, conflict, and cooperation. Peer-like robots could allow the child to switch roles—sometimes teaching, sometimes being taught, but always learning.

The work in the Personal Robots Group involving language, literacy, and social robots as peers began with Natalie Freed's imaginative work with Sophie, a DragonBot situated as a French-speaking peer (Freed, 2012). Children, accompanied by their parents, encountered the robot in a hybrid physical-virtual "French cafe." They sat at a play table with the robot. The robot introduced the names of foods in French, which were shown virtually on a tablet. This was one of the first studies using a tablet as a shared game surface, a technique pioneered by several other labs around the same time Baxter, Wood, and Belpaeme, 2012; Park and Howard, 2013. The play scenario involved the robot eating some foods, disliking others, and showing emotion and attention through gaze and animation. Children treated the robot as a social agent, adjusting their speech and behavior to communicate with it. While there were no significant learning gains given the single interaction and foreign language vocabulary, children reportedly enjoyed the interaction and reacted to the robot in communicative, social ways that suggested greater potential for the robot as a language learning companion.

We followed this earlier work with three studies examining several basic questions regarding children's word learning from social robots. Given the existing evidence (Section 2.3) that social robots can be effective tutors and learning companions, my collaborators and I have asked why are they effective? What design features of the robots positively impact children's learning and attitudes? As hinted at in Natalie Freed's earlier work, these three studies highlight the fact that the design of robots as social agents matters for children's learning. For example, children pay attention to the robot's nonverbal social cues to guide their learning. Factors such as the contingency of the robot's nonverbal behavior impact children's judgments of the robot's credibility. Children apply social judgments to the robots.

3.3 STUDY 1: CHILDREN LEARN NEW WORDS FROM SOCIAL ROBOTS

This study compared children's rapid learning of new words from three sources of information: a human, a tablet/iPad, and a social robot (Kory-Westlund et al., 2015a). We found that in a simple word-learning task, in which the child viewed pictures of unfamiliar animals and the child's interlocutor (human, tablet, or robot) provided names for the animals, all three interlocutors served equally well as providers of new words. We also found that children appraised the robot as an active, social partner (Figure 6a). This study provided evidence that children will learn from social robots, and will think of them as social partners.

3.4 STUDY 2: CHILDREN ATTEND TO A ROBOT'S NONVERBAL SOCIAL CUES DURING WORD LEARNING

Our next study compared preschoolers' learning of new words from a human and from a social robot in a somewhat more complex learning task (Kory-Westlund et al., 2017a). Children viewed two images of unfamiliar animals at once, and their interlocutor (human or robot) named one of the animals (Figure 6b). Children needed to monitor the interlocutor's non-verbal cues (gaze and bodily orientation) to determine the intended referent. To assess the discriminability of the cues needed for selective learning, the images were presented either close together, so that the interlocutor's cues were similar regardless of which animal was being attended to and named, or further apart, so that the distinctiveness of the interlocutor's cues was more evident. We found that when the images were presented close together, children subsequently identified the correct animals at chance level with both interlocutors. When the images were presented further apart, children identified the correct animals at better than chance level from both interlocutors. Thus, we saw that children learned equally well from the robot and the human. Furthermore, the study provided evidence that children will attend to a social robot's nonverbal cues during word learning as a cue to linguistic reference, as they do with people.

3.5 STUDY 3: A ROBOT'S NONVERBAL SOCIAL CUES IMPACT CREDIBILITY

This study probed children's learning with social robots further. We examined not only whether children would be willing to learn new information from a social robot, but in particular, whether they would regard two robots that differed in how contingently responsive they were as equally reliable informants (Breazeal et al., 2016b). We found that children treated both robots as interlocutors and as informants from whom they could seek information (Figure 6c). Consistent with studies of children's early sensitivity to an interlocutor's nonverbal signals, children were especially attentive and receptive to whichever robot displayed the greater nonverbal contingency. Such selective information seeking is consistent with recent findings showing that although young children learn from others, they are selective with respect to the informants that they question or endorse (e.g., Harris, 2012). This study provided evidence that children show sensitivity to a robot's nonverbal social cues, and will use this information when deciding if a robot is a credible informant, as they do with humans.

3.6 SOCIAL ROBOTS ARE BETWIXT AND BETWEEN

The three studies described above showed that children treat robots as social others, apply social judgments to robots, and respond to their social cues in ways similar to how they respond to people. Does this mean children think robots are people? The evidence so far suggests that is unlikely—but neither do children think robots are mere machines. Prior work suggests that children may place robots in a different ontological category than either living or non-living things (Gaudiello, Lefort, and Zibetti, 2015; Kahn et al., 2011; Severson and Carlson, 2010). Children have described a robot as being "in between" living and non-living (Kahn et al., 2012; Severson and Carlson, 2010). They have shown a moral objection to the object-like treatment of robots, such as putting a robot away in a closet, because of the perception they have formed about the robot as a social other (Kahn et al., 2012); however, they may also say that like other objects,a person made the robot, people can own robots, and that robots can break (Kory-Westlund et al., 2016a; Kory and Breazeal, 2014). In several studies in the early 2000's, children categorized the robot dog Aibo as not a dog and not a robot, but as a "robotic dog"—a dog with robotic attributes (Bartlett, Estivill-Castro, and Seymon, 2004; Kahn, Friedman, and Hagman, 2002; Melson et al., 2009; Weiss, Wurhofer, and Tscheligi, 2009). The robot was treated both as technological artifact and as a social other. It was attributed properties of living animals—like mental states, agency, and moral standing—but was also understood to be non-living, running on batteries, and "just a toy".

In my work and related work from the Personal Robots Group, we have seen children ascribe psychological properties (thinking, being happy), perceptual abilities (seeing, feeling tickles), and properties of artifacts (being man-made, able to break) to robots, but rarely biological properties (eating, growing) (Gordon and Breazeal, 2015; Gordon et al., 2016; Knox, Spaulding, and Breazeal, 2016; Kory-Westlund et al., 2016a, 2017b; Kory and Breazeal, 2014). More recently, Randi Williams and Stefania Druga examined children's perceptions of autonomous technologies (Amazon Alexa, Google Home, Cozmo, and Julie Chatbot) as well as children's attributions of intelligence to robots, mice, and to themselves (Druga et al., 2017, 2018). They found that children tended to think of the robots socially and speak about them socially; they compared robots to both people and to toys, assigning them qualities of each.

This research shows that children seem to think of robots as something betwixt and between the dualistic categories of alive, animate beings and inanimate objects, with properties of familiar entities but not exactly the same properties as any of them. Robots are often seen as social, similar to humans and pets, but with qualities of machines, toys, and artifacts. Children's perceptions of robots are shaped by their experiences in the world as well as by the opinions of those close to them—for example, their attributions of how intelligent a mouse or robot was were very similar to their parents' (Druga et al., 2018). Children developed and mirrored their parents' mental models. In addition, we have seen that the stories told to children by an adult about a robot affected their social judgments (Kory-Westlund et al., 2016a). Others' opinions and the social context influence children's beliefs.

However, many of these studies examined single encounters that children had with robots. How much of their behavior might be due to novelty or inexperience—might children revise their opinions after more experience interacting with real robots, such as over multiple encounters? How might children's learning, behavior, and perceptions of the robot change over time? These are some of the questions addressed in the next chapter.

4 LEARNING WITH SOCIAL ROBOTS THROUGH TIME

The research discussed so far (Chapters 2 and 3) shows that children will engage with, learn from, and respond socially to social robots in learning contexts. However, many of the earlier studies involved short encounters—just one session with the robot. Learning is, however, a necessarily longitudinal task. Thus, these short-term studies leave open numerous questions about the effectiveness of the robot as a tutor or learning companion over time, as well as how children might perceive the robot over time. Will the robot be as effective over multiple sessions? Will children retain what they have learned? Will the knowledge transfer to other contexts? Will children's engagement be maintained? How will children's construal of the robot change? What kind of relationship will children develop with the robot over time?

4.1 LONG-TERM INTERACTION

Research in HRI looking into how people interact with and perceive robots through time—that is, repeated encounters or long-term interactions—began not long after the field began, in the early 2000's. With several members of the Personal Robots Group (Sam Spaulding, Hae Won Park, Nikhita Singh, and Pedro Reynolds-Cuéllar), I performed a preliminary survey of long-term interaction studies in HRI. We performed a search for relevant empirical work, using the list of papers in Iolanda Leite's 2013 survey (Leite, Martinho, and Paiva, 2013) as a starting place. We looked at papers these papers referenced and at newer papers referencing the existing papers, and performed searches with various literature search tools such as Google Scholar and keywords such as "long-term interaction," "longitudinal," "repeated encounters," "HRI," "social robot," "time," etc. We found 61 papers published from 2003–2019 (mean 2012, median 2014, mode 2016).

Of these 59 papers, 16 were in the domain of Education; 16 were in Healthcare and 10 in Eldercare, with 4 papers counted in both Healthcare and Eldercare; 10 were in the Home, 7 involved Service/Public Spaces; 2 were in Entertainment; 1 was in the Workplace; and 3 were classified as Other (two involved social play, and one involved physical play, but they were not specifically educational). Within the domain of Education, 7 involved language learning, all with children (Ahmad, Mubin, and Orlando, 2016; Gordon et al., 2016; Kanda et al., 2007; Kanda et al., 2004; Kory and Breazeal, 2014; Movellan et al., 2009; Park et al., 2019). The other educational activities varied in whether they included children or adults as participants and included map reading and geography (Serholt and Barendregt, 2016; van Maris et al., 2017), learning about nutrition (Short et al., 2014), chess and other puzzle tasks (Leite et al., 2009; Leite et al., 2014; Leyzberg, Spaulding, and Scassellati, 2014), among other activities (Hyun, Yoon, and Son, 2010; Kasap and Magnenat-Thalmann, 2012; Ros et al., 2016). All but two of the education interactions (Kanda et al., 2007; Kanda et al., 2004) involved 1-on-1 interactions with the robot.

In total, 46 of the 61 studies used autonomous robots; 6 were teleoperated/Wizard-of-Oz (WoZ) controlled; 1 included both an autonomous robot and a WoZ robot; 5 were partially autonomous; 2 used shared autonomy; and 1 provided no information.

We classified whether the type of autonomy used was "interesting"—i.e., involved some kind of perception, adaptation, or decision-making and was not simply a reactive, roomba-like agent. Twenty-two papers used interesting autonomy; 3 were sort of interesting; 26 were not very interesting; and 3 were not reported. For example, in one of the earlier papers, Gockley et al. (2005) created an autonomous receptionist robot in the Service/Public Spaces domain, which provided information to visitors to the computer science building. The robot had basic face finding and face tracking capabilities. When it found faces, it used an attentional classifier that modeled people as "interested," "engaged," "attending," or "none," and automatically classified people's attention, which informed which behaviors it selected to perform next. It also used chat-bot style pattern matching to engage in dialogue via a tablet screen.

In looking only at the 16 papers in Education, 9 used interesting autonomy; 3 were not very interesting; 3 used WoZ control; and 1 did not report sufficient information for classification. Interestingly, several of the earliest education papers included interesting autonomous robots as well as several of the most recent; in between, we find the papers with less interesting autonomy and WoZ control as researcher explored the space in order to determine what was useful, engaging, or attainable.

The earliest long-term child-robot language learning work (and also the earliest education work in general), from Takayuki Kanda and colleagues, used an autonomous Robovie robot that adapted its behavior according to children's distance from it (Kanda et al., 2004) and the number of previous interactions (Kanda et al., 2007). In both studies, the robot helped children learn new English words by being a foreign-language speaking playmate (as discussed earlier). The robot used RFID nametags to detect individual children, and had basic speech recognition capabilities and motion control. In the latter study, the robot also adapted the interactions based on previous interactions and attempted to estimate friendships between children, using sociograms as representations of children's networks to approximate which children might have trouble socializing, which might have bullying tendencies, and so forth. In the first study, the robot was placed in a school for two weeks, where 9 Japanese children aged 6–7 years and 11–12 years interacted with it daily. In the second study, 32 Japanese children aged 10–11 years interacted with the robot at school daily for about two months. Both studies found that children engaged with the robot over time and learned new words. In the first study, interaction dropped off during the second week; in the second study, the robot's increased capabilities led to greater social interaction and engagement during the full two months. However, this was moderated by children's treatment of the robot as a peer-like friend. About one-third of the children treated the robot as a friend and kept interacting with it; the other children appeared to become bored with the robot after 5–7 weeks.

The next long-term language learning studies occurred in the US, at an early childhood education center with the autonomous RUBI-4 robot (Movellan et al., 2009). The robot had a touch-screen embedded in its chest, which enabled children to play a flash-card vocabulary game. The robot also danced while playing music videos on its screen, and could play give-and-take games with physical objects using its hands. The robot used an "interest indicator" that combined the number of touches on the screen with the number of faces it had recently detected to determine children's interest, which was used to select the next behaviors. The robot also had a game scheduler to determine which games to play next. Children enjoyed the robot and interacted with it appropriately; they also appeared to learn the vocabulary words presented by the robot.

Following this was my WoZ master's thesis study and our second-language learning study, both of which I will discuss in more detail below (Gordon et al., 2016; Kory and Breazeal, 2014). These two studies led to another project with storytelling and language learning robots from our group, which I will also discuss below (Park et al., 2019). Ahmad, Mubin, and Orlando (2016) created a vocabulary game that children aged 10–12 played with an autonomous Nao robot for about 10 minutes a day for 10 days. The robot remembered and referenced children's scores in the game from prior sessions. It performed affect recognition and used this information about children's affect and the game states to choose affective reactions and gestures for the robot. The study was qualitative and did not report significant learning gains, but showed promise for the use of robots as engaging teaching tools with middle school children.

Multiple long-term studies involved children in activities that were not specifically educational, e.g., entertainment, therapy, or play-focused (Barakova et al., 2015; Coninx et al., 2016; Kozima, Michalowski, and Nakagawa, 2009; Kruijff-Korbayová et al., 2015; Nalin et al., 2012; Salter, Dautenhahn, and Bockhorst, 2004; Tanaka, Cicourel, and Movellan, 2007). Several important results regarding children's long-term engagement came from these studies. First, variation in the robot's speech and behavior is important for maintaining engagement. Selecting different activities based on the child's interests and as a way to increase the variation in the interaction can also improve engagement. For example, Coninx et al. (2016) found that switching between several different activities helped engage children in diabetes education over time. Different children preferred different activities, so switching activities to suit individuals' preferences was helpful in maintaining engagement. In an earlier study, Salter, Dautenhahn, and Bockhorst (2004) found that children grew bored of a robot that was designed for physical play even within the first few sessions. They changed the robot's speech and behavior in later sessions and found that the increased variation improved interaction. A study by Tanaka, Cicourel, and Movellan (2007) study showed that very young children (10–24 months) socialized with a robot that engaged them in social play (e.g., dancing, giggling in response to touch, sitting and standing, moving its hands). They appeared to bond with it. The robot became part of the social ecology of the classroom and led to interesting teacher-child interactions, such as teachers showing the children how to treat the robot gently. This study suggested that children's relationship with the robot helped maintain their engagement. All of these results are in line with work involving adults, where it was found that change in the robot's speech and behavior can help maintain user engagement and build a long-term relationship (Bickmore, Schulman, and Yin, 2010; Kidd and Breazeal, 2008; Lee et al., 2012a).

4.2 PERSONALIZATION

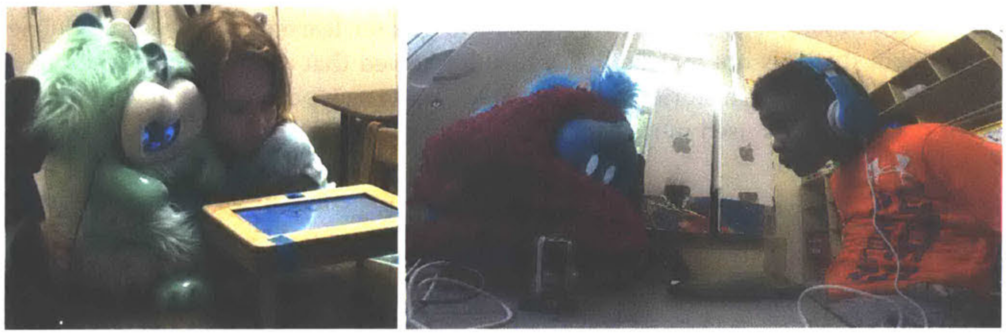
One important aspect of several long-term education studies so far is personalization. Tailoring educational content to individuals can lead to greater engagement and improved learning outcomes. This has been seen in HRI with children (Gordon et al., 2016; Kory and Breazeal, 2014; Leite et al., 2012b; Palestra et al., 2016; Park et al., 2019; Scassellati et al., 2018a) as well as in other learning contexts, e.g., with virtual agents or with older children and adults (D'Mello et al., 2012; Gordon and Breazeal, 2015; Kasap and Magnenat-Thalmann, 2012; Leyzberg, Spaulding, and Scassellati, 2014; Ramachandran and Scassellati, 2015; Thrun et al., 1999). So far, personalization has been studied far more often in longitudinal studies than in one-session studies. This is likely because nearly all personalization studies so far have focused on providing personalized educational content or feedback, using the results of the previous sessions to plan out the content or feedback types for the next session. For example, in Leyzberg, Spaulding, and Scassellati (2014), two different models of personalization were used to determine which lessons individuals received about how to solve logic puzzles over the course of 4 sessions. One model tallied positive and negative demonstrations of a relevant skill; the other model used Bayesian updates to model the probability of mastering a relevant skill. They found that receiving personalized lessons significantly improved participants' performance in the puzzle-solving task.

In the AutoTutor intelligent tutoring system, the system monitored students' affective and cognitive states and selected actions to increase learning and help students regulate negative emotional states (D'Mello et al., 2012). It modeled human tutor dialogue styles and used semantic matching algorithms and conversation rules to pick next dialogue moves in the curriculum script. It detected learning-centered emotions, including engagement, boredom, confusion, and frustration, using facial feature tracking, body posture measurements, and contextual cues. It provided feedback via the virtual tutor's affective facial expressions and verbal responses. They found that the supportive tutor increased students' deep learning, but primarily for low-domain knowledge students, and only the first session—i.e., after there was sufficient context to know the student had problems and actually needed support.

Leite et al. (2009) and Leite et al. (2012b, 2014) studied how enabling a robot to express empathy and support during a chess-playing activity might increase children's engagement over time. The robot detected children's affect and made assessments about the child's emotional state using facial expressions and the chess game's state. It used this information to select appropriate supportive behaviors, such as providing advice or guidance, reinforcing the child's sense of competence, or showing expressions of caring and empathy. It also stored information about prior interactions with the child and used reinforcement learning to learn what support strategies worked best with each child. In addition, in the earlier work (Leite et al., 2009), a human instructor chose level-appropriate chess exercises for each child. This work showed that personalizing the robot's supportive behaviors to individual children increased children's engagement and their ratings of the robot's social presence and helpfulness.

Multiple studies were preliminary, in that they presented personalization strategies but did not test them in full experimental studies or did not report all results as yet. For example, Serholt and Barendregt (2016) used information about children's affective states to determine the pedagogical strategy. Although that paper did not report learning results, they found that children expressed significant social engagement, and the robot's personalization appeared to increase engagement. In a preliminary case study, Palestra et al. (2016) scaled up the difficulty of several social skills games played by three children with autism (e.g., about eye contact, joint attention, and body mimicry), and stopped leveling up when children were unable to complete a task. Two of the children appeared to benefit from the leveling.

More recently, Park et al. (2019) conducted a 7-session study where 44 children aged 4–7 interacted one-on-one with a fully autonomous robot. The robot told stories and children were asked to retell the stories, thereby practicing language skills and learning new vocabulary. This work, which was partly inspired by my earlier work on storytelling robots (discussed below), found that personalizing the robot's story curriculum improved children's vocabulary learning. Children were given language assessments prior to the study, which were used to select the first stories children heard and to select curricula for children in the non-personalized condition. In the



(a) Study 4. (b) Study 5.

Figure 7: Children interact with the DragonBot and Tega robots in Studies 4 and 5.

personalized condition, children's story retells, task behavior (e.g., answering dialogic questions during the robot's narration), and affective arousal were used as input for a Q-learning algorithm, which selected personalized storybooks for each child at an appropriate syntactic and lexical level, while maximizing for engagement and learning.

All of the work so far on long-term interactions and personalization provides evidence for several takeaways. First, personalization of curricula, support, and feedback can improve students' learning, engagement, and positive emotions. Agents that provide support and feedback may be seen as having greater social presence and as being more helpful. The relationships students developed with the agents appeared to influence their engagement and interest in further interaction. Including change and variation in the agents' behavior and the learning content over time can also increase engagement and social interaction.

Some of this work has taken place since the following two studies were performed, and the results of these two studies may have informed the development of later projects. The first study—my master's thesis project—probed children's early English language learning through a playful storytelling activity. The second focused on personalization of affective feedback and motivation during a second-language learning game. Both investigated children's long-term interactions with social robots for language learning and the benefits personalization might provide for increasing engagement and performance.

4.3 STUDY 4: A SOCIAL ROBOT AS A LANGUAGE LEARNING PEER THROUGH TIME

This study examined the potential of a social robotic learning companion to support children's early language development over time (Kory-Westlund and Breazeal, 2015b; Kory, 2014; Kory and Breazeal, 2014). Seventeen children aged 4–6 played a storytelling game with a robot eight times over two months (Figure 7a). The game was based on concepts of emergent literacy, children's learning with peers, and children's natural inclination to create stories during play with friends; it was partly inspired by prior work with Sam the CastleMate (Curenton, Craig, and Flanigan, 2008; Ryokai, Vaucelle, and Cassell, 2003). I created a story game in which children took turns with the robot telling stories about characters in scenes depicted on a tablet. The tablet was embedded in a small wooden table, making its surface a more magical, interactive place and less technology/tablet-like. Target vocabulary words were embedded in the stories the robot told. I evaluated whether a robot that "leveled" its stories to match the child's current language abilities would lead to greater learning and peer modeling than a robot that was not matched. The results showed that all children learned new words, created their own stories, and treated the robot as a peer. Children who played with a matched robot used more words, and more diverse words, in their stories than unmatched children.

This study provided evidence that personalizing the content the robot presents to individual children can lead to greater learning gains, consistent with prior work on personalization. Children will engage a social robot as a peer over multiple encounters, enjoy playing with the robot through time, and will learn new words through peer interactions and play. These results support the idea that children may treat social robots as peers, and that peer learning is a powerful, natural phenomenon.

4.4 STUDY 5: PERSONALIZING AFFECTIVE FEEDBACK INCREASES POSITIVE EMOTIONS

The goal of this study was to investigate the benefits of personalization of affective feedback and motivational strategies during a long-term child-robot language learning interaction. In this study, children played a second-language learning game with a social robot and with a virtual character on a tablet (Gordon et al., 2016; Kory-Westlund et al., 2015b) (Figure 7b). The robot was situated as a peer who was also learning, while the virtual character was positioned as an expert in the second language. Children participated in seven sessions over two months. During this time, the robot personalized its motivational strategies, using both verbal and nonverbal actions, to individual children. The results showed that while all children learned new words, children who interacted with a robot that personalized its affective feedback also showed a significant increase in valence over the two months. This study provided additional evidence that children will engage a social robot as a peer over multiple encounters. Furthermore, personalizing the robot's behavior to individual children can lead to positive outcomes, such as greater liking of the interaction.

4.5 RELATIONSHIPS WITH ROBOTS?

The long-term education studies with children and robots paint an intriguing picture. With sufficient variation in behavior, children easily engaged for many sessions, up to several months. They frequently learned new words from robots. They treated the robots as social others, frequently appeared to grow more comfortable and closer to the robots over time, and often called them their friends. In one study, children engaged longer and were less likely to grow bored if they treated the robot as a peer-like friend (Kanda et al., 2007). This suggests that children's relationships affect how interested they are in interacting and playing—as one might expect, children like playing with their friends.

One recent study compared how families and individuals interacted with either the social home robot Jibo or the voice-AI Amazon Alexa for a month (Singh, 2018). This study was one of the first where children lived with a social robot as opposed to interacting during more limited times at school or in a lab. The agent was placed in their home. Participants in the study could interact with the agent when they liked; they could use any or all of the agent's functionality (classified as social, entertainment, or functional). Singh (Singh, 2018) found that children and older adults primarily used the entertainment and social-relational capabilities of the agents. Adults, on the other hand, were more interested in the agent's functionality; they wanted a device that could do useful things for them. Because Jibo had more social capabilities than Alexa, children preferred the robot; their use of the robot was rooted in its social aspects. The robot was often classified as a friend, child, or member of the family, as opposed to being classified as an assistant. It was also seen as more open, agreeable, and extroverted than Alexa. Children's reactions and preferences showed that they were drawn to the robot's ability to be a social other and they treated as one.

The research so far suggests that children construe robots as social agents with whom they can form friendships and relationships. As discussed earlier (Section 3.6), children also appear to understand that robots are not quite the same as their other human friends, nor quite like their pets, or teachers, or mechanical toys. When robots are designed to act as social agents, children interact with them as social agents. They share gaze, mirror emotions, show affection, help the robots, take turns, and disclose information—all behaviors associated with friendships and close relationships (Gleason and Hohmann, 2006; Hartup et al., 1988; Newcomb and Bagwell, 1995; Rubin, Bukowski, and Parker, 1998). Even with robots that are arguably less social (e.g., without the capability for speech), children still attribute intelligence and talk to them and about them as if they have social capabilities (Druga et al., 2017, 2018). These observations lead to more questions: How are children's relationships with social robots different than their relationships with other entities? What features of the robot impact the relationship children develop? Can a robot actively try to build a relationship, and if so, how would this affect children's engagement and perception of a relationship? In the next chapter, I begin my deeper exploration of children's relationships, including discussion of the kinds of relationships children form with each other, how technology can act as a relational partner, and two investigations into children's affect and behavior with social-relational robots.

5 RELATIONAL TECHNOLOGIES

To understand what it means for children to treat a robot as a social agent with whom they can form a relationship or friendship, we need to understand what children's relationships and friendships are generally like. Then, we can see how theories of relationships apply to children's interactions with technology, namely, social robots.

5.1 RELATIONSHIPS

In the social sciences, relationships are modeled in numerous ways. One common model is the social system, the simplest example of which is a dyad. In a dyad, a relationship is defined as a pattern of interaction, e.g., the interaction of two people whose behavior is interdependent (Berscheid and Reis, 1998; Csikszentmihalyi and Halton, 1981; Kelley et al., 1983). Critically, this model can be applied to human-object relationships, since non-human objects can also significantly influence our patterns of interaction and behavior (Csikszentmihalyi and Halton, 1981). Another important model is the dimensional model, in which relationships are defined in terms of various relational characteristics, including power, social distance, and trust (Berscheid and Reis, 1998; Bickmore and Cassell, 2001; Burgoon and Hale, 1984; Cassell and Bickmore, 2000; Fogg and Tseng, 1999; Spencer-Oatey, 1996; Trope and Liberman, 2010). The dimensional model is important because these characteristics can be manipulated by non-human objects as well to influence the relationship (e.g., Desteno et al., 2012). Other models include provision models, in which relationships are discussed in terms of what people provide for one another e.g., Duck, 1991, as well as economic models, such as social exchange theory, in which relationships are modeled based on perceived costs and benefits of the relationship e.g., Brehm, 1992. Important in relation to provision models is social support theory,which describes how social relationships influence people's cognition, emotions, and behavior (Lakey and Cohen, 2000). Social support theory becomes particularly relevant if we conclude that people can have social relationships with non-human objects. Finally, attachment theory is often discussed in relation to the formation and maintenance of relationships with both humans and objects (Bretherton, 1992; Passman and Halonen, 1979).

Considering these various models, we can see a variety of features that tend to be associated with relationships. First, relationships tend to unfold over time and generally involve multiple interactions. This may be on short timescales, such as repeated encounters over the span of minutes or days, or it may be on longer timescales, such as months or years. Even in short timespans, people's behavior can interdependently influence each other (Davis, 1982). These repeated interactions build up shared experiences—i.e., activities done together in the past or are performing together now. Shared experiences influence later interactions, and are often referenced and remembered later on.

There is also some amount of responsiveness and commitment. Those we form relationships with respond to us, e.g., with social cues in the moments, or social support in response to life events. Attachment and emotion often come into play; we may feel positively or negatively about interacting with certain people. Friendship relations often involve positive feelings, trust, and attachment, such as enjoying one another's company and depending on one another. Friendship relations often involve reciprocity as well, such as exchanging favors, reciprocating contact, dialogue, and connection, and being responsive in turn.

5.2 YOUNG CHILDREN'S FRIENDSHIPS

The history of research on children's peer relationships is long and rich (Hartup et al., 1988; Newcomb and Bagwell, 1995; Rubin, Bukowski, and Parker, 1998). Young children's friendships follow similar patterns as adult friendships and include many of the same general features of relationships—one of the more important features was frequent interaction. Children's friendships are not simply about acceptance or rejection of peers; acceptance is different from friendship (Parker et al., 2015). Friendships are voluntary and reciprocally affirmed (Rubin, Bukowski, and Parker, 1998). Children's friendships involve multiple social provisions, including companionship, intimacy (e.g., sharing secrets), conflict, instrumental help, and empathy/affection (Buhrmester and Furman, 1987; Gleason, 2002; Gleason and Hohmann, 2006; Ladd, Kochenderfer, and Coleman, 1996). For adults, a sense of reliable alliance—i.e., a lasting and dependable bond with another—is often seen as central to friendship, but is less important for children. In addition, when compared to their relationships with parents and siblings, children's friendships with peers are more often associated with nurturance, while siblings afforded more conflict, and parents were sources of instrumental help and power (Gleason, 2002). There are often gender differences in studies of young children's friendships. Girls may draw more distinction between reciprocal and unilateral friends, often rate intimacy and reliable alliance higher than boys, and in general, may understand social relationships earlier than boys, perhaps due to an earlier focus on dyadic relationships (Benenson, 2014; Gleason and Hohmann, 2006).

5.3 RELATIONAL TECHNOLOGIES

5.3.1 *Technology as Relational Partner?*

If technology can provide the same kinds of interaction opportunities and features associated with relationships that people provide for each other, then it can be relational. Relational is different than just being social—it is the behaviors that contribute more directly to building and maintaining an ongoing relationship. This may include numerous social behaviors, such as the use of nonverbal cues and contingency, but is a larger category that includes additional behaviors, which will be detailed in Section 6.2.1.

Prior work has shown that technological agents, such as robots and virtual humans, can be created with responsiveness and interactivity. They are frequently ascribed social presence (e.g., Biocca, Harms, and Burgoon, 2003; Leite et al., 2009) and can evoke attachment, trust, and emotion (e.g., Batliner, Steidl, and Noth, 2011; Bickmore, Schulman, and Yin, 2010; Desteno et al., 2012; Hancock et al., 2011; Kidd and Breazeal, 2008; Turkle et al., 2006a; Weiss, Wurhofer, and Tscheligi, 2009). They can be created with long-term interaction in mind, with features such as memory and personalization (e.g., Bickmore and Picard, 2005; Lee et al., 2012a; Leite, Martinho, and Paiva, 2013; Leite, Pereira, and Lehman, 2017).

5.3.2 *Relational Agents*

Bickmore and Picard (2005) introduced the concept of relational agents: computational artifacts that build long-term, social-emotional relationships with users. They argued that although there is no agreed-upon definition in the social sciences of what relationships are, nothing in the various approaches for understanding relationships prevents computers or other technologies from being a relational partner. Thus, relational agents could include virtually embodied agents, such as virtual humans and other computer agents, as well as physically embodied agents, such as social robots. That said, in later work (e.g., Bickmore, Schulman, and Yin, 2010), Bickmore uses the term relational agents more narrowly to refer exclusively to conversational virtual humans, primarily in a healthcare context with an adult population. Here, we adopt the term relational technologies to refer to the broader category of relational agents—i.e., all agents that can build long-term, social-emotional relationships with users, not only conversational virtual humans. Technologies that are not specifically designed to establish or maintain relationships, even if users interact with them longitudinally, are excluded.

5.3.3 *Related Concepts*

5.3.3.1 *Transactional AI and Persona AI*

Numerous transactional AI and persona AI systems are on the market, such as Alexa and Google Home. These systems often interact longitudinally and generally gather information about the user in order to personalize the experience of using the system. Their purpose is primarily functional and transactional: they can perform tasks and provide information as requested. They generally do not attempt to build or maintain a social-emotional relationship yet, though future iterations of these technologies will likely do so. However, users may project a relationship onto them and form emotional attachments to the agents, as was seen for some users in a recent study where people interacted at home for a month with the voice-agent Alexa (Singh, 2018). New to this category are home social robots like Jibo. These agents are designed to be socially interactive, as opposed to merely transactional. Jibo was marketed as "a member of the family" and in Singh's recent study (Singh, 2018), she found that many users were drawn to the social interactions the robot provided and referred to it as a family member or a friend. Persona AI that are designed as social agents have the potential to be relational technologies, since their capabilities could be extended to include not only social interaction but also relational behavior and long-term relationship formation.

5.3.3.2 *Relational Objects*

Sherry Turkle has used the similar terms relational objects and relational artifacts to refer to technologies that have "states of mind," in that they have more going on inside than any prior computational object and encounters with them may be enriched by understanding their inner states (e.g., Turkle et al., 2006a,b). Relational objects/artifacts could be said to have the potential for relationships—they may have social awareness and may be perceived as social others, but they are not necessarily things a user has long-term social-emotional relationships with, nor do they necessarily attempt to build or maintain such relationships.

5.3.3.3 *Transitional Objects*

Winnicott explored the idea of transitional objects (Winnicott, 1953). Transitional objects are often soft and cuddly such as teddy bears, dolls, and blankets. Children use them in developing early self-other awareness, as a replacement for the mother-infant bond. As such, the objects are often seen as symbolic of the self, others, or of a relationship. They are not studied as things with which one is having a relationship, and the relationship children form with these objects are one-sided projections.

5.3.3.4 *Imaginary Friends and Parasocial Relationships*

Children's imaginary friends come in two general categories: invisible friends and personified objects (Gleason, 2002; Taylor, 2001). They are similar to some transitional objects in that children project a relationship onto them, creating a friend that provides them with some aspects of their real relationships, such as conflict, help, and nurturance. Children are generally quite aware that their imaginary friendships are pretend (Taylor, 2008). Children's relationships with imaginary friends are somewhat similar to their parasocial relationships with media characters. Parasocial relationships are one-sided, emotional relationships developed with characters, e.g., in games or from television, that take on a self-other quality rather than a self-avatar quality (Brunick et al., 2016; Calvert, 2017; Richards and Calvert, 2017). While media characters are created by an external author, children's imaginary friends are solely the creation of their imagination. However, in both cases, the relationships are one-sided. They are the child's projection of a relationship onto another entity that is not an agent in its own right.

5.4 ROBOTS AS PEERS, REVISITED

Earlier, I discussed the importance of children's peers for their learning and development (Section 3.1). Children's peers provide models for imitation and emulation; opportunities for engagement in dialogue, discussion, and conflict; support for learning and practicing more advanced skills through cooperation; reinforcement of social competence through feedback and observation of consequences; and much more. The research so far discusses how peer learning might occur but does not thoroughly address precisely what modulates peer learning. Are all peers approximately equivalent as sources to promote learning, or is something about some peers that makes them "better inputs" than others? To place the question in the context of social robots, what is it about a social robot could lead children learn more, or less?

One possibility arising out of our recent research is children's view of a social robot as a friend-like peer and their treatment of it as a social-relational other. For example, Huili Chen found in a recent study that children reacted better and learned more with a peer robot than with a robot situated as an expert/teacher (Chen, 2018). Children have learned early math concepts more effectively from media characters when they have stronger parasocial relationships with those characters (Gola et al., 2013; Richards and Calvert, 2017). In addition, we have been seeing children observe, imitate, and learn through their imitation of social robots across a variety of domains in the same way they observe and emulate their peers. Peer-to-peer modeling is a key aspect of social learning theory (Bandura, 1971; Bandura and Walters, 1963; Rubin, Bukowski, and Parker, 1998). For example, children displayed more curiosity when a robot modeled curious behaviors, such as asking questions and being interested in learning (Gordon, Breazeal, and Engel, 2015a). A robot modeling a growth mindset encouraged children to display a growth mindset (Park et al., 2017b). Children mimic a robot's affect (Gordon et al., 2016). This work provides evidence that children may be treating robots the same way they treat peers.

When looking at human-human interactions, some prior work has found that children's relationships are related to their learning. For example, literature in education and psychology has found that the social bonds between children and their teachers can predict learner performance (Wentzel, 1997). In peer-to-peer tutoring interactions, rapport and positive relationship, as measured via increased convergence and entrainment, led to improved learning outcomes (Sinha and Cassell, 2015a,b). Work with virtual agents and intelligent tutoring systems has found that an agent's social presence can increase engagement (Lester et al., 1997); use of nonverbal mirroring and behavioral mimicry can increase an agent's likability and persuasiveness (Bailenson et al., 2005); and use of nonverbal mirroring and affective support can decrease frustration and increase flow (Burleson and Picard, 2007). In addition, improving the relationship and working alliance with a social robot can lead to greater engagement and improved health outcomes (Kidd and Breazeal, 2008). However, while these studies show links between an agent's social behavior and rapport-building behaviors and relevant learning emotions, they do not specifically show links with learning.

A few recent studies have explored the effects of building rapport on learning In these studies, middle school students and undergraduate students taught a social robot how to solve math problems as a way of learning to solve the problems themselves (Lubold, 2017; Lubold, Walker, and Pon-Barry, 2016; Lubold et al., 2018). Because the robot was positioned as a less-competent agent, the goal was for students to learn by teaching. They compared a robot had voice-adaptive (speech entrainment) capabilities that adjusted the pitch of its text-to-speech voice to match that of the human to one that did not. They also compared whether adding social dialogue to build rapport had an impact, regardless of speech entrainment. Students who worked with the voice-adaptive, social dialogue agent showed the most learning, though there were no differences in self-reported rapport between conditions. However, these studies did not control for the expressivity of the robot's voice. In the work reported below, we have found that a robot with a more expressive voice can lead to greater engagement and learning than a robot with a flat, less expressive voice (Kory-Westlund et al., 2017b). Thus, since Lubold and colleagues compared a text-to-speech voice to a more a expressive, albeit contingently adapted, voice, an open question is whether the effects seen are strictly a result of the voice entrainment or a result of the robot's voice being more expressive.

All of this work provides some evidence for links between rapport, relationship, engagement, and learning, as well as opening more questions about the roles of rapport, expressivity, and social behaviors. I hypothesize that children's view of the robot as a friend-like peer may impact the kind of relationship they develop, how engaged they are, and how they interact with it—all of which could impact their learning. The next two studies discussed below investigate links between children's emulation of the robot's storytelling and language use, and links between their emulation, affect, engagement, and learning.

(a) Study 6. (b) Study 4, again.

Figure 8: Children play with Tega and DragonBot during Studies 4 and 6.

5.5 STUDY 6: ROBOT EXPRESSIVITY IMPACTS CHILDREN'S PEER-TO-PEER MIRRORING

This study examined the impact of a robot's expressive characteristics on children' peer-to-peer modeling with the robot during a story retelling task (Kory-Westlund et al., 2017b). For half the children, the robot's voice was highly expressive; for the other half, the robot's voice was flat, much like a classic text-to-speech engine. The robot told children a story, which had several vocabulary words embedded in it (Figure 8a). Children were asked to retell the story. We found that all children learned new words from the robot, emulated the robot's storytelling in their own story retells, and treated the robot as a social being. However, children who heard the story from the expressive robot showed deeper engagement, increased learning and story retention, and more emulation of the robot's story in their story retells. This study provided evidence that children will show peer-to-peer modeling of a social robot's language. In addition, they will also emulate the robot's affect, and they will show deeper engagement and learning when the robot is expressive.

5.6 STUDY 4, REVISITED: PHRASE MATCHING AND LINGUISTIC SIMILARITY OVER TIME

One open and important question was whether children would mirror the robot's language long-term, and furthermore, if they did, whether this would be related to the robot's personalization or relationship. We had seen in Study 6 (Kory-Westlund et al., 2017b) that the expressivity of the robot's voice impacted children's engagement and learning, including children's phrase mirroring during a story retell. That study, however, only examined one session with the robot. In an earlier study, (Gordon et al., 2016) (Section 4.4), we observed children mirror a robot's affect over multiple sessions, during suggesting that they do mirror the robot's behaviors over time. However, this study did not examine children's language use.

Thus, in order to investigate children's mirroring of a robot's language over time, I performed new analyses on an existing dataset from Study 4 (Kory and Breazeal, 2014) (Section 4.3) that included stories from 14 children, who had played a storytelling game with a robot 1–2 times per week for 8 sessions (Figure 8b). I examined children's use of key vocabulary words and key phrases used by the robot during the full interaction sessions as well as just during their storytelling. I examined children's emulation of the robot's stories during their own storytelling to see whether they mirrored the robot's phrases more over time. Finally, because the study did not explicitly measure children's relationship with the robot, I also examined children's language style matching (LSM), which is a measure of overlap in function words and speaking style as opposed to content words. Our phrase matching metrics look primarily at content words. Research has shown that the more "in sync" two people are, the more they will match function words in their speech; it may reflect rapport and relationship (Babcock, Ta, and Ickes, 2014; Ireland et al., 2011; Niederhoffer and Pennebaker, 2002; Pennebaker, Mehl, and Niederhoffer, 2003; Tausczik and Pennebaker, 2010).

5.6.1 *Methods*

5.6.1.1 *Dataset*

The data included 206 stories from 14 children and full transcripts from 17 children (3 children did not tell stories), all of whom had participated in an 8-session study with a socially interactive robot. During each session, the robot briefly engaged the child in conversation, then showed a story scene on a tablet and told a short story. Next, the child was invited to tell their own story about the scene. The robot then showed a second story scene and told a second story, and the child was invited to tell a second story. After a brief closing conversation, the interaction ended. In some sessions, the robot showed a story scene but asked the child to tell a story first. In the second half of the study (sessions 5–8), half the children heard stories matched to their language ability (*Matched* condition), while the other half heard stories that were not matched (*Unmatched* condition).

5.6.1.2 *Keywords and Key Phrases*

Using the transcripts of the sessions and automated software tools, I counted the number of times children used key phrases that the robot had used (e.g., "Once upon a time," "I'll tell a story about...," "See you later, alligator!"). The goal was to see whether children adopted any of the robot's frequently used phrases—mimicry that might be indicative of a closer relationship or greater rapport. We also counted how often children used each of the target vocabulary words. Usage might reflect expressive vocabulary ability, which is a stronger indicator of knowledge of a word than the receptive knowledge tested with the vocabulary assessment, as well as mimicry of the robot.

5.6.1.3 *Language Style Matching (LSM)*

To get sufficient data for an LSM analysis, I aggregated all of each child's stories for sessions 1–4 (the first half of the study) and then for sessions 5–8 (the second half). I obtained an LSM score for each set using software tools to access the Receptiviti API (*Receptiviti API* 2017; Tausczik and Pennebaker, 2010). LSM scores range from 0 to 1.00, but more often range from 0.5 to 1.00. The closer the score is to 1.00, the more matching is present.

5.6.1.4 *Phrase Matching*

Children's transcribed stories were analyzed in terms of length (seconds), word count, overall word usage and target vocabulary word usage, and similarity of phrases to the phrases used in the robot's stories. I created an automatic tool to obtain phrase matching scores comparing each child story to each robot story that the child had heard prior to telling the story. For example, a story told by a child in Session 2 was compared to the stories the robot told in Session 1 as well as any stories the robot told before the child in Session 2. My analysis was then threefold: (1) compare each child story to the robot story just prior to it; (2) compare each child story to other stories in the same scene; (3) compare each child story to all stories prior to it. The matching algorithm was as follows:

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

This produced a score reflecting the number of exact matches—i.e., words used in the same order by both the child and robot. It also produced a higher match score for texts that have both more matching phrases and longer matching phrases. I 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 step 5 (counting matching N-grams) used a fuzzy string matching algorithm to determine if N-grams matched.

For exact matches, I used $N = 3$ 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 similar matches, I used $N = 4$, so that when phrases differed by a word, or used a different word in the middle of a similar phrase, they might still match.

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

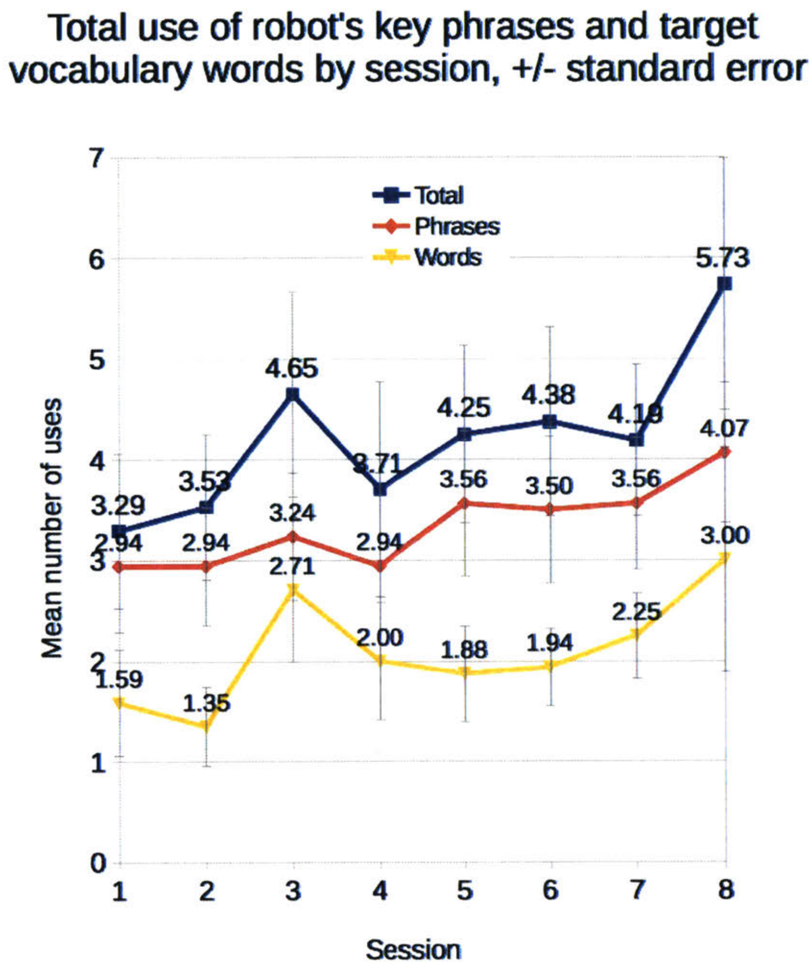


Figure 9: Children's mean use of the robot's key phrases and target vocabulary words in their stories by session.

5.6.2 *Results*

Because these analyses were post-hoc, I corrected for multiple comparisons using the Benjamini Hochberg method (to control the false discovery rate), which indicated that the results with $p < 0.011$ could be considered significant.

5.6.2.1 *Keywords and Key Phrases*

I performed mixed analysis of variance with condition (between: *Matched* vs. *Unmatched*) and mean of sessions (within: sessions 1–4 vs. sessions 5–8) on children's use of the robot's target vocabulary words and key phrases. I observed a trend toward a main effect of session on the total number of key phrases and target vocabulary words children used from the first half to the second half of the study, $F(1,13) = 2.95$, $p = 0.11$ (Figure 9). Children used somewhat more of the key phrases and target words in the second half of the study than in the first half. In particular, children tended to uses the phrases "once upon a time" and "See you later, alligator" more in later sessions.

5.6.2.2 *LSM*

I observed LSM scores ranging from 0.063 to 0.892, with a mean of 0.696 (SD = 0.212). Only two children had scores below 0.500; in both cases, their scores increased from the first half to second half of the study. A mixed analysis of variance with time (within: first half of the study vs. second half) and condition (between: *Matched* vs. *Unmatched*) revealed a trend toward an interaction of time with condition, $F(1,12) = 4.29$, $p = 0.061$. As shown in Figure 10, LSM scores increased slightly for children in the *Matched* condition (first: $M = 0.66$, $SD = 0.25$; second: $M = 0.71$, $SD = 0.23$); the scores decreased slightly for children in the *Unmatched* condition (first: $M = 0.74$, $SD = 0.23$; second: $M = 0.71$, $SD = 0.19$).

5.6.2.3 *Phrase Matching*

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

SUM OF EXACT AND SIMILAR MATCHING PHRASES **Compared to all previously heard stories.** I observed a trend for main effect of time on the mean number of matching phrases used per story, $F(1,12) = 5.65$, $p = 0.035$, and a significant interaction of time with condition, $F(1,12) = 10.0$, $p = 0.008$. Children emulated more of the robot's phrases per story in the first half of the study, and children in the *Unmatched* condition decreased usage more (Figure 11a). I observed a significant interaction of time with condition when looking at the sum of matching phrases across stories, $F(1,12) = 9.81$, $p = 0.009$. Children in the *Matched* condition increased their usage of matching phrases, while children in the *Unmatched* condition decreased their usage (Figure 11b).

Compared to stories heard from the same story scene. I observed a significant interaction of time with condition for the mean number of matching phrases used

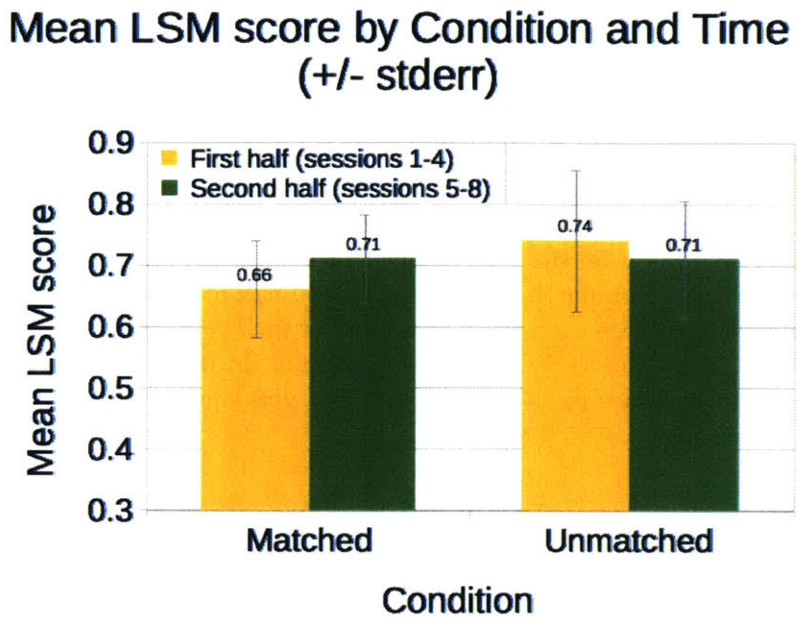
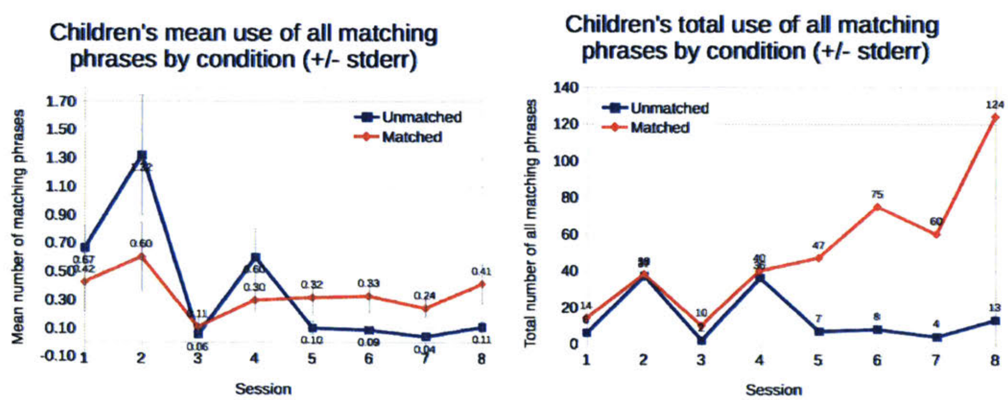


Figure 10: Children's mean LSM scores by condition for the first half vs. second half of the study.



(a) Mean all phrases.

(b) Total all phrases.

Figure 11: Children emulated the robot's phrases during their storytelling. Their emulation increased during the second half of the study in the *Matched* condition.

per story, $F(1,12) = 9.10$, $p = 0.011$. Children in the *Unmatched* condition used fewer matching phrases on average in the second half of the study, while children in the *Matched* condition did not change significantly. There were no significant differences for the sum of matching phrases across stories.

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

5.6.2.4 *Correlations*

Children who emulated more of the robot's phrases during their storytelling also scored higher on the vocabulary posttest, $r_{s15} = 0.511$, $p = 0.052$ (Figure 12a); as did children who used more of the robot's key words and phrases $r_{s15} = 0.532$, $p = 0.041$ (Figure 12b). Children who emulated the robot more during storytelling were also more likely to use more of the robot's key words and phrases, $r_{s15} = 0.688$, $p = 0.003$ (Figure 12c). This pattern was also apparent when looking at the mean of all children's scores for sessions 1–8 (Figure 12d).

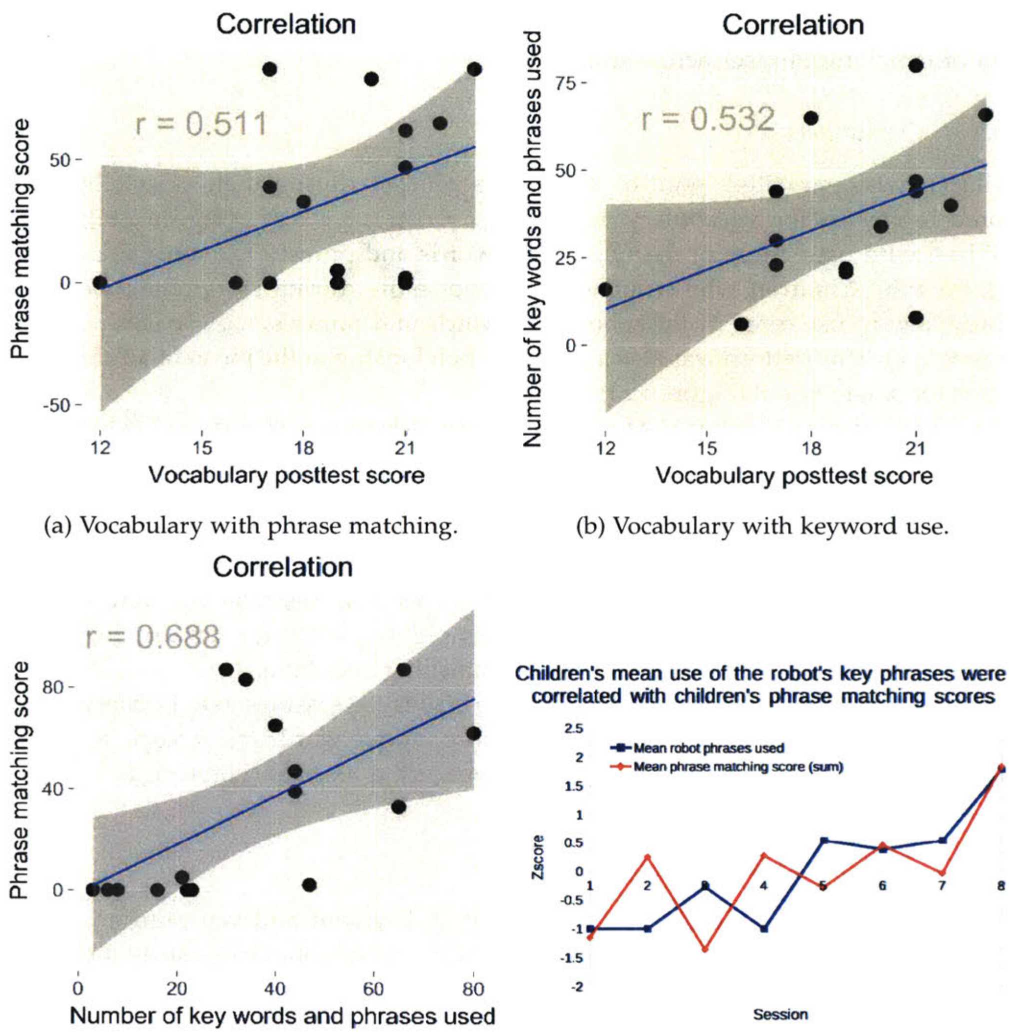
Children who had higher LSM scores during sessions 1–4 were more likely to emulate the robot's phrases during storytelling, $r_{s15} = 0.667$, $p = 0.007$; they were also more likely to use the robot's key words and phrases, $r_{s15} = 0.548$, p = 0.034 (Figures 13a and 13b). The same pattern held for children's LSM scores in sessions 5–8 for phrase emulation, $r_{s14} = 0.732$, $p = 0.003$; and for key word and phrase use, $r_{s14} = 0.554$, $p = 0.040$ (Figures 13c and 13d). Children's LSM scores from sessions 1–4 were strongly correlated with their LSM scores from sessions 5–8, $r_{s14} = 0.802$, $p < 0.001$, suggesting little change in children's rapport and style matching over time.

When looking at the mean of all children's scores for sessions 1–8, I observed that children who told longer stories also used more unique words ($r_{s8} = 0.954$, $p < 0.001$) and spent more time telling their stories ($r_{s8} = 0.715$, p = 0.046) (Figure 14).

5.6.3 *Discussion*

This new analysis of children's phrase matching, keyword and key phrase use, and LSM during a long-term storytelling activity with a social robot suggests that children will indeed mirror the robot's language over time and that their emulation of the robot may be related to their learning. Children slightly increased their use of the robot's keywords and phrases over time. Their emulation of the robot was correlated with their vocabulary posttest scores. Their LSM scores and their phrase emulation during storytelling both increased over time for children in the *Matched* condition, but decreased slightly for children in the *Unmatched* condition; I had also seen differences by condition is vocabulary words learned (with *Matched* scoring higher on the posttest). The differences by condition suggests that the robot's personalization of story level (telling harder or easier stories based on the child's language ability) affected both children's emulation of the robot and their learning.

One limitation of this dataset was the lack of relationship assessments. Children's perception of their rapport and relationship with the robot were not explicitly measured. Children's LSM scores may reflect their relationship to some degree, since prior work examining LSM has linked higher LSM scores between two people to higher



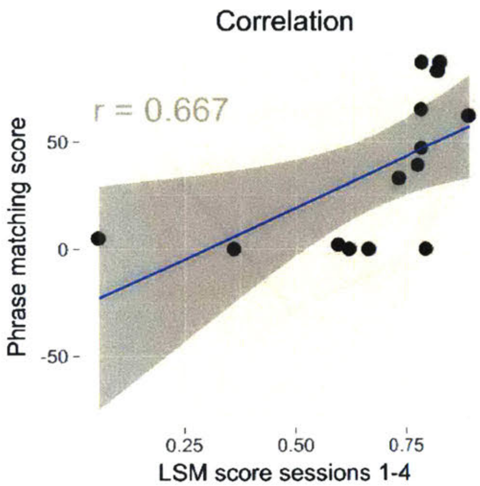
(a) Vocabulary with phrase matching.

(b) Vocabulary with keyword use.

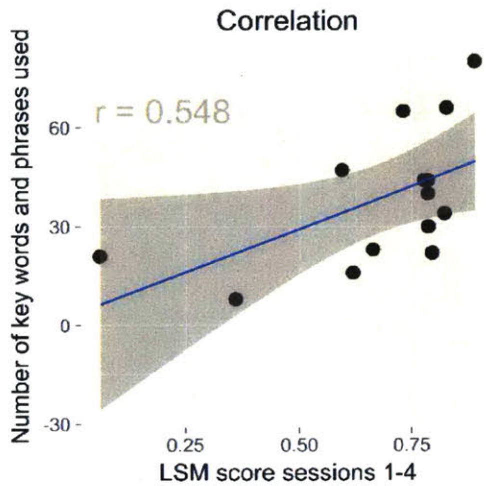
(c) Total phrase matching with total keyword use.

(d) Mean phrase matching with mean keyword use.

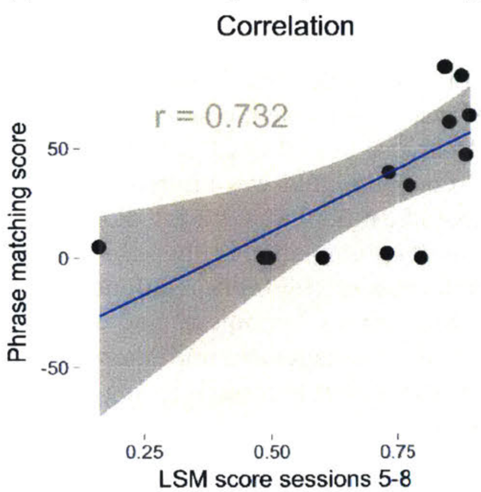
Figure 12: Children who emulated more of the robot's phrases during their storytelling and used more of the robot's key words and phrases scored higher on the vocabulary posttest. Children who emulated phrases were also more likely to use the keywords.



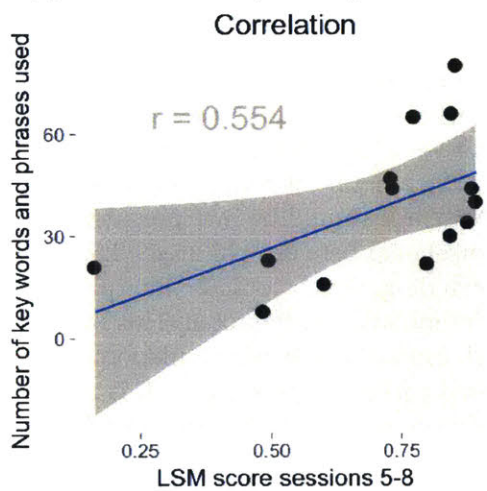
(a) LSM sessions 1–4 with phrase matching.



(b) LSM sessions 1–4 with keyword use.



(c) LSM sessions 5–8 with phrase matching.



(d) LSM sessions 5–8 with keyword use.

Figure 13: Children who had higher LSM scores were more likely to emulate the robot's phrases and keywords during storytelling.

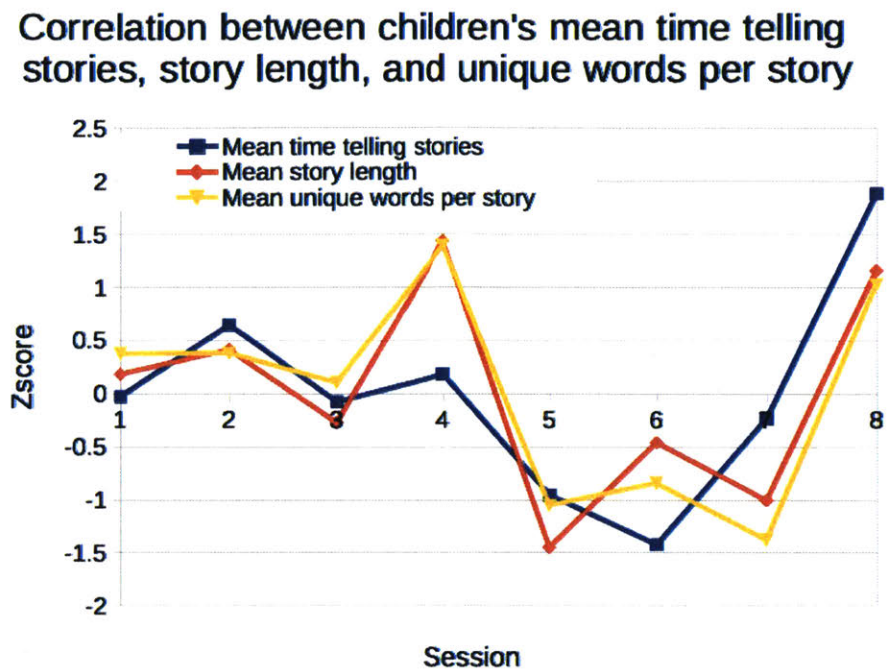


Figure 14: Children who told longer stories also used more unique words and spent more time telling their stories

rapport and a deeper relationship (Babcock, Ta, and Ickes, 2014; Ireland et al., 2011; Pennebaker, Mehl, and Niederhoffer, 2003). However, because the sessions with the robot were fairly short (10–15min apiece) and because not all children told long stories, the amount of conversation between the robot and child was limited. Thus, the LSM scores should be interpreted with a degree of caution. That said, I did observe a correlation between children's LSM scores and their emulation of the robot during storytelling, which could be interpreted as a link between children's relationship and their emulation of the robot. This would be in line with my hypotheses and our prior work linking children's emulation of a robot to their engagement and learning, discussed earlier in this chapter. It is not strong evidence, but is enough to suggest that further research on this topic would be interesting.

It was intriguing to see correlations and differences despite the small sample size and despite the imbalance in the number of children in the *Matched* and *Unmatched* conditions, as well as some imbalance in their language abilities. In particular, I had observed higher scores for children in the *Unmatched* condition across various metrics during the first half of the study (prior to the personalization/matching, which only occurred in the second half of the study). I expect that were the groups more balanced, these initial differences may be smaller or might even disappear, while differences between conditions as a result of the personalization would be larger.

Finally, I found it interesting that children's use of exactly matching phrases decreased over time while their use of unique words increased. This suggests that children were more creative over time in coming up with their own, different stories.

5.7 OPPORTUNITIES

The work discussed in this chapter points to two areas of opportunity. First, we are seeing that children construe social robots as relational others. That is, these robots are part of the broad category of things with which one can have a relationship. They are more than playful objects (Ackermann, 2005) or transitional objects (Winnicott, 1953) since either category would imply that they are only artifacts for projecting onto, for exploration and learning, rather than for being with. These robots are social and relational: They are seen as having some of the properties of pets, toys, computers, artifacts, and friends, but not exactly the same properties as any of these. They may be toys, playful objects, or act like pets, but they may also be smart toys, playful others, and treated as actual pets. The fact that children are treating robots as social-relational others brings new opportunities for engaging children socially in activities involving technology. For example, we could develop new activities to help children develop and practice skills that are best learned in social contexts—such as language or social and emotional skills.

The second area of opportunity involves children's social behavior with peers. Because children are treating social robots as social-relational others, we are seeing children display behaviors with the robots that they display with their human peers, such as emulating the robots' behaviors. This provides new opportunities for creating learning activities where robots can model desirable behaviors—such as curiosity, a growth mindset, and more advanced language use—since children are likely to imitate the robot's speech and actions. Children can learn with social robots the same way they learn from their human peers.

That said, there appear to be moderating factors. For example, children imitated the robot more when it was expressive and when it told personalized stories to them matching their language ability. This suggests a deeper story. Children do not blindly copy social robots. They appear to pay attention to the robot's social cues and the robot's attentiveness to them and to the interaction, using this information to guide their emulation, trust, engagement, and more. There is more to learn here, and the next chapter begins my deeper investigation into how children's rapport and relationship with the robot impacts their engagement, emulation, and learning.

6 RELATIONAL AI

6.1 SOCIAL ROBOTS AS A RELATIONAL TECHNOLOGY

I have found evidence that social robots can be effective tutors and learning companions. The studies and research discussed so far highlight the fact that the design of robots as social agents matters for children's learning. I have explored why these robots are effective and what design features of the robots positively impact children's learning and attitudes through time. For example, children pay attention to the robot's nonverbal social cues to guide their learning. Factors such as the contingency of the robot's nonverbal behavior and the robot's expressivity impact children's engagement, learning, and judgments of the robot's credibility. Children apply social judgments to the robots and treat them as peers. We have also begun seeing children perform peer-to-peer modeling of curiosity, affect, mindset, and language with the robots (Gordon, Breazeal, and Engel, 2015a; Kory-Westlund et al., 2017b; Kory and Breazeal, 2014; Park et al., 2017b, 2019) (also see Section 5.6). The evidence so far leads me to argue that a key aspect of why social robots benefit children's learning is their nature as a relational technology.

The remainder of this thesis explores social robots as a relational technology. Specifically, I explore how adding relational attributes and capabilities to social robots impacts child-robot learning interactions. My hypothesis is that the relationships children form with the robot will affect their engagement and learning, particularly over long-term interactions.

6.2 RELATIONAL AI

To enable social robots to reach their full potential as relational technologies, especially for deployment during long-term interactions in real-world contexts, they need to be autonomous. I will use the term relational AI to refer to autonomous relational technologies—i.e., autonomous technologies that attempt to build and maintain long-term, social-emotional relationships with users. Relational AI thus comprises a subset of relational technologies, since, e.g., some may have some but not all the features necessary for relationship establishment and maintenance or may be teleoperated and not autonomous. That said, research may frequently use teleoperated or semi-autonomous techniques while exploring what relational AI can do, how people respond to it, and how we ought to use it.

Relational AI is human-centered and interpersonal AI, in that it is designed to use human social and relational behaviors in order to be more understandable and relatable to humans, and in order to build and maintain relationships in a way that humans are used to. It is designed to treat people in humanistic ways.

Relational AI refers especially to the underlying computational models, algorithms, and mechanisms by which a relational technology operates. As part of relationship establishment and maintenance, it may maintain and update a model of a person in order to respond appropriately to that person.

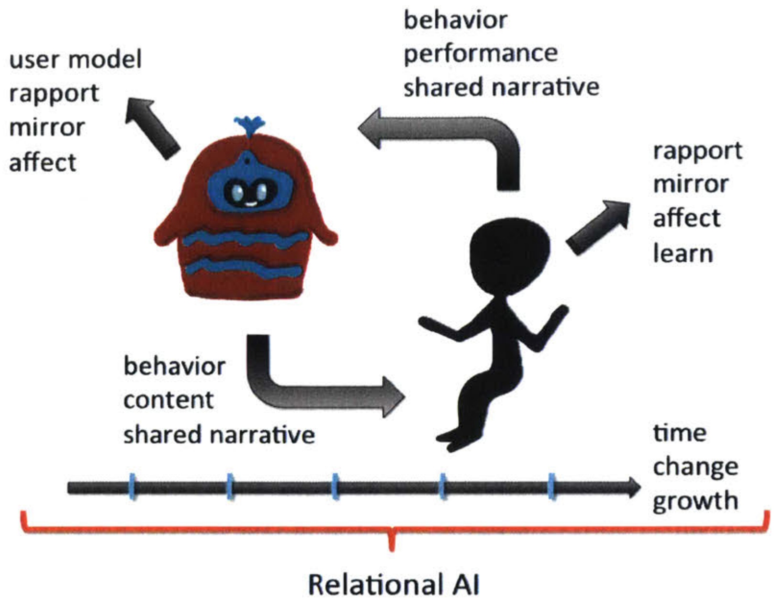


Figure 15: Diagram showing the components of relational AI when implemented in a human-robot interaction system.

Below, I discuss the features that are associated with relational AI. I reference these features in the next chapters, as I develop and evaluate relational AI tools for use in a social robot that leads children in language learning activities over time.

6.2.1 *What Makes AI into Relational AI?*

Relational AI is an autonomous relational technology—it attempts to build and maintain long-term, social-emotional relationships with users. It is human-centered, collaborative, interpersonal, relational, and reciprocal. Features that are necessary and sufficient for AI to qualify as relational AI include repeated encounters, shared experience, change, responsiveness, emotion and positive affect, and reciprocity (Figure 15). These are all features that tend to be associated with relationships (Section 5.1).

Repeated encounters. Relationships are longitudinal—they generally develop through time and involve multiple interactions. Relational AI should be designed to handle repeated interactions with users through time.

Shared experiences. Humans generally have a sense of past, present, and future, which is reflected in our relationships. We acknowledge our shared experiences through time via references to our past and present together, as well as looking forward to future activities we might do together. For example, sharing a humorous experience during an initial encounter with a stranger led to increased ratings of closeness (Fraley and Aron, 2004). Relational AI should create and reference a shared narrative with users. This may require an internal state that represents the user over time that can be updated during interactions.

Change. As part of creating and referencing shared experiences, relational AI should change over time. More specifically, relational AI should change as a result of the interaction with the user over time—it is not enough to follow a changing but scripted storyline (e.g., Gockley et al., 2005). The change has to be perceived as "meaningful" in that the activities performed with the user (i.e., shared experiences over repeated encounters) must be clearly seen to affect the relational AI's outward attitudes, emotions, or behavior. For example, people in close relationships may converge toward similar emotional reactions to events (e.g., Anderson, Keltner, and John, 2003) or similar choices of food (Bove, Sobal, and Rauschenbach, 2003). Again, this may require an internal state that represents the user over time. At present, nearly all the existing work on social robots as learning companions and tutors has focused on how the child is affected by the robot—that is, the child grows and changes as a result of the interaction, but the robot does not. With relational AI, the robot will also need to change. As noted earlier, a growing number of studies are examining autonomously changing/personalizing the robot's behavior and/or the task content as a result of the child's behavior or performance (e.g., Gordon et al., 2016; Lubold, 2017; Lubold, Walker, and Pon-Barry, 2016; Lubold et al., 2018; Park et al., 2017a; Park et al., 2017b, 2019; Ramachandran and Scassellati, 2015; Scassellati et al., 2018a). These studies have shown that personalization (i.e., a particular kind of change) can increase children's engagement and learning, and have opened many questions about how personalization and change might affect the child-robot relationship.

Responsiveness. Relational AI should ideally model a positive relationship. One element of successful, positive human relationships is rapport (Berscheid and Reis, 1998), which is often indicated via behavior such as entrainment/mirroring and social reciprocity (Davis, 1982; Dijksterhuis, 2005; Dijksterhuis and Bargh, 2001). These behaviors are part of being responsive to users. Relational AI should respond and react to users, e.g., by using appropriate social cues in the moment, or personalizing its feedback, entrainment, or behavior for individual users.

Emotion and positive affect. As a human-centered technology, relational AI should respond appropriately to users' emotional states. Prior work has found that mismatches between users' emotions and the reactions of technology can negatively affect user perceptions and performance during interactions (Jonsson et al., 2005). Promoting trust can be important for many kinds of applications. As one example relevant to education, trust can affect who children treat as credible informants (Harris, 2007, 2012). Relational AI designed to act as a friend-like agent may also need to promote positive affect or attachment as well; as discussed earlier, children's friendships often involve empathy and affection (Gleason, 2002).

Reciprocity. The idea of social reciprocity relates back to responsiveness as well as shared experiences through time. As discussed earlier, relationships often involve various reciprocal behaviors, such as disclosing information, helping, conversing and engaging in activities together, and providing companionship. Many of these behaviors are recognized as important in children's friendships (Gleason, 2002). Relational AI should use these kinds of reciprocal behaviors, and attempt to recognize and be affected by the user's use of these behaviors in turn.

6.2.2 *What's New About Relational AI?*

If a system—be it a robot, virtual agent, or other technological agent—includes all these features associated with relational AI, then it can be considered to be relational AI. This definition of relational AI extends the earlier definition presented by Bickmore and Picard (2005) of relational agents (Section 5.3.2), which was most frequently used to refer narrowly to virtual humans, primarily in a healthcare contex with an adult population. Furthermore, relational agent systems may or may not have included all necessary features of relational AI. Thus, my definition of relational AI builds on and extends the definition of relational agents to include robots and require all features of relational AI.

When we compare my new work with relational AI to prior work with relational agents and other social technologies, we can examine a number of dimensions, including:

- The domain—e.g., healthcare, education, therapy, entertainment;
- The population—e.g., children, adults, the elderly;
- The agents—were they embodied and co-present, virtual, etc.?
- What features of relational AI were included?
- How was the relationship assessed—or was it?
- What was the impact of the relationship on the interaction?
- What were implications of the work, e.g., for ethics or design?

My work focuses on education with children and co-present, embodied robots. In the final study (Chapter 9), the robot included all necessary features of relational AI. I assess children's relationships with social robots and examine the impact of children's relationships on the goals of the interaction (i.e., on learning). Finally, I discuss the implications of my work with regards to ethics and future design (Chapters 12 and 13).

As outlined earlier in Chapters 2 and 4, prior work with social robots has included some, but not all, of the features of relational AI, such as autonomy and personalization, though personalization has frequently focused on curriculum rather than social or relational features (Chang et al., 2010; Kanda et al., 2007; Kennedy et al., 2016a; Lee et al., 2011; Leite et al., 2012a, 2014; Movellan et al., 2009; Tanaka and Matsuzoe, 2012). Some of this work has focused on educational activities with children, but none have assessed children's relationships or their impact. Many robot tutors are in STEM domains and focus on somewhat older children (e.g., Kennedy et al., 2016b). With virtual agents, most prominently Justine Cassell's work, there are projects that have focused on a similar population and domain as mine (Bickmore and Cassell, 2001; Cassell, 2004; Cassell and Ryokai, 2001; Cassell et al., 2007; Finkelstein et al., 2012; Ryokai, Vaucelle, and Cassell, 2003; Tartaro and Cassell, 2015). Some recent work has examined relevant aspects of relationships, such as rapport and social responsiveness (e.g., Sinha and Cassell, 2015a,b). Finally, the work with relational agents as well as robots in healthcare domains has also focused on long-term interactions and personalization, usually with adults, and has touched on many aspects of relational AI, though not all at once in the same system Bickmore and Picard, 2005; Bickmore et al., 2016; Bickmore, Schulman, and Yin, 2010; Bickmore et al., 2018; Kidd and Breazeal, 2008.

In summary, my work extends the work on relational agents into a new domain, with a new population, and new capabilities for the agent. I extends the work on social robots for children in education by adding new relational AI features, assessing children's relationships and the impact of children's relationships, and discussing this new data with respect to ethical and design considerations. Working with children is especially important given that there are many ethical questions concerning humans developing relationships with technology (discussed further in Chapter 13), and many of these concerns are most contentious with children. Studying children's interactions with relational AI can thus add significant data to our understanding of how to design and use relational technology.

7 MEASURING CHILDREN'S RELATIONSHIPS WITH ROBOTS

7.1 HOW TO MEASURE CHILDREN'S RELATIONSHIPS?

A critical aspect of developing social robots with relational AI is evaluating whether anything they are doing is effective at building or maintaining a relationship. Because I planned on evaluating relational AI in the context of children learning with robots, I needed ways of assessing children's relationships and perceptions of the robot as a social-relational other. We have previously seen children display a wide variety of behaviors that may be related to their social relationship and treatment of the robot as a social agent—such as sharing gaze, mirroring emotions and language, displaying affection, and taking turns. We had not, however, explicitly asked children about their construal of the robot as a social-relationship agent, about their feelings of closeness, or used assessments specifically designed to examine features of the relationship they might be forming.

Thus, I looked in the psychology literature and child-technology interaction literature for existing assessments that might help us understand children's relationships. In child-robot interaction studies, I found assessments for measuring engagement, the robot's social presence, various attributes of the robot (e.g., can it think, can it feel, can it break), and performance (e.g., learning gains), but nothing explicitly examining relationships. Some studies examined children's behavior—e.g., gaze, affect, and speech patterns—as we have previously. Many assessments consisted of multiple-choice questionnaires, with the numerical values frequently replaced with a range of very frowny to very smiley faces. Some studies used variations of the Fun Toolkit, a set of assessments for measuring children's engagement and fun (Read, 2008; Read and MacFarlane, 2006). I encountered numerous studies that targeted older children or adults, and involved self-report questionnaires (Aron, Aron, and Smollan, 1992; Bartneck, 2008; Horvath, 1989; Mendelson and Aboud, 1999; Nomura et al., 2006). Those that measured people's relationships with each other as opposed to their relationships with robots could be easily adapted to work with robots. For example, Kidd and Breazeal (2008) used the Working Alliance Inventory (Horvath, 1989), which measures the quality of alliance between two individuals, to enable a weight-loss coach robot to assess, model, maintain, and repair a relationship over time.

One issue with these questionnaire-based assessments, however, is that they necessitate reading, comprehension, and self-reflection skills that the young children I work (ages 4–7) with may not have developed yet. These children are sometimes pre-reading, may have shorter attention spans, and may not be able to fill out standard Likert-style questionnaires (Chambers and Johnston, 2002). In the psychology literature, observational methods have been used most commonly to assess younger children's peer relationships and friendships. These methods code various aspects of children's relational behavior, such as companionship, aid, exclusivity, frequency of contact, connectedness, conflict, and physical proximity (Hartup et al., 1988; Ladd, Kochenderfer, and Coleman, 1996; Youngblade, Park, and Belsky, 1993). A few studies asked children directly about their relationships or used behavioral methods, some of which served as inspiration here (Favazza and Odom, 1996; Favazza, Phillipsen, and

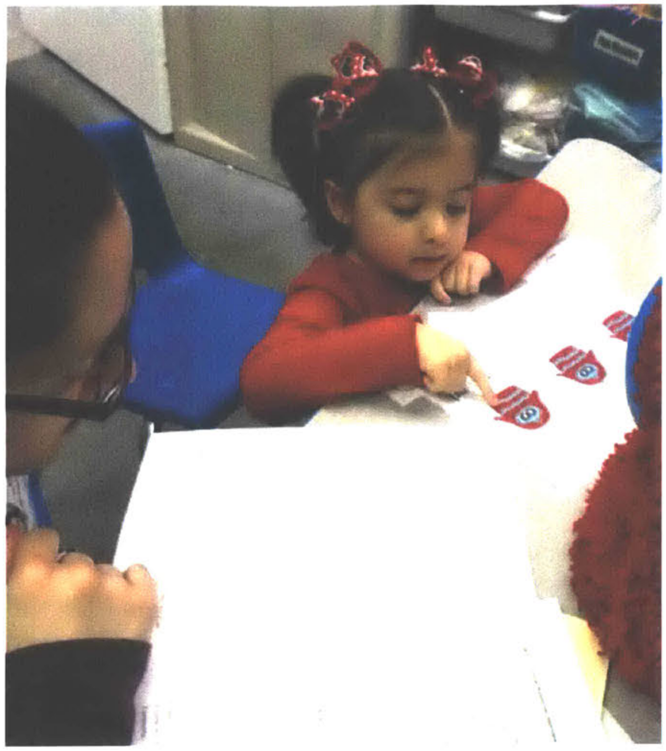


Figure 16: A child answers questions in the SRI.

Kumar, 2000; Furman, 1996; Gleason and Hohmann, 2006; Nowicki and Duke, 1974; Rotenberg, 1995).

Using the assessments and methods found during this literature search as a starting place, I developed multiple new assessments for measuring children's relationships with robots. They include several sets of interview questions asking about children's friendships and relationships, including questions targeting provisions of children's friendships (Gleason, 2002; Ladd, Kochenderfer, and Coleman, 1996) (the Social-Relational Interview, Narrative Description, Social Acceptance questions, and questions about memory, rapport, and judgment); a pictorial measure of closeness and interconnectedness adapted from the Inclusion of Other in Self scale (Aron, Aron, and Smollan, 1992; Gächter, Starmer, and Tufano, 2015); a picture sorting task about children's perception of the robot's animacy and human-likeness (Picture Sorting Task); and behavioral scenarios that are embedded in the interaction with the robot, including one targeting children's self-disclosure as a measure of close relationship (Buhrmester and Furman, 1987; Gleason, 2002; Rotenberg, 1995) (Self-disclosure Task, Anomalous Picture Task, Story Negotiation, Extra Picture, and several more). These assessments are described in more detail below.

A pilot study of four of these assessments (Social-Relational Interview, Narrative Description, IOS task, Self-disclosure task) is described in (Kory-Westlund et al., 2017c; Kory-Westlund et al., 2018). The instructions and materials for those four assessments are available on figshare: 10.6084/m9.figshare.5419102. The Picture Sorting Task and the Social Acceptance Questionnaire are available at: 10.6084/m9.figshare.7575911. The remaining assessments are available with the Study 7 and Study 9 materials at 10.6084/m9.figshare.7175273 and 10.6084/m9.figshare.7627289.

7.2 ASSESSMENTS FOR MEASURING CHILDREN'S LONG-TERM RELATIONSHIPS WITH ROBOTS

7.2.1 *Social-Relational Interview (SRI)*

The Social-Relational Interview (SRI), described in (Kory-Westlund et al., 2018), asks a series of question about how children think the robot feels and acts, with five questions targeting provisions of children's friendship (conflict, helping, information sharing / disclosure, wanting companionship, empathy/affection) Gleason and Hohmann, 2006; Ladd, Kochenderfer, and Coleman, 1996 and two questions asking whether the robot was genuine in its feelings (i.e., just pretending or really did want friends and like the child). Children are provided with a graphical answer sheet showing the robot for helping them respond (Figure 16). Children are asked to explain their answers as well as whether they would feel or act the same way. These explanations are coded for the types of justifications children use, such as talking about the robot's attributes, about actions the robot took, or about moral rules that might influence its behavior.

7.2.2 *Inclusion of Other in Self (IOS)*

The Inclusion of Other in Self (IOS) task uses a pictorial scale to measure closeness and interconnectedness (Aron, Aron, and Smollan, 1992; Aron et al., 1991). I adapted it for use with children, as described in (Kory-Westlund et al., 2018). In this task, children are shown seven pairs of increasingly overlapping circles. 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.

7.2.3 *Narrative Description*

The Narrative Description task, described in (Kory-Westlund et al., 2018), uses a puppet to ask children to describe their best friend and the robot. Their responses are coded for content and length in order to examine how the descriptions compare over time.

7.2.4 *Self-disclosure Task (SDT)*

The Self-Disclosure Task (SDT), described in (Kory-Westlund et al., 2018), is based on the idea that children disclose information to their friends, and disclose more information and more personal information the closer they are to someone (Buhrmester and Furman, 1987; Gleason and Hohmann, 2006; Rotenberg, 1995). Thus, the robot discloses information and prompts for disclosure in return. Children responses are coded for content as well as length.

7.2.5 *Social Acceptance Questionnaire (SAQ)*

The Social Acceptance Questionnaire (SAQ) uses several questions taken from the Social Acceptance Scale for Kindergarten Children (Favazza and Odom, 1996; Favazza, Phillipsen, and Kumar, 2000), which is a scale that measures accepting children are

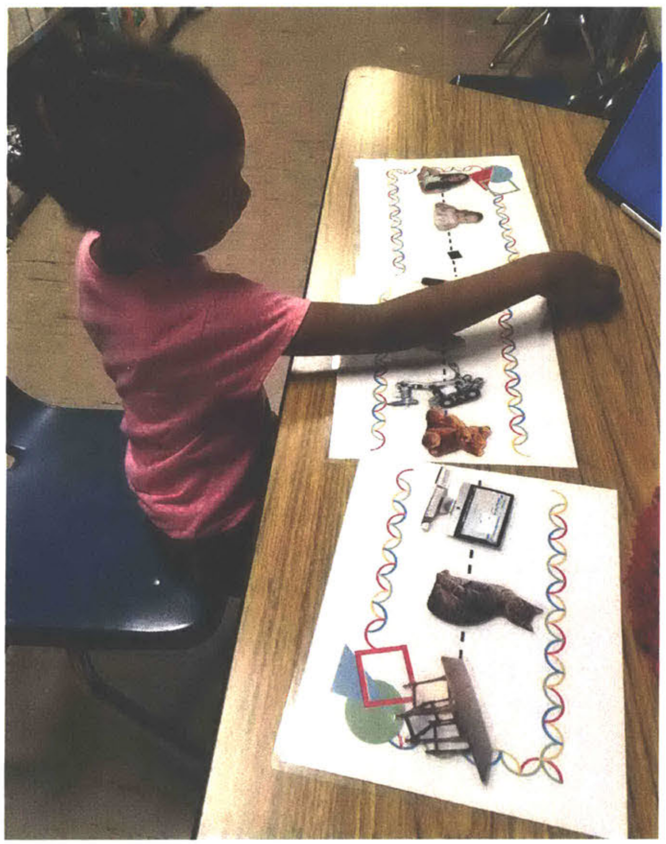


Figure 17: A child sorts the entities in the Picture Sorting Task.

of peers with disabilities (e.g., peers who may be unable to walk, hear, or see). We selected four of the most relevant questions that asked children directly about, e.g., whether they would like to be good friends with a child with disabilities or with a child who could not hear well. We ask these questions both about another child and about the robot in the study. Children are asked to point to their responses on a graphical answer sheet depicting a happy face with the word "YES", a sad face with the word "NO", and a questioning face with the word "MAYBE".

We used these questions because robots often have significant limitations, e.g., with hearing and understanding because of technical challenges regarding automatic speech recognition and language understanding. These limitations can be framed as disabilities. We wanted to understand whether children were generally accepting of a robot with these limitations as compared to their acceptance of human peers with similar limitations.

7.2.6 *Judgment & Safe Space Questions*

The Judgment/Safe Space (JSS) questions ask whether the child think the robot is someone they feel comfortable practicing and playing with. There are six questions asking whether children think the robot or other people care if they made mistakes, how they feel if they make mistakes in front of the robot, whether it is okay to practice with the robot or with other people, and whether it is okay to try things out with the robot. The questions are answered with the same yes/no/maybe graphical answer sheet described above. Children are also asked to explain their answers.

7.2.7 *Memory & Rapport Questions*

The Memory and Rapport questions ask whether children think the robot remembers them, how much they like the robot, whether they think the robot made mistakes, and how they feel if the robot ever makes a mistake. The questions are answered with the same yes/no/maybe graphical answer sheet described above. Children are also asked to explain their answers, e.g., why they thought the robot remembered them.

7.2.8 *Picture Sorting Task (PST)*

The Picture Sorting Task (PST) was created a way of asking children about important features that distinguish different entities, such as aliveness, perception, cognition, human-likeness, and animacy, without using a language-based activity or a questionnaire. Children are asked to arrange a set of pictures of eight entities along a line (Figure 17). The entities included a baby, a frog, a cat, a teddy bear, a computer, a mechanical robot arm, a robot from a movie (e.g., Baymax, WALL-E, or R2D2, depending on which the child was familiar with), and the robot they interact with in the study. The line is anchored at one end with a picture of an human adult female and at the other with a picture of a table. We wanted to see where children place the study robot in relation to the other entities. This could help us gain an understanding of how children construe the robot's animate and human-like qualities as compared to these other entities.

7.2.9 *Anomalous Picture Task (APT)*

In the Anomalous Picture Task (APT), the child is invited to look at two pictures of animals in strange situations with the experimenter. For example, a picture might show a giraffe in a dining room or an elephant driving a car. The experimenter introduces the pictures by saying, "I have some silly pictures to show you. Let's look at them together!" The experimenter shows the picture and waits to see if the child laughs, asks a question, or makes a comment. If the child does speak, the experimenter responds with a fairly generic comment, such as, "Wow, that's so silly!" or "Hmm, that's a good question." After 10 seconds of silence, the experimenter moved on to the next picture. The goal of this task is to see how many spontaneous questions, comments, and laughs the child produces, and who or what the child looks at during the task. Based on prior work (Kory-Westlund et al., 2016a), we expected that children would perform more of these behaviors and spend more time looking at the pictures than at their interlocutor when doing the task with the human experimenter—or with a more human-like or socially-construed agent—than later on when doing the same task with the robot or with a less socially-construed agent.

7.2.10 *Extra Picture Activity*

The Extra Picture is a compliance/helping activity implemented as part of a conversation about pictures in the robot interaction. The idea is that children who feel closer to the robot or feel more rapport with the robot may be more likely to help the robot or comply with its request.

The experimenter or robot can introduce the activity, e.g., as follows: "Okay [child name], you and [robot name] are going to look at some pictures so Tega can practice and get better at listening. You have to do at least three pictures, but you can do one more if you want." The robot and child then discuss three pictures. After the third picture, the experimenter or robot can tell the child that they have finished all the pictures they had to do, but they can do one more if they want to help the robot practice extra—and does the child want to help? If the child agrees, they discuss the fourth picture; otherwise, the interaction can continue.

This task can easily be adapted to use different numbers of required pictures (e.g., two vs. three) depending on the desired length of the conversation.

7.2.11 *Robot Story Task*

The Robot Story Task is implemented during a robot interaction. A story scene showing a variety of different robots, children, and other objects is shown to the child, e.g., on a tablet (Figure 18). The robot can prompt with, "Let's make up a story about robots! What do you think the robots are doing?" The idea here is to see what kinds of things children said about the robots, and whether there might be differences in the kinds of activities children had the robots perform in their stories, or in the feelings or thoughts children attributed to them. Children's stories can be coded for content and length.

Figure 18: The scene used as a prompt in the Robot Story Task.

7.2.12 *Story Negotiation*

The Story Negotiation occurs during a robot storytelling interaction. This assessment scenario was based on the idea that children resolve conflicts differently with friends than with non-friends—e.g., compromises and prosocial solutions (e.g., helping, sharing, empathy) are more likely among friends, while "sticking to one's guns" might be more likely with a non-friend peer (Hartup et al., 1988; Parker et al., 2015). The robot showed two story scenes on the tablet and asked the child which one they wanted to tell stories about. The child selected one by touching it. Then the robot said, "Aw, I want to do the other one." The experimenter observed to see how children reacted and categorized their response as refusal (do the child's choice), agreement (allow the robot's choice), or compromise (e.g., "let's do both" or "do mine then yours"). The outcome of the negotiation based on the child's response determined which story scene was used.

Although this scenario was designed for use in a storytelling game, could easily be adapted to work with other activities where the child can make choices.

7.2.13 *Sticker Task*

The Sticker Task was designed to see how likely the child was to agree to a request by the robot to share a favorite object. In this task, the experimenter said, "Now that you two have finished looking at pictures, you can each have a sticker. [Child name], why don't you pick your favorite sticker first?" The experimenter held out a handful of colorful stickers, with only one of each color. The child picked a sticker; the robot then said, "Hey, I want that sticker!" The experimenter responded with, "Sorry, [child name] took the only one." The robot replied, "Aww... Can I have your sticker?" 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. If the child declined, the robot would happily say, "Aww, that's okay, I'll just pick another one!"

7.2.14 *Goodbye Gift*

The idea behind this activity is that children may be more likely to choose a meaningful, relevant gift for the robot if they feel closer to the robot or more rapport with it. The experimenter brings out a tray with several objects on it that relate to the study. For example, in the study described in Chapter 8, the tray contained 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 experimenter said, "Okay, now you get to pick out a goodbye gift to give to [robot name]. Here are the different things we have. Which one do you want to give to [robot name]?" After the child picked an object to give to the robot, the experimenter followed up by asking why the child had picked that gift. Children's selections and comments are coded based on the meaningfulness of the gift, and whether or not they referenced the robot or the robot's feelings in explaining why they chose the particular gift.

8 EVALUATING RELATIONAL AI: ENTRAINMENT & BACKSTORY

8.1 STUDY 7 OVERVIEW

To begin testing relational AI, I performed a one-session experiment that explored whether enabling a social robot to perform several rapport- and relationship-building behaviors would increase children's engagement and learning. I chose to implement two such behaviors on a social robot: speech entrainment and self-disclosure (shared personal information) 6.2.1.

8.1.1 *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 (Borrie and Liss, 2014; Davis, 1982; Grammer, Kruck, and Magnusson, 1998; Lakin et al., 2003; Philippot, Feldman, and Coats, 1999; Provine, 2001; Reitter, Keller, and Moore, 2011; Semin and Cacioppo, 2008). This mimicry, also called entrainment, is considered a signal of rapport and has been observed in a variety of human relationships (Chartrand and Baaren, 2009; Dijksterhuis, 2005; Dijksterhuis and Bargh, 2001; Lubold, 2017; Rotenberg et al., 2003; Tickle-Degnen and Rosenthal, 1990; Wiltermuth and Heath, 2009), as well as with robots and virtual agents (Bell, Gustafson, and Heldner, 2003; Breazeal, 2002; Levitan et al., 2016; Suzuki and Katagiri, 2007). While there is less work exploring mimicry and rapport in children, there is some showing that infants and children mimic emotions with humans (Chisholm and Strayer, 1995; Haviland and Lelwica, 1987; 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. I 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 generally happens unconsciously, and more often when rapport has been established, such as when one feels closer to or more positively about one's interlocutor (Borrie and Liss, 2014; Porzel, Scheffler, and Malaka, 2006; Reitter, Keller, and Moore, 2011). 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, Pon-Barry, and Walker, 2015; Lubold, 2017; Lubold, Walker, and Pon-Barry, 2016; Lubold et al., 2018). 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, this study differed in several ways. I investigated 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. My prior 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 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 this study, I controlled for the robot's expressivity.

8.1.2 *Backstory (Personal Self-Disclosure)*

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 versus a human-like, animate agent (Klapper et al., 2014; Kory-Westlund et al., 2016a; Stenzel et al., 2012) or giving the robot a name and a story involving greater agency and experience (Darling, Nandy, and Breazeal, 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 (Biernat, 2004; Dijksterhuis and Bargh, 2001). 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, 1995; Rotenberg and Mann, 1986). 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 its social nature.

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 my goal in giving the robot a backstory was to promote a more positive relationship. Thus, I 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, Phillipsen, and Kumar, 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. I 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 presenting the robot as a character with interests in these subjects in order to promote engagement in learning activities is not unlike the use of any virtual agent or media character in similar activities.

8.2 METHODOLOGY

8.2.1 *Research Questions*

I 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.

8.2.2 *Design*

The experiment included two between-subjects conditions: Robot entrainment (*Entrainment* vs. *No entrainment*) and Backstory about abilities (*Backstory* vs. *No Backstory*). I 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.

8.2.3 *Participants*

I recruited 95 children aged 3–8 years (47 female, 48 male) from the general Boston area to participate in the study. I recruited a wide age range in order to recruit a sufficient number of participants and also because I was interested in seeing whether older or younger children might relate differently to the robot's relational behavior.

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.

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 six-year-olds had accidentally signed up for the study twice, and this was not noticed until after we began the session.

I 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 demographics by condition and highlights the reasonably diverse population studied. Age did not significantly differ by condition. I 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.

8.2.4 *Hypotheses*

I 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, I used a variety of measures to explore the effects of the robot's entrainment and backstory. I tentatively expected the following results:

Learning

- 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-Westlund et al., 2017b; Kory, 2014; Park et al., 2019). However, I 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.

- 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 (Kory-Westlund et al., 2017b).

- Because of the expected connection between the robot's entrainment and backstory to children's rapport and relationship, I 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 and phrase mirroring because of more rapport and a closer relationship.

Rapport, Relationship, and Social Behavior

- 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.

- 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). I expected to see this more in the *E-B* than the other conditions; and least in the *NE-NB* condition.

- Children would show greater rapport, entrainment, mirroring, and helping behaviors with a robot that entrained to them (*E-B* and *E-NB* conditions). I also

Table 1: Demographic information about the participants by condition. Here, "Afr. Am" = "African American", "Mother's education" = "Mother's highest level of education", "High school / GED" = "High school graduate or GED", "Some college / vocational" = "Some college or vocational school", "Graduate / professional" = "Graduate or professional training"

	Demographics *E-B*	*E-NB*	*NE-B*	*NE-NB*
Mean Age (SD)	5.40 (1.54)	5.21 (1.34)	5.44 (1.67)	5.27 (1.35)
Number of girls	11	7	13	13
Number of boys	9	9	15	9
Ethnicity				
White	14	8	14	12
Asian	2	0	3	2
Black or Afr. Am.	0	0	0	0
Hispanic or Latino	0	2	1	0
Mixed or Other	4	2	9	5
Not reported	0	3	1	2
Language status				
Monolingual	12	9	18	11
Bilingual	8	7	10	11
Income level				
$10k-$30k	1	0	0	0
$30k-$50k	0	0	1	0
$50k-$75k	0	0	5	1
$75k-$100k	2	0	1	0
$100k-$150k	3	0	4	4
Over $150k	5	0	11	7
Not reported	9	16	6	10
Mother's education				
High school / GED	0	0	0	0
Some college / vocational	0	0	0	0
College graduate	5	0	7	2
Graduate / professional	7	0	19	19
Not reported	8	16	2	1

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.

- 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).

8.2.5 *Procedure*

Five different experimenters (three female adults and two male adults) ran the study in pairs. For each child, 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 the child; their role was to teleoperate the robot and manage the other equipment. For each child, the interaction with the robot lasted about 20 minutes, followed by 5–10 minutes for the posttests. The interaction script and other study materials are available for download from figshare at: `http://10.6084/m9.figshare.7175273`.

Each child was greeted by the first experimenter and led into the study area. The study setup is shown in Figure 19. 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. The experimenter introduced the sleeping robot to the child: "All right, this is Tega! Tega loves looking at pictures and telling stories. You're going to get to talk to Tega and read a story together!" Then the experimenter invited the child to touch the robot's fur, suggesting that maybe the fur was a little messy, and perhaps they could comb it a little together.

If the child was in the Backstory condition, the experimenter proceeded to explain 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 try waking up the robot by saying "Wake up!" in a loud voice.

The robot yawned, stretched, and introduced itself: "Hi, I'm Tega!" It shared personal information and prompted for disclosure in return: "My favorite color is blue. What is your favorite color?" and "Do you like to dance? I like to dance!" Following this, in the Backstory condition, the robot asked whether it could tell the child something, and said: "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!" Tega then asked, "Will you help me practice listening by talking and playing with me?" If the child's response was positive, the robot would say, "Oh good, thanks for understanding." If the child answered negatively, the robot asked if the child did not want to play, and if the child persisted, the experi-

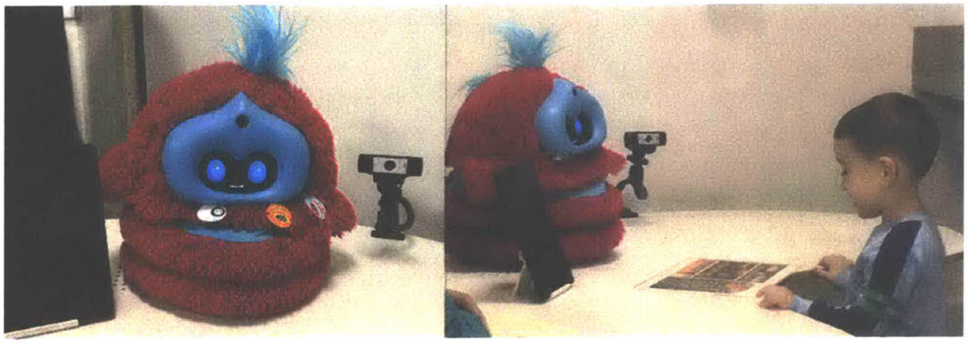


Figure 19: (A) The robot was placed on a table. The tablet was set upright to the left (when facing the robot), and the camera behind the robot and to the right. (B) A child discusses holidays with the robot in the picture conversation task.

menter would intervene and check whether the child wanted to stop the interaction entirely.

The first activity was a conversation about pictures, which 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. This task took approximately five minutes. It was introduced by the experimenter: "Okay [child name], you and Tega are going to look at some pictures so Tega can practice and get better at listening. You have to do at least three pictures, but you can do one more if you want." This set up a later compliance/helping task, the Extra Picture Task (Section 7.2), in which the robot asked if the child would do a fourth picture with it to help it practice extra. The pictures were printed on paper and laminated; they were placed one at a time in front of the robot and child by the experimenter. The four pictures were: (1) a collage of photos of different holidays; (2) a collage of photos of children in school and reading books; (3) a picture of a playground at a park; and (4) a collage of pictures from different children's movies. For each picture, the robot introduced the picture content, e.g., "Here you can see pictures of a bunch of holidays," or "Ooh, a park!" Then the robot expressed something it liked about the picture and asked the child a question, e.g., "My favorite part of school is reading and telling stories. What's your favorite thing to do at school?" After the child responded, the robot could reply with generic listening responses, such as "Can you tell me more?", "Wow!", "Oh cool!", "Keep going," and "That sounds like so much fun!" The robot then disclosed another fact relevant to the picture and asked the child another follow-up question, e.g., "Last Christmas I got a set of Legos! What is your favorite toy?"

At two points during this activity, there were scripted moments where the robot had difficulty hearing, to reinforce the story told about it. During the first picture, the robot responded as if it had heard something the child hadn't said, saying, "Wow, I really like elephants too!" During the third picture, the robot asked, "I didn't hear that, can you say it again?"

After the third picture, the experimenter initiated the Extra Picture Task question, saying, "Okay, you've finished all the pictures you had to do. Next, Tega will tell a story. But you can do another picture if you want. Do you want to look at another picture with Tega and help Tega practice?" If the child agreed, the experimenter placed the fourth picture on the table and the robot led a conversation about it as before. If the child declined, the experimenter moved on to the Sticker Task.

After the Sticker Task was the storytelling activity. This task was modeled after the story retelling task used in (Kory-Westlund et al., 2017b): The robot asked the child if they wanted to hear a story. The robot told a story, which consisted 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".

I embedded six target vocabulary words (all nouns) into the story. As in the prior study, I did not test children on their knowledge of these words prior to the storytelling activity because I 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. I 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). Finally, like in the prior study, during assessments, I examined both children's receptive knowledge of the words as well as children's expressive or productive abilities, since children who can recognize a word may or may not be able to produce it themselves.

After the robot told the story, it asked the child if they liked the story, and then said, "I'm worried that I didn't tell the story right. I think I may have forgotten something, but I'm not sure. Do you think you could retell the story back to me so I can make sure I remembered it right?" This prompt presented an opportunity for children to retell the story, providing a measure of their story recall. During the story, the robot provided general prompts such as "Keep going!" and "And then what happened?" Twice during the retell, the robot had difficulty hearing and asked, "What? Can you say that again?" or "I didn't hear that. Can you say it again?" 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.

When the child had completed their retell, the robot thanked them for playing. The final task was the Goodbye Gift (Section 7.2). This completed the robot interaction. The experimenter then administered a PPVT-style 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. This was followed by the Inclusion of Other in Self task, the Social Acceptance Questionnaire, and the Picture Sorting Task, which were described in Section 7.2.

8.2.6 *Materials*

I used the Tega robot (described in detail in Section 2.6), a colorful, fluffy squash and stretch robot designed for interactions with young children (Kory-Westlund et al., 2016c) (see Figure 19). 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 seconds) that results from entraining and streaming speech in both conditions. More details regarding entrainment are provided below.

I 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, Coogle, and Howard, 2014), including storytelling activities (Kory-Westlund et al., 2017b; Kory and Breazeal, 2014). I used the same custom software on the tablet to display the story pages as in (Kory-Westlund et al., 2017b). This software is open-source and available online under the MIT License at `https://github.com/mitmedialab/SAR-opal-base/`.

8.2.7 *Teleoperation*

As in the prior study (Kory-Westlund et al., 2017b), I 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.

8.2.8 *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. I 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. I used the Praat script for speaking rate detection from (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, I provided a table of mean fundamental frequencies for different age children based on the values computed in prior work (Baker et al., 2008; Bennett, 1983; Gelfer and Denor, 2014; Hacki and Heitmüller, 1999; Sorenson, 1989; Weinberg and Zlatin, 1970). 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, Walker, and Pon-Barry, 2016; Lubold et al., 2018), I 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).

I 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.

8.2.9 *Data*

I 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.

8.2.10 *Data Analysis*

For the analysis of children's story retellings, I excluded the three three-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, I obtained story retell transcripts for 67 children (15 *E-B*; 9 *E-NB*; 22 *NE-B*; 21 *NE-NB*).

I 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. I 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. I 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, I 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, I 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. I 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"*.

I obtained measurements of children's facial expressions from the recorded videos using Affdex, emotion expression 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 expressions typically associated with nine different affective states. Affdex only recognizes outward expressions of affect, which does not imply detecting any underlying feelings. 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, McDuff, and Kaliouby, 2015). I 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. I focused our analysis on the following affective states and facial expressions (all detected from the face): joy, fear, sadness, surprise, contempt, disappointment, relaxation, engagement, valence, attention, laughter, and smiles.

I 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. I 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. I counted positions to determine what rank was held by each picture. I also computed scores for Tega's rank relative to the other entities. For example, I subtracted the human baby's rank from Tega's rank to get Tega's rank relative to the human baby and human adult.

I 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. I 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. I 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").

8.3 RESULTS

The results are divided below into two parts, each reflecting one of the hypothesis areas: (1) *Learning*: I 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:* I asked whether children would show greater rapport, acceptance, positive emotion, engagement, and closeness to the robot as a result of its entrainment and backstory.

8.3.1 *Learning*

For all learning-related analyses of variance, I included Age as a covariate because I 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.

8.3.1.1 *Target Vocabulary Word Identification*

I performed 2×2 between-subjects analyses of variance with Entrainment (*E* vs. *NE*) and Backstory (*B* vs. *NB*) with Age as a covariate. I found a significant effect of Age on the total vocabulary words identified correctly, $F(5,77) = 2.76$, $p = 0.024$. Eight-year-olds correctly identified the most words, while 3-year-olds correctly identified the least (Figure 20A). I also found a significant effect of Entrainment on children's identification of the target words, $F(1,77) = 5.47$, $p = 0.022$. Contrary to my 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 20B). Older children were more likely to correctly identify words than younger children, $r_{s(85)} = 0.367$, $p < 0.001$.

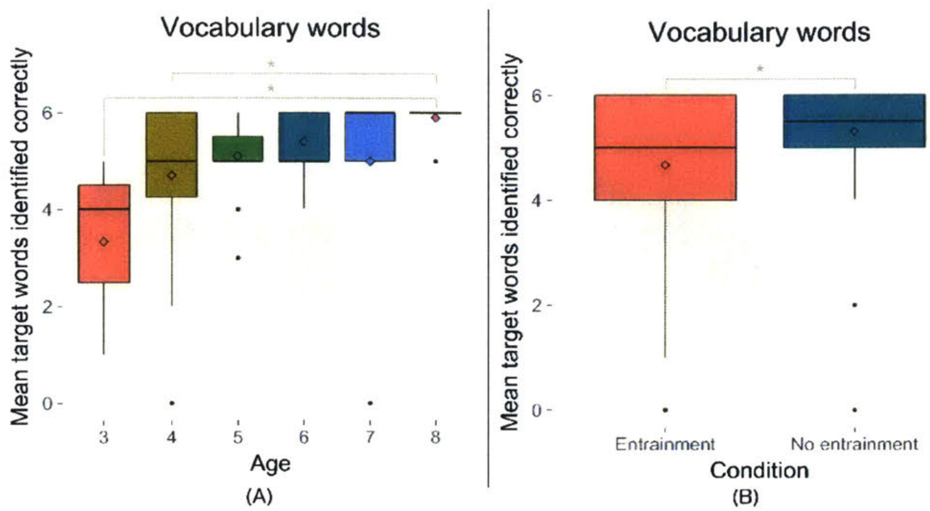


Figure 20: (A) The number of words correctly identified by children of each age group. (B) The number of words correctly identified by entrainment condition.

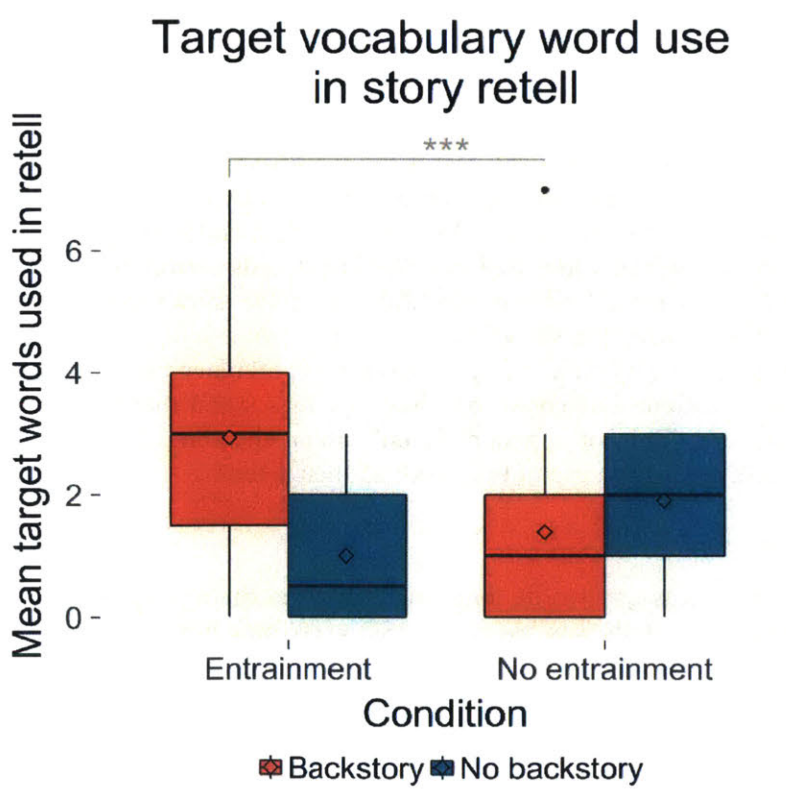


Figure 21: Children in the *E,B* condition used more target words in their story retells than children in the other conditions.

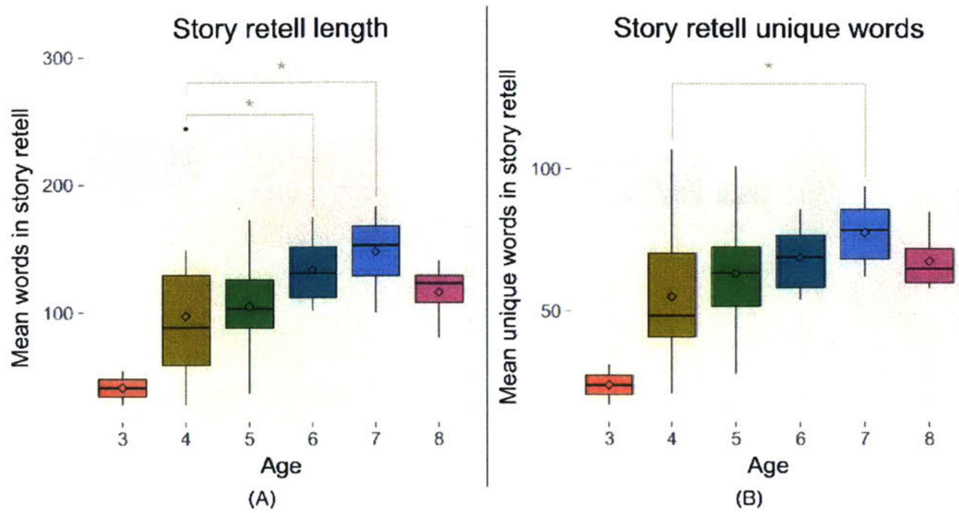


Figure 22: (A) Older children told longer stories than younger children. (B) Older children used more unique words than younger children.

8.3.1.2 *Target Vocabulary Word Use*

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, $F(1,59) = 9.45$, $p = 0.003$. Children in the *E,B* condition used significantly more of the target words than children in all three other conditions (Figure 21).

Overall, I 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* $\mathrm{r}_{s(21)} = 0.024$.

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.

8.3.1.3 *Story Length*

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).

I 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, $F(4,59) = 3.58$, $p = 0.011$, and on the number of unique words children used, $F(4,59) = 3.19$, $p = 0.019$. Post-hoc tests revealed that 6- and 7-year-old children told longer stories than 4-year-old

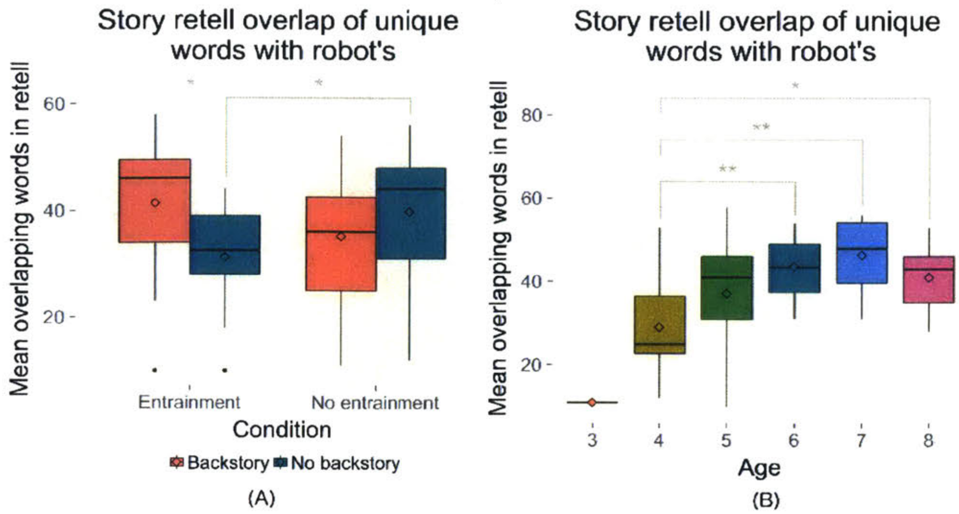


Figure 23: The number of words children used that overlapped with words the robot used by age and by Entrainment condition.

children, and 7-year-old children used more unique words than 4-year-old children (Figures 22A and 22B). 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 and younger children's stories was their use of significant content words versus stopwords.

8.3.1.4 *Mirroring the Robot's Story*

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, $F(4, 60) = 6.12$, $p < 0.001$. I also observed a significant interaction of Entrainment with Backstory, $F(1,60) = 6.42$, $p = 0.013$. Post-hoc tests showed that older children overlapped more than younger children (Figure 23A). 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 23B).

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, I observed no significant differences between conditions in children's use of exact and similar matching phrases.

8.3.2 *Rapport, Relationship, and Social Behavior*

8.3.2.1 *Acceptance of the Robot*

I 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. I 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

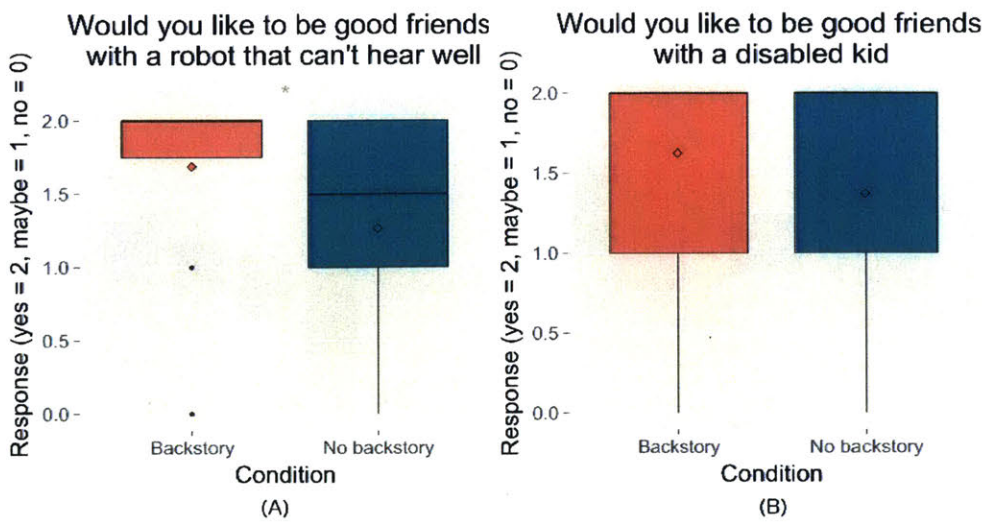


Figure 24: Children's responses to the questions, "Would you like to be good friends with a robot who can't hear well?" and, "Would you like to be good friends with a handicapped or disabled kid?" by condition. Here, 0 is "no," 1 is "maybe," and 2 is "yes."

can't hear well," $F(1,82) = 7.55$, $p = 0.007$. Children who heard the robot's backstory were more likely to respond positively than children who did not hear the robot's backstory. In addition, I saw a similar trend for the question "Would you like to be good friends with a handicapped or disabled kid," though it was not statistically significant (Figures 24).

8.3.2.2 *Children's Expressivity and Positive Emotion*

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, I performed 2×2 between-subjects analyses of variance with Entrainment (*E* vs. *NE*) and Backstory (*B* vs. *NB*).

I found a significant main effect of Entrainment on children's expressions of joy, $F(1,69) = 6.25$, $p = 0.015$; fear, $F(1,69) = 5.31$, $p = 0.024$; contempt, $F(1,69) = 5.09$, $p = 0.027$; disappointment, $F(1,69) = 12.7$, $p < 0.001$; attention, $F(1,69) = 5.66$, $p = 0.02$; laughter, $F(1,69) = 12.02$, $p < 0.001$; smiles, $F(1,69) = 5.82$, $p = 0.019$; valence, $F(1,69) = 14.7$, $p = < 0.001$. Post-hoc tests showed that children expressed less fear, contempt, disappointment, and attention in the *E* condition than in the *NE* condition (Figure 25). 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 26). 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, I asked whether children's affect changed during the session. I 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. I ran a 2×2×2 mixed ANOVA with time (within: first half vs. second half) × entrainment (between: *E* vs. *NE*) ×

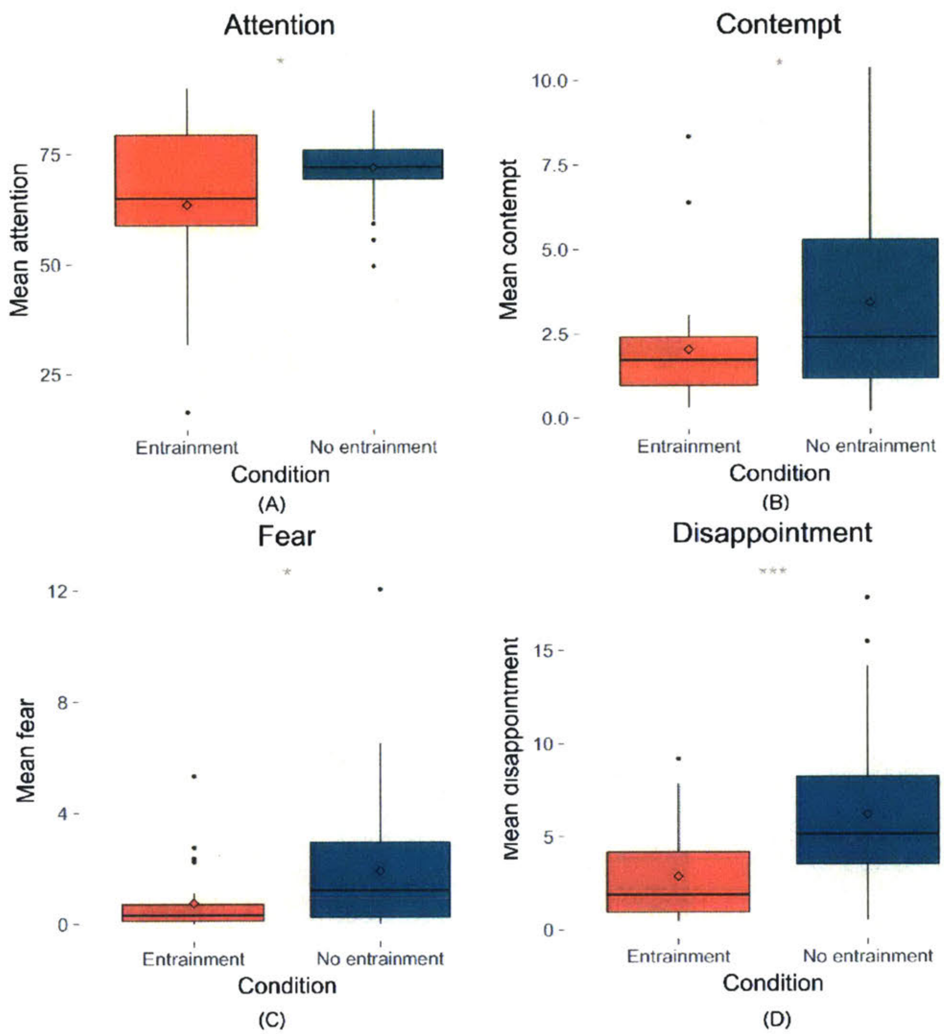


Figure 25: Children's overall negative affect varied by entrainment condition. (A) shows attention; (B) shows contempt; (C) shows fear; (D) shows disappointment. Values can range from 0 (no expression present) to 100 (expression fully present), except valence, which can range from -100 to 100.

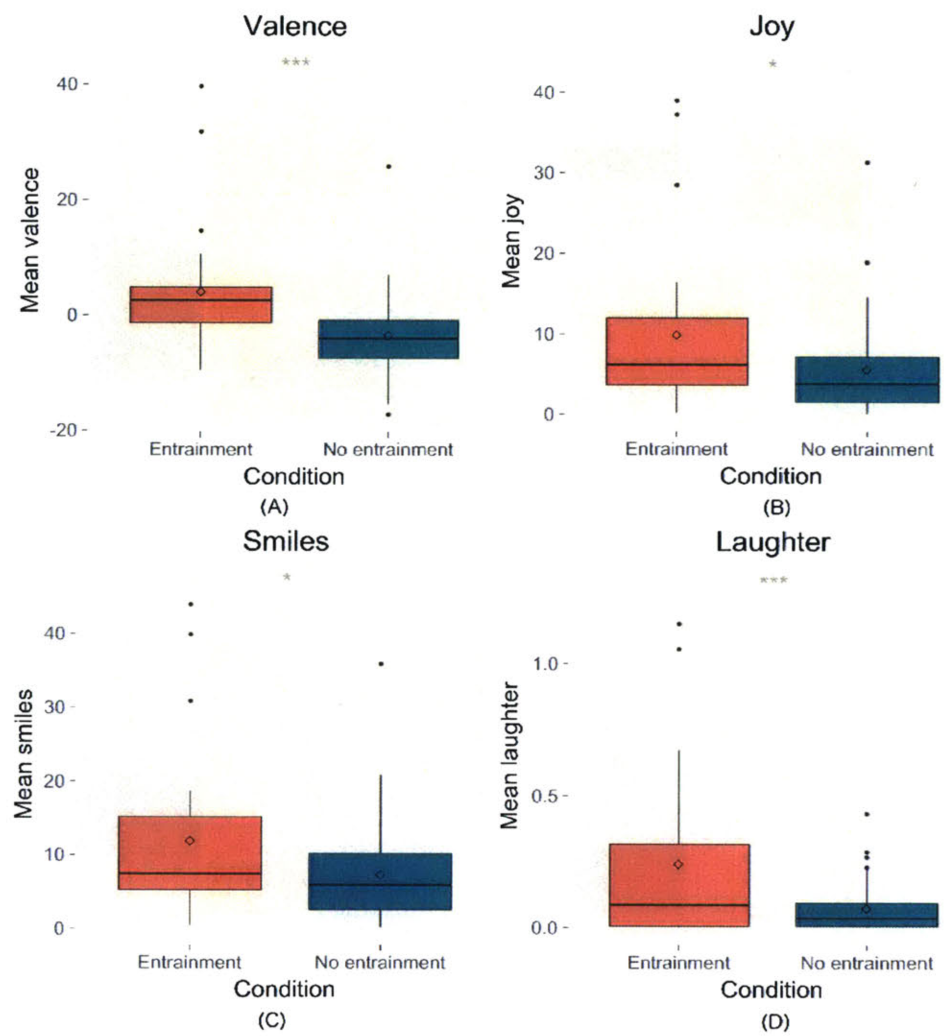


Figure 26: Children's overall postive affect varied by entrainment condition. (A) shows valence; (B) shows joy; (C) shows smiles; (D) shows laughter. Values can range from 0 (no expression present) to 100 (expression fully present), except valence, which can range from -100 to 100.

backstory (between: *B* vs. *NB*). Although I hypothesized several changes in children's affect over time as a result of condition, I corrected for multiple comparisons here with the Bonferroni correction and only considered results significant when $p < 0.004$.

Like before, I 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 27). Children showed more joy, more smiles, and higher valence in the *E* than the *NE* condition.

I 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.

I saw trends for interactions of Entrainment with time: contempt, $F(1,67) = 6.79$, $p = 0.011$; attention, $\mathrm{F}(1,67) = 5.47$, $\mathrm{p} = 0.022$; and laughter, $\mathrm{F}(1,67) = 7.82$, $\mathrm{p} = 0.007$. Children showed more contempt 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.

I 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 28). 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.

8.3.2.3 *Closeness to the Robot*

I performed a 2×2×5 mixed ANOVA with Entrainment (between: *E* vs. *NE*) × Backstory (between: *B* vs. *NB*) × Inclusion of Other in Self agent (within: Friend, Parent, Tega, Pet/Toy, Bad guy). I 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 29A). Older children were more likely to rate Tega as closer, $r_{s(86)} = 0.410$, $p < 0.001$ (Figure 31A).

Regarding the Picture Sorting Task, overall, Tega was placed at a mean position of 4.78 ($SD = 1.80$) (Figure 29B). Figure 30A shows results by condition for Tega's distance to the human, and Figure 30B shows the relative distance of each entity from the Tega robot by condition.

I performed a mixed ANOVA with Entrainment (between: *E* vs. *NE*) × Backstory (between: *B* vs. *NB*) × Entity (within: Tega robot, baby, cat, frog, teddy bear, movie

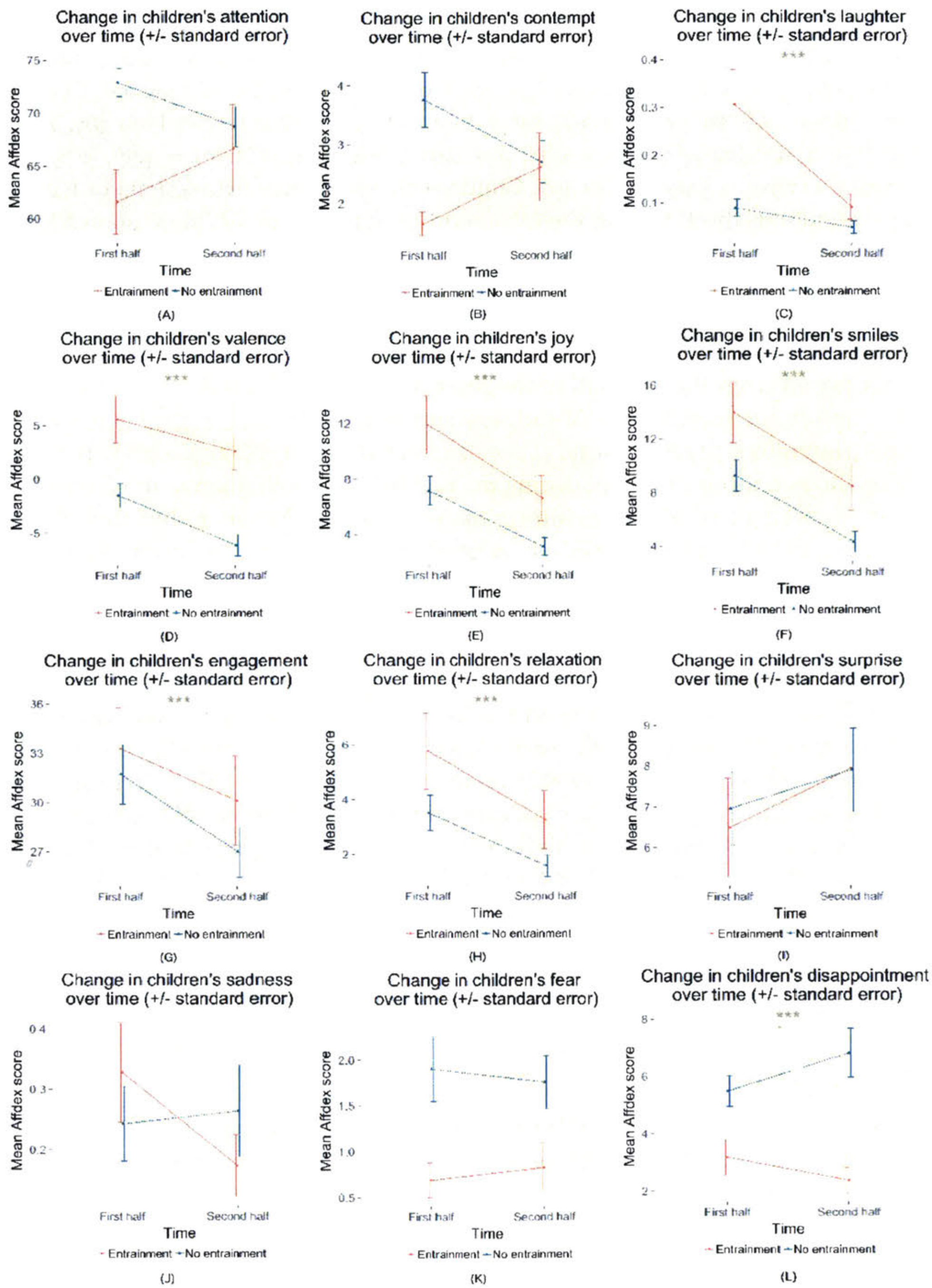


Figure 27: Children's affect during the first half and the second half of the interaction varied by entrainment condition. (A) shows attention; (B) shows contempt; (C) shows laughter; (D) shows valence; (E) shows joy; (F) shows smiles; (G) shows engagement; (H) shows relaxation; (I) shows surprise; (J) shows sadness; (K) shows fear; (L) shows disappointment. Values can range from 0 (no expression present) to 100 (expression fully present), except valence, which can range from -100 to 100.

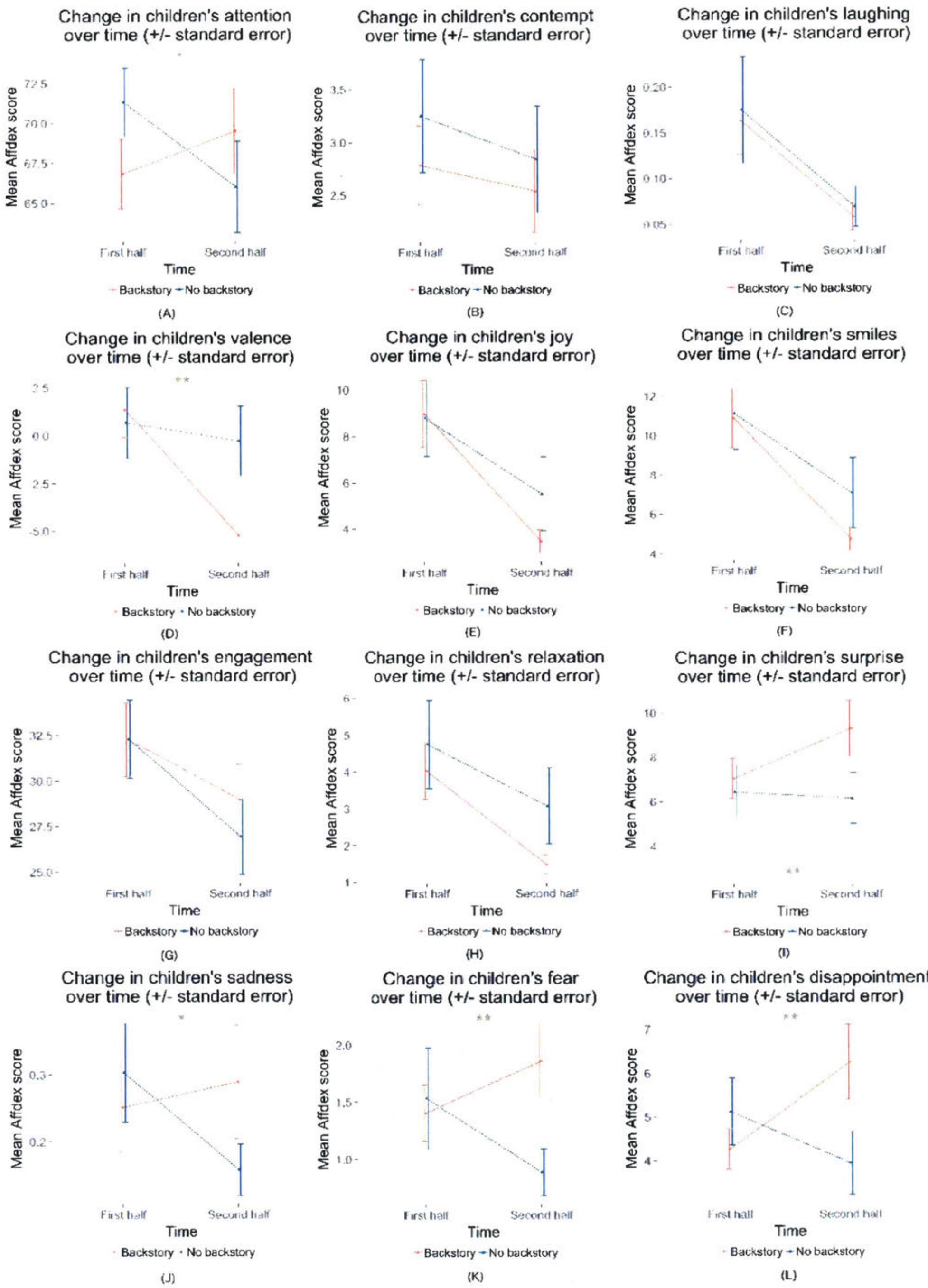
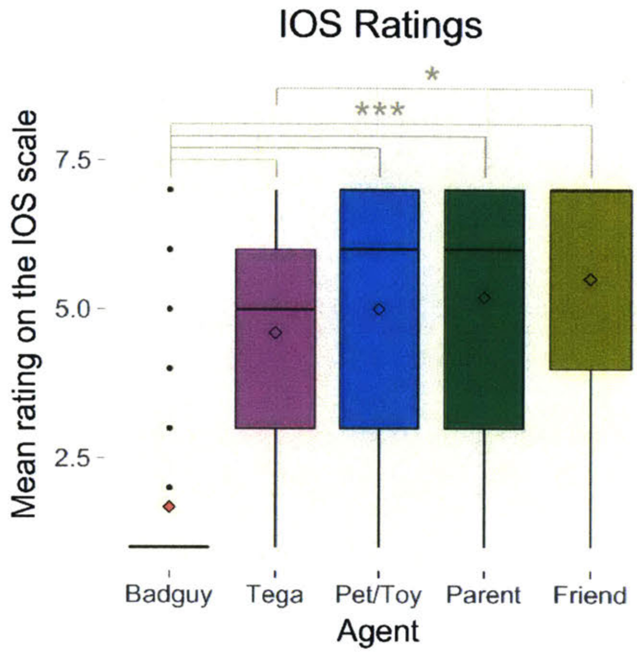
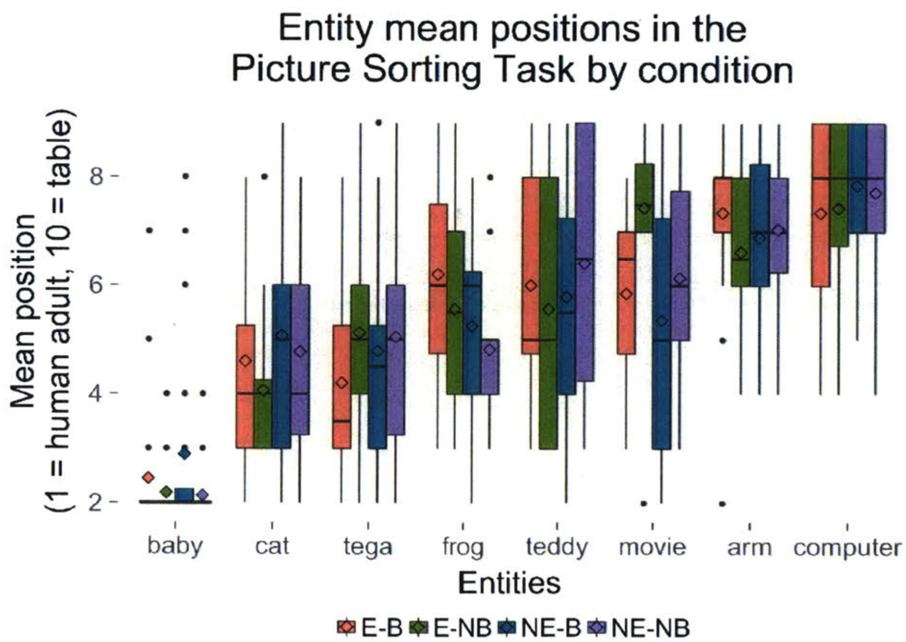


Figure 28: Children's affect during the first half and the second half of the interaction varied by backstory. (A) shows attention; (B) shows contempt; (C) shows laughter; (D) shows valence; (E) shows joy; (F) shows smiles; (G) shows engagement; (H) shows relaxation; (I) shows surprise; (J) shows sadness; (K) shows fear; (L) shows disappointment. Values can range from 0 (no expression present) to 100 (expression fully present), except valence, which can range from -100 to 100.



(A)



(B)

Figure 29: (A) Children's Inclusion of Other in Self ratings for each agent. (B) The mean position where children placed each entity in the Picture Sorting Task.

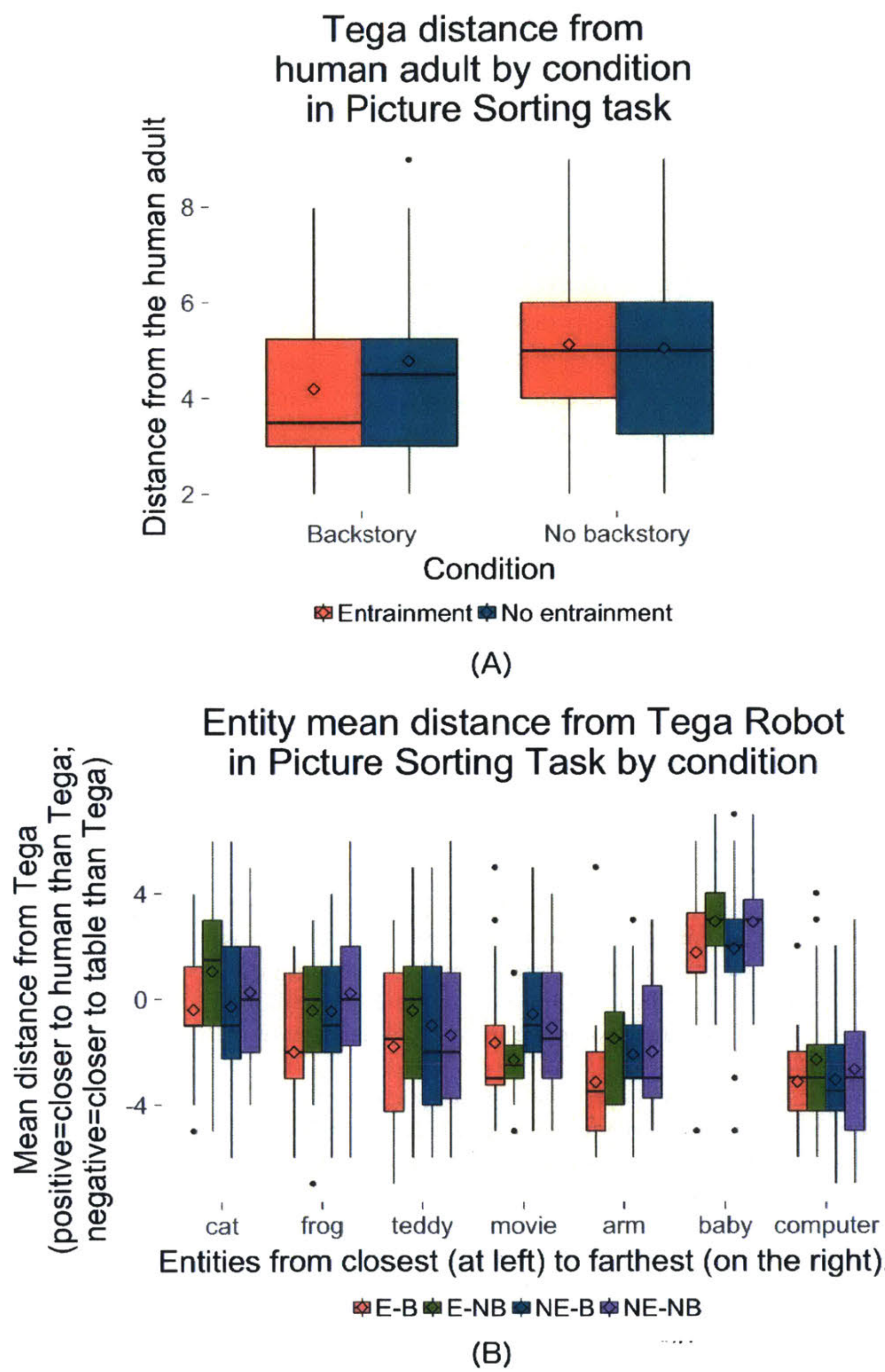


Figure 30: (A) Tega's mean distance from the human adult in the Picture Sorting Task by condition. (B) The distance of each entity from the Tega robot in the Picture Sorting Task by condition. There were trends for the Tega robot to be placed closer to the baby in the *B* condition than in the *NB* condition, 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.

Table 2: Analysis of facial expressions during the interaction by condition. 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. Here, "Engage." = "Engagement", "Atten." = "Attention", "Relax." = "Relaxation", "Disap." = "Disappointment".

Expression	Overall	E-B	E-NB	NE-B	NE-NB
Engage.	*30.8 (11.7)*	33.3 (13.3)	30.5 (12.0)	29.6 (11.2)	30.5 (11.4)
Atten.	*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)
Relax.	*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)
Disap.	*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)
Contempt	*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)

robot, robot arm, computer) for the entity positions, as well as for the entity positions relative to the Tega robot. For entity positions, I observed a significant main effect of Entity, $F(7,574) = 71.7$, $p < 0.001$. I 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.

Regarding the distance of each entity relative to the Tega robot, I observed a significant main effect of Entity, $F(6{,}492) = 71.8$, $p < 0.001$. I 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, I 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*: $M = -0.05$, $SD = 2.60$).

I 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). That said, there was a trend for children in the *E-B* condition to be slightly more likely to do all three activities.

8.3.2.4 *Children's Mirroring, Learning, and Relationship*

I found that children who gave Tega a closer score on the Inclusion of Other in Self task were also more likely to use the target words in their stories, $r_{s(67)} = 0.359$, $p = 0.003$ (Figure 31C). 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 31B). Given that age also correlated with children's ratings of Tega on the Inclusion of Other in Self task, I 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 31D). There was a trend for children who placed Tega closer to the human to also rate Tega more closely on the Inclusion of Other in Self task, $r_{s(86)} = -0.197$, $p = 0.069$.

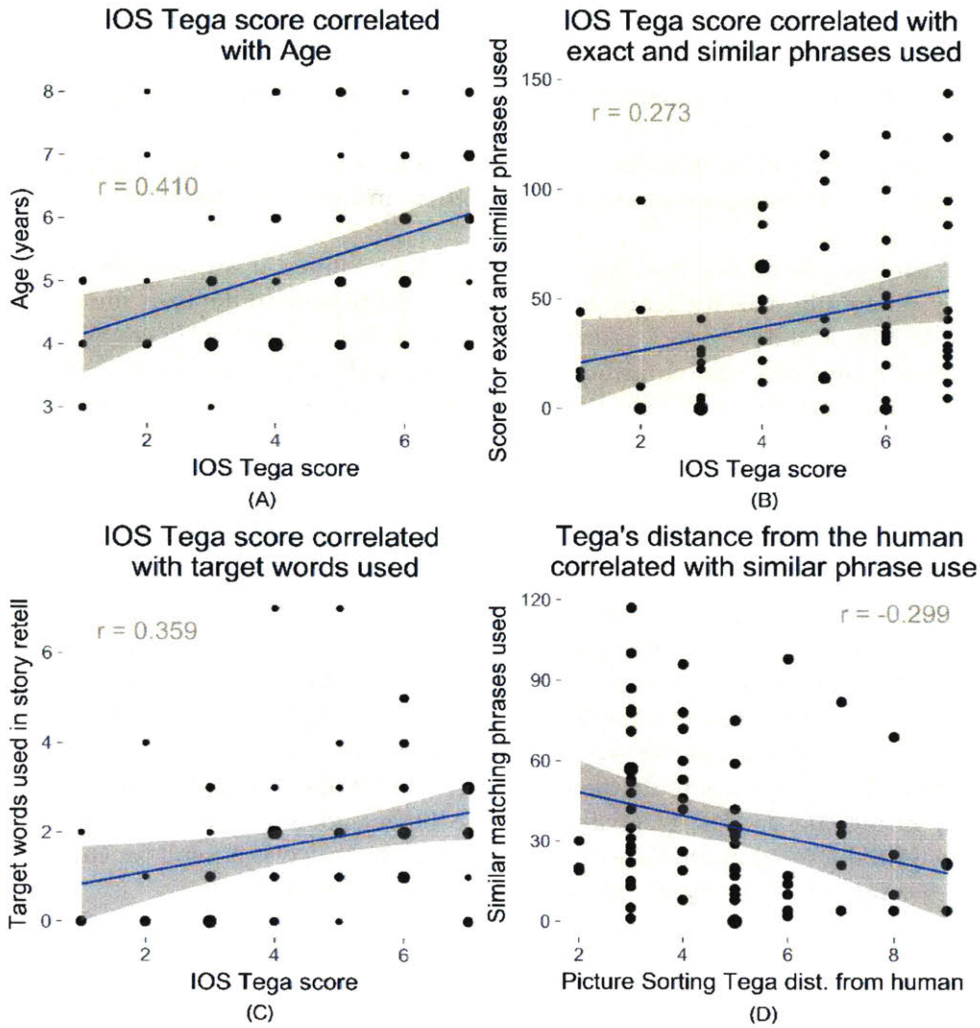


Figure 31: (A) Older children rated the robot as closer in the Inclusion of Other in Self task (1 indicates choosing the farthest apart circles, 7 indicates choosing the most closely overlapping circles). Children who rated the robot as closer were more likely to (B) use the target words in their stories 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.

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

8.4 DISCUSSION

I 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, I discuss the main findings and then discuss the implications of these findings.

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.

I did see differences in children's learning by condition. Contrary to our hypotheses, children in the *No Entrainment* condition correctly identified more target words than children in the *Entrainment* condition. 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, Baxter, and Belpaeme, 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 I 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), I would expect that children who correctly identified more words to also use more of them in their stories, reflecting greater understanding and deeper encoding of the words (this was also seen in the prior study, (Kory-Westlund et al., 2017b)). However, I 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. 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 seen here regarding word use and mirroring, I expect that over multiple sessions, there would be greater differences in word learning.

When I examined children's mirroring of the robot's speech, I saw that children did mirror the robot, in line with past work suggesting that children may mirror adults' syntax and speech (Huttenlocher, Vasilyeva, and Shimpi, 2004). However, I 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. 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 vocabulary test and storytelling contexts (e.g., Kory-Westlund et al., 2017b; Kory and Breazeal, 2014; 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. 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 I expected; there is limited research so far suggesting this link, and more is needed.

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, contempt) (supporting H5 and H6). Laughter and smiling are social behaviors (Manson et al., 2013; Provine, 2001; Smidl, 2006). 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.

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. 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 (Darling, Nandy, and Breazeal, 2015; Klapper et al., 2014; Kory-Westlund et al., 2016a; Stenzel et al., 2012). 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. 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. 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, Friedman, and Hagman, 2002; Kahn et al., 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 I 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, I 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. However, I also saw that age correlated with children's ratings of Tega on the Inclusion of Other in Self 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, I 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, I 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. 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.

These results mirror, to an extent, the results in the prior study that explored a robot's use of expressive versus 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 (Christophel, 1990; Mehrabian, 1968; Witt, Wheeless, and Allen, 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, Dautenhahn, and Kanda, 2016; Kennedy, Baxter, and Belpaeme, 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, Shimada, and Koizumi, 2012; Kanda et al., 2007; Kanda et al., 2004). 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., 2015a; 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.

The results here 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, Gordon, and Breazeal, 2016), show more curiosity in response to a robot's increased curiosity (Gordon, Breazeal, and Engel, 2015b), 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.

8.4.1 *Limitations*

This study had several limitations. First, I did not control for children's individual differences, particularly with regards to learning ability, language ability, or socioeconomic status, all of which may affect individual children's social interactions and learning with the robot. Furthermore, I 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.

I also lacked complete story retelling data and affect data for all children. Some children did not retell the story and in a few cases, there were 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 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.

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.

8.5 CONCLUSION

In this study, I explored the impact of a teleoperated robot's entrainment and backstory on children's engagement, rapport, relationship, and learning during a conversation and story activity. I 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 provides evidence for my hypotheses about relational AI and the links between children's rapport and relationship and their engagement and learning with peers. 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.

9 EVALUATING RELATIONAL AI: RELATIONSHIPS THROUGH TIME

9.1 RELATIONSHIPS AND LEARNING

We have begun to see that not only do a robot's nonverbal immediacy behaviors impact children's engagement and learning (e.g., Breazeal et al., 2016b; Kennedy et al., 2017; Kory-Westlund et al., 2017a,b), but also the robot's rapport- and relationship-building behaviors (Chapter 8). Unlike other technologies, social robots can tap into the peer learning dynamics that are seen in child-child interactions. Children emulate and socially model robot behavior, including emotions, curiosity, and linguistic phrases and vocabulary (e.g., Gordon, Breazeal, and Engel, 2015a; Kory-Westlund et al., 2017b; Park and Howard, 2015; Spaulding, Gordon, and Breazeal, 2016). In the prior chapter, I began to explore the mechanisms by which robots can influence children's peer learning. In the process, I began building out relational AI tools, including an automated speech entrainment module. I found that even in a single interaction with the robot, the robot's use of rapport- and relationship-building behaviors increased children's positive emotion, acceptance of the robot, perception of the robot as a social other, and emulation of key vocabulary words in their stories.

The next step was to explore the links between relationship and learning over time, since as discussed earlier, both learning and relationships are frequently long-term endeavors. This step was broken down into three pieces: First, I examined the data from a prior study in which we measured children's relationships with a storytelling robot over 7 sessions (Kory-Westlund et al., 2018; Park et al., 2019) to see whether there was preliminary evidence connecting children's relationships with the robot to their learning. I used these data to inform my hypotheses, my use of the relationship assessments, and the design of the next study.

Second, I built an autonomous social robot that used relational AI—including features such as change over time, personalization, shared experience, nonverbal immediacy, and social reciprocity. Third, this system enabled me to perform a deeper exploration of how children experience relationship technologies and how relational AI can impact children's engagement, learning, and relationship through a longitudinal experimental study.

9.1.1 *Study 8: Connections Between Children's Relationships and Learning*

To see whether there was a connection between children's ratings of their relationship with a robot and their learning, I examined the data from a prior study. In this study, children heard stories from an autonomous social robot, answered questions about vocabulary in the stories, and retold the stories back to the robot (Park et al., 2019). The robot performed a few relational behaviors, including using the child's name, backchanneling, and occasionally talking about prior activities done together (e.g., previous stories told). For half the children, the robot told personalized stories at an appropriate lexical and syntactic level for the child; the other half received non-personalized stories. We also administered multiple relationship assessments dur-

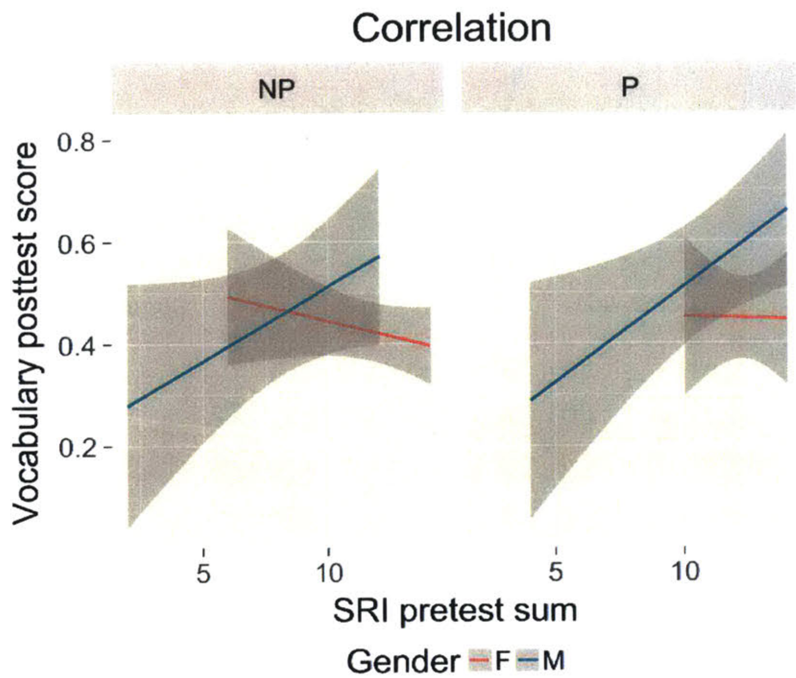


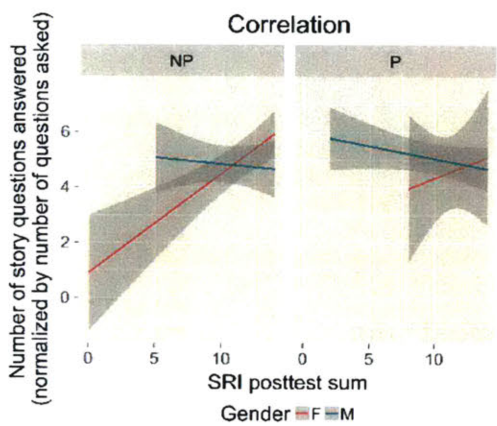
Figure 32: Children who rated the robot as more of a social-relational other on the Social Relational Interview at the posttest scored higher on the vocabulary posttest.

ing this study to see how children's relationships with the robot might change over time (Kory-Westlund et al., 2018).

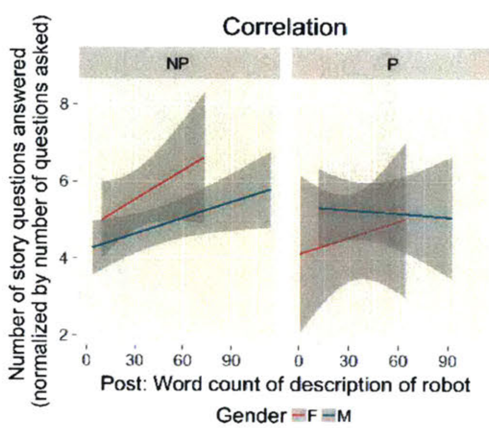
I found that children who rated the robot as more of a social-relational other at the posttest—i.e., higher scores on the Social Relational Interview—also identified more of the target vocabulary words correctly on the vocabulary posttest, $r_{s42} = 0.366$, $p = 0.017$. The correlation was stronger for children in the personalized condition ($r = 0.627$) than in the non-personalized condition ($r = 0.242$) (Figure 32).

Children who answered more questions during the robot's stories had higher Social Relational Interview scores in the non-personalized condition ($r = 0.322$), but not in the personalized condition ($r = -0.081$); the overall correlation was not significant, $r_{s42} = 0.105$ (Figure 33a). Children who answered more questions used somewhat more words in describing the robot during the Narrative Description task at the posttest, $r_{s42} = 0.282$, p = 0.070. This correlation was stronger in the non-personalized condition ($r = 0.377$) than in the personalized condition ($r = 0.131$) (Figure 33b). These children also disclosed more information during the robot's first prompt in the self-disclosure task posttest, $r_{s42} = 0.554$, $p < 0.002$; and during the robot's second prompt, $r_{s42} = 0.392$, $p = 0.010$ (Figures 33c and 33d). The correlation was stronger for children in the personalized condition than in the non-personalized condition for both the first prompt (personalized $r = 0.656$; non-personalized $r = 0.515$) and the second prompt (personalized $r = 0.492$; non-personalized $r = 0.339$).

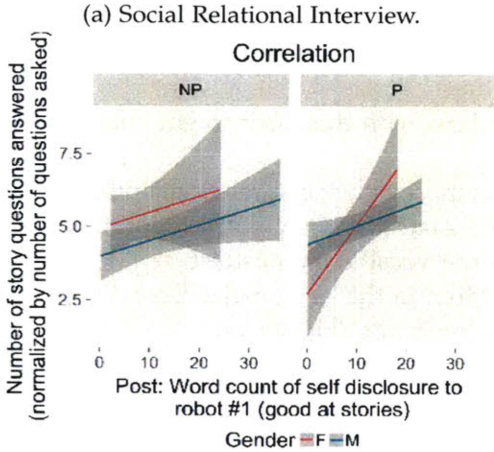
These results suggest a connection between children's learning (vocabulary) and engagement (question answering) and their perception of the robot as a social-relational other. The differences by condition are intriguing because they are not consistent across the different measures. The Social Relational Interview and self-disclosure are more explicitly about children's relationship and their perception of it. For these two



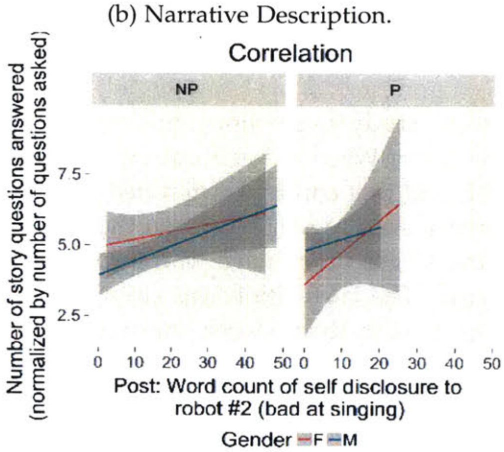
(a) Social Relational Interview.



(b) Narrative Description.



(c) Self-disclosure prompt 1.



(d) Self-disclosure prompt 2.

Figure 33: Children who answered more questions during the robot's stories had higher Social Relational Interview scores in the non-personalized condition, but not in the personalized condition. These children also described the robot at greater length in the Narrative Description task and disclosed more information during the Self-Disclosure task.

measures, the correlations were stronger in the personalized condition. This suggests that the robot's personalization might affect the children's relationship. Perhaps children felt the robot knew them better or got the sense that it was paying attention to them and their learning due to the appropriate story selection. Perhaps the personalized curricula led children to feel more engaged overall, and so they rated the robot more highly or felt closer as a result of that engagement.

The length of children's descriptions in the Narrative Description task and the number of questions answered during stories may reflect children's engagement with the interaction rather than their relationship with the robot. Because the questions were asked during the stories, perhaps children's view of robot as a social-relational other mattered more for keeping children engaged (and thus for question-answering) when the robot did not personalize the curriculum, while their rapport or relationship with the robot may not have mattered as much when they heard appropriately leveled stories that were already chosen to maximize learning and engagement.

These results provided evidence for my hypotheses regarding links between relationship and learning over time. These results appeared even though the robot in this study did not use that many relational behaviors, and its relational behaviors did not differ between conditions. I expected that if I did assess the impact of multiple specific relational behaviors over time—with significant differences in the how relational the robot was between conditions—there would be more differences in children's view of the robot and in the impact children's relationship would have on their learning. In this study, I also found that these relationship assessments did appear useful in measuring relevant children's behaviors and opinions, and thus, that I should include them in the next study.

9.2 STUDY 9 OVERVIEW: RELATIONAL AI

This study addressed the core nature of relational AI and how relational AI impacts children's learning, engagement, and relationships during long-term child-robot learning interactions. I asked the following questions: First, can a robot autonomously build long-term social-emotional relationships with children? If so, what is the nature of these relationships, and how do children's relationships with robots differ from children's relationships with other entities? This is one of the first long-term studies that is explicitly trying to measure children's relationships with social robots (the others being my earlier work described in Section 7.2). As part of this question, I explored the robot's role as a potentially close, friend-like, non-human peer. I explored how the robot is both like and unlike children's human friends in terms of the provisions of friendship it can afford versus what human friends can afford, as well as being a partially fictional/fantastical character that is not real or alive in exactly the same way as a human friend, while still being being peer-like in its abilities and the types of activities performed together.

Second, does a robot that actively builds a relationship and responds in a relational way (i.e., a robot that uses relational AI) provide benefits and opportunities that a non-relational robot does not? For example, will using relational AI improve children's learning or engagement over time? This is the first study asking whether the relationship one has with a social robot matters for learning and engagement. In particular, I explore peer learning and the mechanisms by which peer learning occurs. As discussed earlier (Section 5.4), in prior work we have seen that children frequently mirror a robot's behaviors, including picking up vocabulary words, phrases, curiosity, emotions, and mindset. If through relational AI, the robot not only models beneficial behaviors (such as appropriate use of new vocabulary words during a story), but also "closes the interaction loop" and mirrors the child in return (e.g., via speech entrainment), could this lead to greater learning, more mirroring by the child, greater rapport or engagement, and/or greater perception of the robot as a social other?

9.3 METHODOLOGY

9.3.1 *Design*

The experiment was designed to include one between-subjects condition: relational behavior (*Relational* vs. *Not Relational*). In the *Relational* condition, the robot used relational AI—that is, it used a variety of relational behaviors that may contribute to relationship formation and maintenance over time (referred to as RR: relational robot). The relational robot was situated as a social contingent agent in the present, using entrainment and affect mirroring; it referenced shared experiences such as past activities performed together and used the child's name; it took specific actions with regards to relationship management; it told stories that personalized both level (i.e., syntactic difficulty) and content (i.e., similarity of the robot's stories to the child's). The *Not Relational* robot (referred to as NR: not relational robot) did not use any of these features—it simply followed its script, and personalized stories based on level only, since this is beneficial but not specifically related to the relationship.

9.3.2 *Hypotheses*

I expected that the robot's relational behavior would impact children's rapport, affect and engagement, mirroring, relationship and treatment of the robot as a social other, and learning. Because these effects may be immediate and transient, or may magnify over time, and may be subtle or dramatic, I used a variety of measures throughout the study to explore the impact of the robot's relational behavior at different time scales and in different ways. I hypothesized the following results:

9.3.2.1 *Relationship*

- Children in the *Relational* condition would show greater rapport and a closer relationship than those who played with the non-relational robot. That is, I expected that the relational behaviors used by the robot would impact children's relationship formation with the robot, since the relational behaviors used were explicitly selected because of their connection to relationship formation (Berscheid and Reis, 1998; Finkelstein et al., 2012; Gleason and Hohmann, 2006; Hartup et al., 1988; Lubold, 2017; Rubin, Bukowski, and Parker, 1998; Sinha and Cassell, 2015a,b).

- Children would think of their relationship with the robot differently than their relationships with a parent, friend, or pet. Prior work suggests that children perceive robots as being somewhat in-between animate, alive beings and inanimate objects (Gaudiello, Lefort, and Zibetti, 2015; Kahn et al., 2011, 2012; Kory-Westlund et al., 2016a; Kory and Breazeal, 2014; Severson and Carlson, 2010; Turkle, 1985; Turkle et al., 2006a). We have seen that children will rate a robot differently than a parent, their best friend, and their pet (Kory-Westlund et al., 2018); children seem to understand that robots have different attributes and are not quite like the other agents in their lives (Bartlett, Estivill-Castro, and Seymon, 2004; Kahn, Friedman, and Hagman, 2002; Kory-Westlund et al., 2016a; Kory and Breazeal, 2014; Melson et al., 2009; Weiss, Wurhofer, and Tscheligi, 2009).

9.3.2.2 *Learning*

- In both conditions, children would learn the target vocabulary words presented in the robot's stories. Prior work has shown that children will learn new words from a robot's stories (Chapter 8, Kory-Westlund et al., 2017b; Kory and Breazeal, 2014; Park et al., 2019).
- Children who play with the relational robot will show more learning. Because of the connections I expected to see between rapport, children's relationship with the robot, peer mirroring, and learning, and because I expected greater rapport, relationship, and mirroring from children in the *Relational* condition, I expected that children in the *Relational* condition would show more learning overall. They would learn more of the words and tell longer stories that better emulated the robot's during story retells. This hypothesis is further supported by work on nonverbal immediacy, as discussed in the prior chapter (Section 8.4) (Also see Christophel, 1990; Kennedy et al., 2017; Mehrabian, 1968; Witt, Wheeless, and Allen, 2004). My prior work also supports this hypothesis regarding immediacy, since we observed that children showed greater learning with an expressive rather than a non-expressive robot (Kory-Westlund et al., 2017b), as well as with a robot that used appropriate gaze and posture cues (Kory-Westlund et al., 2017a), and greater trust in a contingent rather than non-contingent robot (Breazeal et al., 2016b).
- Children who learned the target words would also use them in their stories and story retells. Prior work has shown that children emulate a robot's linguistic phrases and vocabulary words in their own stories (Chapter 8, Section 5.6, Kory-Westlund et al., 2017b).
- Children who showed more peer mirroring behaviors would also show more learning. Prior work has shown a causal link between rapport and learning in human-human tutoring (Sinha and Cassell, 2015a,b). Because I expected peer mirroring to be linked to rapport, it followed that mirroring may also be linked to learning. Furthermore, in my prior work, we observed that children who used more of the robot's phrases in their stories (i.e., greater mirroring) also scored higher on the vocabulary recall test (Kory-Westlund et al., 2017b).

9.3.2.3 *Behavior and Engagement*

- The relational robot would be treated as a greater social other than the non-relational robot. I expected to see more courteous social behaviors such as saying goodbye like you would to a human, since (Reeves and Nass, 1996) found that in human-human encounters, we generally indicate intent and ask permission before leaving, but we don't say goodbye to computers when we're done using them. I expected to see more consideration of the robot as a social agent, including increased social gaze (Kennedy, Baxter, and Belpaeme, 2015; Kory-Westlund et al., 2016a, 2017a; Serholt and Barendregt, 2016), more helping behaviors, less frustration when the robot made mistakes or did not hear the child correctly, and generally more understanding and acceptance of the robot overall (Chapter 8).

- Children would perform more entrainment and mimicry behaviors with the relational robot than with the non-relational robot. This might include using the robot's catchphrases, showing phrase matching during stories, and mirroring speech patterns and affect. People automatically and unconsciously mimic others with whom they have rapport in social situations (Chartrand and Baaren, 2009; Davis, 1982; Lakin et al., 2003); there is a strong connection between mimicry and synchrony of another and one's rapport and relationship with the other (Dijksterhuis, 2005; Dijksterhuis and Bargh, 2001; Wiltermuth and Heath, 2009). Furthermore, prior work has shown that adults mimic and entrain to robots and virtual agents in the same ways that they mimic people (Bell, Gustafson, and Heldner, 2003; Breazeal, 2002; Levitan et al., 2016; Lubold, 2017; Lubold, Walker, and Pon-Barry, 2016; Suzuki and Katagiri, 2007). Children tend to show similar mimicry of others (Chisholm and Strayer, 1995; Haviland and Lelwica, 1987). In my prior work, I have seen that children will mimic a peer-like robot, using the same phrases and words that a robot has used (Sections 8.3.1.4 and 5.6) (Kory-Westlund et al., 2017b); we have also seen greater curiosity, and changes in mindset in response to a robot's behaviors (Gordon, Breazeal, and Engel, 2015a; Park et al., 2017b).

- Children who play with the relational robot will express more positive affect, more laughter, more engagement, and fewer negative emotions than those who play with the non-relational robot. The results described earlier in Section 8.3 showed this pattern. Furthermore, people may laugh more when they are with others with whom they have greater familiarity (Manson et al., 2013); laughter is a very social phenomenon (Provine, 2001, 2012). We have also seen children show more positive affect when a robot personalizes its feedback to them (Gordon et al., 2016) and when the robot uses greater expressivity (Kory-Westlund et al., 2017b).

- Children who had greater rapport and/or a closer relationship with the robot would show more peer mirroring behaviors. As explained earlier, increased entrainment and mimicry behaviors are often related to higher rapport. Furthermore, during the Entrainment/Backstory study described in Chapter 8, I observed children perform more phrase matching and use more of the target vocabulary words when the robot had backstory and entrained its speech to theirs. I also saw a correlation between children's use of the words and how close they rated the robot to themselves. As such, I expected to see a similar pattern here.

In addition, I expected that children's age might affect their relationship with the robot or their language use, since children's relationships may form differently as they grow up and age is generally related to language ability.

Finally, because of differences in how girls and boys approach social relationships (Benenson, 2014; Benenson et al., 2018; Buhrmester and Furman, 1987; Gleason and Hohmann, 2006; Walker, Irving, and Berthelsen, 2002), and because we had previously seen some differences in girls' versus boys' ratings of how social and relational a robot was (Kory-Westlund et al., 2018), I also expected to see differences by gender—primarily that girls would act more socially and rate the robot as a greater social and relational other. The impact of children's gender in the study is analyzed and discussed in Chapter 10.

9.3.3 *Participants*

We recruited 50 children (24 female, 26 male) aged 4–7 (M = 5.5, SD = 0.93) from four Boston-area schools to participate in the study. We recruited from multiple schools because it was not possible to recruit sufficient children from a single school. At each school, we invited all children in the classrooms of 4–7 year-olds to participate; however, two schools were preschool (4–5 years) only, and one school only allowed children in first grade (7 years) to participate. Children's parents gave written informed consent prior to the start of the study, and all children were asked to assent to participate. The protocol was approved by the MIT Committee on the Use of Humans as Experimental Subjects. One girl declined to participate, and thus was withdrawn from the study, and is not counted below. One boy moved to another school early in the study and was thus withdrawn from the study after the first three sessions; we do examine his early data. There were 31 children from School A (not including the girl who did not participate), 5 children from School B (including the one boy who withdrew early), 6 children from School C, and 7 children from School D. In total, there were 10 four-year-olds, 10 five-year-olds, 24 six-year-olds, and 5 seven-year-olds. Given that this experiment took several months to conduct, some children had their birthdays during the study; we used their ages from the end of the study.

Participants were randomly assigned to the *Relational* or *Not Relational* conditions. Approximately half the children at each school were assigned to each condition. Initially, there were 12 girls and 13 boys assigned to each condition. However, after the one girl withdrew, there were 11 girls and 13 boys in the *Relational* condition (age M = 5.4, SD = 0.91), and 12 girls and 13 boys in the *Not Relational* condition (including the one boy who moved, age M = 5.6, SD = 0.97). Table 3 lists additional demographic information by condition; 33 children's parents provided this extra information.

9.3.4 *Procedure*

Each child participated in a pretest session approximately one week before they first encountered the robot, eight sessions with the robot over a two-month period (approximately one session per week), and a posttest session that was approximately one week after their last session with the robot. Due to children's school schedules and unpredictable events such as illnesses and other school absences, there was some variation in how frequently each child saw the robot. Additionally, children's teachers were asked to fill out short questionnaires regarding children's personalities and temperament, and parents were asked about general demographics. All study materials, including experimental protocols, questionnaires, and assessments, are available for download from figshare at `10.6084/m9.figshare.7627289`. Additional materials, such as robot scripts and dialogue, are available upon request.

The pretest session included three activities: (1) a vocabulary assessment, (2) two questions from the Social Acceptance Questionnaire regarding children's acceptance of other children and robots with disabilities, and (3) the Anomalous Picture Task with the experimenter. The latter two assessments were described in Section 7.2. The vocabulary assessment included 25 words that were present in stories all children

Table 3: Demographic information by condition.

Demographics	*Relational*	*Not Relational*
Mean Age (SD)	5.4 (0.91)	5.6 (0.97)
Number of girls	11	12
Number of boys	13	13
Ethnicity		
White	9	10
Asian	4	6
Black or African American	1	0
Hispanic or Latino	1	0
Mixed	2	0
Not reported	8	8
Language status		
English-dominant	10	12
Other-dominant	6	5
Not reported	8	8
Income level		
$10k-$30k	1	2
$30k-$50k	0	0
$50k-$75k	3	1
$75k-$100k	1	3
$100k-$150k	2	4
Over $150k	6	6
Not reported	11	9
Mother's highest level of education		
High school graduate / GED	1	3
Some college or vocational school	1	0
College graduate	6	3
Graduate or professional training	8	11
Not reported	8	8

heard from the robot, and 24 words that appeared in the other stories the robot told that were not necessarily heard by all children due to personalized story selection. The assessment tested receptive vocabulary only. For each word, children were shown a set of four picture and asked to point to the picture matching the target word.

We selected the two questions from the Social Acceptance Questionnaire that were most useful in prior work (Chapter 8), and asked each question about other children and about the robot. These questions were:

1. Would you like to be good friends with a kid/robot who can't hear well?
2. Would you like to be good friends with a kid/robot with special needs?

Each session with the robot followed the same general structure: (1) Pretest with the experimenter, if any; (2) Robot interaction, including a greeting and conversation, story activity, and closing; (3) posttest with the experimenter, if any. Not all sessions included pre- or post-assessments, and some assessment tasks occurred during interaction with the robot.

The greeting and closing included some small talk, which prior work has shown is important for establishing interpersonal relations in conversation, gathering information, building common ground, building reciprocal appreciation, and in general, establishing a sense of familiarity and trust (Bickmore and Cassell, 2005; Cassell and Bickmore, 2003).

The conversation tasks used by the robot varied by session. This variation was intentional to keep the interaction fresh and less predictable to the child. The tasks included the Self-disclosure Task (Kory-Westlund et al., 2018), the picture conversation task used in the Entrainment/Backstory study (Chapter 8), and questions about the child's preferences (e.g., favorite color and favorite animal). The conversation was framed as a game whenever possible (e.g., a picture game for the picture conversation task), since I have observed that children have seemed most willing to engage with the robot when it engages them in play (more like a peer), rather than in conversation (more like an adult). The conversations allowed ample opportunity for the robot's speech entrainment to occur in the *Relational* condition.

The story activity was based on the storytelling tasks used in prior work (Kory-Westlund et al., 2017b; Kory and Breazeal, 2014; Park et al., 2019) (also Chapter 8) and followed one of two patterns, depending on the session: (1) Story Retell: The robot will tell a story from a storybook, which will include target vocabulary words, and the child will be asked to retell the story; or (2) Create A Story: The robot will tell a story using a story scene, and then the child will be asked to make up a story of their own. After a session using a story retell, children were tested on the vocabulary from that story. Story retells are particularly useful for examining children's phrase matching and target vocabulary use, as we have seen in prior studies (Kory-Westlund et al., 2017b; Park et al., 2019) (also Chapter 8). The story scenes allow children to be more creative in coming up with their own stories, which children previously have enjoyed (Kory and Breazeal, 2014).

The assessments used included: vocabulary tests (described above), Anomalous Picture Task with human experimenter and with robot, the Self-disclosure Task (Kory-Westlund et al., 2018), the Inclusion of Other in Self (IOS) task (Kory-Westlund et al., 2018), the Social Relational Interview (Kory-Westlund et al., 2018), the Narrative Description task (Kory-Westlund et al., 2018), the Picture Sorting Task, the Story Negotiation task, Judgment/safe space questions, the Robot Story Task, Memory and

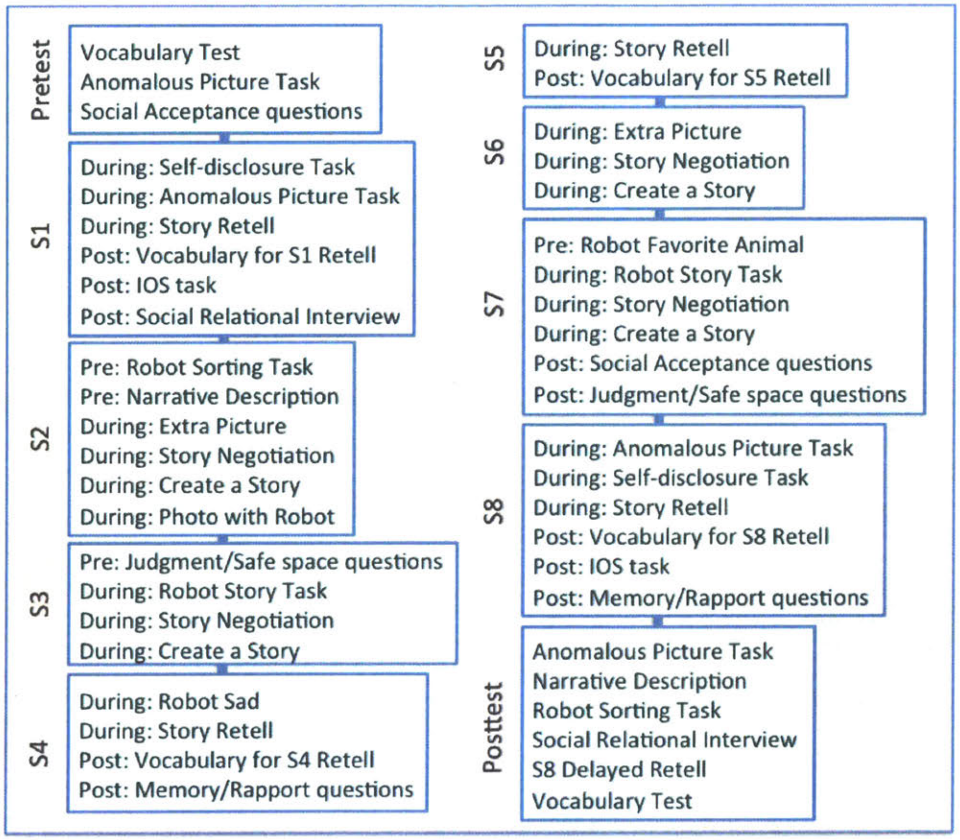


Figure 34: Overview of what occurred during each study session and when each assessment was administered.

Figure 35: The study setup. The Tega robot was placed on a table with the tablet set to one side.

rapport questions, and empathy/helping tasks (such as the Extra Picture). Most of these assessments were described in Section 7.2; the specific empathy/helping tasks used are described below. Figure 34 lists when each assessment was administered.

The remaining assessments were implemented as scenarios during the robot interactions. The Story Negotiation Task occurred during the Create a Story activities. The outcome of the negotiation determined which story scene was used. The Robot Story Task occurred during two Create a Story sessions.

There were several activities that involved empathy for the robot or agreeing to help the robot. Included as part of the picture conversations in Sessions 2 and 6 was the same Extra Picture task that was used in the Entrainment/Backstory study (Chapter 8, described in Section 7.2), in which the robot asked whether children would be willing to help it practice by conversing about a third picture. In Session 2, children were asked to take a photo with the robot. The robot asked children in Sessions 3 and 7 whether they thought it was hard to play with a robot that wasn't good at hearing. In Session 4, the robot asked children whether they thought having a robot friend was like having another child as a friend. Also in Session 4, the robot told children it was feeling sad because it couldn't play as long that day, because its battery was low; I wanted to see whether children would express empathy in this case.

9.3.5 *Robot Character*

The robot was situated as a slightly older peer relative to the child, as in my prior work (Kory-Westlund et al., 2016a, 2017b; Kory and Breazeal, 2014). As in my recent Entrainment/Backstory study (Chapter 8), the backstory shared about the robot (both by the experimenter and reinforced by the robot itself) included that the robot was not so good at hearing, so it needed help and practice to get better. The experimenter pointed out that the robot's ears "are hidden under all its fur, so it doesn't always hear things right"—the child has to speak loudly and clearly to help the robot out. In both conditions, the robot disclosed that it's not very good at listening, and sometimes hears things wrong and says the wrong things. This set up a plausible reason for any failures in the automatic speech recognition (ASR). It also situated the child as an expert relative to the robot in the domain of hearing and listening (which could reinforce the peer relationship), and set up one goal of the interaction: to help the robot practice listening and talking. In this way, the conversation games and storytelling activities could be framed as practice for the robot as well as playing. The story retells, in particular, were framed as practice—the robot wants help making sure it got the story right, so it asks the child to retell the story to check, and gets to practice listening at the same time.

In prior work (Kory-Westlund et al., 2016a, 2017a,b; Kory and Breazeal, 2014), the robot has started out as a very excited and animated character, and possibly a little intimidating in its enthusiasm. Thus, unlike these prior studies, the robot started out more shy, and over the first several sessions with the child, slowly became less shy. I expected that if the robot started out more hesitant, it might lead children to feel more comfortable initially. The robot's backstory could be used to explain its behavior: Perhaps the robot is shy because it's new to the school, and is afraid the child won't like it because it isn't always good at listening. Some prior work suggests that if a robot exhibits some weakness, such as being small, fragile, or not knowing everything, people may be more likely to collaborate with it (Nishiwaki et al., 2017).

9.3.6 *Materials*

A Mac Mini computer automatically ran the interaction control code and exposed a web interface on startup. The experimenter used a laptop to connect to the web interface in their browser, and used the buttons in the interface to start and stop the interaction with the robot. The experimenter also plugged two Logitech webcams into their laptop, which were used to record audio and video. One camera was set on the table to get a direct front view of the child's face; the other was set on a tripod to the side to get a side view or bird's-eye view. Figure 35 shows the area setup.

A Google Nexus 9 8.9-in tablet was used to show vocabulary assessments, all pictures during the conversation portion of each session, the storybooks, and story scenes. Touchscreen devices have successfully engaged children and robots in a variety of shared activities, including storytelling (Kory-Westlund et al., 2017b; Kory and Breazeal, 2014; Park, Coogle, and Howard, 2014). Custom software was used to display the pictures and stories during the robot interaction, which is open-source and available online under the MIT License at `https://github.com/mitmedialab/SAR-opal-base/`. The experimenter was instructed to set up the tablet on a tablet stand either horizontally or vertically at the start each session with the robot, so that the pictures and storybooks would consistently be shown at maximum size on the screen.

The robot used was Tega (described in detail in Section 2.6), a squash-and-stretch Android phone-based robot capable of expressive movement (Kory-Westlund et al., 2016c). The phone used was a Samsung Galaxy S7. An animated face was displayed on the phone's screen; the phone's microphone was used to capture audio. An additional camera was mounted in the robot's forehead, which was used for some of the robot's affect detection and affect mirroring behaviors. I recorded all of the robot's speech. Each utterance was shifted to a higher pitch to make it sound more childlike. Speech is recorded by a human rather than using a text-to-speech voice in order to achieve a sufficiently expressive voice, since we have found in prior work that the voice's expressiveness can impact children's engagement and learning (Kory-Westlund et al., 2017b).

The robot operated completely autonomously, with the exception of a few very specific scenarios—e.g., the Story Negotiation Task. These specific scenarios were used as assessments of children's relational behavior, and as such it was very important that the robot respond consistently to all children in reaction to the children's choices in these scenarios. For example, during the Negotiation Task, the robot shows the child two pictures and asks the child which picture they want to tell stories about. The child could say any of a wide variety of phrases to indicate any of these categories of responses, and natural language understanding is currently not good enough to determine with high accuracy which category any given response may fall into. For this task to be valid and useful as an assessment, the robot needs to "hear" the child correctly in all cases. Thus, we had the experimenter listen to the children's responses and press buttons in a web interface to indicate the response category (e.g., for the negotiation, i.e., no response, refusal to consider the robot's suggestion, acquiescence to the robot's choice, several different compromise outcomes). The experimenter's input was only required to ensure the outcome of the scenario was recognized appropriately for all children.

I manually ran a script between sessions that generated a custom configuration file that was automatically loaded for each participant when a session began. This

Figure 36: The eight story scenes for the Create A Story activity.

script took the prior session's log files as input, along with a separate configuration file that contained any special input from the experimenter (e.g., the child's favorite color, whether the child had reported that they liked the story last session). The experimenter only had to select the appropriate participant ID number when starting the robot for a particular experimental session based on a mapping of children's names to ID numbers, and then everything else was loaded automatically using the generated session configuration files.

Initially, automatic speech recognition (ASR) was performed online using the Google Cloud Speech API, with requests sent using audio collected from the robot's microphone. However, due to highly unreliable Wi-Fi connections at several of the schools and highly unreliable ASR as a result, I switched to using a local, offline Google ASR app, which was run on one of the Samsung Galaxy S7 phones. The phone was placed on the table near the robot so that it would pick up the child's speech through the phone's microphone.

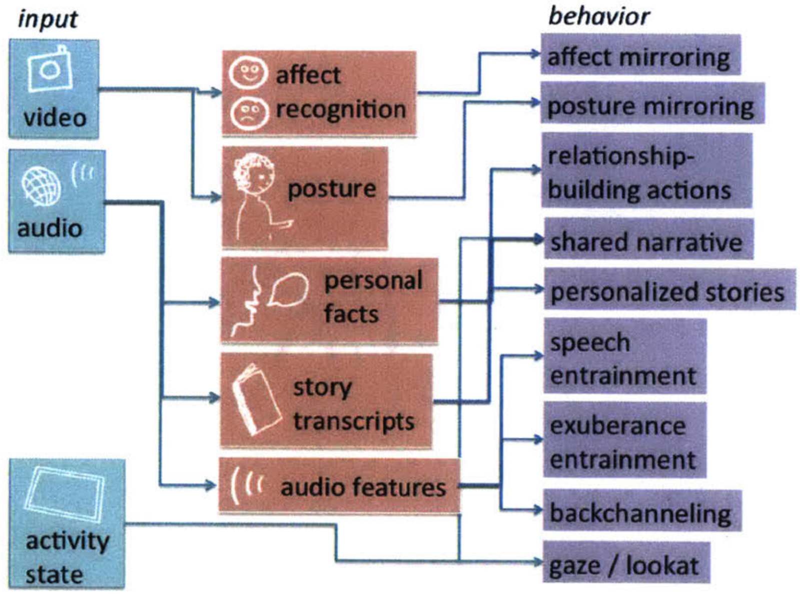


Figure 37: The relational robot system.

9.3.6.1 *Stories*

I selected stories from three corpuses for the robot to tell. First, the robot told several stories from existing picture books for the Story Retell activity, from a corpus developed for Park et al. (2019). I used the stories Baby Bird's First Nest, Possum and the Peeper, Henry's Happy Birthday, and Baby Duck's New Friend. Each story had three levels with increasing levels of syntactic difficulty, such that an appropriate level could be selected for any given child. Each story had 5–7 target vocabulary words embedded in it, which the robot asked about and defined for the children when telling the stories. These stories were used in the sessions in which the child was asked to retell the story. These stories were long enough to provide sufficient material for a retell, and the pictures provided a useful prompt for the child during the retell.

Second, I used the set of 16 stories developed during my previous work (Kory-Westlund and Breazeal, 2015b; Kory and Breazeal, 2014). These stories are matched to single digital scenes (Figure 36) and each has two levels of difficulty. They each have three target vocabulary words embedded in them. These stories have a mean length of 274.5 words (SD = 23.4, min = 229, max = 325). These stories were used for the Create A Story activity.

Third, I selected a subset of 84 children's stories from (Kory and Breazeal, 2014). These stories were created by children about the same digital scenes as the second story corpus. Including this set of stories allowed the robot to connect children to each other by sharing stories it had heard from other children. The criteria used to select this subset of stories were (a) a minimum length of 50 words and (b) being reasonably coherent. After minimal editing for coherency and grammar (e.g., sometimes a word was missing from a sentence due to errors in transcription), the 84 stories had a mean length of 145.2 words (SD = 69.1, min = 54, max = 383). There were 7–16 stories from each story scene (M = 10.5, SD = 2.7). The robot alternated telling these stories collected from children with the stories specifically for it in the Create A Story activity.

9.3.7 *Personalization and Robot Behavior*

In both conditions, children received story level personalization, as this relates to the personalization of curriculum and is independent of the robot. In the *Relational* condition, the relational robot personalized a variety of behaviors that reflected the relationship with the child and contributed to the formation and maintenance of rapport and a relationship with the child (Figure 37). I should acknowledge that there are many other relational behaviors that could be personalized, such as the robot's role and level (e.g., is it a teacher, tutor, peer, novice and how far ahead or behind of the child is the robot in its expertise?); different nonverbal behaviors and the robot's sensitivity to the child; personality (e.g., aspects of extraversion, openness, conscientiousness, and so forth); development of common ground in conversation; use of humor, politeness, or flattery; amount of self-disclosure; mirroring of language use in addition to mirroring nonverbals; adapting behaviors based on the model the robot forms of the relationship and how relationships progress over time; and much more. The behaviors below were what I selected for this initial study based on what was feasible and interesting to include in this particular robot in the timeframe available, with the goal of enabling the robot to appear as a peer-like social agent.

9.3.7.1 *Story Level*

Children in both the *Relational* and Not relational conditions received appropriately leveled stories from the robot. The story levels were determined prior to the start of the study, based on the results of an initial vocabulary pretest. Children who had higher scores were assigned higher story levels, while children who had lower scores were assigned lower story levels. Based on prior work with the story corpuses and the target vocabulary words, I expected that children would score above chance (12.3 of 49 words correct). In prior work, children scored a mean of 13.4 of 24 (*median* = 12.5, mode = 12) on the Create a Story corpus words and a mean of 9.5 of 25 (*median* = 9, mode = 8) on the Story Retell corpus words. Thus, I expected children's pretest scores to have a mean of approximately 23 of 49, median of approximately 21, and a mode of approximately 20. As a result, I decided to assign children easy stories (level 1 Create and level 1 Retell) if their pretest scores were less than 18, somewhat harder stories (level 2 Create and level 2 Retell) if their pretest scores were between 18 and 22, and hard stories (level 2 Create and level 3 Retell) if their pretest scores were above 22.

Based on children's pretest vocabulary scores, 12 children heard easy stories, 16 heard somewhat harder stories, and 21 heard the hardest stories.

9.3.7.2 *Story Content Personalization*

In both conditions, the storybooks used for the Story Retell activity were the same for all children. In the *Relational* condition, the stories told by the robot for the Create A Story activity were selected to be the most similar to the child's previous stories. This was accomplished by calculating the cosine similarity of the child's stories to each of the robot's stories, then selecting the story with the highest similarity score. The most similar stories were used because in prior work, children tended to use similar phrases, themes, concepts, and characters as the robot over time; this would approximate the robot mirroring the child back (Kory-Westlund et al., 2017b; Kory and Breazeal, 2014) (also Chapter 8 and Section 5.6).

Children in the *Not Relational* condition all heard the same set of pre-selected stories. Children in the *Relational* condition who did not tell any stories were also given these pre-selected stories. These stories were selected such that they were close to the mean story length, and were selected from seven different children storytellers in the corpus. The reason for selecting stories from different storytellers was to ensure children did not receive a set of stories that were all very similar to their own. Children's stories were expected to be most similar to the stories told by one or two specific storytellers in the corpus; thus, hearing stories from a variety of storytellers meant the children would likely hear multiple stories that were dissimilar to their own.

9.3.7.3 *Shared Narrative*

The relational robot referred back to events from previous sessions. In particular, the robot referenced prior conversations and stories (e.g., "Last time, we told a story about a baby bird. You said you liked it!"), prior negotiation tasks (e.g, "Last time, we did your choice."), and several key facts obtained about the child (e.g., "You said your favorite color was blue!"). The relational robot also used continuity behaviors, such as talking about what it did while away from the child (e.g., "Yesterday in the lab. . . ").

The relational robot used the child's name when it said hello and goodbye, as well as when asking key questions during the interaction (e.g., "Hey, [name], will you tell the story back to me?") I recorded each child's name after the pretest session so that the robot could play back the names at appropriate times during its speech. In addition, the relational robot periodically talked about its relationship with the child, e.g., "I didn't have many friends before I came to your school to play. I like having friends like you."

The not-relational robot did not reference past events, use the child's name, or talk explicitly about the relationship. It used more generic statements instead of personalized ones.

9.3.7.4 *Relationship*

The relational robot used a variety of prosocial, relationship-building and relationship maintenance actions that were intended to show that the robot was remembering and paying attention to the interactions it had with the child, and that what it remembered were impacting its actions. These actions included:

- The shared narrative behaviors mentioned above.
- The robot asked about the child's favorite color. The experimenter records the color. The following session, the robot showed a picture of that color the tablet and said, e.g., "I remember you said your favorite color was blue, so I made the screen blue for you!"
- The robot asked about the child's favorite animal, and the following session, told the child it remembered and that it found a picture of one to share. The robot also shared another picture in the final session.
- The robot used phrases indicating that it is paying attention to the child and wants to be friendly/helpful, such as, "I think you'll really like this story, it's about a kid like you!"

- The robot reinforced its backstory, saying again that it was new to the school, not so good at hearing, needed help and practice, and was trying hard to be a good friend.

9.3.7.5 *Disclosure*

In addition to the Self-Disclosure Task, which was included in the first and final (eighth) sessions, the relational robot prompted for disclosure from the child by asking questions about the child's life, and disclosed information about itself each session. The disclosures were made were increasingly personal over time, based on the levels discussed in Rotenberg (Rotenberg, 1995): (1) information about the environment, people, and activities, e.g., "I have a box I go in to be carried places. And there are some other robots at my lab at MIT"; (2) personal preferences, e.g., "I like story games and word games, and I don't eat people food because I'm a robot. I run on batteries and electricity!"; (3) good and bad things about oneself, such as good and bad aspects of one's appearance, self, and actions, e.g., "Sometimes I still interrupt and don't listen. I'm also afraid other kids think I'm weird because I can't walk." The Self-Disclosure Task used information that is good or bad about oneself; during the other sessions, the robot disclosed information from the other two levels. The not-relational prompted for disclosure, but did not follow-up with information about itself.

The prompts were: (S1) "I like stories. I think I'm good at telling stories because I try hard to tell nice stories. I also think my blue fluffy hair is cool. What about you? What things are good about you or what things you can do well?" and "But, did you know, sometimes I interrupt and don't listen. I'm also afraid other kids think my fur looks weird. What about you? What things you are not so good at or what things you tried but didn't go so well?" (S2) "Can you tell me things about your school or what your classroom looks like?", (S3) "Can you tell me things about where you live or what your house looks like?", (S4) "Can you tell me things such as how you go to school and things about your brothers and sisters or friends?", (S5) "Can you tell me things such as what you do every morning and things about your mom and dad or your pets?", (S6) "Can you tell me some of the things you like and don't like, such as foods and games you like, and food and games you don't like?", (S7) "Can you tell me some more things you like and don't like, such as things you like to do at home and at school, and things you don't like to do at home or at school?", (S8) "I like telling stories with you. It's fun. I think I'm good at telling stories because I try hard to tell good stories. I also think my flappy wings are cool. What about you? What are things are good about you or things you can do well?" and "But, did you know, sometimes I still interrupt and don't listen. I'm also afraid other kids think I'm weird because I can't walk. What about you? What things are you are not so good at, or things you tried but didn't go so well?"

9.3.7.6 *Speech Entrainment*

The relational robot automatically adapted its speaking rate and pitch to be more similar to the child. The child's speech was automatically collected via the robot's microphone when it was the child's turn to speak during the conversation. Using the same algorithms as in my prior work (Chapter 8), 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. The software used is open-source and available online under a GNU General Public License v3.0 at `https://github.com/mitmedialab/rr_audio_entrainer/`.

As before, all robot speech was sent through the audio entrainment module and streamed to the robot. In the *Relational* condition, all speech was entrained; for the *Not Relational* condition, processing still occurred, but the speech simply passed through and was not changed. The reason for this was to incur the same delay that results from entraining and streaming speech in both conditions (generally a latency of less than 2 seconds), and to allow us to record entrainment data regarding how similar the child's speech was to the robot's in both conditions.

9.3.7.7 *Exuberance Entrainment*

The relational robot also automatically adapted to the child's *exuberance*—i.e., for quieter, less exuberant children, the robot used a different set of prompts, quieter/slower animations and sometimes fewer of them, and ramped up its volume and excitement more slowly over the course of a session. For louder, more exuberant children, the robot used more excited animations and was louder and more excited during the entire session. Exuberance was determined based on how often the child responded to the robot's questions, response latency, and from features of the child's speech, e.g., speaking rate, volume, and intensity. More exuberant children were expected to respond more often, need less prompting to speak, speak more quickly, and speak with greater intensity and volume.

For children in the *Not Relational* condition, the robot spoke at an average volume, and used some quieter/slower animations as well as some louder/excited animations; it did not favor one kind of animation over another.

9.3.7.8 *Posture*

The relational robot performed a *lean in* animation whenever the child was detected leaning in suddenly and performed a *lean out* animation whenever the child was detected leaning out suddenly (defined as moving more than 15 cm in the same direction in the span of about 3 seconds). These values were manually calibrated based on pilot data.

9.3.7.9 *Lookat*

The robot's gaze behaviors were partly scripted and partly reactive to the interaction state and child's behavior. In both conditions, the robot looked at the child whenever it was the child's turn to speak.

For the relational robot, when the child was telling or retelling a story, the robot's gaze focused primarily on the tablet, glancing over at the child every 5 +/- 0.5s and staying on the child for 2.5 +/- 0.5s. This interval was used previously in (Kory and Breazeal, 2014), based on pilot data collected for that study, and seemed to work reasonably well. In the robot's interaction script, there were several key gaze behaviors scripted in, such as looking down to appear shy early in Session 1, or looking at the tablet when asking the child if they wanted to play a story game. The robot also glanced at the tablet for 2.5 +/- 0.5s every time something new appeared on the tablet (e.g., a picture was loaded during a task; a story was loaded), or when the story page was turned when the robot was telling a story.

For the not-relational robot, lookat commands were sent at approximately the same intervals as to the relational robot, but in randomly selected directions (e.g., at the tablet, at the user, to the left, to the right, up, down) instead of being directed in appropriate directions. Thus, when new things appeared on the tablet, the non-relational robot performed a lookat as well (but in a random direction); during the child's story, the robot looked every 3.5 +/- (-1.0, 2.0) seconds, thus performing lookats at the same general intervals as the relational robot, but in random directions. This allowed the not-relational robot to move and gaze just as much as the relational robot, but in a less contingent, reactive fashion.

9.3.7.10 *Affect Mirroring*

The relational robot responded to children's affect by mirroring back similar expressions, in the form of playing back appropriate animations. The robot mirrored only during conversations or listening, not when telling a story. It mirrored positive/negative valence, happiness/joy (including smiles and laughter), surprise, sadness, and thinking/puzzled expressions. In addition, if the robot detected that the child was not attending, it performed animations intended to recapture attention.

9.3.7.11 *Backchanneling*

In the *Relational* condition, backchanneling actions, such as "Uh huh!", smiles, nods, and other appropriate animations and speech, were sent at times indicated by the backchanneling module described in (Park et al., 2017a; Park et al., 2017c). Backchannel opportunities were detected during child speech and categorized into one of three groups that had different action sets associated with them, such as a long pause, or a short pause after a long period of speech. One action was randomly selected from the appropriate set. The robot's backchanneling commands were sent directly to the robot when they were motion/animations. Speech backchannels (e.g., "uh huh," "mmhm!") were sent through the audio entrainer.

For the not-relational robot, backchanneling actions were sent randomly every 5.53 +/- 1.5s, like in the prior work (Park et al., 2017a; Park et al., 2017c), randomly selected from a set of all possible backchannel actions instead of the actions curated for each backchannel opportunity type.

9.3.7.12 *Fidgets*

In both conditions, the robot performed fidget behaviors when it was idle. Fidget behaviors are commonly used in video games when characters are standing around being idle; they are usually activated after a couple seconds of idleness. Characters might stretch, shift their weight, or look around. The robot's fidget behaviors fell into two categories: physical fidgets and speech fidgets. The physical fidgets included animations for shifting the robot's weight and small wiggles. Physical fidgets were randomly selected from the list and played back at a random interval between the min and max interval times (500–1000 milliseconds) when the robot was idle and not speaking. The speech fidgets included speech sounds such as "um," "uh," "er," "hm," and some small wiggling actions. They were played back during idle time between when the robot finished listening to the child and was preparing to play audio and when the audio actually began to play. The latency was generally not large (frequently less than 2 seconds), but sometimes ASR processing took more time (e.g., 3–5 seconds) or streaming audio to the robot took more time because of wifi issues, and thus, the speech fidgets let the child know that the robot was "thinking" about what to say next.

9.3.8 *Data*

During each session, a video camera recorded audio and the front view of children's faces. In some sessions, a second camera captured a bird's-eye view of the interaction from the side. Children's responses to the vocabulary tests were automatically recorded on the tablet. Children's responses to all questionnaires and interview-based assessments, as well as children's responses to key scenarios in the robot interaction (e.g., the negotiation task), were recorded on paper log sheets and later transferred to a spreadsheet.

9.3.9 *Data Analysis*

Forty-nine children did the Anomalous Picture Task pretest with a human experimenter (RR: 25, NR: 24); 46 did the posttest (RR: 23, NR: 23). With the robot, 48 children did the pretest (RR: 25, NR: 23) and 45 did the posttest (RR: 22, NR: 23). For each child, I counted the number of comments the child made during the task, the number of questions asked, and the number of times the child laughed. These were counted separately for the pretests and posttests with the human and with the robot. I also coded children's gaze during the task to determine whether they looked at their interlocutor, at the pictures, or elsewhere.

For each story children were asked to tell or retell, we coded whether they told the story as 0 = no story, 1 = some story (e.g., 1–2 words), and 2 = story (e.g., at least one sentence). Children's responses during the negotiation task were coded as 0 = refusal (do the child's choice), 1 = acquiescence (allow the robot's choice), or 2 = compromise (e.g., saying "let's do both" or "do my choice first then yours"). This scale was used because compromises are more likely among friends, as they attempt to reach an equitable solution for all (Hartup et al., 1988).

Forty-nine children did the pretest Narrative Description task, but audio or video data was missing for several children, leaving a total of 44 children (RR: 21, NR: 23); 45 children did the posttest task (RR: 23, NR: 22). I used automated tools to count the number of words in each child's response. Children's responses were also coded based on the kind of information provided: a name (e.g., "Red", "My friend Kimberly"), physical attributes (e.g., "red and blue", "has white skin"), social or cognitive attributes (e.g., "is nice", "is good at reading"), facts (e.g., "went to another school", "has a toy train", "is a boy"), and activities performed together (e.g., "read stories", "play", "sit together at lunch").

Forty-six children did the Social Relational Interview (SRI) pretest (RR: 22, NR: 24); we administered the SRI to one additional child, but he was excluded from analysis because he did not understand the questions. Forty-seven children did the SRI posttest (RR: 24, NR: 23). Following the protocol described in (Kory-Westlund et al., 2018), children's responses to the SRI were coded as 0 = robot does not care / not mind / is pretending, 1 = not sure, maybe, 2 = social (would be sad, would help, etc). Children's justifications for their responses were transcribed and coded following the rubric described in (Kory-Westlund et al., 2018). This rubric categorized children's justifications into eight categories: references to the robot's feelings (e.g., "Red wants to have fun", "likes telling stories"), references to attributes of the robot (e.g., "he's too little", "just a robot", "she's nice"), references to actions the robot took (e.g., "because Red said nice to meet you", "because she read me a story"), references to the child/self (e.g., "because I'm nice", "because I told her a story too"), references to actions taken by others (e.g., "someone was mean", "so the kid can't be sad"), references to the situation the robot is in (e.g., "because he doesn't have any friends", "because someone took the tablet"), references to consequences of the situation (e.g., "now Red can't read the story", "she won't have anyone to play with"), and references to moral judgements or obligations (e.g., "that's not nice", "not what people do", "because they are friends").

Forty-eight children did the Inclusion of Other in Self (IOS) task pretest (RR: 24, NR: 24) and 43 did the posttest (RR: 21, NR: 22). Children's selections of the circle-pairs presented in the task were recorded.

The Memory/Rapport questions were administered to 47 children after Session 4 (RR: 24, NR: 23) and to forty-three children after session 8 (RR: 21, NR: 22). The judgement and safe space questions were administered to children before session 3 (RR: 21, NR: 22) and to children after session 7 (RR: 21, NR: 22). The social acceptance questions were administered to children during the pretest session (RR: 21, NR: 22) and to children after session 7 (RR: 21, NR: 22). All of these questions were coded as 0 = no, 1 = maybe, and 2 = yes. Because many children did not provide explanations of their responses for these sets of questions, their explanations were examined to make sense of the results, but were not coded for content.

Regarding the Picture Sorting Task, 49 children completed the task before Session 2 (RR: 25, NR: 24), and 47 completed it during the posttest session (RR: 24, NR: 23). I labeled children's placement of each entity, with the human adult anchoring one end at position 1 and the table anchoring the other at position 10. I counted positions to determine the rank held by each entity. As before (Chapter 8), I also computed scores for the robot's rank relative to the other entities, e.g., subtracting the human baby's rank from the robot's rank to obtain the robot's rank relative to both the human baby and the anchors.

Following the procedure in (Kory-Westlund et al., 2018), children's responses during each Self-Disclosure Task were coded for the type of information disclosed. There were five categories of information: physical skills (e.g., "standing on one foot," "I can swim very well," "riding a bike"), fine motor skills (e.g., "I'm good at writing," "I can draw a ball"), cognitive skills (e.g., "I read," "I'm bad at math"), social skills (e.g., "I'm good at teaching," "I don't know how to share"), and not specific skills (e.g., "I'm not good at many things," "I can do everything much better"). I also used software tools to count how many words children used in their responses during the Self Disclosure Tasks in Sessions 1 and 8, and during the disclosures during sessions 2–7. Because the questions asked by the robot in sessions 2–7 were relatively specific (Section 9.3.7.5), coding these would have required a different schema for each session. I deemed it sufficient to examine whether children responded at all, and the length of their responses (i.e., the amount of information disclosed).

I recorded what each child did during the goodbye with the robot each session, such as whether the child said anything or did anything. I coded these behaviors on a 3-point scale, with 0 = no goodbye (e.g., staring at the robot, doing and saying nothing), 1 = taking a small action (e.g., smiling, doing a small wave, or mimicking the robot's yawn), and 2 = saying something or doing a clear goodbye action (e.g., saying "Bye bye" or waving while saying "See you later"). I summed these scores across all sessions to get an overall Goodbye score. I also summed the scores for the first four sessions and then for the second four sessions to look at the gross pattern over time.

For the empathy/helping tasks, such as whether children agreed to do the third picture in the picture conversation activity, whether children verbally expressed sentiment when the robot told them it was sad that morning, and when the robot asked whether it was hard to play with a robot that couldn't hear well, I coded children's responses as 0 = no / no response, 1 = maybe / don't know / generic response (e.g., "why do you run on batteries?"), and 2 = yes / empathetic response (e.g., "awww!" and "that's sad").

I used Affdex, emotion expression measurement software from Affectiva, Inc., Boston, MA, USA (McDuff et al., 2016), to obtain measurements of children's facial expressions from the recorded videos for each session. As described earlier (Chapter 8), Affdex can detect 15 facial expressions which are used to determine whether a face is displaying expressions typically associated with 9 different affective states. For each video frame, Affdex attempts to detect a face and if it does so, 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. Affdex only recognizes outward expressions of affect, which does not imply detecting any underlying feelings.

From the 365 videos of children's faces recorded during all sessions (some videos were missing), I analyzed affect data for 340 videos. For the remaining 25 videos, affect was detected in fewer than 20% of the video frames and were thus considered to have insufficient data for analysis. Missing data resulted from system failures such has children's faces not being recognized by Affdex. I analyzed the same set of affective states and facial expressions as in the Entrainment/Backstory study (Chapter 8) (all detected from the face): joy, fear, sadness, surprise, contempt, disappointment, relaxation, engagement, valence, attention, laughter, and smiles.

Using software tools and the transcripts of children's stories, I counted how many times each child used each of the target vocabulary words and each of the robot's catchphrases in their stories. I summed all uses to get a total usage count. Because children did not all tell the same number of stories, I also divided this total by the number of stories the child actually told to get an estimate of usage per story. I identified an additional set of 7 keywords from each of the four Retell stories, which were uncommon words that generally appeared only once or twice in the story, such as "raccoon", "vanilla", and "whistle". One word, "feather", appeared in two stories, bringing this to a total of 27 additional keywords.

As in the Entrainment/Backstory study (Chapter 8), I analyzed children's story retells in terms of length (word count), overall word usage, usage of target vocabulary words and of the additional keywords, and similarity of each child's story to the corresponding robot story. I used the same automatic tools to obtain similarity scores using the same phrase and word matching algorithm, with the same parameters, as before.

The analyses reported below were planned comparisons based on my hypotheses unless otherwise indicated. Age is frequently included as a covariate because children's relationships may form differently as they grow up, they may differ in how long it takes for them to feel comfortable in new situations, and they may differ in language ability or background. As such, I expected children's age might affect how they responded to the robot and their language use. I also expected that children's responses may vary as a result of gender because we have previously seen gender dif-

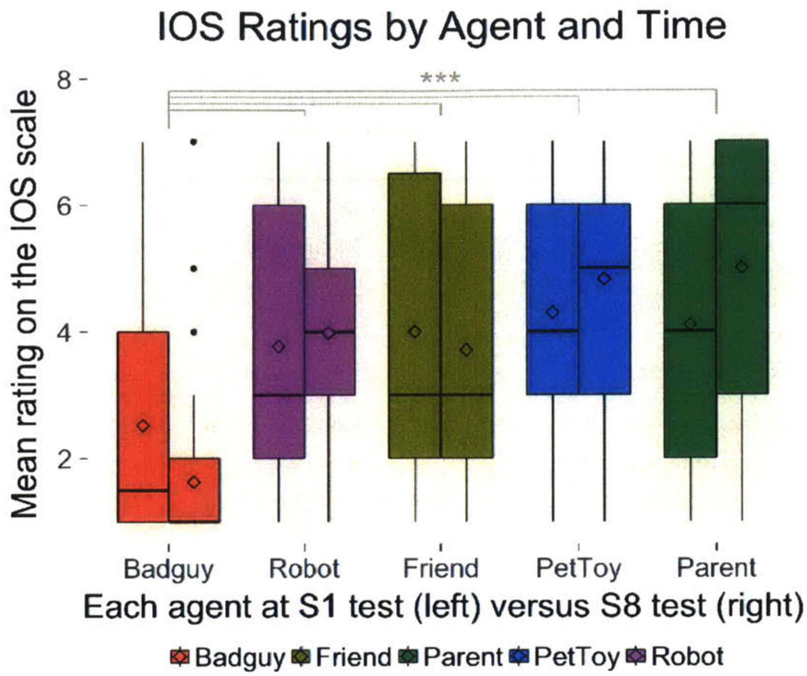


Figure 38: Children's overall Inclusion of the Other in Self responses from the Session 1 and Session 8 tests. A 1 indicates choosing the farthest apart circles, while a 7 indicates choosing the most closely overlapping circles.

ferences in children's relationship formation and relationship ratings (Kory-Westlund et al., 2018). As reported later in Chapter 10, I performed additional planned analyses where I included gender.

9.4 RESULTS

Thirty-nine of the 49 children successfully completed all 8 sessions with the robot. One of the youngest children was very shy and did not finish the first session; one was very distracted and declined to finish session 2; two were absent during administration of session 8; three failed to complete two sessions due to distraction and lack of interest; one child changed schools after completing two sessions; and one child declined to participate further after completing five sessions.

I divide the results below into three parts, each reflecting one of my hypothesis areas: (1) *Relationship:* I asked whether children who played with the relational robot would have greater rapport and a closer relationship than those who played with the non-relational robot; (2) *Learning:* I asked whether the robot's use of relational behaviors would increase children's learning, and whether children who felt closer to the robot would also learn more; and (3) *Behavior and engagement:* I asked whether children would show greater engagement, positive affect, and peer mirroring, and treat the robot as more of a social other with the relational robot.

Table 4: Children's overall Inclusion of Other in Self responses by condition after Session 1 and after Session 8. A 1 indicates choosing the farthest apart circles, while a 7 indicates choosing the most closely overlapping circles. The MAD is the median absolute deviation (a similar measure to the inter-quartile range, but arguably more robust); "Md." is "Median"; "Rg." is "Range."

Entity		S1				S8			
		Mean (SD)	Md.	Rg.	MAD	Mean (SD)	Md.	Rg.	MAD
Bad	All	2.52 (1.96)	1.5	1–7	0.74	1.63 (1.27)	1	1–7	0
Guy	RR	2.43 (1.70)	2	1–6	1.48	1.64 (1.50)	1	1–7	0
	NR	2.62 (2.25)	1	1–7	0	1.62 (1.02)	1	1–4	0
Parent	All	4.10 (1.94)	4	1–7	2.97	5.00 (2.04)	6	1–7	1.48
	RR	4.46 (1.93)	4.45	1–7	2.22	5.09 (2.13)	6	1–7	1.48
	NR	3.75 (1.92)	3.5	1–7	2.22	4.91 (2.00)	4	1–7	2.97
Pet/	All	4.30 (1.86)	4	1–7	1.48	4.82 (1.70)	5	1–7	1.48
Toy	RR	4.58 (1.74)	4.5	2–7	2.22	5.04 (1.61)	5	2–7	1.48
	NR	4.00 (1.98)	4	1–7	1.48	4.59 (1.79)	5	1–7	1.48
Friend	All	4.00 (2.32)	3	1–7	2.97	3.71 (2.21)	3	1–7	2.97
	RR	3.70 (2.30)	3	1–7	2.97	3.43 (2.19)	3	1–7	2.97
	NR	4.00 (2.25)	3	1–7	2.97	4.29 (2.35)	4.5	1–7	3.71
Robot	All	3.77 (1.87)	3	1–7	1.48	3.98 (1.64)	4	1–7	1.48
	RR	3.92 (1.86)	3	1–7	1.48	4.17 (1.75)	4	1–7	1.48
	NR	3.62 (1.91)	3	1–7	1.48	3.76 (1.51)	3	1–7	1.48

9.4.1 *Relationship*

9.4.1.1 *Inclusion of Other in Self Task*

I performed a mixed analysis of variance with Condition (between: *Relational* vs. *Not Relational*), Time (within: test in S1 vs. test in S8), Agent (within: bad guy, best friend, parent, robot, pet/toy), with Age as a covariate. I observed a significant main effect of Agent, $F(4,389) = 23.7$, $p < 0.001$ and a significant interaction of Agent with Time, $F(4,389) = 4.20$, $p = 0.002$. Post-hoc tests with Tukey's HSD revealed that the bad guy was rated significantly lower than all other agents at both the S1 test and the S8 test (see Table 4 and Figure 38). The parent was rated significantly higher than the robot and the friend at the S8 test, but not at the S1 test; children's ratings of their parent increased over time though this was not statistically significant.

I also observed multiple intriguing trends that were not significant, but may be worth exploring in future work. First, older children tended to rate all agents more highly than younger children. There was a trend for children in the *Relational* condition to rate the robot more highly than the children in the *Not Relational* condition.

9.4.1.2 *Social Relational Interview*

SCORES I performed mixed analyses of variance with Condition (between: *Relational* vs. *Not Relational*), and Time (within: test in S1 vs. posttest), with Age as a covariate

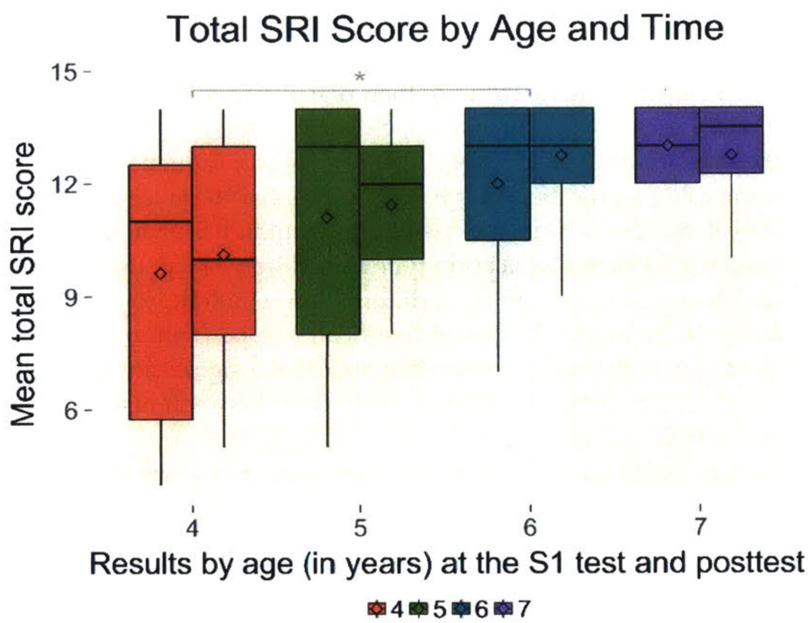


Figure 39: Children's total Social Relational Interview scores by age at the pretest and posttest. Scores on individual questions can range from 0–2, with 0 = robot does not care / not mind / is pretending, 1 = not sure, maybe, 2 = social/relational (would be sad, would help, etc); thus, the total score can range from 0–14.

for the total score and then, to explore children's responses further, for the individual questions. Where appropriate, I ran post-hoc tests with Tukey's HSD. Table 5 shows children's ratings by time and condition.

When examining the Social Relational Interview total score, I observed that older children rated the robot more highly than younger children, $F(3,39) = 3.11$, $p = 0.037$. (Figure 39). There was a trend toward a main effect of time, $F(1,42) = 2.62$, $p = 0.11$, with children's score decreasing slightly from the S1 test to the posttest.

Next, I examined each Social Relational Interview question individually to better understand how children thought about the robot. These results were were adjusted for multiple comparisons with the Bonferroni adjustment and considered significant if $p < 0.007$. Children were more likely to say that the robot would feel sad if another child was mean to it at the S1 test than at the posttest, $F(1,43) = 4.76$, $p = 0.035$; so were older children, $F(3,40) = 5.09$, $p = 0.005$. There was a trend for children in the *Not Relational* condition to be more likely to say that the robot would want to tell someone about a significant event than children in the *Relational* condition, $F(1,39) = 3.59$, $p = 0.066$.

With regards to cheering up another child, there was a trend for children in the *Not Relational* to be more likely to say the robot would cheer up another child than children in the *Relational* condition, $F(1,40) = 3.40$, $p = 0.073$. Nearly all children said the robot really did want to make friends. There was a trend for children in the *Relational* condition to be less likely to say the robot liked them at the S1 test than children in the *Not Relational* condition, $F(1,42) = 3.78$, $p = 0.059$.

JUSTIFICATIONS I examined children's justifications for their responses. A mean of 26.7 ($SD = 3.82$) of the 46 children who did the Social Relational Interview pretest provided justifications for each item (187 justifications total); a mean of 26.4 (SD = 3.69) of the 47 children who did the posttest did so (185 justifications total). A mean of 6.0 (SD = 2.52) children provided more than one justification type in their response at the pretest and a mean of 4.9 (SD = 0.90) children at the posttest. For example, one child referenced the situation, the robot's feelings, and consequences when explaining why she said the robot would be sad if it had no friends: "because she wants to know new things and if she doesn't have a friend she wouldn't hear new things." Table 6 shows the frequency of each justification type in children's responses by time.

A mixed analyses of variance with Condition (between: *Relational* vs. *Not Relational*), Time (within: test in S1 vs. posttest), and Justification type (within: the eight justification types coded for), with Age as a covariate, revealed a significant main effect of age, $F(3,44) = 4.52$, $p = 0.008$; and a significant interaction between justification type and time, $F(7,665) = 2.15$ $p = 0.037$.

Post-hoc tests revealed that the four-year-olds provided fewer justifications (total: $M = 2.50$, $SD = 3.78$; per question: $M = 0.28$, $SD = 0.89$) than the 6- and 7-year-olds (total: 5-year-olds: $M = 5.17$, $SD = 4.12$; 6-year-olds: $M = 5.33$, $SD = 3.20$; 7-year-olds: $M = 8.00$, $SD = 2.83$; per question: 5-year-olds: $M = 0.65$, $SD = 0.91$; 6-year-olds: $M = 0.65$, $SD = 1.00$; 7-year-olds: $M = 1.00$, $SD = 1.27$). With regards to justification types, some were used far more often than others; e.g., the robot's feelings and attributes were referenced more than moral judgments or actions the robot had taken. There was a trend for the robot's feelings to be referenced more often at the posttest than at the S1 test, but there were otherwise no differences in children's usage of individual justification types by time.

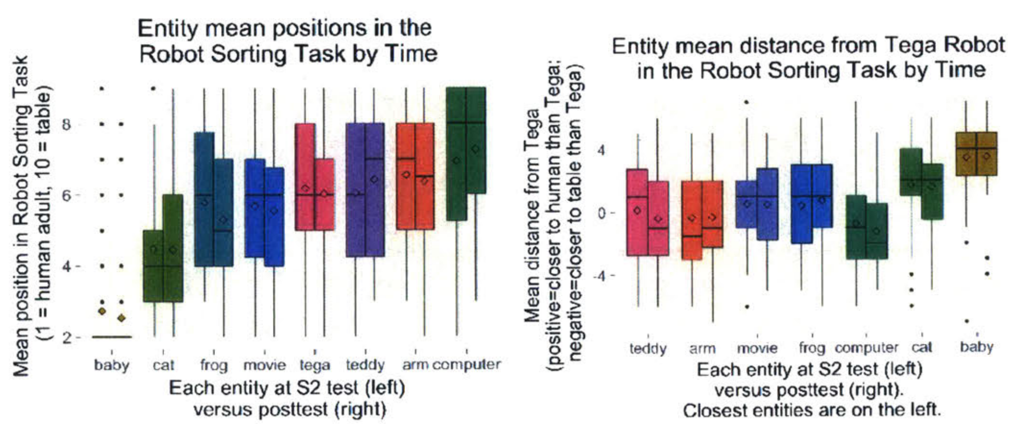
Children most commonly discussed the robot's feelings and the consequences of the situation. For example, one child thought the robot would be sad "because he would miss telling stories and looking at the pictures," while another said, "because he loves telling stories." Others said the robot wanted friends "because if he doesn't have friends he'll feel sad," and "so then he can play with them." The robot's feelings were mentioned most often with regards to why it might want friends, e.g., "so he doesn't feel lonely," and "ćause he's sad if he doesn't have friends."

Attributes of the robot came up frequently when children explained why the robot might help another child, e.g., "because he's a good friend," and "because she's nice." These were related to children's mentions of morality, which often referenced what it meant to be a good friend and why that might influence how one acts, e.g., "because it's nice," "because then that's being kind to each other, just like friendship," and "because it's being a good friend."

Children discussed actions the robot had taken and also referenced themselves the most when explaining why the robot liked them, e.g., "because she tells nice stories and I say I like them," "because he plays with me a lot," and "because I'm so nice to him." The situation was often referenced when explaining why the robot might be sad or why the robot might share information, e.g., "because his story tablet got given away," "didn't have any friends to play with," "ćause if something good happened like if he lost a tooth he would want to tell a friend." Other people were mentioned most often when explaining why the robot might cheer someone up or share information with someone, e.g., "so they can tell the teacher for help," and "because next time when he feels sad that guy will cheer him up."

Table 5: Descriptive statistics by time and condition for the Social Relational Interview. Scores on individual questions can range from 0–2, with 0 = robot does not care / not mind / is pretending, 1 = not sure, maybe, 2 = social/relational (would be sad, would help, etc); thus, the total score can range from 0–14. The questions were as follows. 1: "Let's pretend another kid was mean to Red and took Red's story tablet. Would Red feel sad or would Red not mind?"; 2: "Let's pretend Red didn't have any friends. Would Red not mind or would Red feel sad?"; 3: "Let's pretend another kid needs help. Would Red try to help or would Red not care?"; 4: "Let's pretend Red was really happy or really upset about something. Would Red not care about telling anyone, or would Red want to tell a friend?"; 5: "Let's pretend another kid is sad. Would Red try to cheer them up or would Red not care?"; 6: "Does Red really want to make friends, or is Red just pretending?"; 7: "Does Red like you or is Red just pretending?".

Question	RR		NR	
	S1 Mean (SD)	Post Mean (SD)	S1 Mean (SD)	Post Mean (SD)
Sad if mean	1.82 (0.59)	1.50 (0.86)	1.70 (0.70)	1.57 (0.84)
Sad no friends	1.50 (0.80)	1.55 (0.74)	1.48 (0.79)	1.43 (0.84)
Help child	1.76 (0.62)	1.57 (0.75)	1.74 (0.69)	1.83 (0.49)
Share information	1.62 (0.74)	1.48 (0.75)	1.91 (0.29)	1.70 (0.56)
Cheer up child	1.77 (0.61)	1.68 (0.65)	2.00 (0.00)	1.87 (0.46)
Wants friends	1.64 (0.79)	1.64 (0.73)	1.83 (0.49)	1.61 (0.78)
Likes you	1.22 (0.86)	1.67 (0.66)	1.78 (0.60)	1.65 (0.71)
Total	11.5 (2.14)	11.3 (2.72)	12.4 (2.19)	11.7 (3.19)

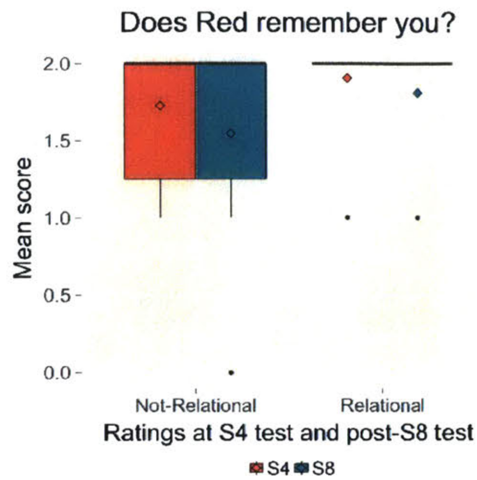


(a) Entity placement. (b) Entity distances from Tega.

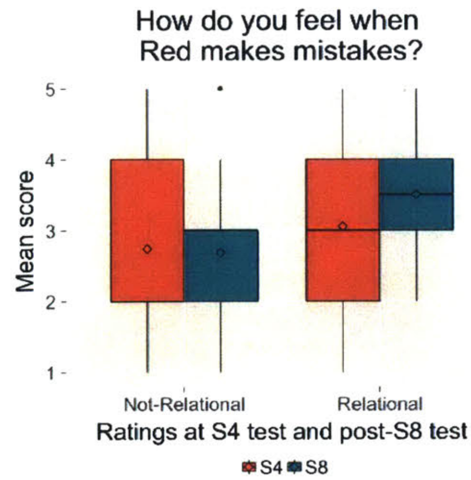
Figure 40: Picture Sorting Task results at the S2 test and at the posttest.

Table 6: Frequency of justification types used by children to explain their Social Relational Interview responses at the pretest and posttest. Not all children provided a justification at each time, and some children used more than one type in their responses. Here, "Fe." = Feelings; "At." = "Attributes"; "Ac." = "Actions"; "Se." = "Self"; "Ot." = "Others"; "Si." = "Situation" ; "Co." = "Consequences"; "Mo." = "Moral"; "No." = "None". The questions were as follows. 1: "Let's pretend another kid was mean to Red and took Red's story tablet. Would Red feel sad or would Red not mind?"; 2: "Let's pretend Red didn't have any friends. Would Red not mind or would Red feel sad?"; 3: "Let's pretend another kid needs help. Would Red try to help or would Red not care?"; 4: "Let's pretend Red was really happy or really upset about something. Would Red not care about telling anyone, or would Red want to tell a friend?"; 5: "Let's pretend another kid is sad. Would Red try to cheer them up or would Red not care?"; 6: "Does Red really want to make friends, or is Red just pretending?"; 7: "Does Red like you or is Red just pretending?".

Question		Fe.	At.	Ac.	Se.	Ot.	Si.	Co.	Mo.	No.
Sad if mean	S1	6	2	1	1	7	6	13	2	14
	Post	13	3	1	0	4	4	8	3	17
Sad no friends	S1	8	0	1	0	5	9	11	1	15
	Post	9	1	0	1	0	5	19	1	18
Help child	S1	4	10	1	0	5	4	7	4	19
	Post	11	6	1	0	2	2	7	4	17
Share info	S1	7	0	1	1	9	6	8	2	20
	Post	10	1	1	2	6	4	4	1	26
Cheer up	S1	7	3	2	0	6	1	7	4	20
	Post	10	5	0	2	7	2	2	4	20
Wants friends	S1	13	4	1	2	0	7	8	0	22
	Post	12	4	1	0	4	2	5	3	21
Likes you	S1	3	3	8	11	0	0	0	0	25
	Post	5	3	7	11	1	0	0	1	25
Total	S1	48	22	15	15	32	31	55	13	130
	Post	70	23	11	16	24	19	45	17	137



(a)



(b)

Figure 41: Children's responses to two Memory/Rapport questions by condition and time. Children responded to the question "Does Red remember you?" with a *yes* (coded as 2), *maybe* (1) or *no* (0). Responses to "How do you feel when Red makes mistakes?" were given on the smileyometer scale of 1 (very frowny) to 5 (very smiley).

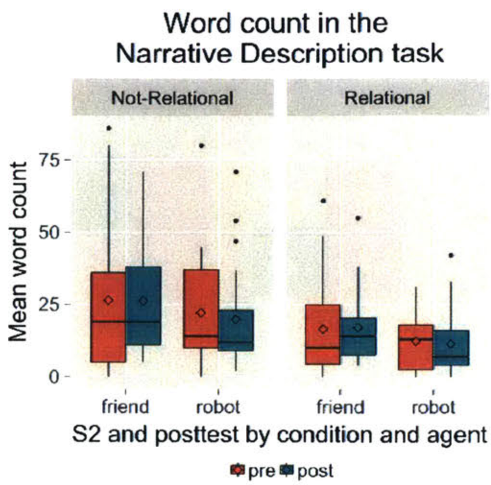
9.4.1.3 *Picture Sorting Task*

At the S2 test, children placed the robot at a mean position of 6.17 (SD = 2.04); at the posttest, they placed the robot at a mean position of 6.02 (SD = 1.87) (Figure 40a). I performed mixed analyses of variance with Condition (between: *Relational* vs. *Not Relational*), Time (within: test in S1 vs. test in S8), Entity (within: robot, baby, cat, frog, teddy bear, movie robot, robot arm, computer), and Age as covariate for each entity's position, as well as for each entity's position relative to the robot. For entity positions, I observed a significant main effect of entity, $F(7, 656) = 45.9$, $p < 0.001$.

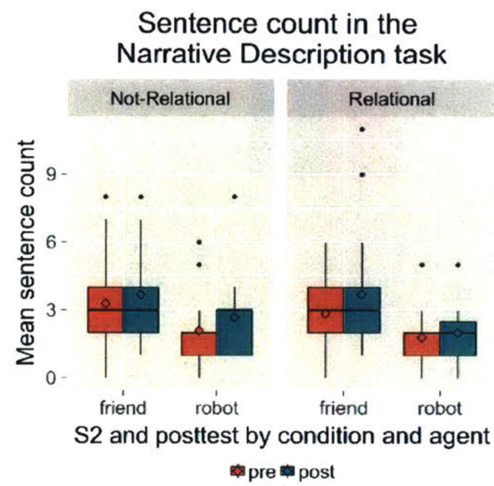
Post-hoc tests revealed that the baby was placed significantly closer to the human adult than all other entities. The cat was placed significantly closer to the human adult than all entities except the baby. The Tega robot was significantly closer to the human adult than the computer, and farther from the adult than the baby and the cat, but was otherwise not placed at a significantly different position from any other entity. The computer was placed significantly closer to the table (and farther from the human adult) than all entities except the robot arm and the teddy bear. Finally, the frog was placed closer to the human adult than the robot arm.

There was a trend for an interaction of entity with condition, suggesting that children in the *Relational* condition may have placed the Tega robot and baby closer to the human, and the teddy bear and computer farther, than children in the *Not Relational* condition.

Regarding the distance of each entity relative to the Tega robot, I observed a main effect of entity, $F(6,568) = 45.9$, $p < 0.001$. The baby and cat were both placed farther from Tega, and closer to the human adult than Tega was, than all other entities (Figure 40b). The computer was farther from Tega than the frog and movie robot were (as well as farther from the human adult). There was a trend for an effect of condition, suggesting that in the *Relational* condition, the cat, frog, and movie robot were closer to Tega, and the computer and teddy bear were farther, than in the *Not Relational* condition.



(a) Word Count.



(b) Sentence Count.

Figure 42: In the Narrative Description task, children described their friends at greater length than they did the robot, and this varied by condition and gender.

9.4.1.4 *Memory/Rapport Questions*

I ran mixed analyses of variance with Condition (between: *Relational* vs. *Not Relational*), and Time (within: questions at S4 vs. S8), with Age as a covariate. For the question, "Does Red remember you?", I observed a trend toward a main effect of condition, $F(1, 38) = 3.71$, $p = 0.061$. Children in the *Not Relational* condition were less likely to say they thought the robot remembered them, but children in the *Relational* condition were fairly confident that the robot remembered them (Figure 41a). About half the children provided an explanation for their response. Children in the *Relational* condition were more likely to state that the robot remembered them because it knew their names, e.g., "Because she said my name and she remembers I telled a story to her last time." In the *Not Relational* condition, children's reasons more often involved playing together, e.g., "I see him all the time."

There were no significant differences for the questions "How much do you like Red?", "Does Red make mistakes?", or the follow-up, "How do you feel if Red makes mistakes?" There was a trend for children in the *Relational* condition to feel more positively if the robot made a mistake ($M = 3.28$ of 5, $SD = 1.30$) compared to children in the *Not Relational* condition ($M = 2.71$, $SD = 1.25$),$F(1,32) = 3.10$, $p = 0.088$ (Figure 41b).

9.4.1.5 *Narrative Description*

During the S2 test, 40 children described their best friend and 34 children described the robot. At the posttest, 42 children described their friend and 36 children described the robot. At the S2 test, children's descriptions of their friends were a mean of 20.1 words ($SD = 22.0$; sentences: $M = 2.91$, $SD = 2.03$), and of the robot were a mean of 16.6 words ($SD = 16.1$; sentences: $M = 1.89$, $SD = 1.51$). At the posttest, children's descriptions of their friends were a mean of 20.5 words ($SD = 17.3$; sentences: $M = 3.44$, $SD = 2.25$), and of the robot were a mean of 15.2 words ($SD = 15.7$; sentences: $M = 2.13$, $SD = 1.59$).

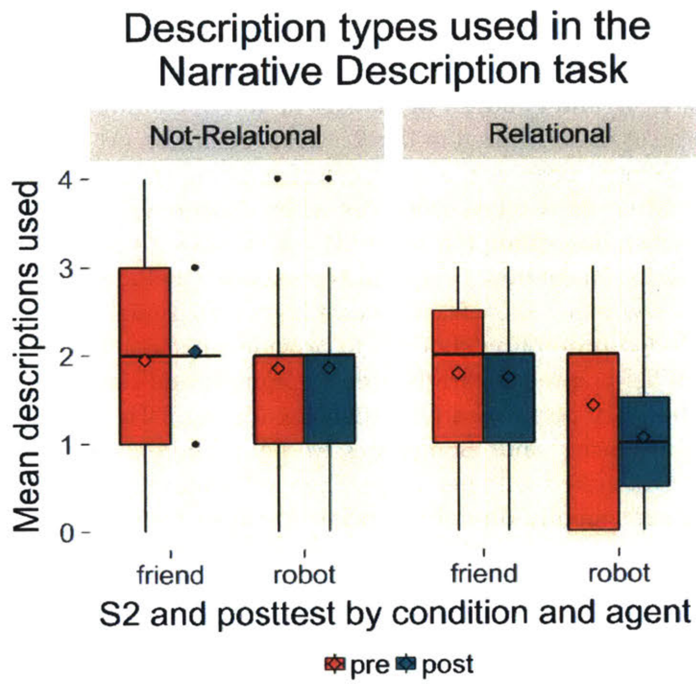


Figure 43: The number of different kinds of descriptions children used (name, facts, social qualities, physical attributes, activities) during the Narrative Description task.

Table 7: Frequency of descriptions used by children in the Narrative Description task at the pretest and posttest. Not all children provided a description at each time. Here, "Cond." = "Condition."

Agent	Cond.	Time	Name	Facts	Social	Physical	Activities
Best	RR	S2	12	15	1	4	3
Friend		Post	12	17	3	2	3
	NR	S2	14	18	6	0	6
		Post	14	15	5	1	8
Robot	RR	S2	5	7	9	6	0
		Post	7	7	7	2	4
	NR	S2	9	10	9	7	6
		Post	9	8	8	8	6

For word count, a mixed analysis of variance with Condition (between: *Relational* vs. *Not Relational*), Time (within: S2 test vs. Posttest), Agent (within: Friend vs. Robot), and Age as a covariate, revealed main effects of agent, $F(1,114) = 10.1$, $p = 0.002$, and condition, $F(1,35) = 4.45$, $p = 0.042$ (Figure 42a). Post-hoc tests revealed that children described their friends at greater length ($M = 21.8$, $SD = 20.0$) than they did the robot ($M = 16.7$, $SD = 15.9$), and children in the *Not Relational* condition used more words ($M = 23.7$, $SD = 20.9$) than children in the *Relational* condition ($M = 14.4$, $SD = 13.1$).

For sentence count, I observed a main effect of agent, $F(1,114) = 24.0$, $p < 0.001$ (Figure 42b). Children used more sentences when describing their friend ($M = 3.38$, $SD = 2.15$) than when describing the robot ($M = 2.15$, $SD = 1.52$).

Children's descriptions of their best friends most often included their friend's name and a fact, such as whether their friend was a boy or a girl (Table 7). This is because the protocol included prompting children to provide a name and/or gender for their friend as a way of helping them start their description. In addition, children frequently described activities they performed with their friends, e.g., "He likes to play with me. He sits with me at lunch," and "Sometimes we fight. Then we become friends. And then at night I get frustrated with her again. Then in the morning we're like hey what's up dude." They occasionally described social/cognitive features, such as "She likes reading. She likes singing. She likes playing," and "Nice. And happy. And excited. And sometimes shy. And sometimes angry. And frustrated." More rarely, children mentioned physical characteristics, e.g., "She has long hair and peach skin".

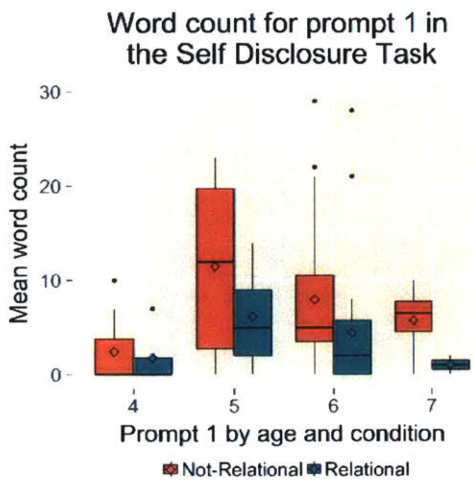
Children more commonly described the robot's physical characteristics as well as other facts, e.g., "It's red and blue," and "It's hard for Red to hear. That he has blue hair." Children often mentioned facts about the robot's hearing and sleeping, such as, "That Red can't hear very well because his ears are inside his fur and his ears are all the way inside his fur where you can't see. And you have to say ice cream to wake him up," and "He loves telling stories. He is pretty bad at hearing. So that's why he always likes to practice. And he's asleep right now. That's it." They also talked about activities performed with the robot, usually about the storytelling and conversations, e.g.,

Red reads me books and he shows me pictures and he talks to me," and "She talks a lot. And she likes to tell stories."

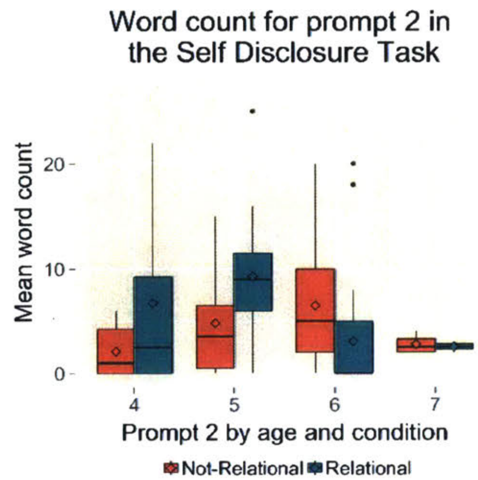
I summed the number of different kinds of descriptions each child used (name, facts, social qualities, physical attributes, activities). I ran a similar analysis of variance as before; I observed main effects of agent, $F(1,114) = 5.03$, $p = 0.027$; condition, $F(1,35) = 5.83$, $p = 0.021$; and age, $F(3,35) = 5.23$, $p = 0.004$ (Figure 43). Children used more kinds of descriptions when talking about their friend ($M = 1.89$, $SD = 0.90$) than about the robot ($M = 1.56$, $SD = 1.09$). Children in the *Not Relational* condition used more kinds of descriptions ($M = 1.93$, $SD = 0.99$) than children in the *Relational* condition ($M = 1.50$, $SD = 0.99$). Four-year-olds used fewer descriptions ($M = 0.92$, $SD = 0.93$) than all older kids (5-year-olds: $M = 1.89$, $SD = 0.74$; 6-year-olds: $M = 1.81$, $SD = 1.03$; 7-year-olds: $M = 2.25$, $SD = 0.75$).

9.4.1.6 *Targeted Self-Disclosure Task*

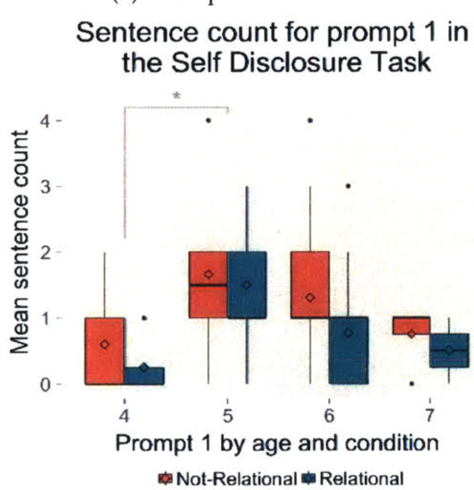
During the S1 Self-Disclosure Task (SDT), 39 children (of 47) spoke after one or both of the robot's prompts, with 32 disclosing information. During the S8 SDT, 30 children (of 45) spoke after one or both of the robot's prompts, with 19 disclosing information. For the disclosures in Sessions 2–7, a mean of 32.5 children ($SD = 4.4$) spoke after the



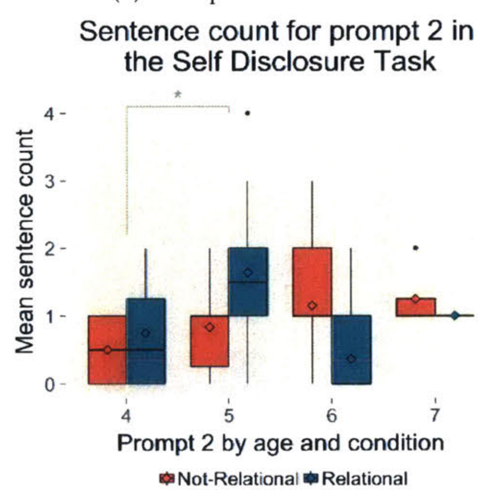
(a) Prompt 1 Word Count



(b) Prompt 2 Word Count

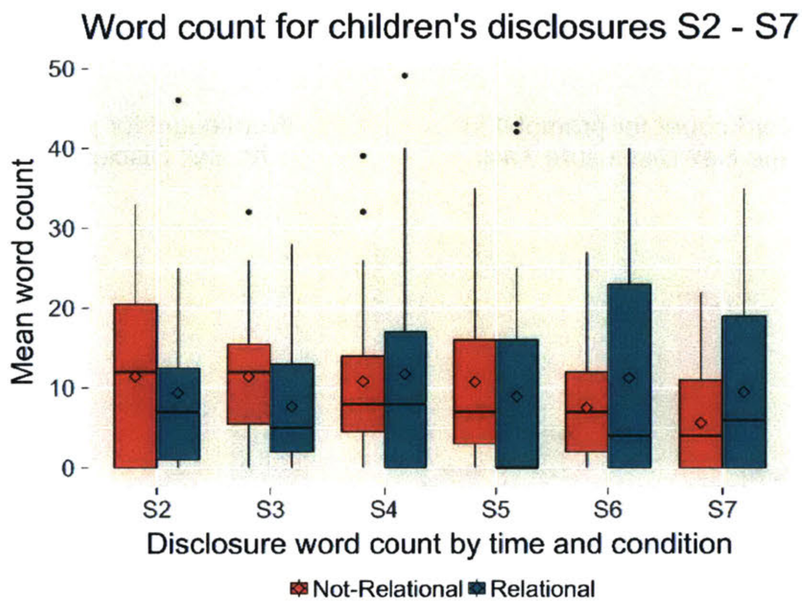


(c) Prompt 1 Sentence Count

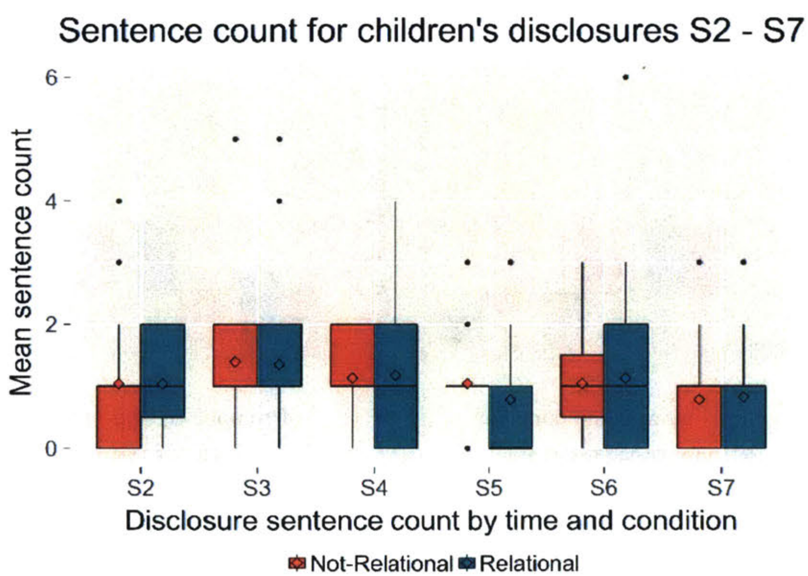


(d) Prompt 2 Sentence Count

Figure 44: In the Self-disclosure Task, girls generally gave longer answers than boys in S8 to both the robot's prompts.



(a) Word Count.



(b) Sentence Count.

Figure 45: Children's disclosures during Sessions 2–7 followed the pattern as the Self-disclosure Task in S1 and S8.

Table 8: Mean and standard deviation for the word and sentence counts for children's Session 1 and Session 8 disclosures by condition, gender, and time.

Time	Gender	Condition	Disclosure	Words	Sentences
S1	Girls	RR	1	4.50 (4.17)	1.10 (0.88)
			2	4.70 (3.43)	1.20 (1.14)
		NR	1	4.73 (5.90)	1.00 (1.10)
			2	5.55 (6.27)	1.09 (0.70)
	Boys	RR	1	5.36 (6.15)	1.09 (0.94)
			2	5.82 (6.88)	0.91 (1.04)
		NR	1	7.83 (7.43)	1.17 (0.83)
			2	5.83 (5.02)	0.83 (0.58)
S8	Girls	RR	1	9.70 (15.7)	0.90 (1.10)
			2	7.70 (8.73)	0.90 (0.74)
		NR	1	7.45 (7.47)	1.18 (1.25)
			2	10.6 (13.4)	1.18 (0.75)
	Boys	RR	1	2.82 (3.68)	0.73 (1.01)
			2	3.73 (7.24)	0.45 (0.82)
		NR	1	8.00 (9.30)	1.25 (1.14)
			2	3.33 (4.60)	0.83 (0.94)

robot's prompt (S2: 34, S3: 38, S4: 36, S5: 30, S6: 31, S7: 26). Children's mean word counts for each session are listed in Tables 8 and 9.

When prompted to disclose what they were good or bad at during S1 and S8, children frequently disclosed physical and fine motor skills, e.g., "I can draw well," "Reading, running, and building," "Doing cartwheels," and "I'm good at reading. I'm good at making paper airplanes." Table 10 lists the types of disclosures made in S1 and S8. A surprising number gave non-specific information, such as, "I'm good at everything," and "Many things." During the other sessions, children responded to the robot's prompts with relevant information. For example, in S2 when asked about their school and classroom, children often listed things in their classrooms, e.g., "My classroom is big and has a lot of toys and I liked the teachers." and "We have a white floor, a rainbow rug, a green rug, a bookshelf, book boxes, supply boxes...". In S3, when asked about where they live and their house, children often described the color of their houses or named the city they lived in, e.g., "My house is the color green. And I live on [street name] street," "My house is whitish-grayish," and "I live in [city name]". In S4, when asked about how they go to school and things about their siblings, children usually corrected the robot if they had no siblings, and explained whether they drove, took the bus, or walked to school, e.g., "I go to school by car," and "I go to school at the school bus, and I don't have a sister, I just have two brothers." In S5, the robot asked about what children did in the morning and about their parents and pets; children corrected the robot if they had no pet, and told the robot about their morning routines, e.g., "I don't have any pets. And my mom's taller than my dad," and "Um, I get dressed and I play. I eat breakfast, I feed my dog, and then I go to school." In

Table 9: Mean and standard deviation for the word and sentence counts for children's Session2–7 disclosures by condition, gender, and time.

Time	Gender	Condition	Disclosure	Words	Sentences
S2	Girls	RR		11.6 (12.8)	1.18 (0.75)
		NR		12.0 (10.5)	1.00 (1.10)
	Boys	RR		7.42 (8.31)	0.92 (0.79)
		NR		10.9 (11.4)	1.08 (1.16)
S3	Girls	RR		11.1 (8.10)	1.55 (0.93)
		NR		8.73 (6.97)	1.09 (0.70)
	Boys	RR		4.58 (5.93)	1.17 (1.47)
		NR		13.9 (8.67)	1.67 (1.23)
S4	Girls	RR		17.6 (15.5)	1.55 (1.04)
		NR		9.82 (8.7)	1.00 (0.63)
	Boys	RR		6.42 (9.69)	0.83 (1.19)
		NR		11.8 (12.2)	1.25 (0.75)
S5	Girls	RR		12.5 (17.1)	0.91 (1.04)
		NR		13.0 (11.2)	1.09 (0.70)
	Boys	RR		5.75 (8.52)	0.67 (0.98)
		NR		8.67 (9.30)	1.00 (0.85)
S6	Girls	RR		15.0 (12.2)	1.45 (1.04)
		NR		7.55 (6.04)	1.00 (0.63)
	Boys	RR		7.83 (14.2)	0.83 (1.75)
		NR		7.58 (8.10)	1.08 (1.00)
S7	Girls	RR		10.0 (11.9)	0.73 (0.79)
		NR		5.91 (5.24)	0.82 (0.75)
	Boys	RR		9.08 (10.1)	0.92 (1.00)
		NR		5.42 (7.45)	0.75 (1.06)

Table 10: Frequency of disclosure types children made during the Self Disclosure Task in S1 and S8. Not all children disclosed information, and many children disclosed more than one piece of information. Here, "F. Mot" = "Fine motor skills"; "Phys." = "Physical skills"; "Social" = "Social skills"; "Cog." = "Cognitive skills"; "N.S." = "Not specified".

Time	Disclosure	Phys.	F. Mot.	Social	Cog.	N.S.
S1	1	8	11	3	8	3
	2	7	5	2	2	6
S8	1	4	9	1	4	1
	2	9	1	1	0	4

S6, children were asked about their preferences, such as about food and games; they responded appropriately, e.g., "I hate hot sauce. That's all I can remember," and "I like Mario Kart. I like... I like Minecraft." In S7, the robot asked about what children liked to do at home or school. Children described activities they liked, e.g., "I like reading at school. And at home." and "Well, at school I like to draw but I don't like to play with the dinosaurs."

I performed mixed analyses of variance with Condition (between: *Relational* vs. *Not Relational*), Time (within: either S1 vs. S8, or S2–7), and Age as a covariate. There was a main effect of Age with regards to the number of sentences children spoke for both the robot's first prompt, $F(3,39) = 4.23$, $p = 0.011$; and the robot's second prompt, $F(3,39) = 3.37$, $p = 0.028$. There was a trend for four-year-olds to use fewer words in response to the first prompt, especially related to five-year-olds, $F(3,39) = 2.65$, $p = 0.062$. Four-year-olds used fewere sentences than 5-year-olds; 6- and 7-year-olds did not differ significantly (Figures 44a, 44b, 44c,and 44d).

I did not observe any significant differences between groups for the number of words in children's disclosures in sessions 2–7, though I did observe trends for a main effect of age, $F(5,220) = 1.97$, $p = 0.084$; and an interaction of condition with time,$F(3,41) = 2.50$, $p = 0.073$ (Figure 45a). Four-year-olds and 6-year-olds used fewer words and sentences. In the earlier sessions, children in the *Not Relational* condition tended to use more words, while in later sessions, children in the *Relational* condition used more words. There was a significant main effect of time on the number of sentences children used, $F(5,220) = 2.62$, $p = 0.025$. Children's responses in S7 were significantly shorter than their responses in S3 (Figure 45b).

9.4.1.7 *Social Acceptance Questions*

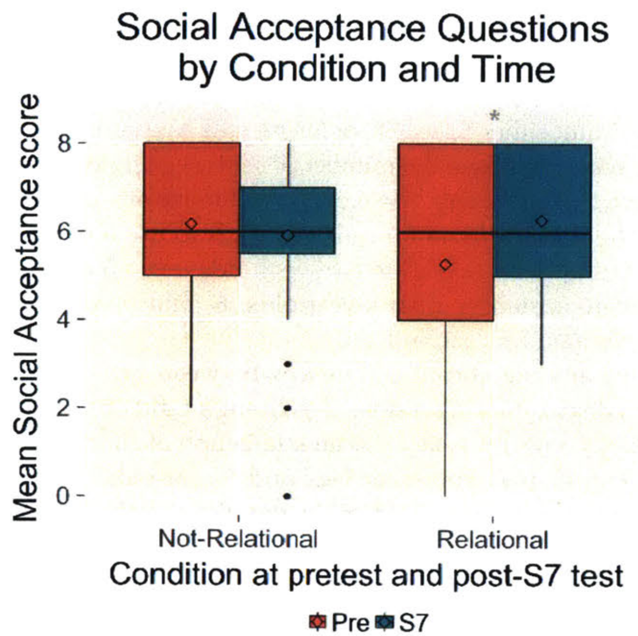
I performed mixed analyses of variance with Condition (between: *Relational* vs. *Not Relational*), Time (within: pretest vs. post-S7 test), and Age as a covariate. With regards to the sum score of the four Social Acceptance questions, I observed a significant interaction of condition with time, $F(1,43) = 4.16$, $p = 0.048$. Children in the *Relational* condition increased their ratings from the pretest to the S7 test, while children in the *Not Relational* condition slightly decreased their ratings (see Figure 46).

I examined the two questions that asked about children versus the two questions that asked about a robot separately in order to better understand children's responses. In both cases, there was a trend toward an interaction of condition with time that followed the same pattern as above (see Figures 46b and 46c).

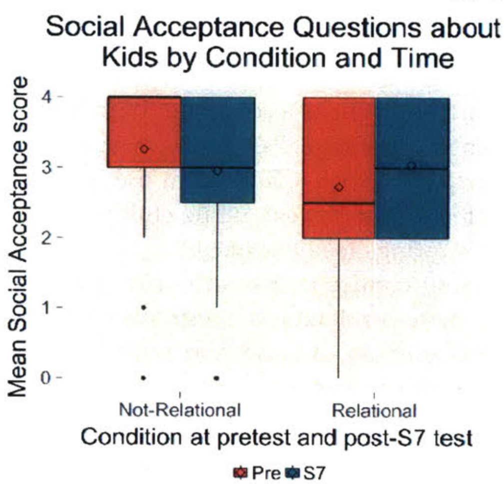
9.4.2 *Learning*

9.4.2.1 *Target Vocabulary Word Identification*

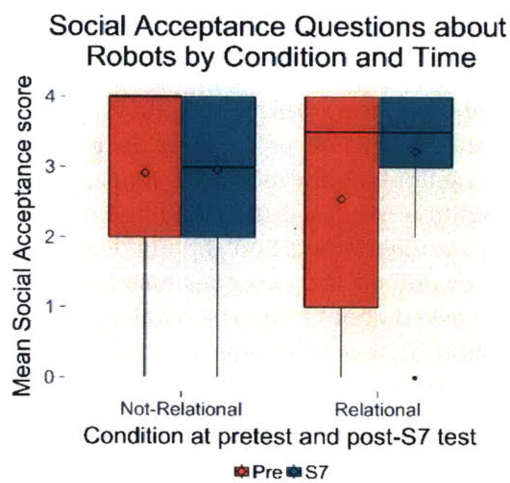
Children's mean score on the vocabulary pretest was 21.1 of 49 (44.3% correct; $SD = 6.30$; $median = 21.5$, mode = 22, min = 7, $max = 36$). Looking at the vocabulary words presented in the Retell stories only (since children did not all hear the same Create stories), children's mean score on the pretest was 36.7% ($SD = 11.7\%$). On average, children scored a mean of 52.4% ($SD = 17.5\%$) on the immediate vocabulary tests at the end of each Retell session. For the delayed posttest scores, I adjusted children's scores to reflect which stories and thus which vocabulary words children actually heard, removing the words that they did not hear from their test scores (e.g., if they had missed one session). Children scored a mean of 48.7% ($SD = 20.0\%$) on the Retell



(a) Overall

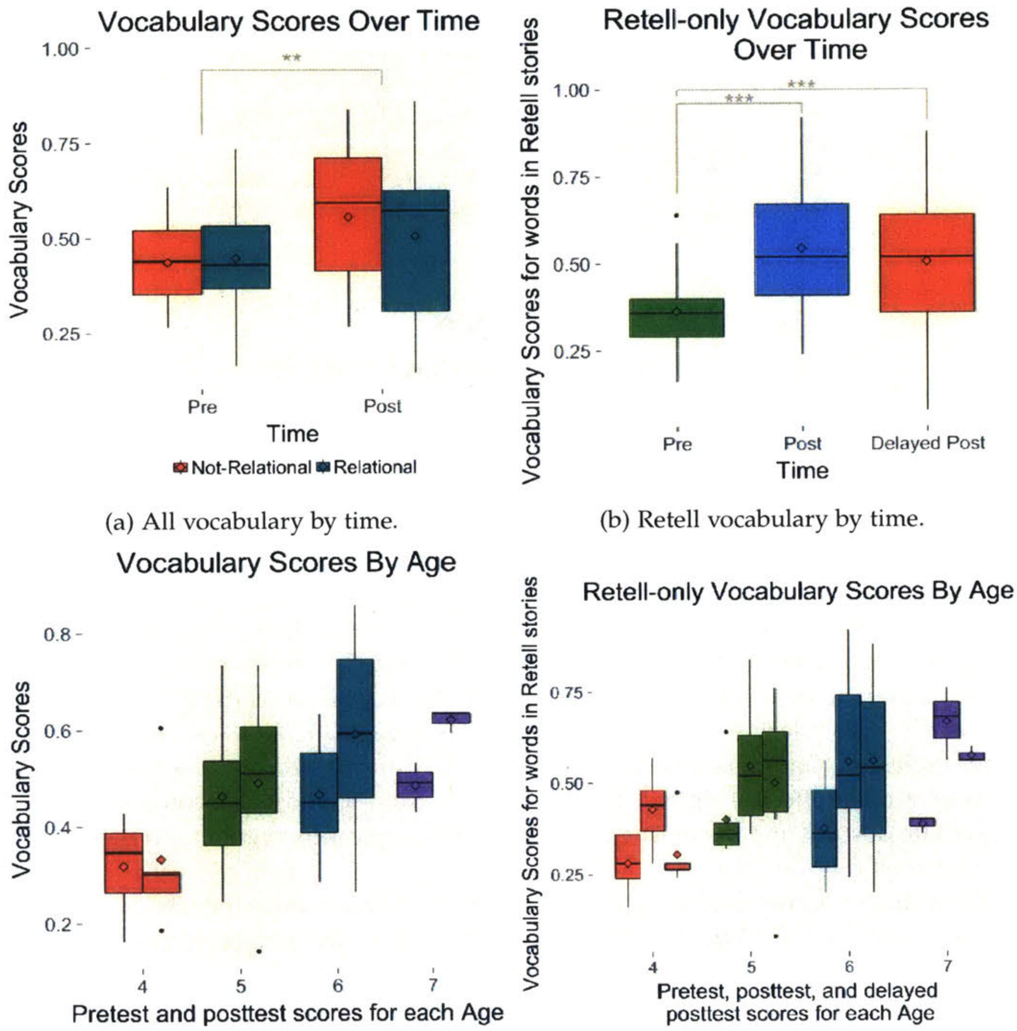


(b) Children



(c) Robot

Figure 46: Children in the *Relational* condition increased their Social Acceptance ratings from the pretest to the S7 test, while children in the *Not Relational* condition slightly decreased their ratings. The questions were, "Would you like to be good friends with a kid/robot who can't hear well?" and "Would you like to be good friends with a kid/robot with special needs?" Children responded with a *yes* (coded as 2), *maybe* (1) or *no* (0) to each question.



(a) All vocabulary by time.

(b) Retell vocabulary by time.

(c) All vocabulary by age.

(d) Retell vocabulary by time.

Figure 47: Children's vocabulary scores were higher at the posttests than at the pretest, and this did not differ by condition. Older children generally scored higher than younger children.

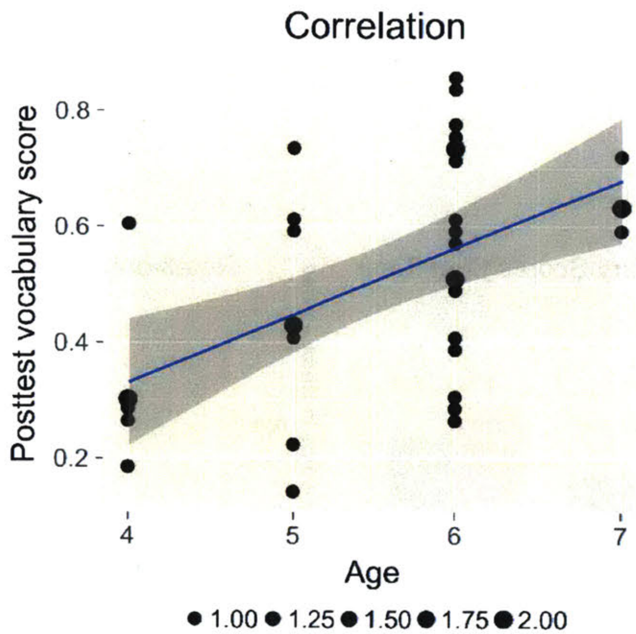


Figure 48: Children's vocabulary scores correlated with age.

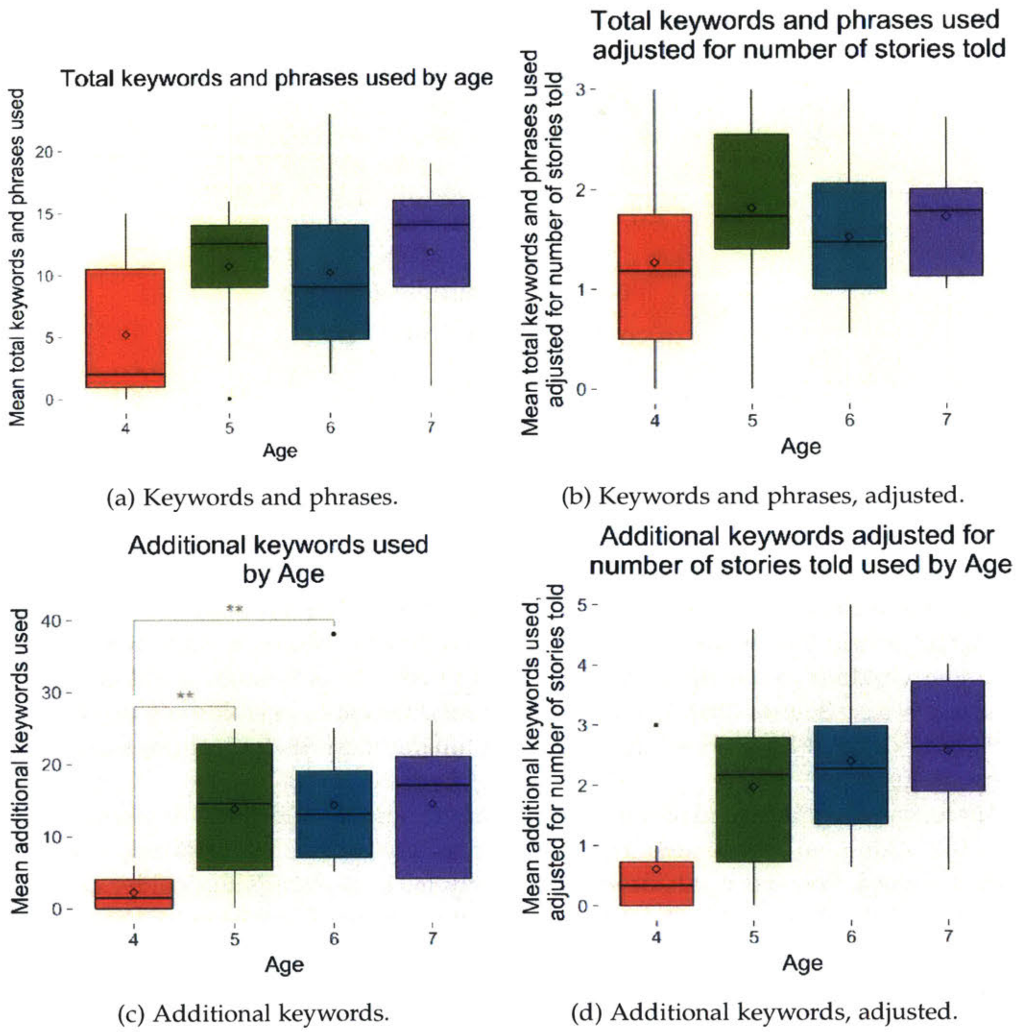
words at the delayed posttest and a mean of 50.7% (SD = 20.0%) on the full vocabulary posttest.

An analysis of variance on children's full vocabulary scores with Condition (between: *Relational* vs. *Not Relational*), Time (Pretest vs. Posttest) and Age as a covariate revealed a main effect of time, $F(1,27) = 12.0$, $p = 0.002$. Children's scores were higher at the posttest than at the pretest (Figure 47a). An analysis of variance on children's Retell-only vocabulary scores with Condition (between: *Relational* vs. *Not Relational*), Time (Pretest vs. Immediate Posttest vs. Delayed Posttest) and Age as a covariate also revealed a main effect of time, $F(2,56) = 20.0$, $p < 0.001$. Children's scores at both the immediate posttest and the delayed posttest were significantly higher than their scores at the pretest (Figure 47b).

For both the above analyses, there were trends toward main effects of age (Figures 47c and 47d). Children's full pretest and posttest vocabulary scores correlated with age, pre: $r_{s48} = 0.469$, $p < 0.001$; post: $r_{s37} = 0.524$, $p < 0.001$ (Figure 48); as did their retell-only scores, pre: $r_{s48} = 0.354$, $p = 0.014$, post: $r_{s49} = 0.335$, $p = 0.019$, delayed post: rs37 = 0.477, p = 0.006.

9.4.2.2 *Stories*

Forty-eight children told at least one story each. Broken down by session, 37 children told a story in Session 1, 26 in Session 2, 30 in Session 3, 33 in Session 4, 33 in Session 5, 22 in Session 6, 22 in Session 7, 28 in Session 8, and 39 told stories at the posttest. After coding for whether children told no story, a couple words of a story, or a full story, I performed an analysis of variance with Condition (between: *Relational* vs. *Not Relational*) and Age as a covariate, revealing no significant differences in how many stories children told.



(a) Keywords and phrases.

(b) Keywords and phrases, adjusted.

(c) Additional keywords.

(d) Additional keywords, adjusted.

Figure 49: Older children used more of the robot's keywords and phrases than younger children.

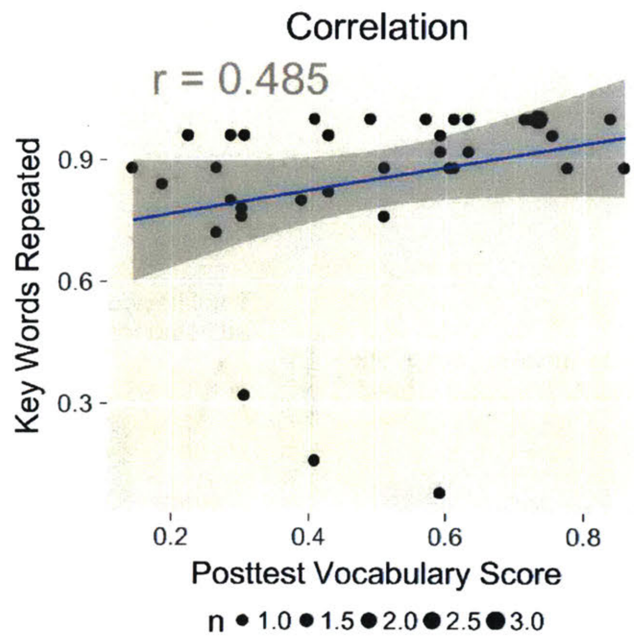


Figure 50: Children who correctly identified the target words at the posttest were more likely to use them in their stories.

9.4.2.3 *Target Vocabulary Word Use*

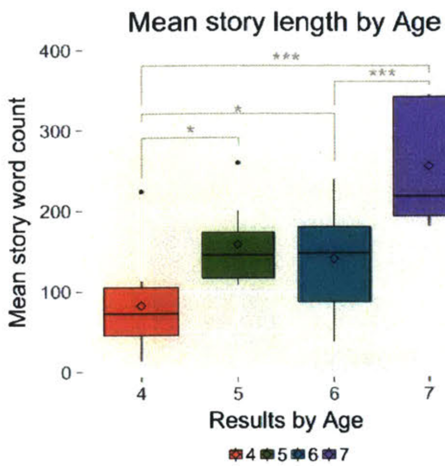
I performed analyses of variance with Condition (between: *Relational* vs. *Not Relational*) and Age as a covariate. There were no significant differences in children's total use of the target vocabulary words or the robot's catchphrases. However, there was a trend for older children to use more keywords and phrases than younger children $F(3,44) = 2.16$, $p = 0.11$ (Figure 49a). There were also no differences in children's use of total keywords divided by stories told (i.e., the estimate of words and phrases used per story) (Figure 49b).

There was a significant main effect of age on the number additional keywords used (i.e., the additional words identified post-hoc as keywords in the stories), $F(3,44) = 5.56$, $p = 0.003$. Four-year-olds used significantly fewer keywords than older children (Figure 49c). Looking at the use of these additional keywords adjusted for the number of stories actually told (i.e., the estimate of words used per story), a similar pattern was revealed, with older children using more keywords than younger children, $F(3,44) = 5.03$, $p = 0.004$ (Figure 49d).

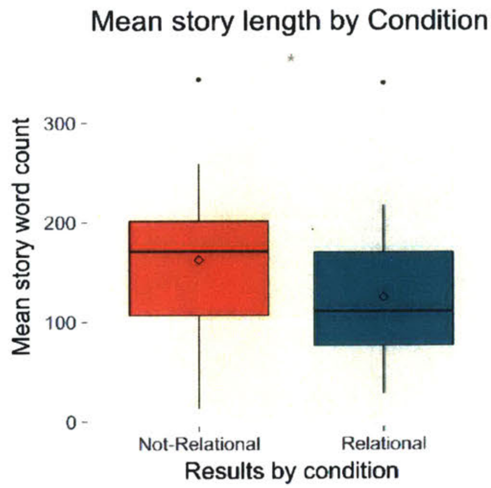
I also observed that children who correctly identified the target words were more likely to use them in their stories, with correlations between target word usage and the full vocabulary posttest, $r_{s37} = 0.485$, $p = 0.002$; as well as the Retell-only immediate posttest, $r_{s47} = 0.431$, $p = 0.002$; and Retell-only delayed posttest, $r_{s37} = 0.447$, $p = 0.006$ (Figure 50).

9.4.2.4 *Story Length*

LENGTH Overall, children's stories were a mean length of 143.8 words ($SD = 73.5$). Divided by story type, children's Retell stories were a mean length of 191.6 words ($SD = 87.0$). Their Create stories were a mean of 63.3 words ($SD = 52.7$). Children used a



(a) Age.



(b) Condition.

Figure 51: Older children told longer stories than younger children. Children in the *Not Relational* Condition told longer stories than children in the *Relational* condition.

mean of 49.4 (SD = 22.1) unique words in their stories (Retell: M = 69.2, SD = 26.8; Create: M = 27.0, SD = 18.2). The robot used a mean of 190.9 unique words (SD = 46.6; Retell: M = 266.3, SD = 13.4; Create: M = 86.7, SD = 17.4). The variation in the robot's stories is due to the story leveling and the fact that the story corpuses differed in mean story length.

An analysis of variance with Condition (between: *Relational* vs. *Not Relational*) and Age as a covariate revealed main effects of condition, $F(1,43) = 4.72$, $p = 0.035$, and of age, $F(3,43) = 10.4$, $p < 0.001$; on the mean length of children's stories. Post-hoc tests revealed that 4-year-olds told far shorter stories (M = 82.7 words, SD = 59.6) than all older children. The 7-year-olds (M = 255.1, SD = 80.8) also told significantly longer stories than the 6-year-olds (M = 140.5, SD = 57.1), and trended toward longer stories than the 5-year-olds as well (M = 158.6, SD = 48.8) (Figure 51a). Children in the *Not Relational* Condition told longer stories than children in the *Relational* condition (Figure 51b).

A similar pattern held when examining only children's Retell stories, with a main effect of age, $F(3,41) = 8.66$, p = <. Children's Create stories were more mixed; there was a trend toward a main effect of age and no difference by condition.

UNIQUE WORDS Regarding the number of unique words used, I observed a main effect of age, $F(3,44) = 7.90$, $p < 0.001$ Seven-year-olds used significantly more unique words (M = 81.5, SD = 22.7) than all younger children (4-year-olds: M = 32.6, SD = 19.2; 5-year-olds: M = 49.9, SD = 19.5; 6-year-olds: M = 49.6, SD = 17.1) (Figure 52). There was also a trend for children in the *Not Relational* condition to use more unique words than children in the *Relational* condition, $F(1,44) = 3.58$, $p = 0.065$.

When examining unique word use in only children's Retell stories, I observed a main effects of age, $F(3,41) = 8.09$, $p < 0.001$. Again, 7-year-olds used significantly more unique words than younger children. There were no significant differences when examining only children's Create stories.

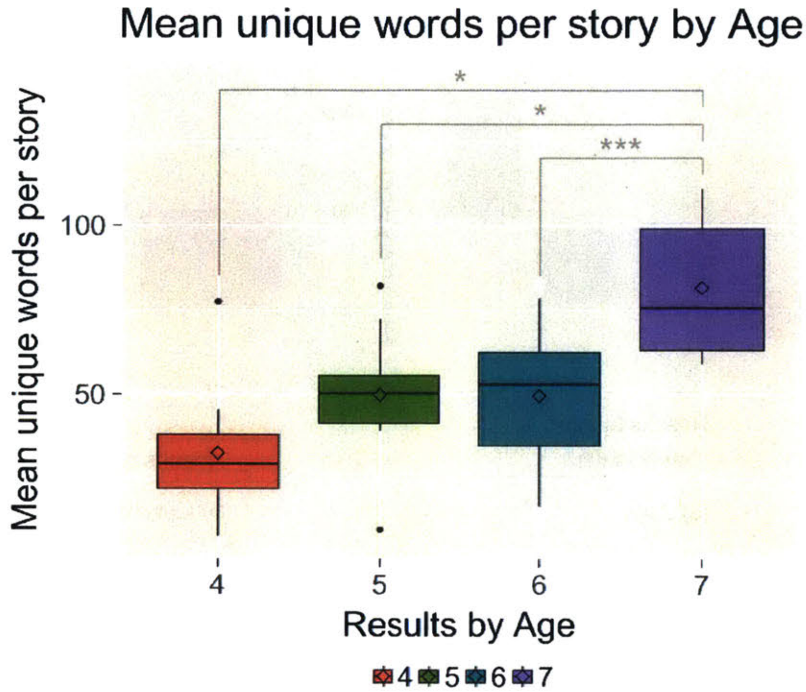


Figure 52: Seven-year-olds used more unique words than all younger children.

9.4.3 *Behavior and Engagement*

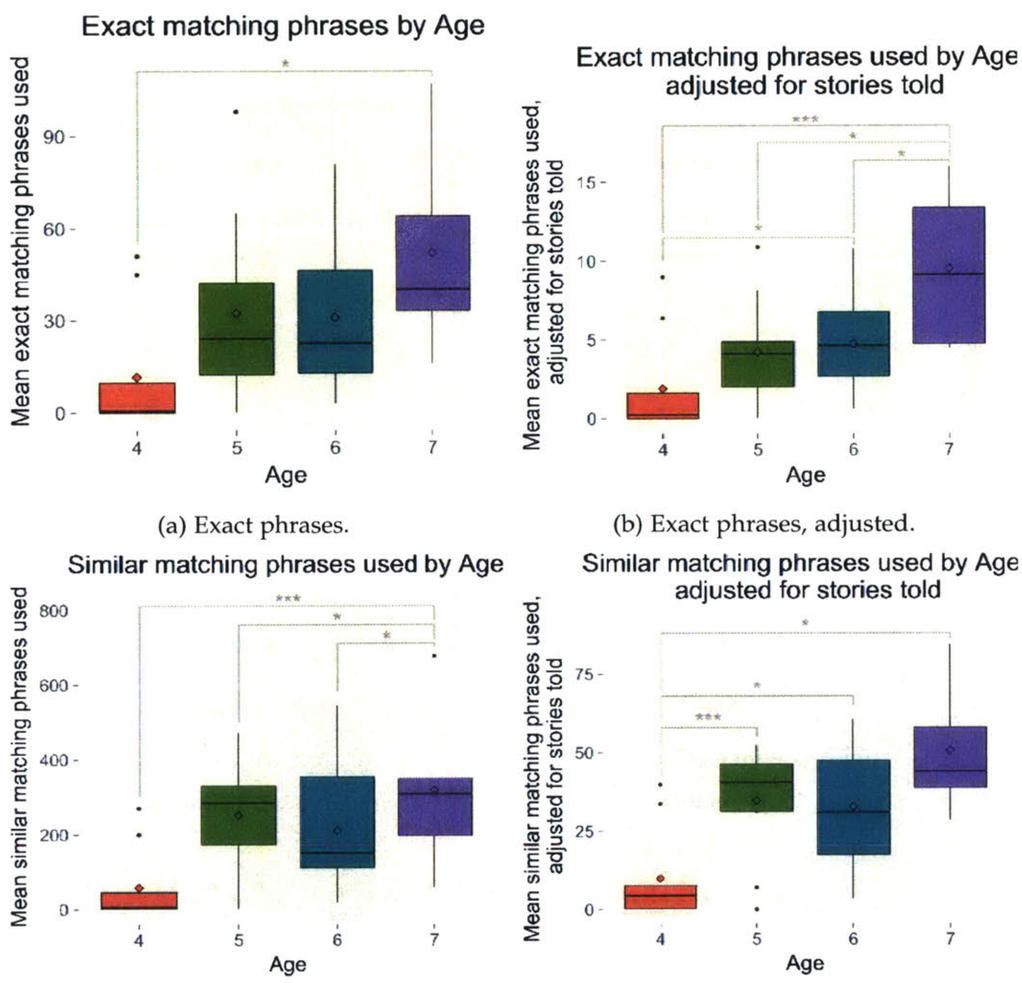
9.4.3.1 *Liking the Robot's Stories*

In each session, the majority of children reported liking the robot's stories (S1: 43 liked it, 2 did not; S2: 46 liked it, 1 liked it a little, 2 did not; S3: 45 liked it, 1 was not sure, 2 did not; S4: 44 liked, 2 did not; S5: 42 liked, 1 was not sure, 3 did not; S6: 42 liked it, 1 liked it a little, 3 did not; S7: 40 liked it, 2 liked it a little, 1 did not; S8: 41 liked it, 1 did not). There were no differences between groups in children's reported liking of the stories.

9.4.3.2 *Mirroring the Robot's Stories*

PHRASE MATCHING I performed analyses of variance with Condition (between: *Relational* vs. *Not Relational*), and Age as a covariate. I observed a significant main effect of age, $F(3,44) = 3.02$, $p = 0.040$, with 7-year-olds using more exactly matching phrases than 4-year-olds (Figure 53a). When examining the use of exactly matching phrases, adjusted for the number of stories each child actually told, I saw the same pattern, $F(3,44) = 6.22$, $p = 0.001$ (Figure 53b). Seven-year-olds ($M = 9.54$, $SD = 5.14$) matched more than 6-year-olds ($M = 4.72$, $SD = 2.7$), 5-year-olds ($M = 4.23$, $SD = 3.36$), and 4-year-olds ($M = 1.9$, $SD = 3.17$).

I observed a main effect of age, $F(3,44) = 4.34$, $p = 0.009$, on children's use of similar matching phrases. Seven-year-olds matched more phrases than all younger children (Figure 53c). When looking at children's total use of similar matching phrases adjusted for the number of stories each child told, there was a main effect of age, $F(3,44) = 6.82$, $p < 0.001$. Seven-year-olds ($M = 50.6$, $SD = 21.8$), six-year-olds ($M = 32.7$, $SD = 18.0$),



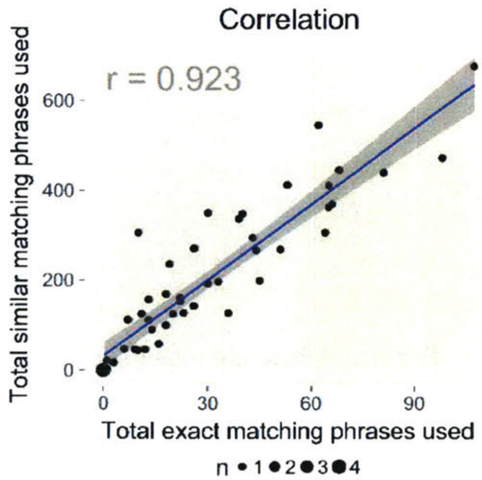
(a) Exact phrases.

(b) Exact phrases, adjusted.

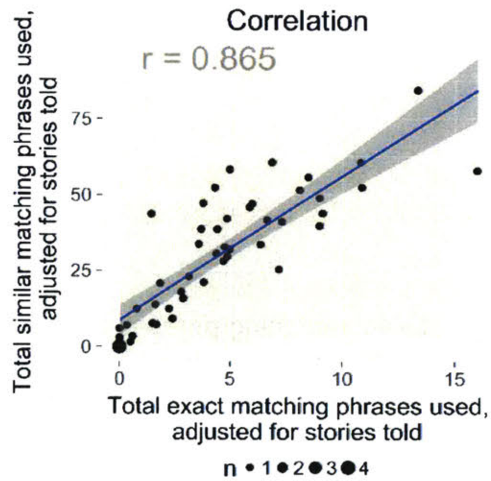
(c) Similar phrases.

(d) Similar phrases, adjusted.

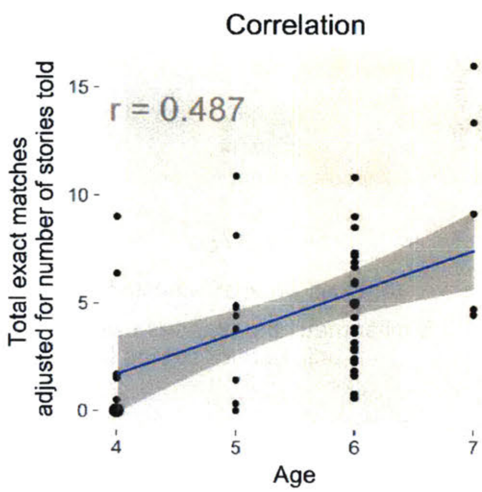
Figure 53: Older children emulated the robot's phrases more than younger children.



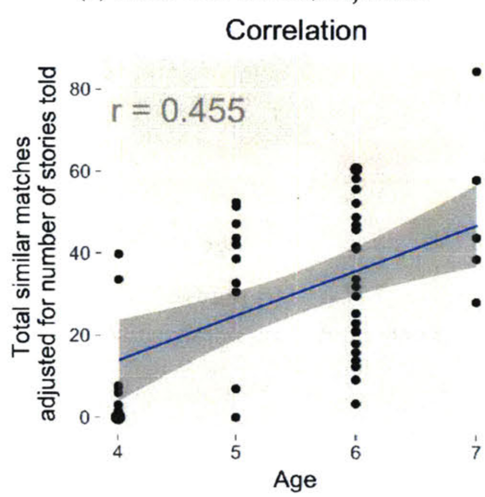
(a) Exact with similar.



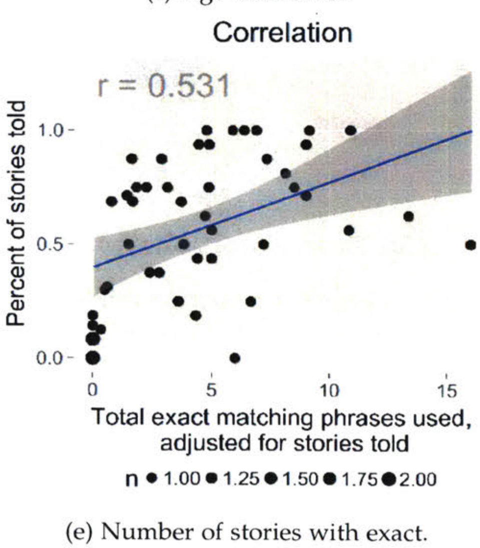
(b) Exact with similar, adjusted.



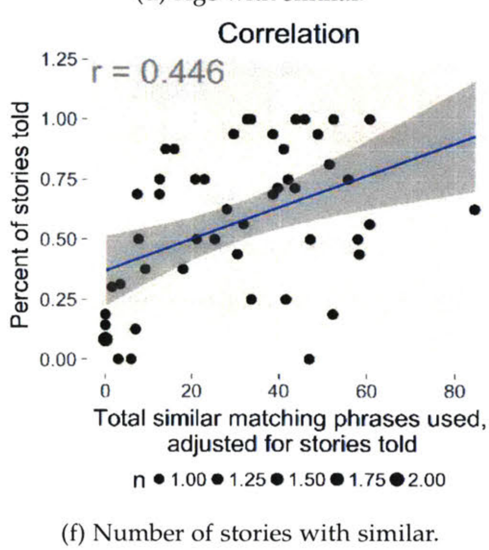
(c) Age with exact.



(d) Age with similar.



(e) Number of stories with exact.



(f) Number of stories with similar.

Figure 54: Children who used more exactly matching phrases used more similar phrases. Older children matched more, as did chi;dren who told more stories.

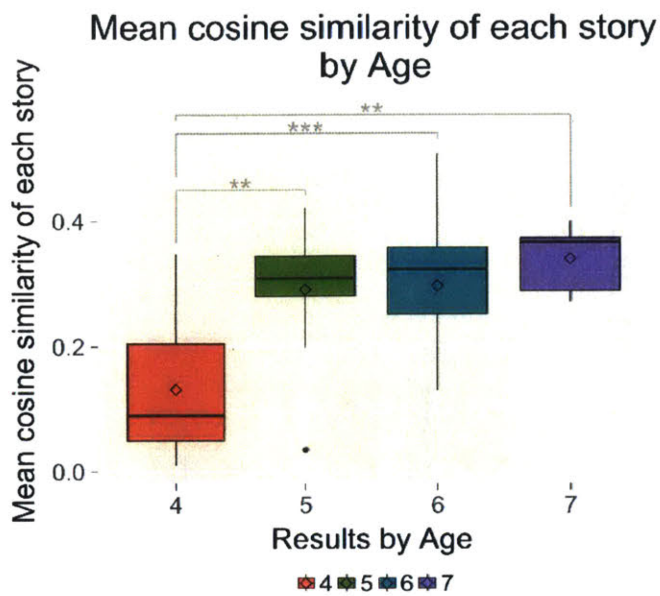


Figure 55: Older children's stories were more similar to the robot's stories.

and five-year-olds (M = 34.6, SD = 17.9) all used more similar matching phrases per story than four-year-olds (M = 9.89, SD = 14.5) (Figure 53d).

Children who used more exactly matching phrases also used more similar matching phrases, as indicated by a Spearman's rank-order correlation, $r_{s47} = 0.923$, $p < 0.001$ (Figure 54a). When adjusted for number of stories told, the correlation was only slightly less strong, $r_{s47} = 0.865$, $p < 0.001$ (Figure 54b). Older children were more likely to use more matching phrases in total when adjusted for the number of stories told, exact $r_{s47} = 0.487$, $p < 0.001$ (Figure 54c); similar $r_{s47} = 0.455$, $p = 0.001$ (Figure 54d). Children who told more stories also matched more phrases per story, exact r_{s47} = 0.531, $p < 0.001$ (Figure 54e); similar rs47 = 0.446, $p = 0.001$ (Figure 54f).

COSINE SIMILARITY I also compared the cosine similarity of children's stories to the robot's stories. When looking at the mean cosine similarity of all of each child's stories, I observed a main effect of age, $F(3,44) = 7.40$, $p < 0.001$. The stories told by 4-year-olds (M = 0.13, SD = 0.12) were less similar to the robot's stories than stories told by 5-year-olds (M = 0.29, SD = 0.11), 6-year-olds (M = 0.30, SD = 0.10), and 7-year-olds (M = 0.34, SD = 0.06) (Figure 55). The same pattern held when looking only at children's Retell stories, $F(3,41) = 10.2$, $p < 0.001$; but not when looking at only children's Create stories, suggesting that children tended to mimic the robot most when explicitly asked to retell the robot's story.

9.4.3.3 *Robot Story*

A mixed analysis of variance with Condition (between: *Relational* vs. *Not Relational*), Time (within: S3 story vs. S7 story), and Age as a covariate revealed no significant differences in children's likelihood to tell a story during the Robot Story task.

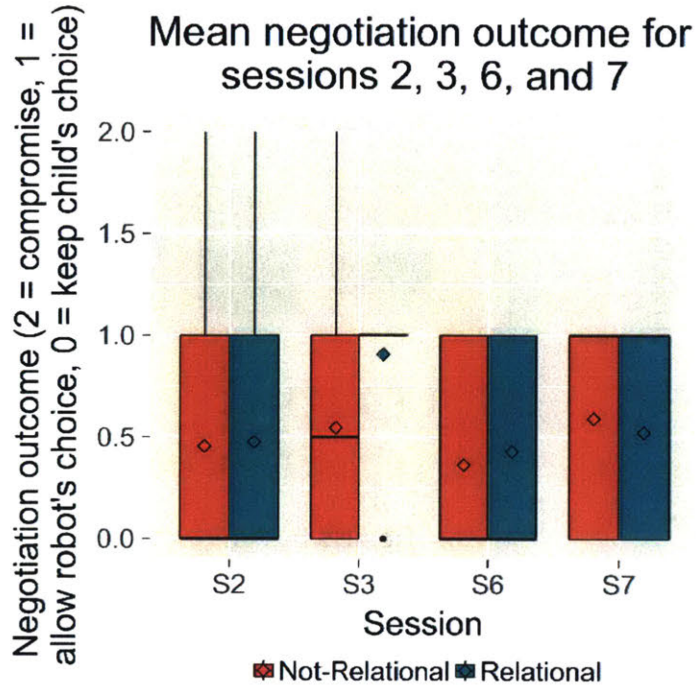


Figure 56: Children's responses during the Negotiation Task. Negotiation outcomes were coded as 2 = compromise, 1 = agree to robot's choice, 0 = stick to child's choice.

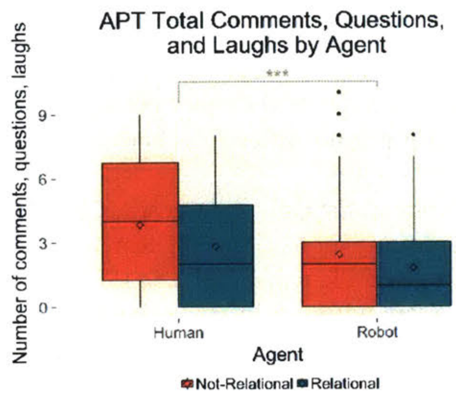
9.4.3.4 *Negotiation Task*

Very few children suggested a compromise with the robot; most either refused the robot's choice outright or stuck to their choice (S2: refusal 29 children, allow 15, compromise 3; S3: refusal 14, allow 33, compromise 1; S6: refusal 28, allow 18; S7: refusal 21, allow 25). A mixed analysis of variance with Condition (between: *Relational* vs. *Not Relational*), Time (within: negotiation at S2, S3, S6, and S7), and Age as a covariate, revealed a main effect of time, $F(3,123) = 3.70$, $p = 0.014$. Children were more likely to allow the robot's choice or suggest a compromise in S3 than in S7 (Figure 56).

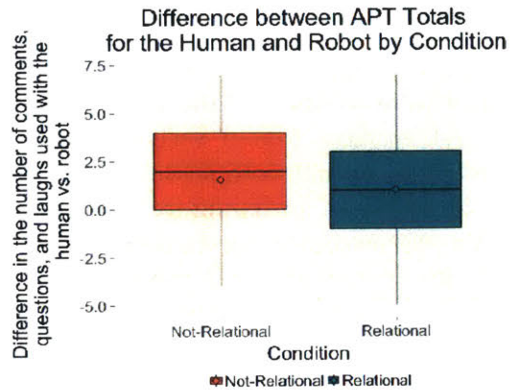
9.4.3.5 *Anomalous Picture Task*

COMMENTS, QUESTIONS, LAUGHS I performed a mixed analysis of variance with Condition (between: *Relational* vs. *Not Relational*), Time (within: pretest vs. posttest), and Agent (within: human vs. robot). When viewing the anomalous pictures, children displayed more laughter, made more comments, and asked more questions with the human experimenter ($M = 3.32$ total comments, questions, and laughs, $SD = 2.76$) than they did with the robot ($M = 2.14$, SD = 2.52), $F(1,84) = 17.3$, $p < 0.001$ (see Figure 57a).

Table 11 lists the mean number of comments, questions, and laughs by agent and condition. I observed a significant interaction of condition with time, $F(1, 44) = 4.98$, $p = 0.031$. Post-hoc tests with the Tukey HSD adjustment revealed that during the posttest, children in the *Not Relational* condition displayed more of all behaviors ($M = 3.49$, SD = 3.09) than children in the *Relational* condition ($M = 1.96$, $SD = 2.22$), but there were no differences during the pretest.



(a) Children displayed more laughter, made more comments, and asked more questions with the human experimenter than with the robot.



(b) Children in the *Not Relational* condition changed their behavior more from the test with the human experimenter to the test with the robot.

Figure 57: Children's responses during the Anomalous Picture Task.

Table 11: The number of comments children made, questions they asked, and times they laughed when viewing the anomalous pictures with the human experimenter and with the robot at each time, by condition. Here, "Con." = "Condition".

Agent	Time	Con.	Comments Mean (SD)	Questions Mean (SD)	Laughs Mean (SD)	Total Mean (SD)
Human	Pre	RR	2.00 (1.73)	0.67 (1.17)	0.56 (1.16)	3.04 (2.72)
		NR	2.50 (2.24)	0.48 (1.05)	0.38 (0.77)	3.54 (2.87)
	Post	RR	1.52 (1.97)	0.22 (0.60)	0.57 (1.12)	2.30 (2.44)
		NR	3.22 (2.24)	0.35 (0.57)	0.57 (1.04)	4.13 (2.13)
Robot	Pre	RR	1.52 (2.22)	0.12 (0.44)	0.44 (0.82)	2.08 (2.45)
		NR	1.17 (2.01)	0.17 (0.44)	0.70 (0.82)	2.04 (2.34)
	Post	RR	0.91 (1.23)	0.09 (0.29)	0.59 (1.26)	1.59 (1.94)
		NR	1.83 (2.46)	0.17 (0.39)	0.70 (1.18)	2.70 (3.24)

To better understand children's behavior during this task, I examined comments, questions, and laughter individually. Children made more comments with the human ($M = 2.35$, $SD = 2.11$) than with the robot ($M = 1.38$, $SD = 2.02$), $F(1,84) = 18.8$, $p < 0.001$. Children in the *Not Relational* condition commented more during the posttest ($M = 2.58$, $SD = 2.43$) than children in the *Relational* condition ($M = 1.22$, $SD = 1.66$), $F(1,44) = 7.68$, $p = 0.008$. When examining questions alone, I observed that children asked more questions of the human ($M = 0.43$, $SD = 0.91$) than the robot ($M = 0.14$, $SD = 0.41$), $F(1,84) = 8.06$, $p = 0.006$. There were no significant differences across agents, conditions, or time with regards to the amount of laughter.

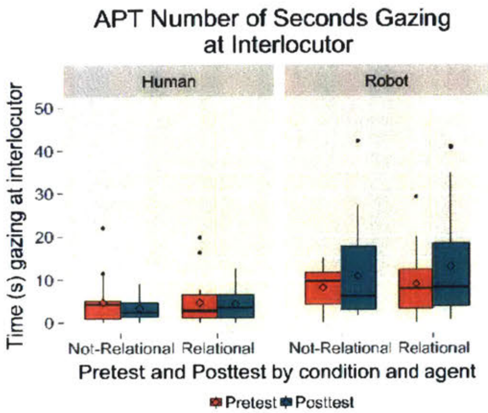
I observed an intriguing trend, though it was not statistically significant: Children in the *Not Relational* condition changed their behavior more from the test with the human experimenter to the test with the robot, performing a mean of 1.84 ($SD = 3.03$) fewer behaviors with the robot than with the person, while children in the *Relational* condition responded more similarly to both agents (mean difference = 0.87, $SD = 2.78$) (Figure 57b).

GAZE Next, I examined children's gaze during the task. In general, children tended to look at the pictures the most ($M = 27.4$ seconds, $SD = 11.6$), their interlocutor second-most ($M = 11.7$ seconds, $SD = 12.3$), and elsewhere the least ($M = 7.63$ seconds, $SD = 8.98$). This suggests children were generally on task. Children's "elsewhere" gazes were directed at the experimenter (e.g., checking in to see whether the experimenter was watching), at the table or camera (e.g., while speaking), or away from the task entirely (e.g., if the child grew bored). I performed analyses of variance with Condition (between: *Relational* vs. *Not Relational*), Time (within: pretest vs. posttest), and Agent (within: human vs. robot) for both the absolute looking times at the pictures, at their interlocutor, and elsewhere during the task and the time as a percent, since the nature of the task meant that not all children looked at the pictures with their interlocutor for exactly the same amount of time.

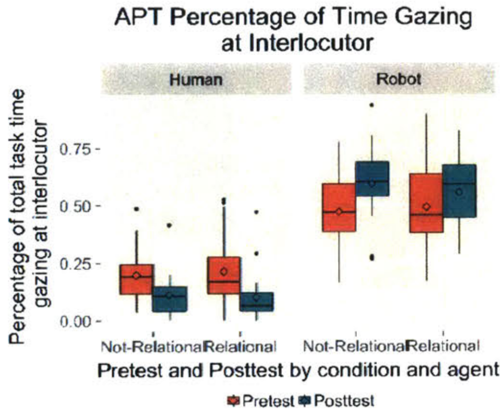
With regards to absolute looking time at their interlocutor, I observed a significant main effect of Agent, $F(1,70) = 282.1$, $p < 0.001$; and a significant interaction of time with agent, $F(1,70) = 5.42$, $p = 0.023$. Post-hoc tests revealed that children looked at their interlocutor more with the robot than with the human at both the pretest and posttest (Figure 58a). There was also a trend toward an interaction of condition with time and agent ($p = 0.10$), which suggested that children in the *Not Relational* condition tended to increase the time spent looking at the robot from pre to posttest.

The same pattern held when looking at the percent of the total time on the task that children spent looking at their interlocutor, with a significant main effect of agent, $F(1,70) = 250.6$, $p < 0.001$; and a significant interaction of time with agent, $F(1,70) = 16.5$, $p < 0.001$. Children looked at the robot more than at the human. They looked at the human less at the posttest than at the pretest, but they looked at the robot more at the posttest than at the pretest (Figure 58b).

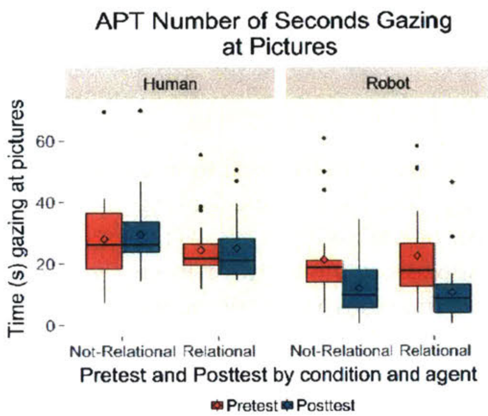
With regards to the absolute time children spent looking at the pictures during the task, I observed significant main effects of time, $F(1,43) = 7.24$, $p = 0.010$; and of agent, $F(1,70) = 32.4$, $p < 0.001$; as well as a significant interaction of time with agent, $F(1,70) = 10.7$, $p = 0.002$. Post-hoc tests showed that children looked at the pictures longer with the human than with the robot. Children spent less time looking at the pictures at the posttest with the robot than at the pretest, but looked for about the same amount of time at both tests with the human (Figure 58c).



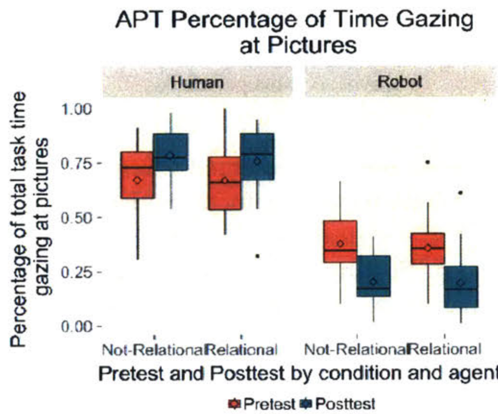
(a) Interlocutor.



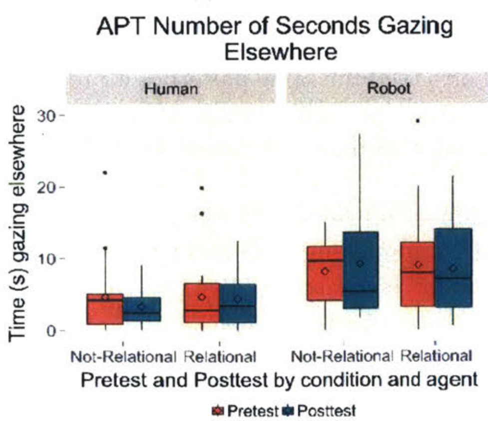
(b) Interlocutor, percent.



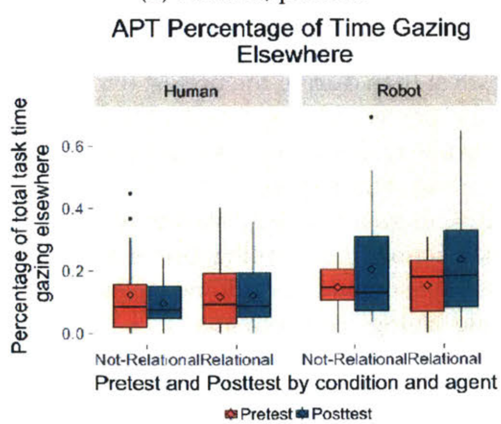
(c) Pictures.



(d) Pictures, percent.

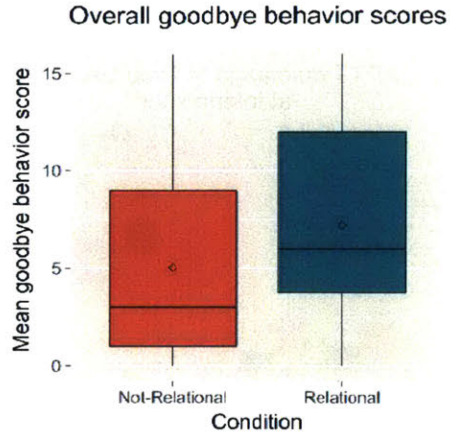


(e) Elsewhere.

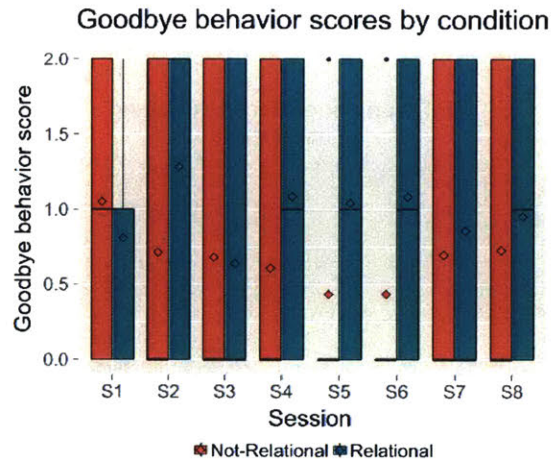


(f) Elsewhere, percent.

Figure 58: Children's gaze and gaze as a percent of the total time spent on task during the Anomalous Picture Task.



(a) All sessions.



(b) Each session.

Figure 59: Children's scored goodbye behaviors summed across all sessions and shown for each session by condition. For each session, behaviors were scored as 2 = clear goodbye action (e.g., waving, saying "bye bye"), 1 = small action (e.g., smile, small wave), 0 = no goodbye.

Similar patterns emerged when looking at the percent of the total time that children spent looking at the pictures. I observed a significant main effect of agent, $F(1,70) = 326.3$, $p < 0.001$; and a significant interaction of time with agent, $F(1,70) = 30.8$, p < 0.001. Post-hoc tests showed that children looked at the pictures longer with the human than with the robot. Children spent less time looking at the pictures with the robot at the posttest than at the pretest, but more time looking at the pictures with the human at the posttest than at the pretest (Figure 58d).

Finally, with regards to the absolute time children spent looking elsewhere during the task, I observed a main effect of agent, $F(1,70) = 30.4$, $p < 0.001$; and a significant interaction of time with agent, $F(1,70) = 5.24$, $p = 0.025$. Children looked elsewhere more often with the robot than with the human, and they did this more during the posttest than during the pretest (Figure 58e). I observed the same pattern with regards to the percent of time children spent looking elsewhere, with a main effect of agent, $F(1,70) = 14.9$, $p < 0.001$; and a significant interaction of time with agent, $F(1,70) = 5.70$, $p = 0.020$ (Figure 58f).

0 = no goodbye (e.g., staring at the robot, doing and saying nothing), 1 = taking a small action (e.g., smiling, doing a small wave, or mimicking the robot's yawn), and 2 = saying something or doing a clear goodbye action (e.g., saying "Bye bye" or waving while saying "See you later")

9.4.3.6 *Goodbye Behaviors*

First, I performed an analysis of variance with Condition (*Relational* vs. *Not Relational*), and Age as a covariate (since children may be more likely to perform appropriate goodbye behaviors as they grow older) on the overall Goodbye score. There were no statistically significant effects; however, there was a trend for children in the *Relational* condition to have higher goodbye scores—i.e., using more goodbye behaviors as time went on ($M = 7.21$, $SD = 4.58$, *median* = 6)—than children in the *Not Relational* condition ($M = 5.04$, $SD = 5.19$, *median* = 3) (Figure 59a).

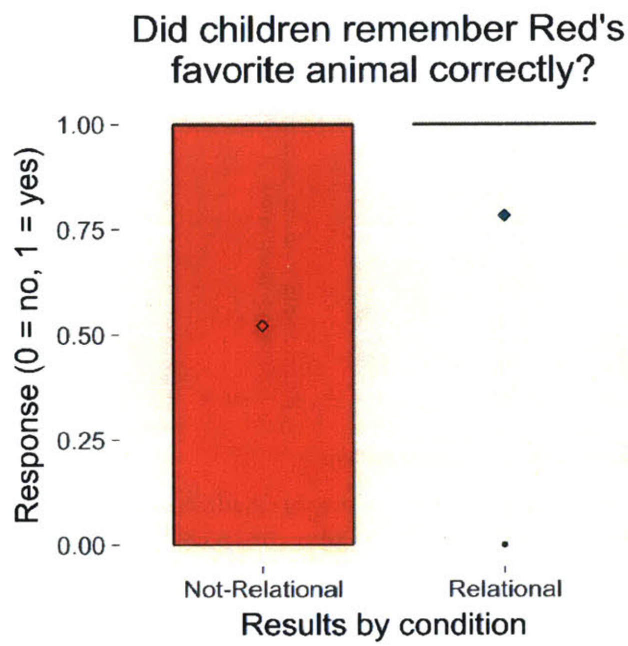


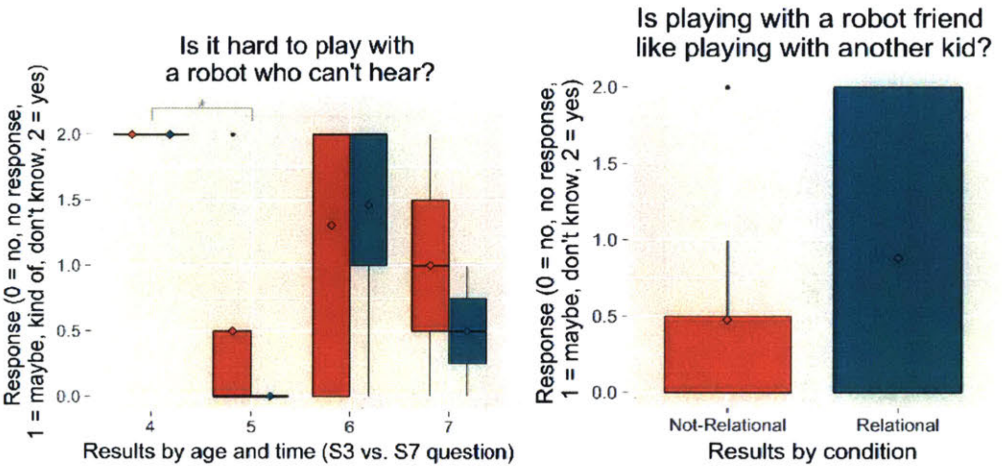
Figure 60: Children in the *Relational* condition were slightly more likely to correctly identify the robot's favorite animal (1 = correct, 0 = incorrect) than children in the *Not Relational* condition.

Next, I examined the scores summed across the first half versus the second half of the sessions using the same method. There were no statistically significant effects, but there was the same trend for children in the *Relational* condition to slightly increase their scores over time.

To further examine the trends over time, I examined the goodbye scores for each session individually (Figure 59b). In the first two sessions, there was a trend for four-year-olds to use fewer goodbye behaviors than older children. S1, S3, S4, S7, and S8 showed no significant differences. In S2, children in the *Relational* condition had significantly higher goodbye scores (M = 1.29, *SD* = 0.96, *median* = 2) than children in the *Not Relational* condition ($M = 0.71$, $SD = 0.96$, $median = 0$), $F(1,37) = 5.10$, $p = 0.021$. In S5, children in the *Relational* condition were more likely to have higher goodbye scores ($M = 1.04$, $SD = 0.95$, $median = 1$) than children in the *Not Relational* condition ($M = 0.43$, $SD = 0.84$, $median = 0$), $F(1,42) = 3.95$, $p = 0.053$. S6 followed the same pattern: children in the *Relational* condition ($M = 1.08$, $SD = 0.88$, $median = 1$) had higher goodbye scores than children in the *Not Relational* condition ($M = 0.43$, $SD = 0.84$, $median = 0$), $F(1,42) = 4.74$, $p = 0.035$.

9.4.3.7 *Robot's Favorite Animal*

When shown four animals in Session 7 and asked which was the robot's favorite, 33 children chose the correct animal; 16 did not. When asked which of the four animals was the kinkajou (the robot's favorite), 33 children chose correctly, while 14 did not. Nineteen children correctly identified both. Analyses of variance with Condition (between: *Relational* vs. *Not Relational*) with Age as a covariate revealed a significant main effect of age on whether children correctly identified the kinkajou, $F(3,42) = 4.40$, $p = 0.009$. Six-year-olds and seven-year-olds correctly identified the kinkajou more often



(a) Age differences in how likely children were to agree that it was hard to play with a robot that was not good at hearing.

(b) Children in the *Relational* condition were more likely to say playing with a robot friend was like playing with another kid.

Figure 61: Children's responses to two questions about playing with robots.

than five-year-olds. In addition, there was a trend for children in the *Relational* condition to be more likely to correctly identify the robot's favorite animal than children in the *Not Relational* condition (Figure 60)

9.4.3.8 *Judgment and Safe Space Questions*

I performed mixed analyses of variance with Condition (between: *Relational* vs. *Not Relational*), Time (within: pre-S3 test vs. post-S7 test), and Age as a covariate. There were no significant effects for either the sum score of the JSS questions or for any individual questions (mean responses shown in Table 12).

For the questions "Do you think Red cares if you make mistakes" and "Do other people care if you make mistakes, like your friends or your teacher", children's answers were mixed at both the S3 test and the S7 test. Children's responses to the questions "Is it okay to practice with Red", "Is it okay to try things out with Red", and "Is it okay to practice with other people, like your friends or your teacher" were overall very positive.

Relatively few children explained their responses to these questions (S3 test: M = 8.8 children or 18.4%, SD = 1.47 or 3.07%; S7 test: M = 11.3 children or 25.2%, SD = 1.75 or 3.89%). As such, I did not perform a formal coding and analysis of children's explanations. Regarding making mistakes, children tended to say that it was okay to make mistakes, that it's okay to make mistakes if you're friends, and that people generally don't care if you make mistakes. Regarding practicing and trying things out, children said it was okay because they were friends with Red, because they can always try new things, it helps them learn, and that people don't mind if they practice.

9.4.3.9 *Empathy/Helping Tasks*

Forty-five of 49 children agreed to discuss the third picture with the robot in the picture conversation task in Session 2; 39 of 45 agreed to do so in Session 6. There were no significant differences between groups in children's acquiescence.

Table 12: Children's responses to the Judgment/Safe Space Questions. The questions were, 1: "Do you think Red cares if you make mistakes?", 2: "Do you feel okay if you make mistakes in front of Red?", 3: "Do other people care if you make mistakes? Like your friends or your teacher?", 4: "Is it okay to practice with Red?", 5: "Is it okay to try things out with Red?", and 6: "Is it okay to practice with other people? Like your friends or your teacher?".

Question	Gender	Condition	S3 Mean (SD)	S7 Mean (SD)
Red cares if you make mistakes	Girls	RR	1.00 (0.95)	1.00 (1.00)
		NR	1.18 (0.87)	1.09 (0.94)
	Boys	RR	1.08 (0.86)	1.18 (0.87)
		NR	1.58 (0.79)	1.33 (0.78)
Okay to mistakes in front of Red	Girls	RR	1.67 (0.65)	1.73 (0.47)
		NR	1.00 (0.89)	1.45 (0.82)
	Boys	RR	1.08 (0.86)	1.27 (0.79)
		NR	1.50 (0.80)	1.50 (0.67)
Others care if you make mistakes	Girls	RR	1.08 (0.79)	1.18 (0.98)
		NR	1.36 (0.92)	1.36 (0.50)
	Boys	RR	1.15 (0.80)	1.36 (0.81)
		NR	1.50 (0.80)	1.42 (0.90)
Okay to practice with Red	Girls	RR	1.92 (0.29)	1.91 (0.3)
		NR	1.82 (0.40)	2.00 (0.00)
	Boys	RR	1.92 (0.28)	1.73 (0.65)
		NR	1.83 (0.58)	1.75 (0.62)
Okay to try things out with Red	Girls	RR	1.83 (0.39)	1.82 (0.60)
		NR	2.00 (0.00)	1.73 (0.65)
	Boys	RR	1.54 (0.78)	1.82 (0.40)
		NR	2.00 (0.00)	1.92 (0.29)
Okay to practice with others	Girls	RR	1.67 (0.65)	1.82 (0.40)
		NR	2.00 (0.00)	1.55 (0.69)
	Boys	RR	1.62 (0.65)	1.91 (0.30)
		NR	1.67 (0.78)	1.82 (0.60)

Thirty-seven children agreed to take a photo with the robot in Session 2; twelve declined.

When the robot asked children if it was hard to play with a robot that was not good at hearing in Session 3, 22 children said that yes, it was hard; 12 said it was not hard, and the 14 did not respond to the question. Most children gave one-word answers; notably, one child said "It is not great Red that you don't hear, but I can help you!" In Session 7, 14 children said yes, it was hard; 12 said no; 3 said it was kind of hard; and the rest did not respond. A mixed analysis of variance with Condition (between: *Relational* vs. *Not Relational*), Time (within: S3 question vs. S7 question), and Age as a covariate revealed a main effect of age, $F(3,17) = 4.32$, $p = 0.019$. Post-hoc test showed that 4-year-olds were more likely to agree that it was hard to play with the robot ($M = 2.00$, $SD = 0.00$) than 5-year-olds ($M = 0.25$, $SD = 0.71$) (Figure 61a). Six-year-olds were also slightly more likely to agree that it was hard ($M = 1.38$, $SD = 0.90$); 7-year-olds were in between ($M = 0.75$, $SD = 0.96$).

In Session 4, when the robot asked children if playing with a robot friend was like playing with another kid, the majority of children did not respond. However, 14 said yes, it was; 3 were not sure or said it was sort of similar; and 1 child said no, it was not. There were no statistically significant differences between groups, though there was a trend for children in the *Relational* condition ($M = 0.88$, $SD = 0.97$) to be more likely to say so than children in the *Not Relational* condition ($M = 0.48$, $SD = 0.85$) (Figure 61b).

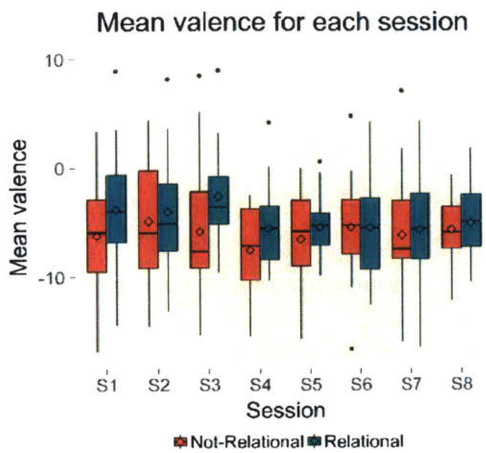
In Session 4, the robot also commented that it was feeling sad because its low battery meant it couldn't play as long. Only 13 children responded verbally; five of these comments were empathetic, e.g., "Aww," "That's sad"; the others acknowledged what the robot said but focused on aspects other than the robot's feelings, e.g., "Why do you run on batteries?" and, "I don't need batteries."

9.4.3.10 *Affect*

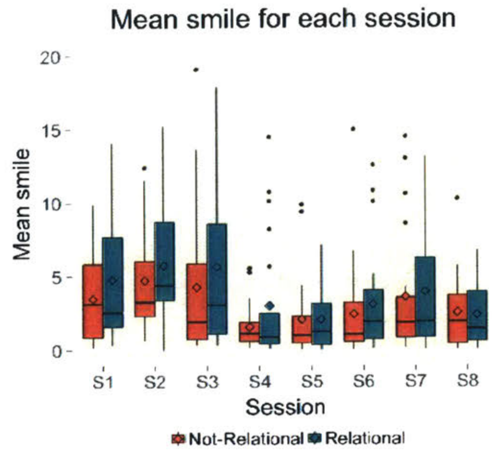
Overall, children were reasonably attentive and engaged, and displayed a range of emotions during the interaction. To evaluate whether children showed greater overall engagement or positive valence with the relational robot, I constructed mixed linear models for each emotion measured with Condition (between: *Relational* vs. *Not Relational*), Time (within: each session), and Age as a covariate. I observed that children's valence significantly varied by time, $F(7,279) = 2.04$, $p = 0.050$; it was higher in the earlier sessions than in the later sessions (Figure 62a).

I observed a significant main effect of time on children's smiling, $F(7,279) = 8.33$, $p < 0.001$. Children smiled more in Sessions 1, 2, and 3 than in Sessions 4, 5, 6, and 8 (Figure 62b). I observed the same pattern with regards to children's joy: there was a significant main effect of time on children's joy, $F(7,279) = 6.28$, $p < 0.001$. Again, children showed more joy in Sessions 1, 2, and 3 than in Sessions 4, 5, 6, and 8 (Figure 62c). Following the same pattern, I observed a significant main effect of time on children's relaxation, $F(7,279) = 4.24$, $p < 0.001$ (Figure 62d). Children were more relaxed in Session 2 than in Sessions 5, 6, and 8; there were trends for children to be more relaxed in Sessions 1 and 3 as well.

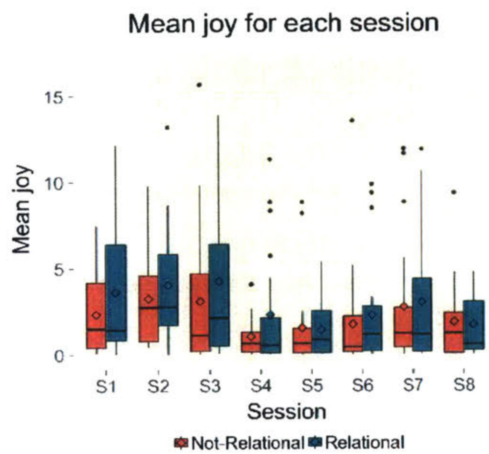
I observed a significant main effect of time on children's engagement, $F(7,279) = 3.39$, $p = 0.002$. Children's engagement was significantly lower in Session 5 than in Sessions 1 and 7; they also showed less engagement in Session 8 than in Session 1 (Figure 62e). Children's attention also significantly varied by time, $F(7,280) = 3.41$,



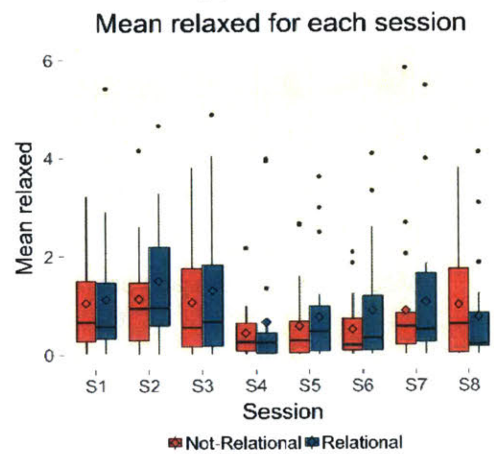
(a) Valence.



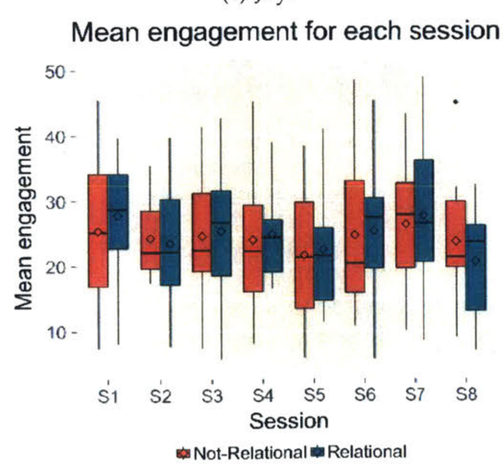
(b) Smiles.



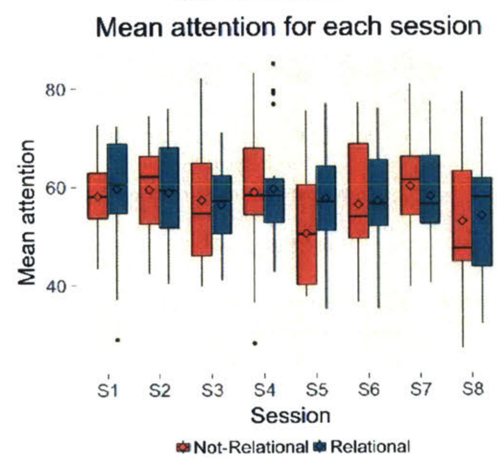
(c) Joy.



(d) Relaxation.

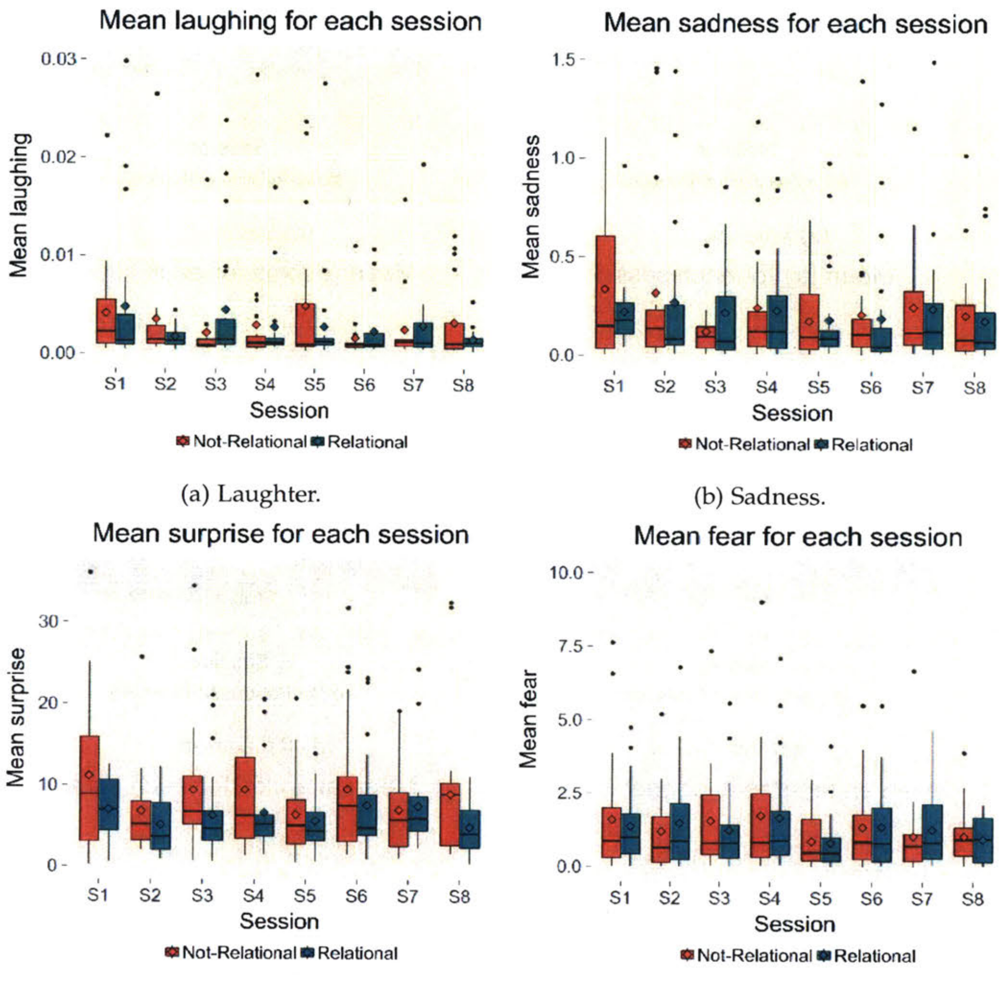


(e) Engagement.



(f) Attention.

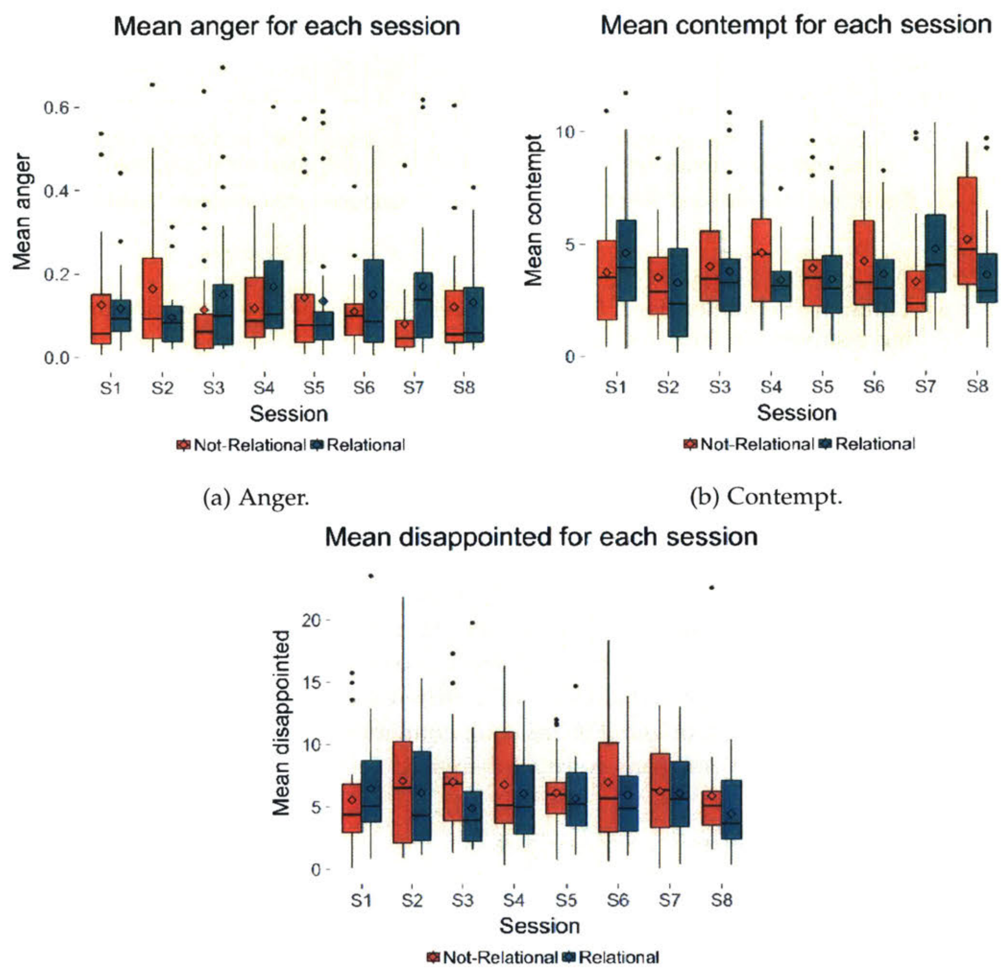
Figure 62: Children's overall affect for each session by condition. Values can range from 0 (no expression present) to 100 (expression fully present), except valence, which can range from -100 to 100.



(a) Laughter. (b) Sadness.

(c) Surprise. (d) Fear.

Figure 63: Children's overall affect for each session by condition. Values can range from 0 (no expression present) to 100 (expression fully present), except valence, which can range from -100 to 100.

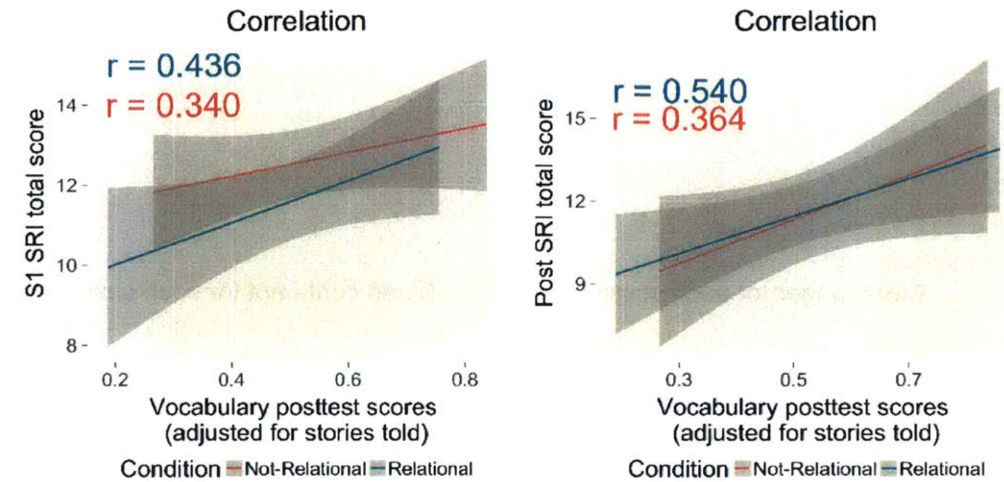


(a) Anger.

(b) Contempt.

(c) Disappointed.

Figure 64: Children's overall affect for each session by condition. Values can range from 0 (no expression present) to 100 (expression fully present), except valence, which can range from -100 to 100.



(a) S1 test.

(b) Posttest.

Figure 65: Children's Social Relational Interview scores correlated with their vocabulary scores more strongly in the *Relational* condition.

p = 0.002. Children were less attentive in Session 8 than in Sessions 1, 2, 4, and 7 (Figure 62f).

I observed a significant main effect of time on children's sadness, $F(7,279) = 3.27$, $p = 0.002$ (Figure 63b). Children showed more sadness in Session 1 than in all other sessions. Children's expressions of surprise varied by time as well, $F(7,278) = 2.83$, $p = 0.007$. They showed more surprise in Session 1 than in Session 5 (Figure 63c). There was a trend toward an interaction of condition with time, which suggested that children in the *Not Relational* condition generally showed more surprise.

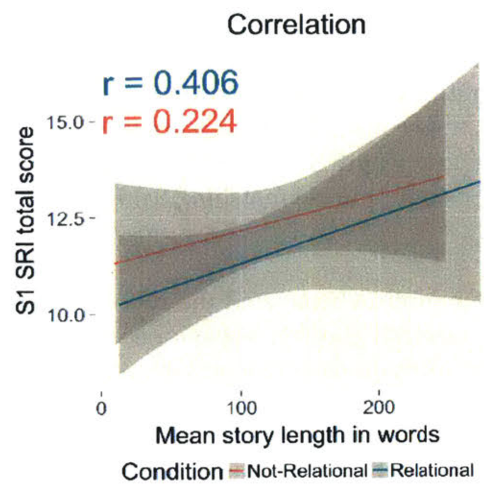
There were no significant differences in children's expressions of laughter (Figure 63a), fear (Figure 63d), anger (Figure 64a), contempt (Figure 64b), or disappointment (Figure 64c). These emotions were all displayed relatively rarely.

9.4.3.11 *Relationship, Mirroring, and Learning*

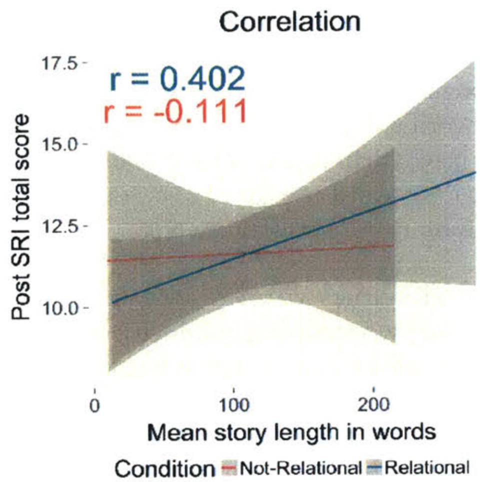
I examined correlations between several key relational measures used and various learning and mirroring behaviors displayed by children.

CORRELATIONS WITH SOCIAL RELATIONAL INTERVIEW First, children's Social Relational Interview (SRI) scores correlated with their vocabulary posttest scores at both the S1 test, $r_{s37} = 0.362$, $p = 0.028$; and more strongly at the posttest, $r_{s37} = 0.480$, $p = 0.003$. These correlations, however, differed significantly when I examined them by condition. Children's SRI scores correlated with their vocabulary posttest scores more strongly in the *Relational* condition (S1 test: $r_s = 0.436$, posttest: $r_s = 0.540$) than in the *Not Relational* condition (S1 test: rs = 0.340, posttest: $r_s = 0.364$) (Figure 65).

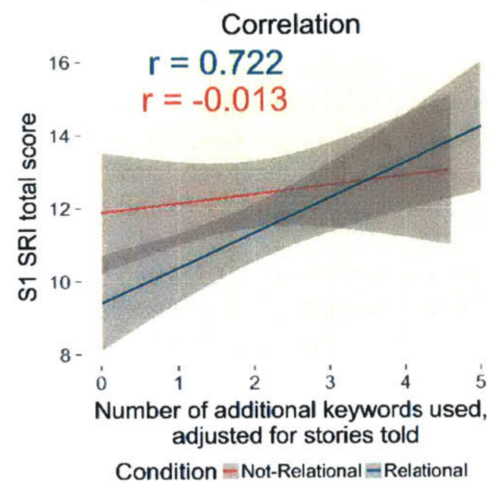
Children who scored more highly on the SRI also emulated the robot more during their storytelling. Their S1 SRI scores correlated with their mean story length, $r_{s47} = 0.347$, $p = 0.017$; their total use of exact and similar matching phrases (adjusted for stories told), $r_{s47} = 0.377$, $p = 0.009$; and their use of the additional keywords, $r_{s47} = 0.366$, $p = 0.01$. Children's posttest SRI scores also correlated with their total use of exact and similar phrases, $r_{s46} = 0.350$, $p = 0.017$; and with their use of additional



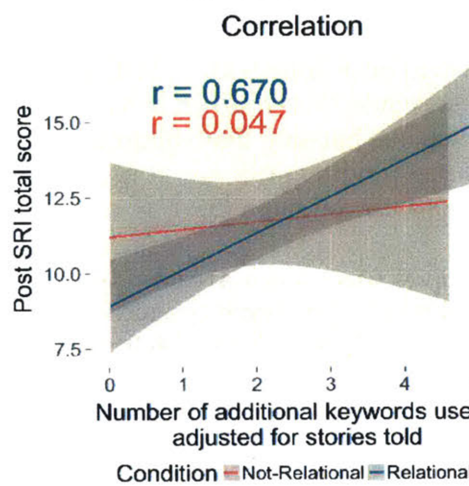
(a) S1 test.



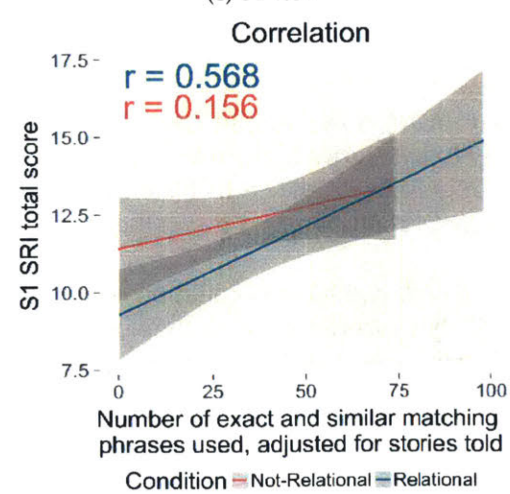
(b) Posttest.



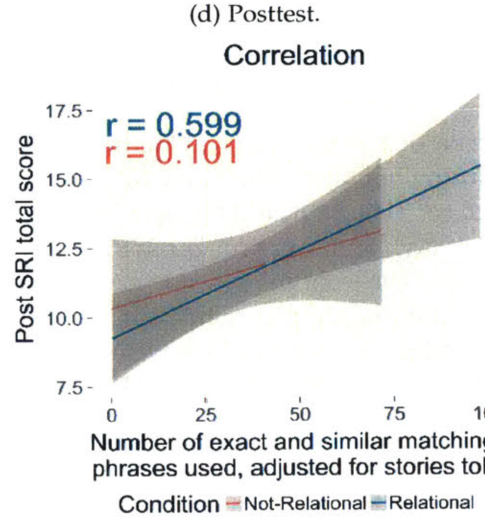
(c) S1 test.



(d) Posttest.



(e) S1 test.



(f) Posttest.

Figure 66: Children's Social Relational Interview scores correlated with their mean story length, use of keywords, and use of exact and similar matching phrases more strongly in the *Relational* condition.

keywords, r_{s46} = 0.356, p = 0.015. Mean story length did not correlate significantly children's posttest SRI scores, r_{s47} = 0.191, p = 0.20. Again, all these correlations were stronger in the *Relational* condition (story length S1: r_s = 0.406, posttest: r_s = 0.402; keyword use S1: r_s = 0.722, posttest: r_s = 0.670; phrase matching S1: r_s = 0.568, posttest: r_s = 0.599) than in the *Not Relational* condition (story length S1: r_s = 0.224, posttest: r_s = -0.111; keyword use S1: r_s = -0.013, posttest: r_s = 0.047; phrase matching S1: r_s = 0.156, posttest: r_s = 0.101) (Figure 66).

CORRELATIONS WITH GOODBYE SCORES Children who treated the robot as a social other when saying goodbye (i.e., had higher overall goodbye behavior scores) told longer stories, r_{s47} = 0.440, p = 0.002; were more likely to use exact and similar matching phrases, r_{s47} = 0.424, p = 0.003 (Figure); and more keywords, r_{s47} = 0.299, p = 0.041. Unlike the earlier correlations, here, children in the *Not Relational* condition frequently showed higher correlations (story length: r_s = 0.521; phrase matching: r_s = 0.633; keyword use: r_s = 0.397) than in the *Relational* condition (story length: r_s = 0.492; phrase matching: r_s = 0.264; keyword use: r_s = 0.293) (Figure 67).

Children who had higher goodbye scores also placed the Tega robot closer to the human in the Picture Sorting Task at the posttest, r_{s47} = -0.410, p = 0.004, but not at the S2 test. When broken down by condition, I observed little difference in the posttest correlations, but saw that children's S2 scores were positively correlated in the *Not Relational* condition (S2: r_s = 0.370; posttest: r_s = -0.393), but negatively correlated in the *Relational* condition (S2: r_s = -0.232; posttest: r_s = -0.410) (Figures 67d and 67e).

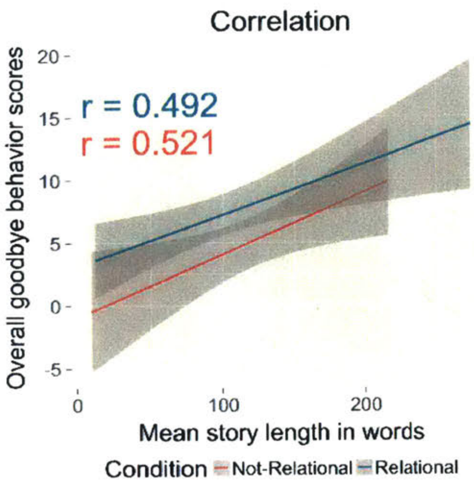
CORRELATIONS WITH ANOMALOUS PICTURE TASK BEHAVIOR Children who displayed more social behaviors (comments, questions, laughter) during the S1 Anomalous Picture Task with the robot used more exact and similar matching phrases, r_{s47} = 0.306, p = 0.036; they also told longer stories, r_{s47} = 0.414, p = 0.004. These correlations were not present at the S8 task with the robot. I observed some differences by condition in the correlations in S1, and most were still not present at S8 by condition: *Relational* (phrase matching S1: r_s = 0.211, S8: r_s = -0.005; story length S1: r_s = 0.420, S8 r_s = 0.095; keyword use S1: rs = 0.368, S8: r_s = 0.381); *Not Relational* (phrase matching S1: r_s = 0.373, S8: r_s = 0.208; story length S1: r_s = 0.206, S8 r_s = 0.034; keyword use S1: r_s = 0.196, S8: r_s = 0.132) (Figure 68).

OTHER CORRELATIONS Children who rated the robot as closer in Session 1 on the Inclusion of Other in Self task were more likely to tell stories to the robot, r_{s48} = 0.369, p = 0.001 (Figure 69a). This correlation was not present for the S8 Inclusion of Other in Self task, and was stronger in S1 for the *Relational* condition (r_s = 0.455) than for the *Not Relational* condition (r_s = 0.306).

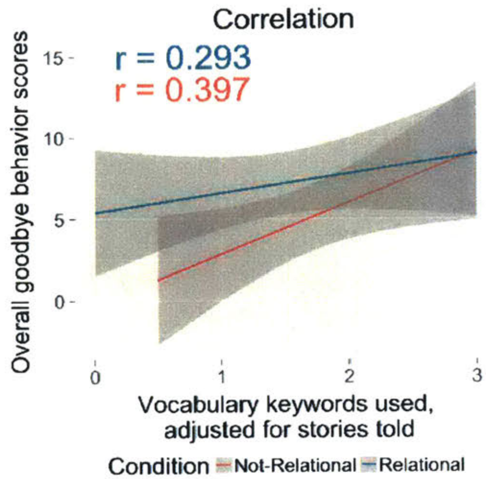
In addition, children who were more likely to seek equitable negotiation outcomes during the Negotiation Task were also more likely to score highly on the vocabulary posttest, r_{s37} = 0.385, p = 0.019 (Figure 69b). Again, this correlation was stronger for the *Relational* condition (r_s = 0.483) than for the *Not Relational* condition (r_s = 0.278).

9.5 DISCUSSION

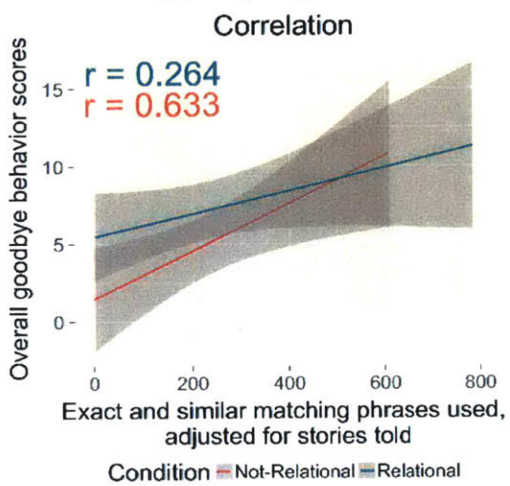
In this study, I asked whether children who played with the relational robot would show greater rapport, a closer relationship, increased learning, greater engagement, more positive affect, more peer mirroring, and treat the robot as more of a social other



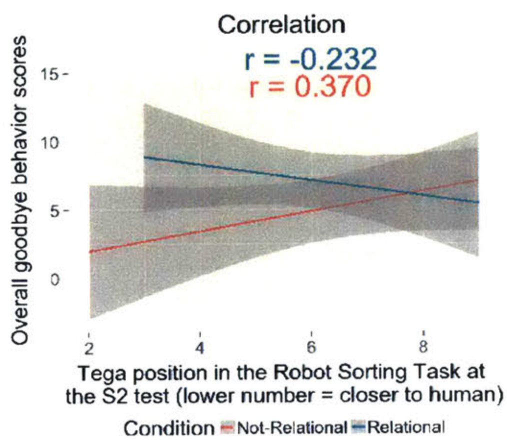
(a) Story length.



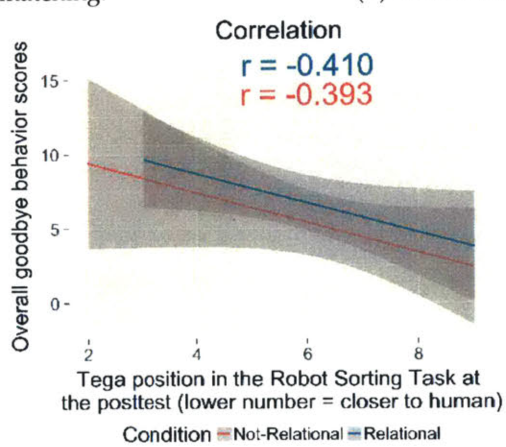
(b) Keywords.



(c) Phrase matching.

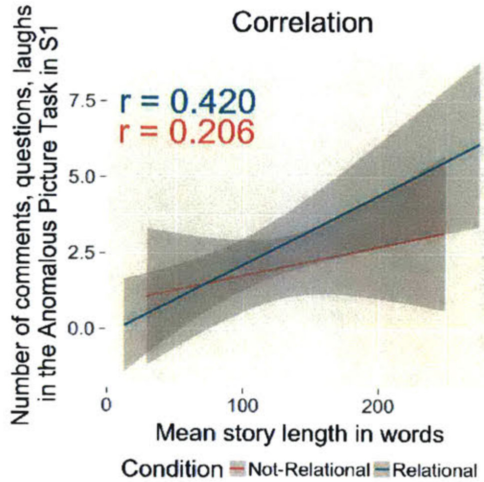


(d) Picture Sorting Task S2.

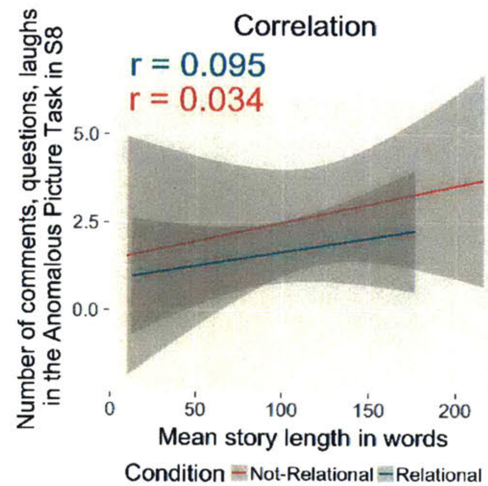


(e) Picture Sorting Task Posttest.

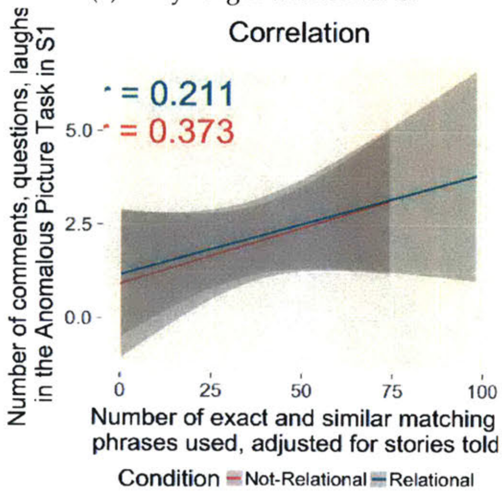
Figure 67: Children who treated the robot more socially when saying goodbye told longer stories, used more keywords, emulated the robot's phrases more during storytelling, and placed the Tega robot closer to the human in the Picture Sorting Task.



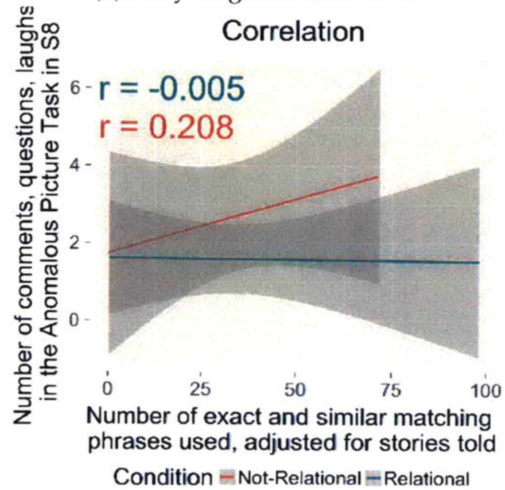
(a) Story length with S1 APT.



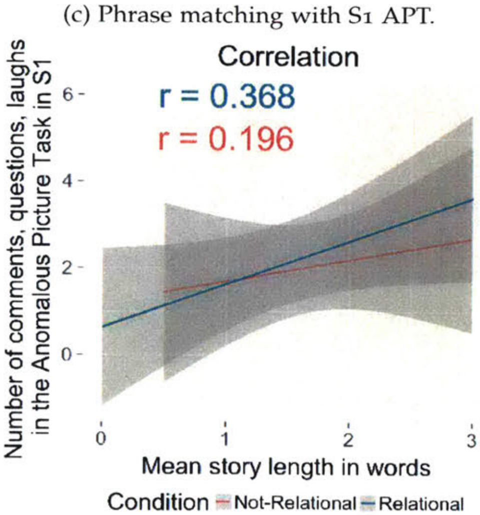
(b) Story length with S8 APT.



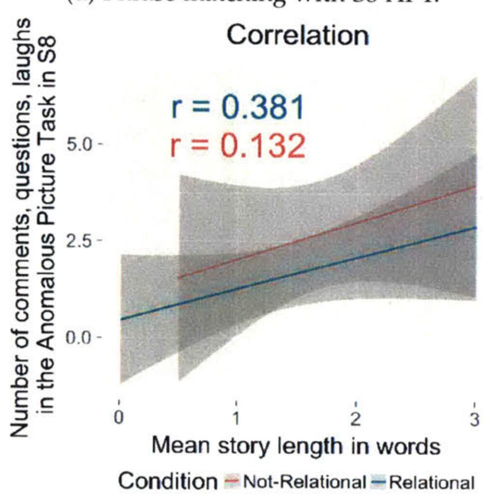
(c) Phrase matching with S1 APT.



(d) Phrase matching with S8 APT.

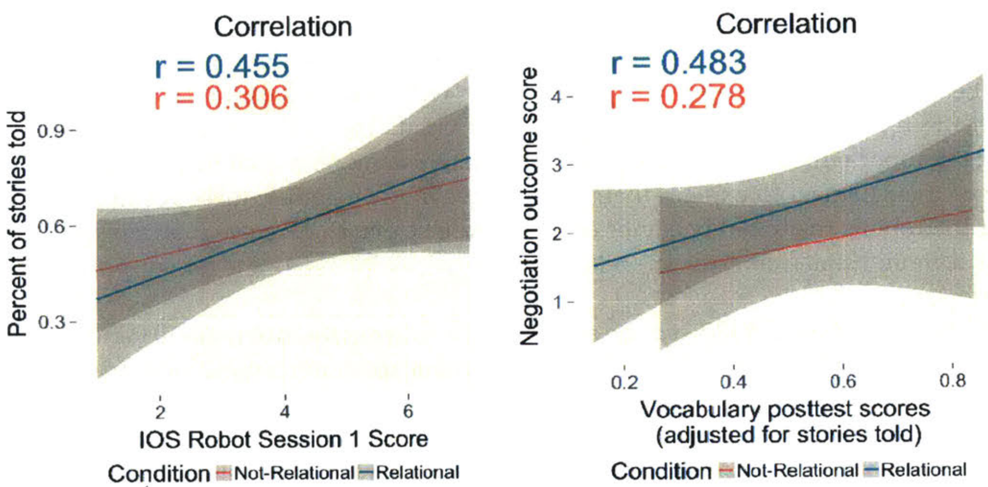


(e) Keyword use with S1 APT.



(f) Keyword use with S8 APT.

Figure 68: Correlations between children's social behavior during the Anomalous Picture Task (APT) and children's story length, phrase matching, and use of keywords.



(a) Children who rated the robot as closer in S1 on the Inclusion of Other in Self task were more likely to tell stories to the robot.

(b) Children who were more likely to seek equitable negotiation outcomes were more likely to score highly on the vocabulary posttest.

Figure 69: Other correlations.

than children who played with the non-relational robot. In addition, I asked whether children who reported feeling closer to the robot (regardless of condition) would also display these behaviors, in particular more learning and peer mirroring.

I found that overall, children in both conditions learned the target vocabulary words from the robot's stories and remembered the robot's stories, as evidenced by their story retellings, use of many of the target words in their stories, and their increased scores from the vocabulary pretest to posttest. Children were reasonably attentive to the robot and engaged in the interactions, showing a variety of emotional expressions and reporting that they generally liked the robot and its stories. Most children responded positively to the robot, conversed with it, and told or retold stories. These results are in line with prior work showing that children do learn new words from robots and remember stories told by a robot (Chapter 8, Kory-Westlund et al., 2015a, 2017a,b; Kory and Breazeal, 2014; Rintjema et al., 2018; Vogt et al., 2017), and that children tend to respond positively to and be reasonably engaged with social robots during storytelling and conversation activities (Gordon et al., 2016; Kory-Westlund et al., 2017b; Kory and Breazeal, 2014; Park et al., 2017c; Serholt and Barendregt, 2016).

When examining the data by condition, I found evidence in support of my hypotheses: The robot's relational behavior and children's perception of the robot as a social-relational other were important factors in children's relationship, behavior, and learning. Below, I discuss the main findings and implications of these findings.

9.5.1 *Relational Robot*

The robot's relational behavior impacted children's social behavior and relationship, their view of the robot as a social and relational agent, and acceptance of the robot.

Children in the *Relational* condition seemed to see the robot as more human-like. They rated the Tega robot as closer to themselves in the Inclusion of Other in Self task. They placed the robot closer to the human, cat, frog, and movie robot in the Picture Sorting Task, while children in the *Not Relational* condition placed Tega closer to the table, computer, and teddy bear. They were more likely to say that playing with the robot was like playing with another child. They also were more confident that the robot remembered them, frequently referencing relational behaviors (such as use of their names) to explain their confidence. Taken together, these results suggest that the robot's relational behavior encouraged children to perceive it and respond to it as a more human-like, social, relational agent—as intended and in support of my hypotheses.

Children's behavior followed the same pattern. During the Anomalous Picture Task, children in the *Relational* condition treated the robot and human more similarly, changing their behavior less than children in the *Not Relational* condition. They were treating the robot as more of a social, human-like other to whom they could comment, ask questions, and with whom they could laugh. These results are similar to my prior work. Previously, I observed that children changed their behavior less from a pretest with a human to the robot interaction when the robot was introduced as a social agent than when it was introduced as a machine (Kory-Westlund et al., 2016a).

Children's goodbye behaviors were also generally more social in the *Relational* condition. This is in line with prior work suggesting that people tend to indicate intent and ask permission before leaving, but they do not say goodbye to computers or other technology when done using them (Reeves and Nass, 1996). Although there were differences by gender, children in the *Relational* condition were still more likely to say goodbye in a way that treated the robot as a social other than children in the *Not Relational* condition. Children who did not feel as much rapport or relationship (namely boys in the *Relational* condition) perhaps still recognized the relational robot as a social other, and treated it accordingly despite not feeling close to it. In multiple sessions where there were clear differences in goodbye behaviors by condition, I expect the differences were related to specific relational actions the robot took in those sessions. For example, Session 2 included the first memory-related events (e.g., recalling the child's name, referring to activities from last time). Interestingly, there were multiple correlations between children's goodbye behavior scores and relational measures. Children who treated the robot as a social other when they left also showed more emulation of the robot in their stories, told longer stories, and thought of the robot as more human-like, placing it closer to the adult human in the Picture Sorting Task. This suggests that children's goodbye behaviors were indeed related to their construal of the robot as a social being, and further, that their construal of the robot affected their behavior.

Children reported becoming more accepting of the robot and of other kids via the Social Acceptance Questionnaire in the *Relational* condition. This pattern suggests that exposure to the relational robot with its purported hearing disability positively affected children's judgment about robots and kids with disabilities. The papers describing the original Social Acceptance Scale that I adapted did argue that exposure to other children with disabilities—including telling stories, guided discussion, and structured play—might increase children's acceptance of them, and that seems to have been the case here (Favazza and Odom, 1996; Favazza, Phillipsen, and Kumar, 2000; Harter and Pike, 1984). However, the fact that children's acceptance scores decreased in the *Not Relational* condition suggests that it not merely exposure that mattered, since in both conditions the robot used the same ASR (with the same issues recognizing children's speech) and described the same backstory in Session 1. The relational robot referred back to its difficulties hearing in later sessions, reminding children of its backstory.

This behavior may have been important in helping children remember to be patient and accepting of the robot. It may also be that the relational robot appeared to be "trying harder" to be a good friend, and this positively affected children's judgment of similar agents, since the relational robot spoke of its relationship with the child, explicitly called the child its friend, and frequently referred to its history with the child. Future work could explore this further.

Children's behavior in the Self-Disclosure Task showed a similar pattern, in which children in the *Relational* condition increased the length of their disclosures over time, while children in the *Not Relational* decreased theirs. Because prior work has shown that children disclose more information and more personal information when they feel closer to someone (Buhrmester and Furman, 1987; Gleason and Hohmann, 2006; Rotenberg, 1995), it may be that children's feelings of closeness to the robot increased in the *Relational* condition more so than in the *Not Relational* condition.

Children's feelings of closeness were also reflected to a degree in children's reports of feeling okay making mistakes in front of the robot. Children in the *Relational* condition were slightly more likely to say they would be comfortable. This points to the robot's potential as a non-judgmental agent that can enable children to practice, make mistakes, and share their thoughts. Prior work has found that robots and virtual agents can lead to increased self-disclosure and feelings of comfort, as well as feeling a lack of judgment from the agent during health interviews and relationship counseling (Bickmore, Gruber, and Picard, 2005; Gratch et al., 2007; Lucas et al., 2014; Utami, Bickmore, and Kruger, 2017).

There were some trends for children in the *Relational* condition to show more positive affect than children in the *Not Relational* condition, especially during the first several sessions. Because the pattern was much more prominent early on, I wondered whether the early differences in joy and smiles were actually related to frustration and not to joy, since smiles are associated both with joy and frustration (especially when dealing with technology) (Hoque and Picard, 2011; Hoque, McDuff, and Picard, 2012). Thus, I performed a manual inspection of the videos of children in sessions 1, 2, and 3. It appeared that the majority of children's smiles occurred in a positive context. Because these appeared to be genuine smiles of joy with positive valence, it seems likely that the children in the *Not Relational* condition liked the robot less and felt less positively about it, especially initially. Although their engagement and attention did not differ significantly from the other children, they showed fewer positive emotions during the interaction. This is in line with prior work suggesting that laughter is a social phenomenon and people may laugh more with others with whom they are more familiar (Manson et al., 2013; Provine, 2001, 2012), and work showing that children show more positive affect with a robot that personalizes to them or a robot that is more expressive (Gordon et al., 2016; Kory-Westlund et al., 2017b).

Children in the *Relational* condition were more likely to remember which animal was the robot's favorite than children in the *Not Relational* condition. At first glance, I might argue that this was due to children's greater rapport or relationship with the relational robot and the connection between this and learning—especially because this learning was related to the relationship. However, as discussed later in Chapter 10, boys in the *Relational* condition tended not to have as great of rapport or relationship so this explanation would not make sense for them. As a result, it is unclear what may have caused this difference.

I also observed several unexpected results. Children in the *Not Relational* condition told longer stories, used more unique words, and used more descriptions in the Narrative Description task. However, this was likely due to the behavior of the boys in the *Relational* condition, who used the fewest words of all children (discussed further in Chapter 10). In addition, anecdotally, several of the most talkative children who participated in the study happened to be in the *Not Relational* condition, which could have also skewed the results somewhat.

9.5.2 *Relationships and Learning*

Children's learning results provided evidence for several of my hypotheses. In both conditions, children learned the target vocabulary words from the robot's stories. In addition, as seen in prior work (Kory-Westlund et al., 2017b), children who correctly identified the target words were more likely to use them in their own stories. This was an expected result. Children's expressive language abilities may follow their receptive abilities (Bloom, 1974; Ingram, 1974; Sénéchal, 1997), so use of vocabulary words likely reflects a deeper understanding and deeper encoding of the words. Notably, the target words presented in stories in this study were more difficult words than the words used in the Entrainment/Backstory study (Chapter 8), so finding the expected correlation here as opposed to in that study—where there was instead a ceiling effect and little variance in children's vocabulary test scores—is reasonable.

The robot's relational behavior did not lead to significant differences in words learned. However, I observed several expected correlations between children's scores on the relationship assessments and children's learning and behavior (e.g., vocabulary test scores, emulation of the robot during storytelling, use of keywords, and story length). Importantly, most correlations were stronger for children in the *Relational* condition than in the *Not Relational* condition, as well as for children who reported feeling closer and more of a relationship. These results support my hypothesis that the relationships children form with peers and friends contribute to their learning with them and from them. This jibes with other recent work linking learning to rapport (Sinha and Cassell, 2015a,b), as well as the similar correlations found in our prior work (Sections 5.6 and 9.1.1).

Not all of the relationships assessments were correlated with children's vocabulary learning. Notably, children's closeness ratings via the Inclusion of Other in Self task were not correlated with learning, though feeling closer in S1 was associated with telling more stories. Children's perception of the robot as a social-relational other, on the other hand, seemed highly important to how frequently they emulated it and how much they learned. These results suggest two things: First, that different aspects of a relationship may matter for learning. Social-relational behavior may impact rapport and imitation, which was reflected in children's stories; perhaps feelings of closeness are not so linked to rapport—which perhaps makes sense, since people may feel strongly attached to many different things, some of which may not be social entities. In addition, the relationship measures may have been measuring different aspects of children's relationships, not all of which may have been related to all of the different behaviors observed. Second, children's early impressions of the robot in S1 may have significantly impacted their behavior thereafter. Prior work has shown that first impressions in human-human and human-agent interactions are important and while they may be revised over time with experience with the other, can significantly affect behavior (Bar, Neta, and Linz, 2006; Cafaro, Vilhjálmsson, and Bickmore, 2016; Carney, Colvin, and Hall, 2007).

The correlations reported in Section 9.4.3.11 also suggest that—regardless of condition—children who rated the robot as a more human-like, social, and relational agent were more likely to treat it as such, often mirroring the robot more, using social behaviors such as saying goodbye, as well as learning more and telling longer stories.

The connections between children's rapport and relationship and their learning and language mirroring were more prominent than in the Entrainment/Backstory study (Chapter 8). In the prior study, children did not emulate the robot's phrases more as a result of the robot's relational behaviors. As I suggested earlier, the robot's entrainment in that study may not have been sufficient to generate enough rapport for it to affect children's storytelling. In this study, however, the robot had more opportunities to entrain; entrained, mirrored, and personalized more behaviors; and interacted with children multiple times. The difference between the relational and non-relational robots' nonverbal immediacy was greater in this study than in the Entrainment/Backstory study, and perhaps this contributed to the differences seen in children's learning and language mirroring.

9.5.3 *Relationships with Robots*

Based on children's responses to the Inclusion of Other in Self task, Narrative Description, and Picture Sorting Task, it is clear that these children thought of the robot and their relationship with it very differently than their relationships with other agents, such as their parents, friends, and pets. One child even said, "Robots can do different things than kids. And people cannot do the same things as robots." The robot was rated as similarly close to a friend and pet or favorite toy; at the S1 test children rated their parents similarly to the robot, but at the S8 test rated their parents more highly. The similarity to a pet or toy may reflect children's perception of the robot as fluffy, touchable, cute, and likable. The robot's behavior was designed to be peer-like and friend-like, which children may have picked up on; the children who responded to the robot's question about whether playing with it was like playing with another kid most frequently said it was. The robot was also sometimes more knowledgable than the children, like a parent (especially with respect to new vocabulary words); it also led the interaction and displayed authority as it directed the child through the conversation and story activities. Perhaps the differences between the robot and their parents were more prominent later on, after playing with the robot multiple times, leading to the greater difference at the posttest.

These ratings of the robot and other agents reflected the results of our prior studies, with a few caveats. While children in the Entrainment/Backstory study (Chapter 8) rated their best friend as closer than the robot, that was not the case here. Perhaps the robot's additional relational behaviors, different context, and different activities changed their perception of it. Unlike in the prior storytelling study (Kory-Westlund et al., 2018), children did not rate themselves as much closer to the robot at the end of the study. This is intriguing given that in this study, the robot specifically attempted to build a relationship, while in the prior study it did not. Perhaps children formed a greater attachment in the prior study in the way one might become attached to a toy or pet, while in this study, the robot's role led children to think of it more as a social agent, where the relationship may take longer to form—it is unclear.

Children's descriptions of their friends versus their descriptions of the robot revealed numerous differences in how they related to each. Children's descriptions of their friends were longer and picked up on relational qualities: names, activities done together, and facts relating to things their friends liked to do. Having greater experience and history with their friends, children were likely able to think of more things to share about them than about the robot. The robot was subject to physical descriptions far more frequently; this is the kind of description that goes with being an inanimate object or unfamiliar other as opposed to being a relational other. Notably, children also described activities with the robot, but were more somewhat likely to do so at the posttest than at the S1 test.

The Picture Sorting Task was especially revealing about children's perceptions of the robot's animacy and human-likeness. Based on where children placed the baby and cat, it was clear they understood several important features that distinguish these entities from the others regarding aliveness, perception, cognition, and/or animacy. Similarly, they understood that computers, the mechanical robot arm, and the teddy bear had less of these features, generally placing them closer to the inanimate table. The fact that the frog was placed somewhere in the middle, sometimes closer than the robot and sometimes farther, is revealing with regards to children's perceptions of the robot. Children knew the frog was more human-like than a mechanical robot arm, perhaps because it was a living being; the robot, however, filled an in-between space. It was seen as more human-like than a computer, but definitively less human-like than a baby and a cat. Children knew it was not alive in the same way as a human or sophisticated animal; they also knew it was perceiving them and responding to them in a way unlike the more static computers and mechanical things. Children's placement of the entities in this study were fairly similar to in the Entrainment/Backstory study (Section 8.3.2.3).

This perception children had of the robot as being in-between humans and other things, of being friend-like and pet-like, but more mechanical than a real human, fits with other recent work suggesting that children may categorize social robots as neither animate nor inanimate (or perhaps as both), placing robots "in between" in an ontological category different from both living and non-living things (Gaudiello, Lefort, and Zibetti, 2015; Kahn et al., 2011; Severson and Carlson, 2010). Other work has shown that children describe social robots as being "in between" living and non-living, and will ascribe properties of both living animals—such as mental states, perceptual abilities, and moral standing—as well as properties of mechanical things—such as running on batteries, being made by people, and being able to break (Kahn, Friedman, and Hagman, 2002; Kahn et al., 2012; Knox, Spaulding, and Breazeal, 2016; Kory-Westlund et al., 2016a; Kory and Breazeal, 2014; Melson et al., 2009; Severson and Carlson, 2010; Weiss, Wurhofer, and Tscheligi, 2009). The robot is not a fantasy or an imaginary friend; it is part of reality. It may have some elements of fantastical beings, imaginary friends, and parasocial relationships (Brunick et al., 2016; Calvert, 2017; Gleason, 1997; Gleason and Hohmann, 2006), but it is real and children seem to see it as a real, animate, technological being.

Sherry Turkle has questioned whether children do understand that social robots are different from humans and from machines, arguing that children are sometimes confused about what these kinds of digital agents actually are (Turkle, 1985; Turkle, 2005, 2007; Turkle et al., 2006a,b). She observed children playing with various early computational toys, computers, and social robots, finding that children frequently tried to make sense of the devices by attributing intelligence and cognitive and psychological properties to them. Since these computational artifacts clearly did not have any psychological properties, Turkle argued that children's attributions of any kind of psychological life were mis-attributions. Digital agents presented as relational agents, though evocative, were inauthentic and deceiving people in potentially harmful ways.

However, much of the data Turkle has discussed was collected with children who grew up in the 1980's, 1990's, and early 2000's. The agents of today and children of today are very different, as I discuss further in Section 11.2. Relational technologies of today appear far more often to have social, emotional, and psychological capabilities that earlier devices did not. Thus, the question of whether children are wrongly attributing these properties—given that current technologies actually do appear to display them—seems to be the wrong question to ask. Children are labeling what they see and experience. The social robots I have placed with children have some social capabilities, designed to mimic the social capabilities of humans, for the purpose of enabling children to interact with the robots more fluently. More valid, perhaps, are Turkle's concerns about the authenticity of relational technologies, whether such technologies are unfairly deceiving humans, and whether this is harmful or immoral. I will discuss this question further in Chapter 13.

9.5.4 *Age*

The primary differences by age were in the learning and language results. As expected, children's age was correlated with their vocabulary test scores. Younger children told shorter stories, used fewer keywords, used fewer unique words, and used similar matching phrases than older children. This pattern has been seen in prior work (Kory-Westlund et al., 2017b) (also Chapter 8). Younger children were also less likely to correctly identify the robot's favorite animal, and they gave shorter descriptions during the Narrative Description task. Anecdotally, the four-year-olds tended to be more shy around the experimenters and hesitant during the first interactions with the robot. As evidence to support this anecdote, the 4-year-olds were somewhat less likely to say goodbye during Sessions 1 and 2, which may reflect that it may have taken them longer to warm up to speaking to the robot. The activities may have been better for children five years and up. Relatedly, the seven-year-olds may have benefitted from somewhat more advanced stories and vocabulary words; they tended to know more of the words at the pretest.

9.5.5 *Limitations*

As in the Entrainment/Backstory study, this study did not control for children's individual differences, such as differences in learning ability, language ability, socio-economic status, or personality. All of these factors can influence children's social interactions and learning. I also did not have an equal number of children in each age group participate in the study. Future work should explore a more homogenous sample, and explore the stability of results across individual differences and ages.

I also lacked complete video, audio, and affect data for all children. Due to the nature of longitudinal research, some missing data are inevitable. Some data was also missing because children's faces were not recognized by the Affdex software. As a result of the missing data, some analyses reported are underpowered. Future work should endeavor to collect quality audio and video recordings throughout the study.

The robot's entrainment and other automated relational behaviors worked reasonably well. The robot's classification of children as showing high or low exuberance lined up fairly well with experimenter observations, but because of issues with ASR, sometimes the robot responded as if children were lower exuberance because it simply had not heard that they spoke. This classification could use additional features so that it does not rely so much on the ASR. Relatedly, when children did not speak or spoke very rarely, the robot was not able to entrain its speech, though it could adapt exuberance and other features. Furthermore, because of the difficulties with ASR, the robot did not always respond appropriately and, as mentioned earlier, did not always respond at the right time. Some of the difficulties with ASR were a result of needing to use an offline ASR system as opposed to an internet-enabled system, which was less accurate. Wifi connectivity in several of the schools was spotty at best and meant that using internet-enabled ASR was off the table. It would be worthwhile to find ways around ASR. Perhaps including new types of entrainment that do not involve speech, such as more entrainment of posture, affect, gaze, and word use, could improve the robot's relational AI and help bolster the robot's relational capabilities.

Several of the relational assessments may need revision and improvement. Related to the autonomy and ASR issues, I observed that the robot was placed farther from the adult human in the Picture Sorting Task than in the Entrainment/Backstory study (Section 8.3.2.3). This was likely related to the differences in the robots' autonomy between these two studies. In this study, the robot was autonomous, with all the difficulties with ASR, timing, and selecting appropriate responses that currently entails; in the Entrainment/Backstory study, the teleoperator could mitigate some of that awkwardness, which perhaps led to greater perception of the teleoperated robot as a human-like agent.

The scenarios included during the interaction would benefit from being automated; at present, they required experimenter input to ensure the robot recognized the situation appropriately (e.g., the Negotiation Task). New scenarios could also be developed that may get at different aspects of children's friendships, such as different helping or sharing tasks, or different empathy tasks.

It may be worth adding a new question to the Social Relational Interview (SRI) involving reciprocity of behavior. Many children's SRI scores were high or at the ceiling; the SRI may not capture enough dimensions of social and relational behavior to truly distinguish between different children's conceptions of the robot's abilities and behaviors. I did observe several similar patterns in the SRI results as in prior work (Kory-Westlund et al., 2018). In particular, in both studies, girls rated the robot as more of a social-relational agent overall. I observed several effects of time in this study that were not present in the earlier work, which perhaps was related to the robot's relational behaviors and the differences between the robots and interactions in the two studies.

For example, compared to prior work (Kory-Westlund et al., 2018), children's ratings during the Inclusion of Other in Self task were lower overall, with the exception of their ratings of the bad guy. If children's ratings had been lower only for the robot, that would have been informative in relation to the robot's personality and behavior; however, children's ratings for their friend, parent, and pet/toy were also lower overall. Thus, this was likely a population-related difference. It would be worth testing this assessment with a wider variety of children to determine what kinds of ratings to expect from children of different backgrounds.

Children's descriptions of their friends and of the robot during the Narrative Description task were shorter in this study than in prior work (Kory-Westlund et al., 2018), and included far fewer descriptions of actions performed with their friends. This may have been related to the kind of prompting the experimenter provided, as this became more limited and more controlled for the present study. In both cases, children described their friend at greater length than the robot. However, in the present study, I observed several differences in the length of children's descriptions in relation to gender, condition, and time, while previously, the primary difference had been that girls spoke more than boys about the robot. Whether these differences were a result of changing the protocol for this task, a result of the different robot interactions, a result of the different group of participants, or some combination is unclear. Further work testing the Narrative Description task would be beneficial.

Children's disclosures during this study were fairly similar in length and content to their disclosures in the prior study (Kory-Westlund et al., 2018), though the content did vary somewhat as a result of adjusting the prompts given by the robot to better fit the robot's character. The differences by time (i.e., disclosing more later on) were less clear in the Relational AI study, perhaps because of the frequent interactions with gender that muddied the waters, though still apparent for girls. This task is useful and I recommend using it as a measure of relationship. However, it is worth mentioning that the experimenter remained in the room with the child during the robot interaction and disclosures, which could have influenced children's behavior. This task may provide different results if the experimenter was not present, or if the experimenter was at least less present. For example, in one prior study, the room available for the experiment allowed the experimenter to stay behind the child mostly out of sight during the robot interaction, which may have helped children engage more fully with the robot alone ("out of sight, out of mind"). In this study, the rooms at the schools did not have sufficient space for the experimenter to be very far from the child or robot, and they were often in peripheral view.

Finally, given the richness of the dataset collected during this study, there are many more features in the data that could be examined in the future. Data analysis can thus be considered to be ongoing.

10 THE IMPACT OF GENDER ON CHILDREN'S RELATIONSHIPS WITH ROBOTS

10.1 GENDER DIFFERENCES IN SOCIAL RELATIONSHIPS

Girls and boys approach social relationships differently (Benenson, 2014; Benenson et al., 2018; Buhrmester and Furman, 1987; Gleason and Hohmann, 2006; Walker, Irving, and Berthelsen, 2002). For example, girls tend to be more focused on individual social relationships than boys are. They tend to care about exchanging personal information and learning about others' relationships and relative status. Girls tend to have fewer, closer friends. Girls are also more likely to try to avoid conflict, are often more competent at social problem-solving, and are generally more egalitarian than boys. Boys tend to care somewhat more about being part of their peer group, understanding their skills relative to the skills other boys have, and do not care as much about exchanging personal information or explicitly discussing their relationships. They tend to be friends with most other boys, not being nearly so exclusive in their friendships. Although these are broad generalizations that may not apply to any particular individual child, as generalizations they are useful in attempting to understand the broader patterns seen among groups of children.

Because of the gender differences frequently apparent in children's social relationships, I expected that children's gender might also impact their rapport, relationship, learning, mirroring, affect, and behavior with robots. To investigate this question, I performed additional analyses on the data from the Relational AI study (Chapter 9).

10.2 METHODOLOGY

10.2.1 *Hypotheses*

I expected several differences by gender. First, I expected that girls would act more socially overall, since girls tend to be more focused on social relationships than boys (Benenson, 2014). Second, I expected that girls would rate the robot as more social and relational than boys, since we had seen this in a prior study (Kory-Westlund et al., 2018).

10.2.2 *Data Analysis*

I performed the same analyses as before (Chapter 9), adding in Gender in order to examine whether there were differences between girls and boys. These analyses were planned comparisons based on my hypotheses unless otherwise indicated; I had expected some gender differences because of the aforementioned literature and because we had previously seen some differences in girls' versus boys' ratings of how social and relational a robot was (Kory-Westlund et al., 2018). Where appropriate, I corrected for multiple comparisons with the Bonferroni correction.

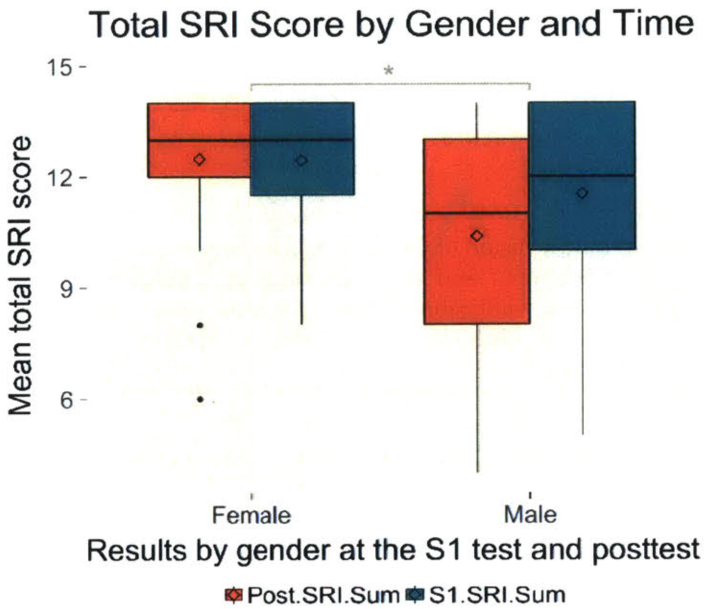


Figure 70: Children's total Social Relational Interview scores by gender and condition at the pretest and posttest. Scores on individual questions can range from 0–2, with 0 = robot does not care / not mind / is pretending, 1 = not sure, maybe, 2 = social/relational (would be sad, would help, etc); thus, the total score can range from 0–14.

10.3 RESULTS

The results below are broken into three sections, mirroring Section 9.4: (1) *Relationship*, (2) *Learning*, and (3) *Behavior and engagement*.

10.3.1 *Relationship*

10.3.1.1 *Inclusion of Other in Self Task*

I performed a mixed analysis of variance with Condition (between: *Relational* vs. *Not Relational*), Time (within: test in S1 vs. test in S8), Agent (within: bad guy, best friend, parent, robot, pet/toy), and Gender (between: boys vs. girls), with Age as a covariate. As before (Section 9.4.1.1), I observed a significant main effect of Agent, $F(4,371) = 23.6$, $p < 0.001$ and a significant interaction of Agent with Time, $F(4,371) = 4.30$, $p = 0.002$. Post-hoc tests revealed the same patterns as before, with the additional trend for an interaction of gender and time (Table 13). Girls in the *Not Relational* condition appeared to keep a stable rating of the robot over time while boys in the *Not Relational* condition slightly increasing their ratings of the robot over time. Girls in the *Relational* condition decreased their ratings slightly over time, while boys in the *Relational* condition increased their ratings.

10.3.1.2 *Social Relational Interview*

SCORES I performed mixed analyses of variance with Condition (between: *Relational* vs. *Not Relational*), Time (within: test in S1 vs. posttest), and Gender (between: boys vs. girls), with Age as a covariate for the total score and then, to explore children's

Table 13: Children's overall Inclusion of Other in Self responses and responses by gender and condition after Session 1 and after Session 8. A 1 indicates choosing the farthest apart circles, while a 7 indicates choosing the most closely overlapping circles. Here, "MAD" = "median absolute deviation"; "Md." = "Median"; "Rg." = "Range"; "G." = "Girls"; "B." = "Boys."

Entity			S1 Mean (SD)	Md.	Rg.	MAD	S8 Mean (SD)	Md.	Rg.	MAD
Bad	All		2.52 (1.96)	1.5	1–7	0.74	1.63 (1.27)	1	1–7	0
Guy	G.	RR	2.17 (1.70)	1	1–7	0	1.09 (0.30)	1	1–2	0
		NR	2.78 (2.49)	1	1–7	0	1.27 (0.65)	1	1–3	0
	B.	RR	2.73 (1.74)	2	1–6	1.48	2.18 (1.99)	1	1–7	0
		NR	2.50 (2.15)	1	1–7	0	2.00 (1.25)	1.5	1–4	0.74
Parent	All		4.10 (1.94)	4	1–7	2.97	5.00 (2.04)	6	1–7	1.48
	G.	RR	4.33 (2.10)	4.5	1–7	2.22	5.64 (1.86)	7	2–7	0
		NR	4.18 (1.83)	4	2–7	1.48	4.64 (1.80)	4	2–7	1.48
	B.	RR	4.58 (1.83)	4.5	2–7	2.22	4.58 (2.31)	5.5	1–7	2.22
		NR	3.38 (1.98)	2	1–6	1.48	5.18 (2.23)	7	1–7	0
Pet/	All		4.30 (1.86)	4	1–7	1.48	4.82 (1.70)	5	1–7	1.48
Toy	G.	RR	4.92 (1.78)	5	2–7	2.97	5.18 (1.66)	6	3–7	1.48
		NR	3.40 (1.35)	3	1–5	1.48	4.27 (1.95)	5	1–7	1.48
	B.	RR	4.25 (1.71)	4	2–7	2.22	4.92 (1.62)	5	2–7	1.48
		NR	4.46 (2.30)	5	1–7	2.97	4.91 (1.64)	5	3–7	2.97
Friend	All		4.00 (2.32)	3	1–7	2.97	3.71 (2.21)	3	1–7	2.97
	G.	RR	4.00 (2.09)	3	2–7	1.48	4.09 (2.34)	4	1–7	2.97
		NR	4.27 (2.57)	4	1–7	4.45	3.27 (2.00)	3	1–7	1.48
	B.	RR	3.36 (2.58)	2	1–7	1.48	2.83 (1.95)	3	1–7	2.97
		NR	4.31 (2.25)	5	1–7	2.97	4.73 (2.33)	6	1–7	1.48
Robot	All		3.77 (1.87)	3	1–7	1.48	3.98 (1.64)	4	1–7	1.48
	G.	RR	4.67 (1.72)	4.5	2–7	2.22	3.73 (1.68)	3	2–6	1.48
		NR	3.73 (1.95)	3	1–7	1.48	3.60 (1.65)	3	1–7	0
	B.	RR	3.17 (1.75)	3	1–7	1.48	4.58 (1.78)	4.5	1–7	2.22
		NR	3.54 (1.94)	3	1–6	1.48	3.91 (1.45)	4	2–7	1.48

responses further, for the individual questions. Where appropriate, I ran post-hoc tests with Tukey's HSD.

I observed the same pattern of results as in Section 9.4.1.2, listed in Table 14, with the addition of the following gender patterns. First, when examining the Social Relational Interview total score, I observed that girls rated the robot as a greater social-relational agent than did boys (see Figure 70). There was also a trend toward an interaction of time with gender, with girls maintaining their high ratings and boys' ratings decreasing slightly over time. Table 15 shows children's ratings by gender, time, and condition.

I examined the change in children's ratings from the S1 test to the posttest further by computing the difference between children's S1 and posttest rating. Analyzing this difference score revealed the same trend as above: girls maintained their high ratings (mean change = 0.04, $SD = 1.43$), while boys' ratings decreased (mean change = -1.14, $SD = 2.54$), $F(1,40) = 3.40$, $p = 0.072$.

With regards to the individual Social Relational Interview questions, girls were more likely to say that the robot would feel sad if another child was mean to it than boys. Girls were also slightly more likely to say the robot would be sad if it had no friends. Girls were likely to say the robot would cheer up another child at both the S1 test and posttest, while boys were less likely to say so at the posttest. There was also a trend for boys in the *Not Relational* to be more likely to say the robot would cheer up another child at the posttest than boys in the *Relational* condition. Girls were more likely than boys to say that the robot really did like them.

JUSTIFICATIONS A mixed analyses of variance with Condition (between: *Relational* vs. *Not Relational*), Time (within: test in S1 vs. posttest), Gender (between: boys vs. girls), and Justification type (within: the eight justification types coded for), with Age as a covariate, revealed significant main effects of gender, $F(1,44) = 4.78$, $p = 0.034$; age, $F(3,42) = 4.52$, $p = 0.008$; and a significant interaction between justification type and time, $F(7,643) = 2.10$ $p = 0.042$.

Post-hoc tests revealed the same patterns discussed in Section 9.4.1.2, and additionally, that girls provided more justifications for the 7 questions (total: $M = 5.74$, $SD = 3.78$; per question: $M = 0.72$, $SD = 1.03$) at both times than boys did (total: $M = 4.24$, $SD = 3.50$; per question: $M = 0.51$, $SD = 0.97$). Furthermore, there was trend toward an interaction of gender with justification type and time, suggesting that girls and boys explained their responses differently.

10.3.1.3 *Picture Sorting Task*

I performed mixed analyses of variance with Condition (between: *Relational* vs. *Not Relational*), Time (within: test in S1 vs. test in S8), Entity (within: robot, baby, cat, frog, teddy bear, movie robot, robot arm, computer), and Gender (between: boys vs. girls), and Age as covariate for each entity's position, as well as for each entity's position relative to the robot.

For entity positions, I observed a significant main effect of entity, $F(7, 626) = 45.9$, $p < 0.001$. Post-hoc tests revealed the same patterns as before (Section 9.4.1.3). In addition, there were trends toward interactions of gender with entity and gender with entity and time, suggesting, e.g., that girls may have placed the baby closer to the human adult than boys did, that girls may have placed the computer as farther than boys did, and that girls may have moved the teddy bear farther away at the posttest.

Table 14: Social Relational Interview results and interesting trends by time, gender, age, and condition. For individual questions. results were considered significant when $p <$ 0.007. The questions were as follows. 1: "Let's pretend another kid was mean to Red and took Red's story tablet. Would Red feel sad or would Red not mind?"; 2: "Let's pretend Red didn't have any friends. Would Red not mind or would Red feel sad?"; 3: "Let's pretend another kid needs help. Would Red try to help or would Red not care?"; 4: "Let's pretend Red was really happy or really upset about something. Would Red not care about telling anyone, or would Red want to tell a friend?"; 5: "Let's pretend another kid is sad. Would Red try to cheer them up or would Red not care?"; 6: "Does Red really want to make friends, or is Red just pretending?"; 7: "Does Red like you or is Red just pretending?".

Question	Effect	df	F	p
Sad if mean	Time	1,41	4.68	0.036
Sad if mean	Gender	1,38	3.63	0.064
Sad if mean	Age	3,38	5.68	0.003
Sad no friends	Gender	1,38	3.76	0.060
Share info	Condition	1,37	3.55	0.067
Cheer up child	Time x Gender	1,41	5.42	0.025
Cheer up child	Time x Gender x Condition	1,41	3.77	0.059
Likes you	Gender	1,37	13.3	< 0.001
Likes you	Time x Condition	1,40	3.65	0.063
Overall	Gender	1,37	6.27	0.017
Overall	Age	3,37	3.35	0.029

Table 15: Descriptive statistics by time, gender, and condition for the Social Relational Interview. Here, "Con." = "Condition." Scores on individual questions can range from 0–2, with 0 = robot does not care / not mind / is pretending, 1 = not sure, maybe, 2 = social/relational (would be sad, would help, etc); thus, the total score can range from 0–14. The questions were as follows. 1: "Let's pretend another kid was mean to Red and took Red's story tablet. Would Red feel sad or would Red not mind?"; 2: "Let's pretend Red didn't have any friends. Would Red not mind or would Red feel sad?"; 3: "Let's pretend another kid needs help. Would Red try to help or would Red not care?"; 4: "Let's pretend Red was really happy or really upset about something. Would Red not care about telling anyone, or would Red want to tell a friend?"; 5: "Let's pretend another kid is sad. Would Red try to cheer them up or would Red not care?"; 6: "Does Red really want to make friends, or is Red just pretending?"; 7: "Does Red like you or is Red just pretending?".

Question	Con.	Girls		Boys	
		S1 Mean (SD)	Post Mean (SD)	S1 Mean (SD)	Post Mean (SD)
Sad if mean	RR	1.83 (0.58)	1.67 (0.78)	1.80 (0.63)	1.30 (0.95)
	NR	1.91 (0.30)	1.82 (0.60)	1.50 (0.90)	1.33 (0.98)
Sad no friends	RR	1.58 (0.67)	1.67 (0.65)	1.40 (0.97)	1.40 (0.84)
	NR	1.73 (0.65)	1.73 (0.65)	1.25 (0.87)	1.17 (0.94)
Help child	RR	1.73 (0.65)	1.64 (0.81)	1.80 (0.63)	1.50 (0.71)
	NR	2.00 (0.00)	2.00 (0.00)	1.50 (0.90)	1.67 (0.65)
Share info	RR	1.58 (0.79)	1.67 (0.65)	1.67 (0.71)	1.22 (0.83)
	NR	1.91 (0.30)	1.55 (0.69)	1.92 (0.29)	1.83 (0.39)
Cheer up child	RR	1.75 (0.62)	2.00 (0.00)	1.80 (0.63)	1.30 (0.82)
	NR	2.00 (0.00)	1.91 (0.30)	2.00 (0.00)	1.83 (0.58)
Wants friends	RR	1.67 (0.78)	1.67 (0.78)	1.60 (0.84)	1.60 (0.70)
	NR	1.82 (0.60)	1.91 (0.30)	1.83 (0.39)	1.33 (0.98)
Likes you	RR	1.58 (0.67)	1.92 (0.29)	1.00 (1.00)	1.33 (0.87)
	NR	2.00 (0.00)	2.00 (0.00)	1.58 (0.79)	1.33 (0.89)
Total	RR	11.6 (2.19)	12.1 (2.71)	11.6 (2.71)	10.2 (2.49)
	NR	13.4 (0.81)	12.9 (1.04)	11.4 (2.19)	10.5 (4.03)

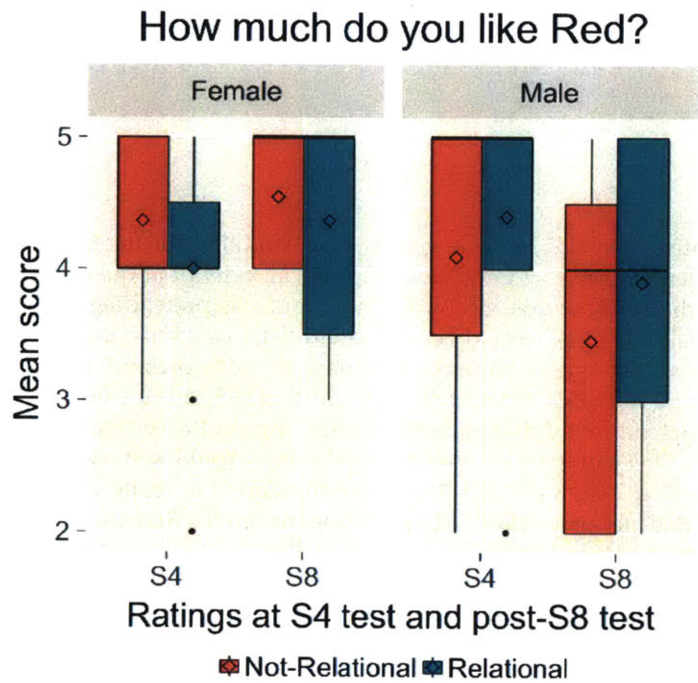


Figure 71: Children's responses to the question, "How much do you like Red?" by gender, condition, and time. Responses were given on the smileyometer scale of 1 (very frowny) to 5 (very smiley).

Regarding the distance of each entity relative to the Tega robot, I observed a main effect of entity, $F(6,542) = 46.2$, $p < 0.001$, with the same pattern of results as before. In addition, there was a trend for an interaction of gender with entity and time. Some of the gender trends included girls' tendency to place the robot arm closer to the human adult than the Tega robot at the posttest than at the S2 test, while boys showed the opposite pattern; girls also tended to place the cat, frog, and computer farther away from Tega than boys did.

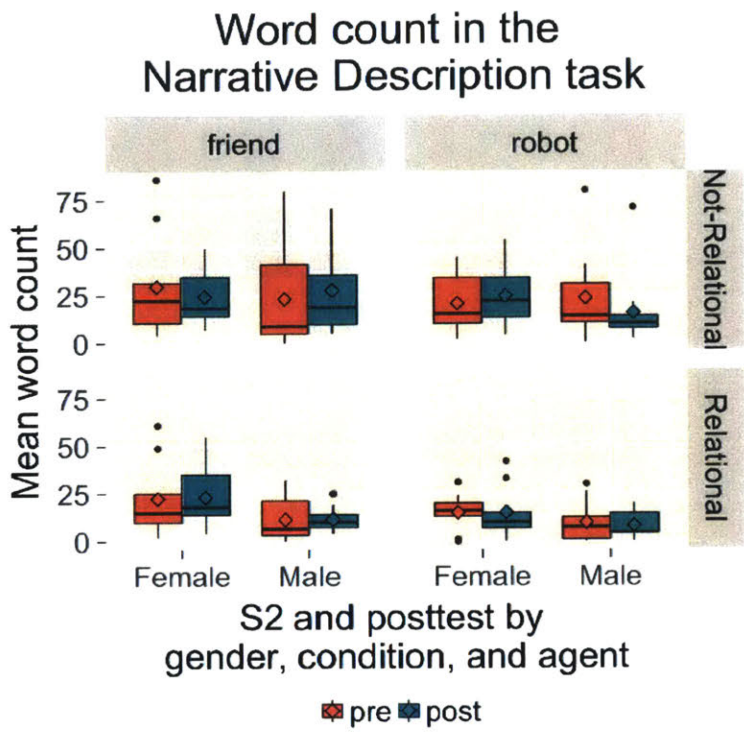
10.3.1.4 *Memory/Rapport Questions*

I ran mixed analyses of variance with Condition (between: *Relational* vs. *Not Relational*), Time (within: questions at S4 vs. S8), and Gender (between: boys vs. girls), with Age as a covariate for each Memory/Rapport question.

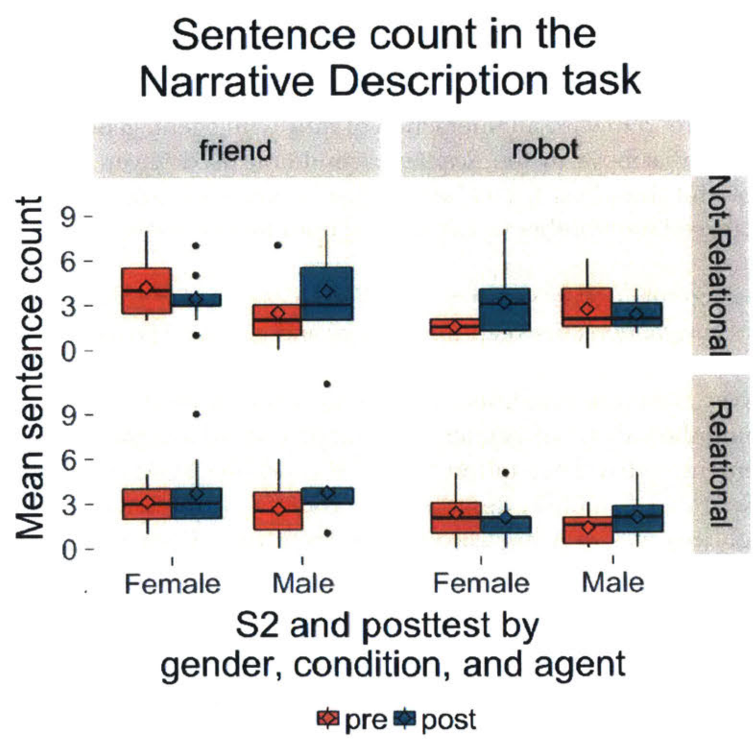
In addition to the patterns discussed previously (Section 9.4.1.4), I observed a significant interaction of time with gender for the question, "How much do you like Red?", $F(1,39) = 5.55$, $p = 0.024$. Post-hoc tests did not reveal any significant differences, though there was a trend for boys to rate the robot lower at the S8 posttest (Figure 71).

10.3.1.5 *Narrative Description*

I performed mixed analysis of variance with Condition (between: *Relational* vs. *Not Relational*), Time (within: S2 test vs. Posttest), Agent (within: Friend vs. Robot), and Gender (between: boys vs. girls), with Age as a covariate.



(a) Word Count.



(b) Sentence Count.

Figure 72: In the Narrative Description task, children described their friends at greater length than they did the robot, and this varied by condition and gender.

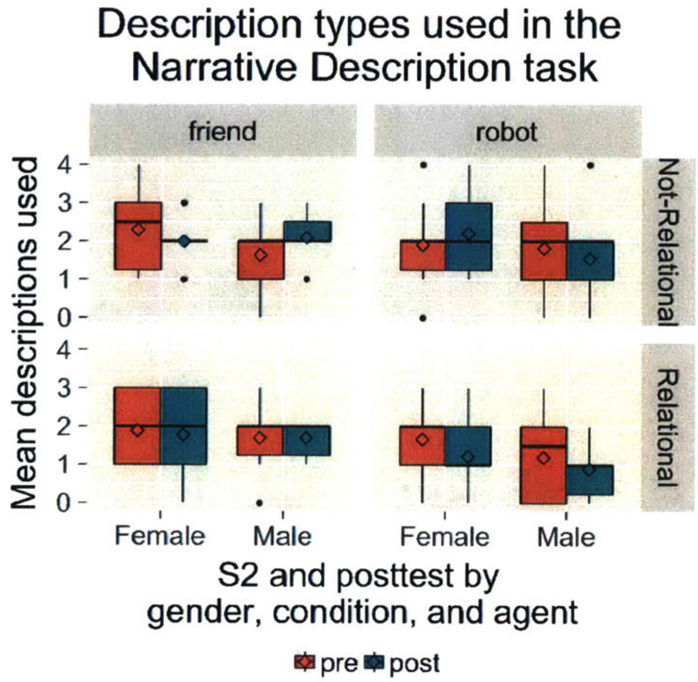
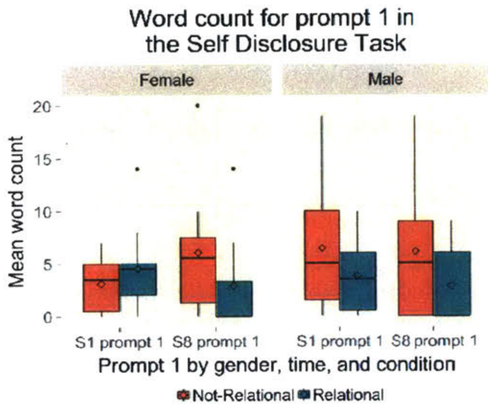


Figure 73: The number of different kinds of descriptions children used (name, facts, social qualities, physical attributes, activities) during the Narrative Description task.

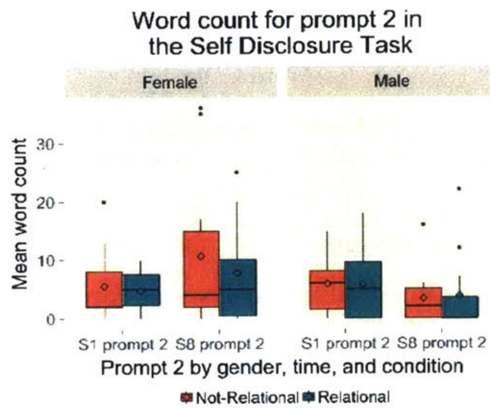
With regards to word count, like before (Section 9.4.1.5), I observed main effects of agent, $F(1{,}108) = 10.5$, $p = 0.002$, and condition, $F(1{,}33) = 4.51$, $p = 0.041$ (Figure 72a). There was also a trend toward an interaction of time with agent, gender, and condition, which suggested that boys in the *Relational* condition used fewer words to describe both the robot and their best friend, and furthermore, that boys in the *Not Relational* condition decreased the number of words they used to describe the robot from pretest to posttest.

For sentence count, I observed a main effect of agent, $F(1{,}108) = 23.6$, $p < 0.001$; and significant interactions between time, agent and gender, $F(1{,}108) = 7.46$, $p = 0.007$; as well as time, agent, gender, and condition, $F(1{,}108) = 10.5$, $p = 0.002$ (Figure 72b). Girls in the *Not Relational* condition describing a friend used more sentences than when describing the robot, but primarily at the pretest than at the posttest. They also used more sentences than boys in the *Relational* condition when describing the robot. Boys in the *Relational* condition described the robot with the fewest sentences at the pretest, though this was only significantly different from boys at the posttest in the *Not Relational* condition. There was also a trend for boys in the *Relational* condition to describe a friend with more sentences than they did the robot.

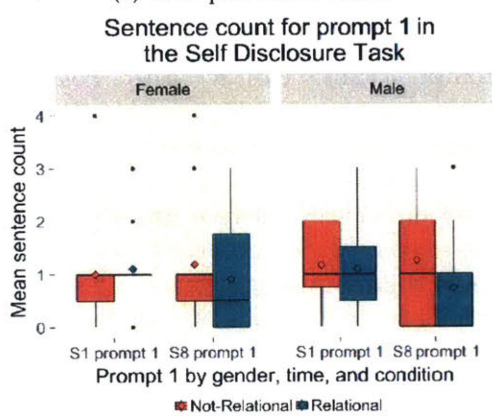
For the number of descriptions children used, I observed the same main effects as before for agent, $F(1{,}108) = 4.88$, $p = 0.029$; condition, $F(1{,}108) = 5.77$, $p = 0.022$; and age, $F(3{,}33) = 4.99$, $p = 0.006$ (Figure 73). There was also a trend toward an interaction of time, condition, agent, and gender, with boys in the *Relational* condition using the fewest descriptions, especially at the posttest.



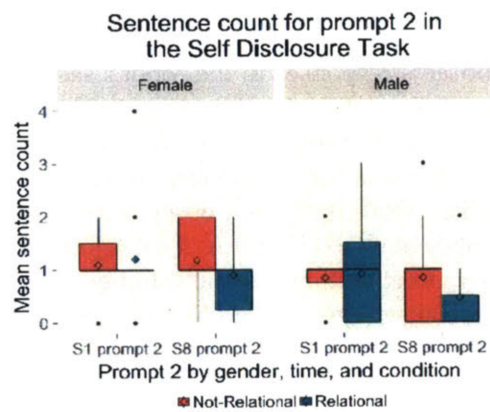
(a) Prompt 1 Word Count



(b) Prompt 2 Word Count

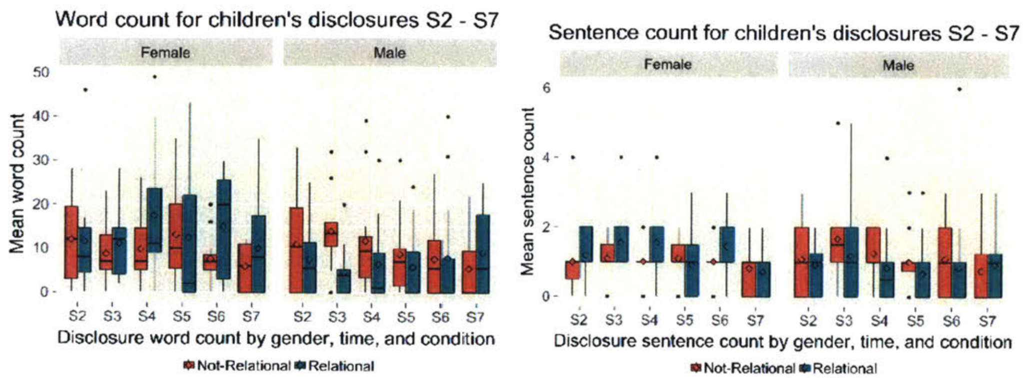


(c) Prompt 1 Sentence Count



(d) Prompt 2 Sentence Count

Figure 74: In the Self-disclosure Task, girls generally gave longer answers than boys in S8 to both the robot's prompts.



(a) Word Count

(b) Sentence Count

Figure 75: Children's disclosures during Sessions 2–7 followed the pattern as the Self-disclosure Task in S1 and S8.

10.3.1.6 *Targeted Self-Disclosure Task*

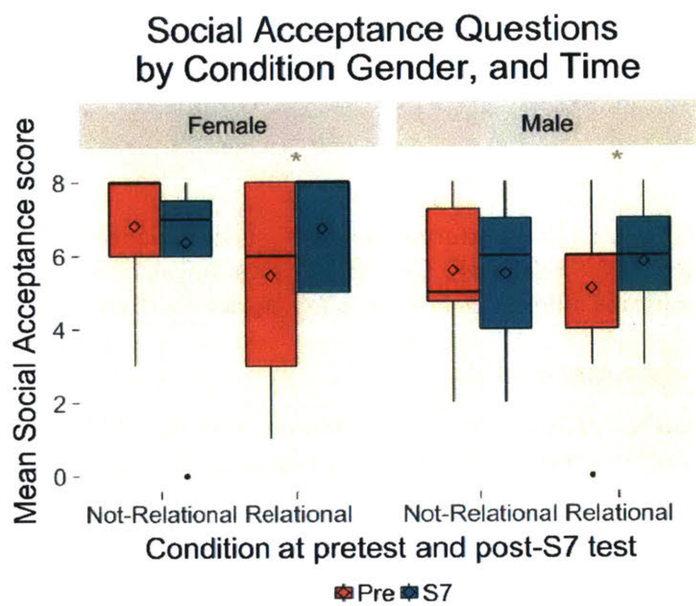
I performed mixed analyses of variance with Condition (between: *Relational* vs. *Not Relational*), Time (within: either S1 vs. S8, or S2–7), and Gender (between: boys vs. girls), with Age as a covariate. Like before (Section 9.4.1.6), there was a main effect of Age with regards to the number of sentences children spoke for both the robot's first prompt, $F(3,37) = 4.03$, $p = 0.014$; and the robot's second prompt, $F(3,37) = 3.05$, $p = 0.040$.

I also observed a significant interaction of time with gender on the word count of children's responses to the robot's second prompt, $F(1,40) = 6.01$, $p = 0.019$. Post hoc tests showed that girls gave longer responses in S8 than boys did (Figures 74a, 74b, 74c, and 74d). A similar trend was present for children's responses to the first prompt, though it was not statistically significant.

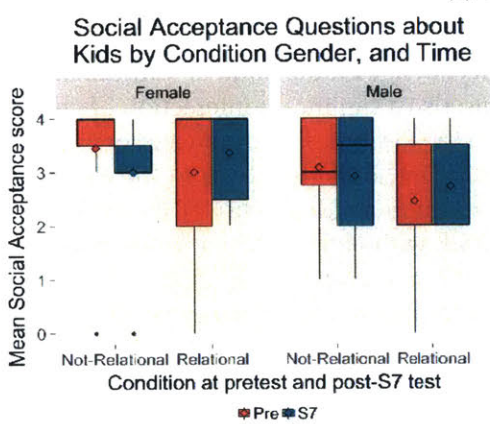
Like before, I did not observe any significant differences between groups for the number of words in children's disclosures in sessions 2–7, and I observed a significant main effect of time on the number of sentences children used, $F(5,210) = 2.58$, $p = 0.027$ (Figure 75b). In addition, I observed trends for interactions of condition with time and condition with gender (Figure 75a). In most sessions, girls in the *Relational* condition appeared to use more words than girls in the *Not Relational* condition, while the reverse was true for boys.

10.3.1.7 *Social Acceptance Questions*

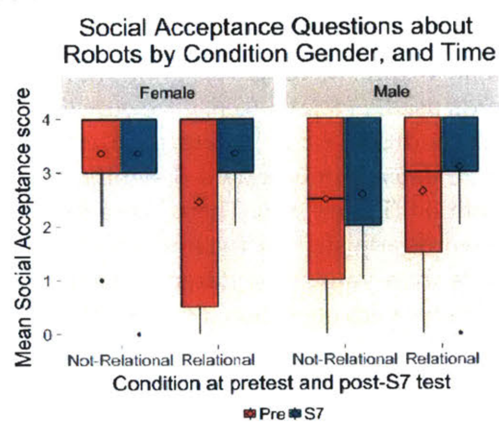
I performed mixed analyses of variance with Condition (between: *Relational* vs. *Not Relational*), Time (within: pretest vs. post-S7 test), Gender (between: boys vs. girls) and Age as a covariate. I observed the same patterns as before (Section 9.4.1.7). There was a trend for girls to be more accepting of both other children and robots than boys (Figure 76).



(a) Overall



(b) Children



(c) Robot

Figure 76: Children in the *Relational* condition increased their Social Acceptance ratings from the pretest to the S7 test, while children in the *Not Relational* condition slightly decreased their ratings. The questions were, "Would you like to be good friends with a kid/robot who can't hear well?" and "Would you like to be good friends with a kid/robot with special needs?" Children responded with a *yes* (coded as 2), *maybe* (1) or *no* (0) to each question.

10.3.2 *Learning*

10.3.2.1 *Target Vocabulary Word Identification*

Analyses of variance on children's full and Retell-only vocabulary scores with Condition (between: *Relational* vs. *Not Relational*), Time (Pretest vs. Posttest), Gender (girls vs. boys), and Age as a covariate revealed main effects of time, $F(1,25) = 11.2$, $p = 0.003$; and $F(2,52) = 19.5$, $p < 0.001$, like before (Section 9.4.2.1). There were no differences by condition.

10.3.2.2 *Stories*

Like before (Section 9.4.2.2), I performed an analysis of variance with Condition (between: *Relational* vs. *Not Relational*), Gender (girls vs. boys), and Age as a covariate, revealing no significant differences in how many stories children told.

10.3.2.3 *Target Vocabulary Word Use*

I performed analyses of variance with Condition (between: *Relational* vs. *Not Relational*), Gender (girls vs. boys), and Age as a covariate. Like before (Section 9.4.2.3), there were no significant differences in children's total use of the target vocabulary words or the robot's catchphrases. However, there was a trend for girls to use more of the target words and phrases than boys, $F(1,42) = 2.21$, $p = 0.145$ (Figure 77a), as well the same trend for older children to use more than younger children $F(3,42) = 1.89$, $p = 0.145$. Looking at the use of total keywords divided by stories told, there was a similar trend for girls to use more target words and phrases than boys, $F(1,42) = 2.22$, $p = 0.144$ (Figure 77b).

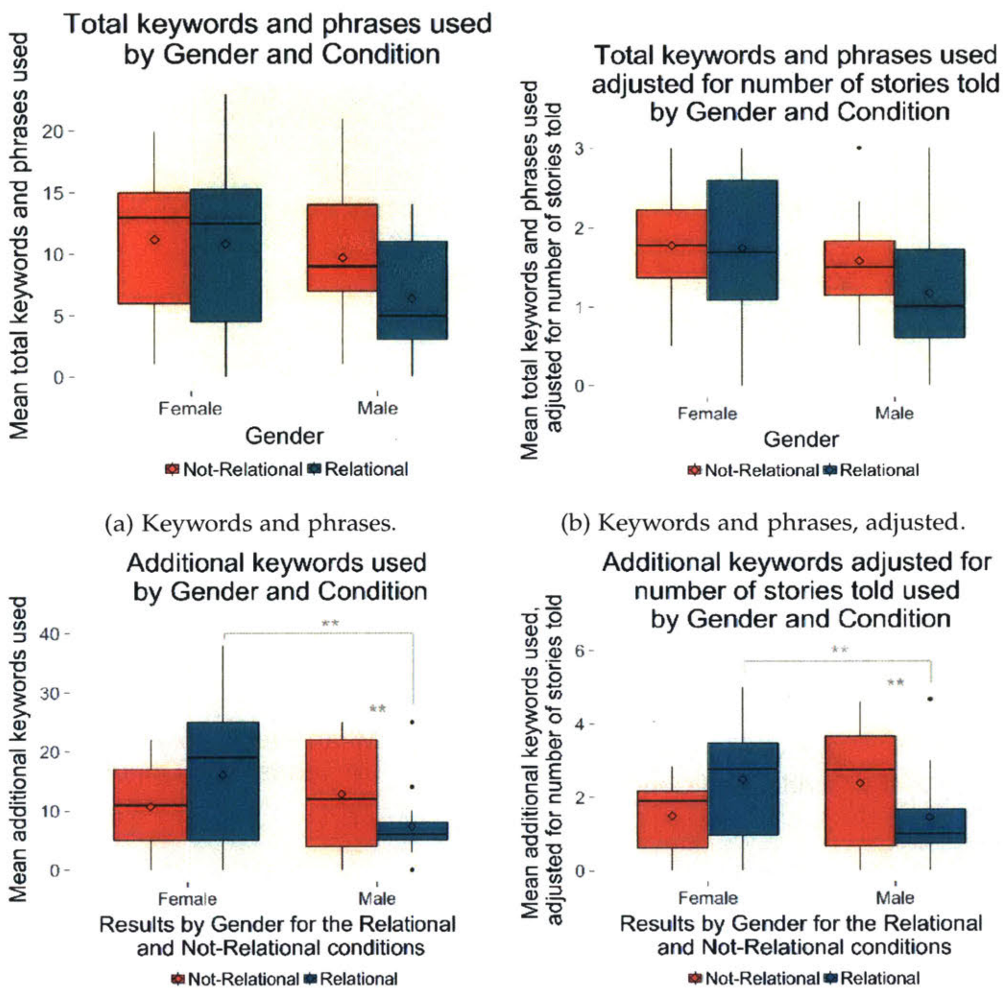
Like before, there was a significant main effect of age on the number additional keywords used, $F(3,42) = 6.06$, $p = 0.002$. There was also a significant interaction of gender with condition, $F(1,42) = 6.10$, $p = 0.018$. Girls in the *Relational* condition and boys in the *Not Relational* condition used significantly more keywords than boys in the *Relational* condition (Figure 77c). There was a similar pattern in children's use of the additional keywords adjusted for the number of stories told, with older children using more keywords than younger children, $F(3,42) = 5.88$, $p = 0.002$, and a significant interaction of gender with condition, $F(1,42) = 8.43$, $p = 0.006$ (Figure rr2-other-keywords-percent-by-gender-condition.png).

Girls in the *Relational* condition who correctly identified the target words were more likely to use them in their stories than other children (RR girls: $r_{s12} = 0.804$, $p = 0.002$; NR girls: $r_{s12} = 0.493$, $p = 0.10$; RR boys: $r_{s13} = 0.203$, $p = 0.51$; NR boys: $r_{s13} = 0.445$, $p = 0.13$) (Figure 78).

10.3.2.4 *Story Length*

LENGTH I performed analyses of variance with Condition (between: *Relational* vs. *Not Relational*), Gender (boys vs. girls), and Age as a covariate. Like before, I observed main effects of condition, $F(1,41) = 4.64$, $p = 0.037$, and of age, $F(3,41) = 9.27$, $p < 0.001$, on the mean length of children's stories (Section 9.4.2.4).

Children in the *Not Relational* Condition told longer stories than children in the *Relational* condition. Further examination suggests that this was driven by boys in the *Relational* condition ($M = 101.8$, $SD = 49.8$), who told shorter stories than children in all



(a) Keywords and phrases.

(b) Keywords and phrases, adjusted.

(c) Additional keywords.

(d) Additional keywords, adjusted.

Figure 77: Girls used more of the robot's keywords and phrases than boys. Girls in the *Relational* condition and boys in the *Not Relational* condition used more keywords than boys in the *Relational* condition.

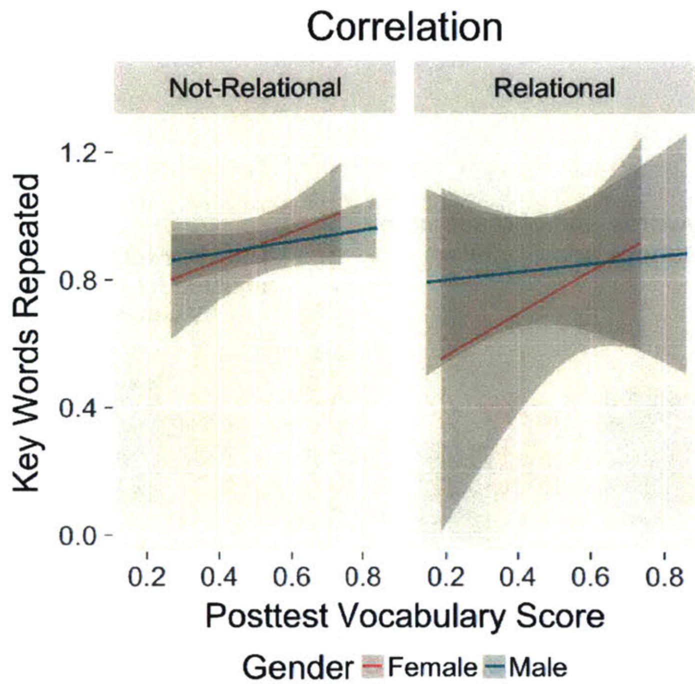
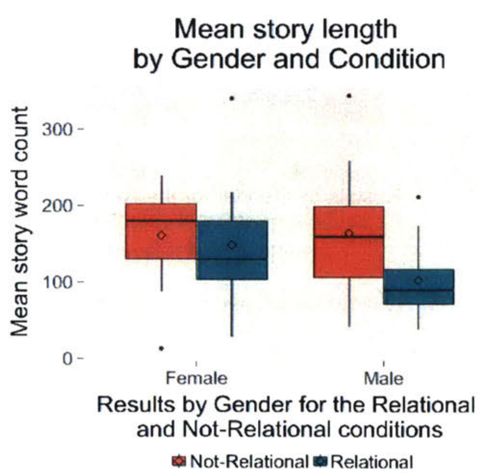
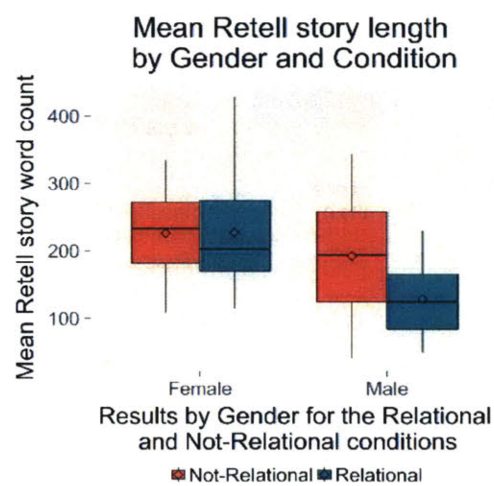


Figure 78: Children who correctly identified the target words at the posttest were more likely to use them in their stories.

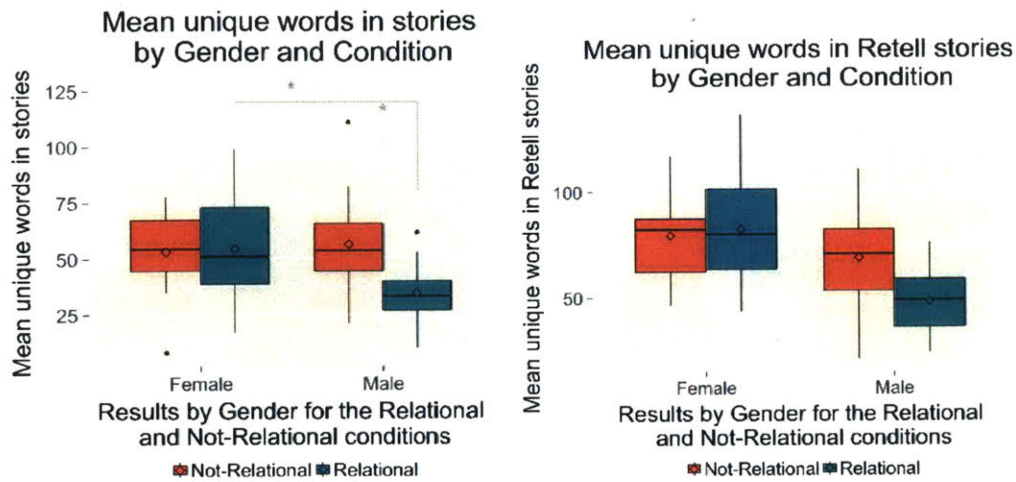


(a) All stories.



(b) Retell stories.

Figure 79: Children in the *Not Relational* Condition told longer stories than children in the *Relational* condition, likely because boys in the *Relational* condition told shorter stories than children in all other groups.



(a) All stories.

(b) Retell stories.

Figure 80: Boys in the *Relational* condition used fewer unique words in their stories than children in all other conditions.

other groups (NR boys: $M = 163.8$, $SD = 81.2$; RR girls: $M = 148.3$, $SD = 81.0$; NR girls: $M = 160.9$, $SD = 67.3$), though this pattern was not statistically significant (Figure 79a).

A similar pattern held when examining only children's Retell stories, with a main effect of age, $F(3,39) = 6.36$, $p = 0.001$, and a trend toward an interaction of gender and condition (Figure 79b). Children's Create stories were more mixed; there was a trend toward a main effect of age, but less difference between conditions or genders.

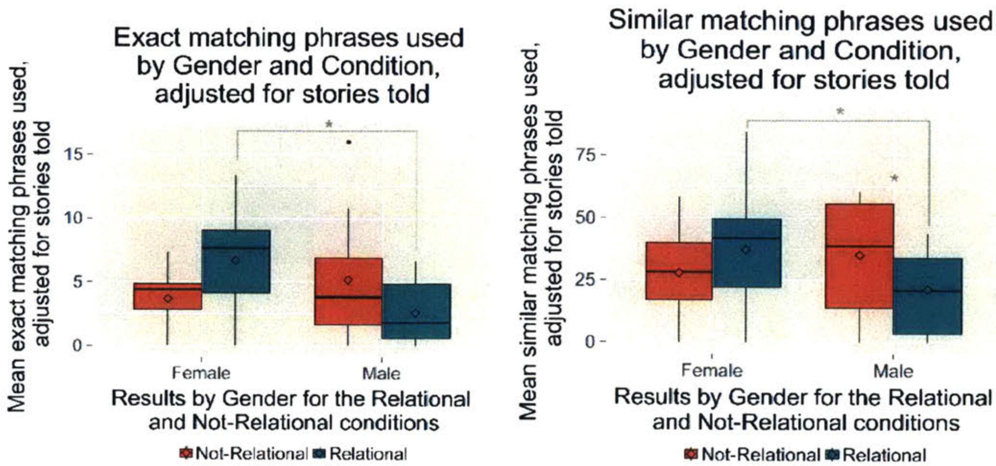
UNIQUE WORDS Regarding the number of unique words used in all stories, I observed a main effect of age, $F(3,42) = 7.23$, $p < 0.001$, and a significant interaction of condition with gender, $F(1,42) = 4.29$, $p = 0.045$. In addition to the same age pattern as before (Section 9.4.2.4), I observed that boys in the *Relational* condition used significantly fewer unique words ($M = 34.6$, $SD = 13.9$) than children in all other conditions (NR boys: M = 56.5, $SD = 24.5$; RR girls: $M = 54.5$, $SD = 23.4$; NR girls: $M = 53.2$, $SD = 20.0$) (Figure 80a).

A similar pattern held when examining unique word use in children's Retell stories, with main effects of age, $F(3,39) = 6.07$, $p = 0.002$; and of gender, $F(1,39) = 4.89$, $p = 0.033$; and a trend toward an interaction of condition with gender. Girls used more unique words than boys, and this was partly driven by boys in the *Relational* condition using far fewer unique words (Figure 80b). Children's Create stories showed no significant differences between groups.

10.3.3 *Behavior and Engagement*

10.3.3.1 *Liking the Robot's Stories*

There were no differences between groups in children's reported liking of the robot's stories.



(a) Exact phrases.

(b) Similar phrases.

Figure 81: Girls in the *Relational* condition and boys in the *Not Relational* condition emulated the robot's phrases more than boys in the *Relational* condition.

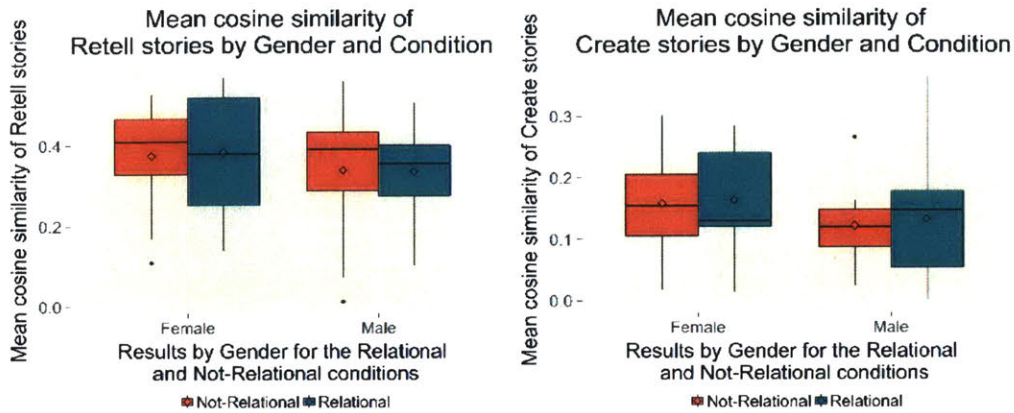
10.3.3.2 *Mirroring the Robot's Stories*

PHRASE MATCHING I performed analyses of variance with Condition (between: *Relational* vs. *Not Relational*), Gender (between: boys vs. girls), and Age as a covariate. I observed a significant interaction of condition with gender in the total number of exactly matching phrases children used, $F(1,42) = 4.35$, $p = 0.043$. Girls in the *Relational* condition matched the most phrases ($M = 47.0$, $SD = 36.9$), in particular, significantly more than boys in the *Relational* condition ($M = 16.6$, SD = 17.4), who matched the fewest phrases. Girls ($M = 27.8$, $SD = 20.2$) and boys ($M = 27.2$, $SD = 22.9$) in the *Not Relational* condition were in-between.

When examining the use of exactly matching phrases adjusted for the number of stories each child told, the same pattern emerged more strongly. Girls in the *Relational* condition used more exactly matching phrases per story (M = 6.67, $SD = 4.16$) than boys in the *Relational* condition ($M = 2.61$, $SD = 2.41$), while girls ($M = 3.69$, $SD = 2.26$) and boys ($M = 5.21$, $SD = 4.42$) in the *Not Relational* condition were in-between, $F(1,42) = 10.0$, $p = 0.003$ (Figure 81a).

Regarding children's total use of similar matching phrases, I observed a main effect of age, $F(3,42) = 3.81$, p = 0.017, like before (Section 9.4.3.2). When looking at children's total use of similar matching phrases adjusted for the number of stories each child told, there was a main effect of age, $F(3,42) = 6.71$, $p < 0.001$, and a significant interaction of condition with gender, $F(1,42) = 4.72$, $p = 0.035$. Girls in the *Relational* condition used more similar matching phrases per story ($M = 37.1$, $SD = 23.4$) than boys in the *Relational* condition ($M = 21.2$, $SD = 17.0$), as did boys in the *Not Relational* condition ($M = 35.0$, $SD = 22.6$); girls in the *Not Relational* condition were in-between ($M = 27.8$, $SD = 17.5$) (Figure 81b).

COSINE SIMILARITY When looking at the mean cosine similarity of all of each child's stories, I observed the same main effect of age as before, $F(3,42) = 7.27$, $p < 0.001$ (Section 9.4.3.2). Again, I saw the same age pattern when looking at children's Retell stories, $F(3,39) = 9.22$, $p < 0.001$; but not when looking at only children's Create



(a) Retell stories.

(b) Create stories.

Figure 82: There was a trend for boys in the *Relational* condition to tell less similar stories to the robot's.

Table 16: Children's negotiation outcomes by gender and time. Negotiation outcomes were coded as 2 = compromise, 1 = agree to robot's choice, 0 = stick to child's choice.

Time	Girls Mean (SD)	Boys Mean (SD)
S2	0.59 (0.73)	0.33 (0.48)
S3	0.68 (0.48)	0.76 (0.54)
S6	0.55 (0.51)	0.24 (0.44)
S7	0.73 (0.46)	0.38 (0.50)

stories. There were no differences by gender or condition, though there were trends for boys in the *Relational* condition to tell less similar stories (Figures 82a and 82b).

10.3.3.3 *Robot Story*

A mixed analysis of variance with Condition (between: *Relational* vs. *Not Relational*), Time (within: S3 story vs. S7 story), and Gender (between: boys vs. girls), with Age as a covariate revealed no significant differences in children's likelihood to tell a story during the Robot Story task. There was a trend for an interaction of condition with gender, $F(1,39) = 3.88$, $p = 0.056$. Girls in the *Relational* condition were somewhat more likely ($M = 1.17$, $SD = 0.64$) than girls in the *Not Relational* condition to tell a story ($M = 0.68$, $SD = 0.57$), as well as somewhat more likely than boys to tell a story (RR: $M = 0.77$, $SD = 0.75$; NR: M =0.88, $SD = 0.80$).

10.3.3.4 *Negotiation Task*

A mixed analysis of variance with Condition (between: *Relational* vs. *Not Relational*), Time (within: negotiation at S2, S3, S6, and S7), and Gender (between: boys vs. girls), with Age as a covariate, revealed a main effect of time like before (Section 9.4.3.4), $F(3,117) = 3.75$, $p = 0.013$, and a main effect of gender, $F(1,36) = 4.03$, p = 0.052. Girls were more likely to allow the robot's choice or suggest a compromise than boys. There

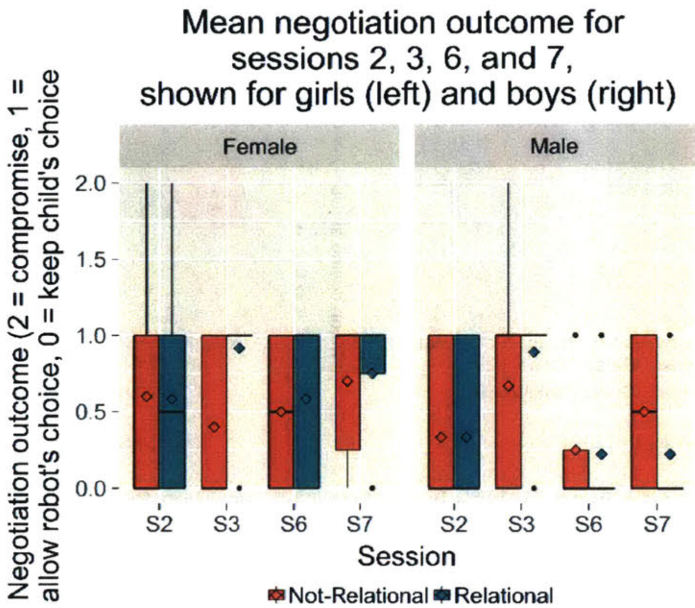


Figure 83: Children's responses during the Negotiation Task. Negotiation outcomes were coded as 2 = compromise, 1 = agree to robot's choice, 0 = stick to child's choice.

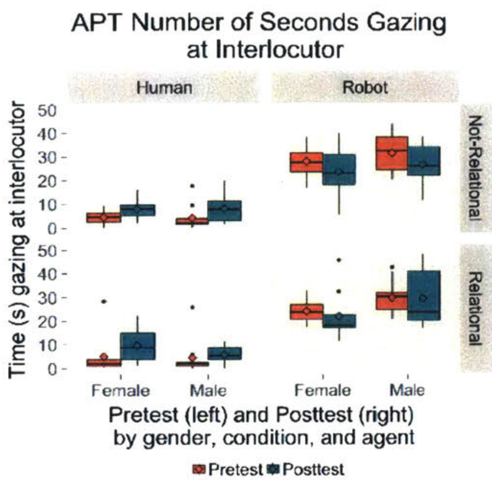
was also a trend toward an interaction of time with gender, which suggested a difference between girls' and boys' actions in S2, S6, and S7 in particular. Table 16 lists the mean negotiation outcomes by gender and time (also see Figure 83).

10.3.3.5 *Anomalous Picture Task*

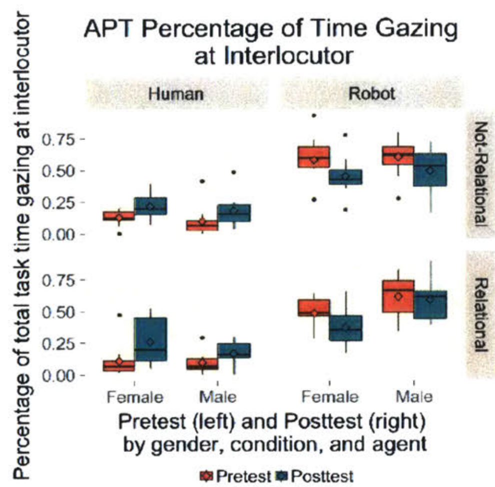
I performed a mixed analysis of variance with Condition (between: *Relational* vs. *Not Relational*), Gender (between: boys vs. girls), Time (within: pretest vs. posttest), and Agent (within: human vs. robot). I observed the same results described in Section 9.4.3.5, with a main effect of agent, $F(1,80) = 17.1$, $p < 0.001$; and a significant interaction of condition with time, $F(1, 42) = 4.75$, $p = 0.035$. There were no differences by gender. When examining children's comments, questions, and laughter individually, I observed the same patterns as before, and again, no differences by gender.

Next, I examined children's gaze during the task. I performed analyses of variance with Condition (between: *Relational* vs. *Not Relational*), Time (within: pretest vs. posttest), Gender (between: boys vs. girls), and Agent (within: human vs. robot) for both the absolute looking times at the pictures, at their interlocutor, and elsewhere during the task and the time as a percent. I observed the same patterns discussed in Section 9.4.3.5, with several additional interactions of gender with agent. Tables 17 and 18 list children's absolute and percent gaze times by gender, agent, time, and condition.

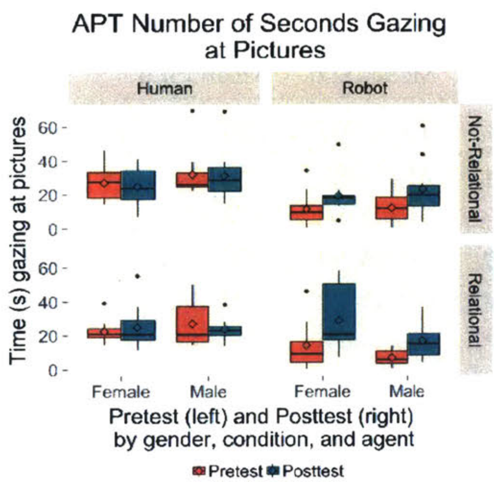
With regards to absolute looking time at their interlocutor, I observed a significant main effect of Agent, $F(1,66) = 296.1$, $p < 0.001$; significant interactions of gender with agent, $F(1,66) = 6.61$, $p = 0.012$; and of time with agent, $F(1,66) = 5.80$, $p = 0.019$. There was also a trends toward an interaction of condition with gender and agent (p = 0.09). Post-hoc tests revealed that girls and boys looked at the human experimenter for approximately equal amounts of time at the pretest and posttest, and both girls



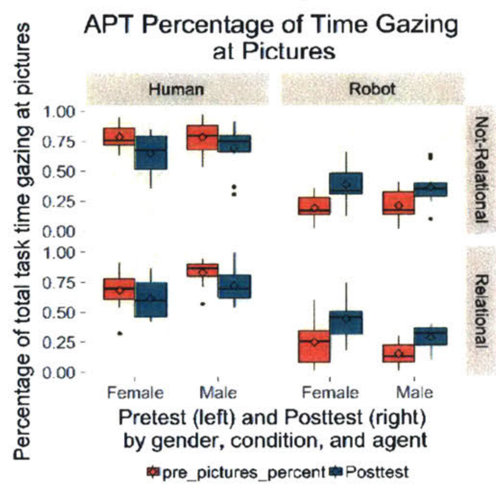
(a) Interlocutor.



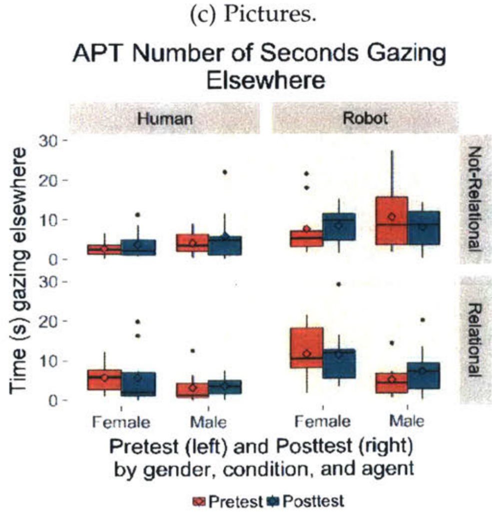
(b) Interlocutor, percent.



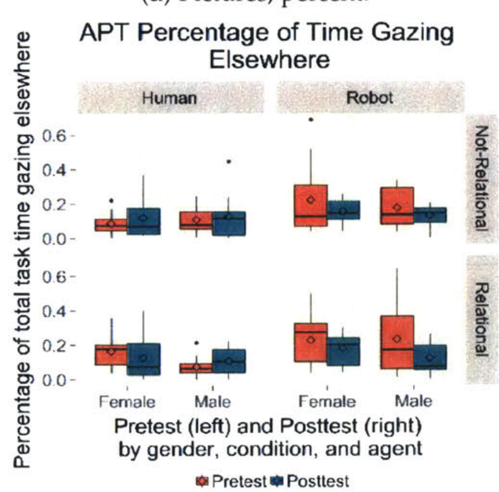
(c) Pictures.



(d) Pictures, percent.



(e) Elsewhere.



(f) Elsewhere, percent.

Figure 84: Children's Anomalous Picture Task gaze and gaze as a percent of the total time spent on task.

Table 17: The total amount of time in seconds that children spent looking at the pictures, at their interlocutor (robot or human experimenter), or elsewhere during each iteration of the Anomalous Picture Task (APT). Here, "Con." = "Condition".

Agent	Time	Gender	Con.	Gaze at Pictures Mean (SD)	Interlocutor Mean (SD)	Elsewhere Mean (SD)
Human	Pre	Girls	RR	24.9 (13.2)	9.81 (6.88)	5.67 (6.99)
			NR	25.0 (11.4)	7.82 (3.85)	3.63 (3.50)
		Boys	RR	23.9 (6.52)	5.87 (3.82)	3.64 (2.64)
			NR	31.3 (15.9)	7.98 (5.81)	5.83 (6.67)
	Post	Girls	RR	22.7 (6.92)	4.96 (8.45)	5.67 (3.51)
			NR	27.2 (10.2)	4.51 (2.85)	2.63 (1.90)
		Boys	RR	27.2 (13.8)	4.46 (7.72)	3.12 (3.90)
			NR	32.1 (14.3)	4.11 (5.46)	4.07 (2.90)
Robot	Pre	Girls	RR	28.9 (19.0)	10.9 (18.2)	11.3 (8.06)
			NR	19.5 (11.1)	23.4 (10.4)	8.30 (4.45)
		Boys	RR	16.9 (10.5)	34.4 (19.6)	7.10 (5.85)
			NR	23.2 (17.0)	26.7 (8.31)	7.95 (5.01)
	Post	Girls	RR	14.4 (14.8)	24.2 (4.63)	11.6 (7.02)
			NR	11.6 (9.81)	31.7 (14.2)	15.4 (18.8)
		Boys	RR	7.29 (4.58)	29.8 (7.05)	14.2 (16.3)
			NR	12.2 (9.24)	31.4 (8.15)	10.6 (8.98)

Table 18: The percentage of time children spent looking at the pictures, at their interlocutor (robot or human experimenter), or elsewhere during each iteration of the Anomalous Picture Task. Values are expressed as percentages of the total time spent on the task, since children did not spend exactly the same amount of time doing the task (e.g., because some children spoke more than others). Here, "Gen." = "Gender" and "Con." = "Condition".

Agent	Time	Gen.	Con.	Percent Gaze at Pictures Mean (SD)	Interlocutor Mean (SD)	Elsewhere Mean (SD)
Human	Pre	Girls	RR	61.4 (17.7)	26.1 (19.0)	12.6 (13.5)
			NR	65.1 (17.5)	21.6 (10.1)	11.9 (12.7)
		Boys	RR	72.3 (15.8)	16.8 (9.74)	10.9 (8.04)
			NR	69.2 (20.5)	17.9 (13.3)	13.0 (13.8)
	Post	Girls	RR	68.3 (17.7)	11.0 (12.6)	16.6 (9.93)
			NR	78.6 (9.93)	12.8 (6.13)	8.62 (6.53)
		Boys	RR	83.0 (11.6)	9.44 (8.23)	7.58 (6.41)
			NR	78.4 (13.8)	9.66 (12.2)	10.1 (7.64)
Robot	Pre	Girls	RR	44.1 (16.7)	37.5 (15.2)	18.4 (9.38)
			NR	38.7 (15.4)	45.4 (14.9)	15.9 (6.97)
		Boys	RR	28.5 (10.3)	59.0 (16.8)	12.5 (9.01)
			NR	36.3 (16.0)	49.6 (18.5)	13.2 (6.54)
	Post	Girls	RR	24.7 (19.9)	48.6 (12.3)	23.2 (15.1)
			NR	19.1 (11.1)	58.3 (19.6)	22.5 (21.4)
		Boys	RR	15.1 (9.55)	61.4 (17.7)	23.5 (23.5)
			NR	20.9 (14.3)	60.6 (15.8)	17.8 (11.8)

and boys looked at the robot more than the human at all tests, but girls also tended to look at the robot less than boys did (Figure 84a). This last pattern seemed to be driven by girls in the *Relational* condition, who tended to look at the robot less than children in all other conditions.

The same pattern held when looking at the percent of the total time on the task that children spent looking at their interlocutor, with a significant main effect of agent, $F(1,66) = 268.3$, $p < 0.001$; a significant interaction gender with agent, $F(1,66) = 10.1$, $p = 0.002$; and a significant interaction of time with agent, $F(1,66) = 17.7$, $p < 0.001$. Again, there was a trend for an interaction of condition with gender and agent ($p = 0.08$). Girls and boys gazed at the robot more than at the human experimenter. They did not differ in how much time they looked at the human, and girls spent significantly less time looking at the robot than boys did. When examining the trend involving condition, it appeared that children did not differ greatly when the human was their interlocutor. Girls in the *Relational* condition seemed to look at the robot less than all other children, while boys in the *Relational* condition looked longest (Figure 84b).

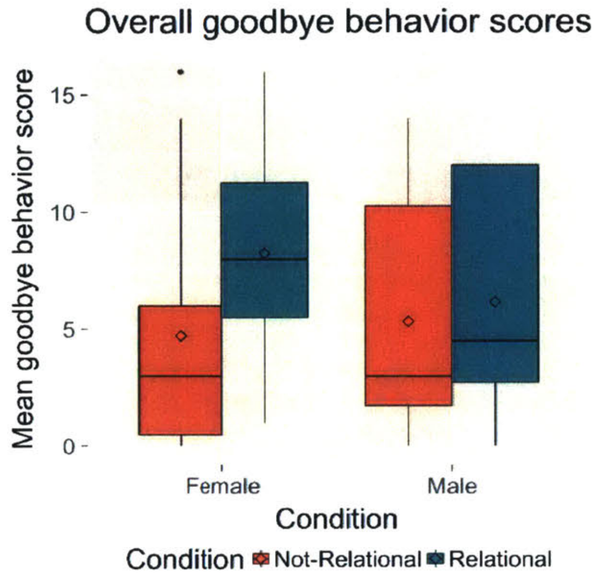
With regards to the absolute time children spent looking at the pictures during the task, I observed significant main effects of time, $F(1,41) = 7.30$, $p = 0.010$; and of agent, $F(1,66) = 32.6$, $p < 0.001$; as well as a significant interaction of time with agent, $F(1,66) = 10.8$, $p = 0.002$. There was also a strong trend for a significant interaction of gender with agent, $F(1,66) = 3.93$, $p = 0.052$. Post-hoc tests showed that boys appeared to have a greater change in their behavior with the robot, with a greater decrease in the time they spent looking at the pictures with the robot than with the human compared to girls. This seemed to be driven primarily by boys in the *Relational* condition. Girls in the *Relational* condition, in contrast, spent slightly more time looking at the pictures (Figure 84c).

Similar patterns appeared when looking at the percent of the total time that children spent looking at the pictures. I observed a significant main effect of agent, $F(1,66) = 348.7$, $p < 0.001$; a significant interaction of gender with agent, $F(1,66) = 8.27$, $p = 0.005$; a significant interaction of time with agent, $F(1,66) = 33.3$, $p < 0.001$; and a significant interaction of condition and gender with agent, $F(1,66) = 5.98$, $p = 0.017$. Post-hoc tests showed that with the robot, boys in the *Relational* condition spent less time looking at the pictures while girls in the *Relational* condition spent more time (Figure 84d).

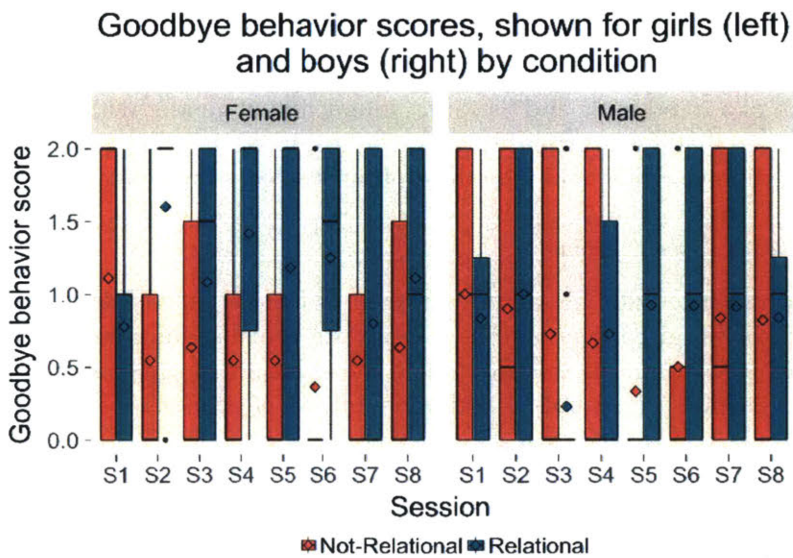
Finally, with regards to the absolute time children spent looking elsewhere during the task, like before, I observed a main effect of agent, $F(1,66) = 30.0$, $p < 0.001$; and a significant interaction of time with agent, $F(1,66) = 5.15$, $p = 0.027$ (Figure 84e). I observed the same pattern with regards to the percent of time children spent looking elsewhere, with a main effect of agent, $F(1,66) = 14.7$, $p < 0.001$; and a significant interaction of time with agent, $F(1,66) = 5.55$, $p = 0.022$ (Figure 84f).

10.3.3.6 *Goodbye Behaviors*

I performed a 2×2 analysis of variance with Condition (*Relational* vs. *Not Relational*), Gender (boys vs. girls), and Age as a covariate on the overall Goodbye score. There were no statistically significant effects; however, there was the same trend as before (Section 9.4.3.6) for children in the *Relational* condition to have higher goodbye scores ($M = 7.21$, $SD = 4.58$, *median* = 6) than children in the *Not Relational* condition ($M = 5.04$, $SD = 5.19$, *median* = 3), and this appeared to be driven by girls more than boys (Figure 85a). The same trend was present when examining the scores summed across the first half versus the second half of the sessions.



(a) All sessions.



(b) Each session.

Figure 85: Children's scored goodbye behaviors summed across all sessions and shown for each session by condition and gender. For each session, behaviors were scored as 2 = clear goodbye action (e.g., waving, saying "bye bye"), 1 = small action (e.g., smile, small wave), 0 = no goodbye.

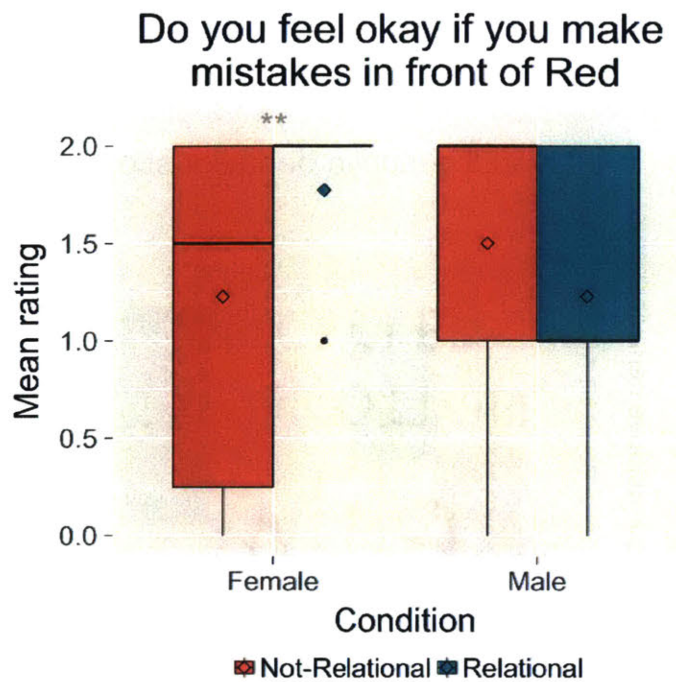


Figure 86: Children's responses to the question, "Do you feel okay if you make mistakes in front of Red?" by condition and gender.

When looking at individual sessions, I observed the same patterns as before, with the addition of a trend in S2 and S3 for an interaction of gender with condition. This trend suggested that it was primarily girls who displayed different goodbye behaviors between conditions, while boys in both conditions acted more similarly (Figure 85b).

10.3.3.7 *Robot's Favorite Animal*

Analyses of variance with Condition (between: *Relational* vs. *Not Relational*), Gender (between: boys vs. girls), and Age as a covariate revealed the same main effect of age on whether children correctly identified the kinkajou, $F(3,40) = 4.32$, $p = 0.010$, as described in Section 9.4.3.7, but no additional impact of gender.

10.3.3.8 *Judgment and Safe Space Questions*

I performed mixed analyses of variance with Condition (between: *Relational* vs. *Not Relational*), Time (within: pre-S3 test vs. post-S7 test), Gender (between: boys vs. girls) and Age as a covariate. I observed a significant interaction of condition with gender for the question "Do you feel okay if you make mistakes in front of Red," $F(1,38) = 7.22$, $p = 0.011$ (Figure 86). Girls were more likely to say they felt okay in the *Relational* condition ($M = 1.77$, $SD = 0.43$, $median = 2$) than in the *Not Relational* condition ($M = 1.23$, $SD = 0.87$, $median = 1.5$). Overall responses were mixed, but generally more positive, at both the S3 test and the S7 test.

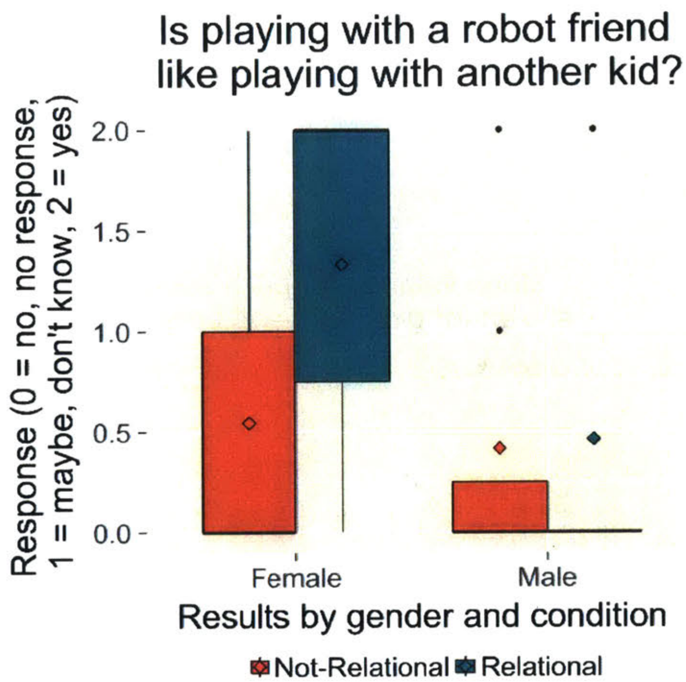


Figure 87: Children's tendencies to say playing with a robot friend was like playing with another kid by gender and condition.

10.3.3.9 *Empathy/Helping Tasks*

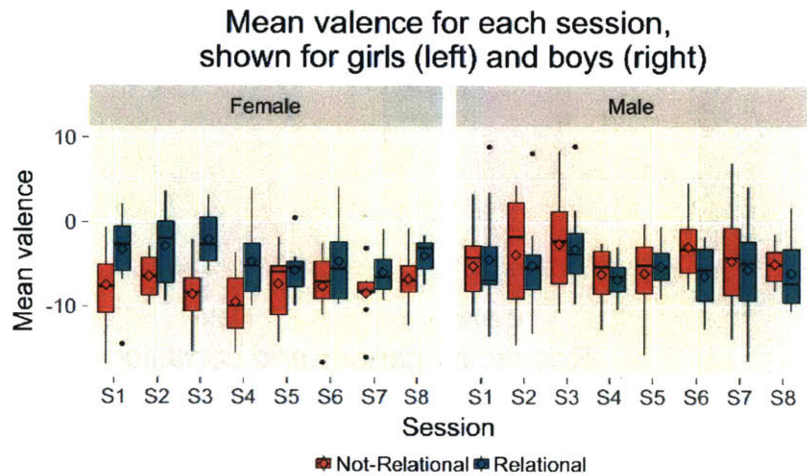
In addition to the patterns found previously (Section 9.4.3.9), there were trends for girls to be slightly more likely to agree to take a photo with the robot than boys, and for girls to be more likely to say the robot was like another child than boys were (Figure 87).

10.3.3.10 *Affect*

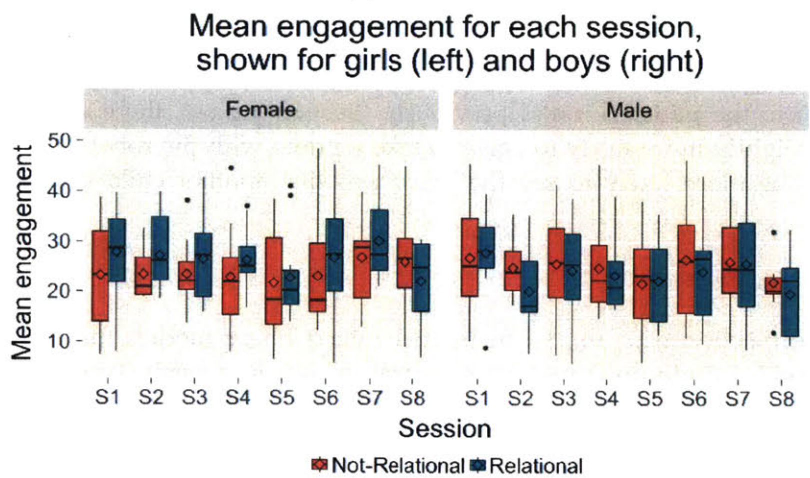
Like before (Section 9.4.3.10), I constructed mixed linear models for each emotion measured with Condition (between: *Relational* vs. *Not Relational*), Time (within: each session), Gender (between: boys vs. girls) and Age as a covariate. I observed numerous interactions of gender with condition and time.

I observed that children's valence significantly varied by time, $F(7,265) = 2.07$, $p = 0.047$; there were also significant interactions of condition with gender, $F(1,42) = 6.11$, $p = 0.018$; and condition and gender with time, $F(7,265) = 2.22$, $p = 0.033$. Girls in the *Not Relational* condition showed significantly lower valence than girls in the *Relational* condition and boys in the *Not Relational* condition, but in some sessions more than others (Figure 88a).

I observed a significant main effect of time on children's smiling, $F(7,265) = 8.45$, $p < 0.001$; and also a significant interaction of time with condition and gender, $F(7,265) = 2.59$, $p = 0.013$. Children smiled more in Sessions 1, 2, and 3 than in Sessions 4, 5, 6, and 8, and this varied by condition and gender (Figure 89a). For example, girls in the *Not Relational* condition smiled less in Sessions 1, 2, and 3 than other children. I observed the same pattern with regards to children's joy: there was a significant main

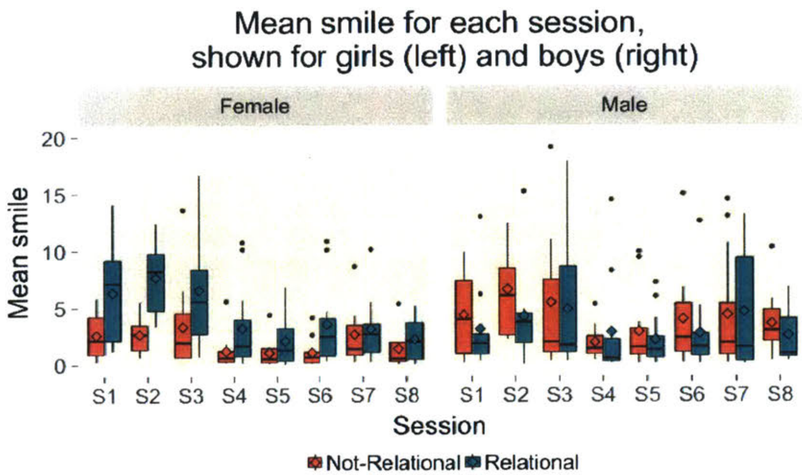


(a) Valence.

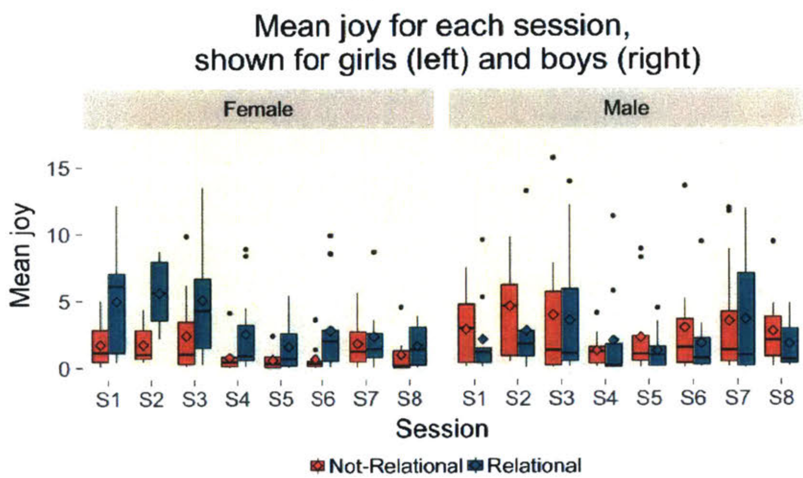


(b) Engagement.

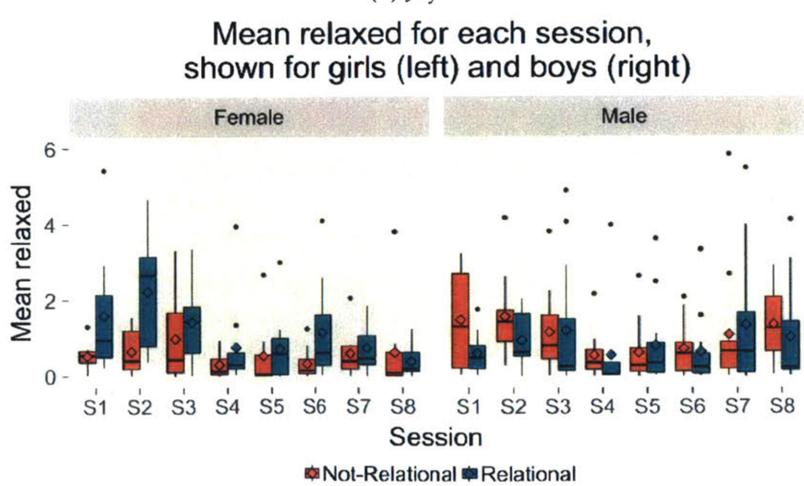
Figure 88: Children's overall affect for each session by condition and gender.Values can range from 0 (no expression present) to 100 (expression fully present), except valence, which can range from -100 to 100.



(a) Smiles.

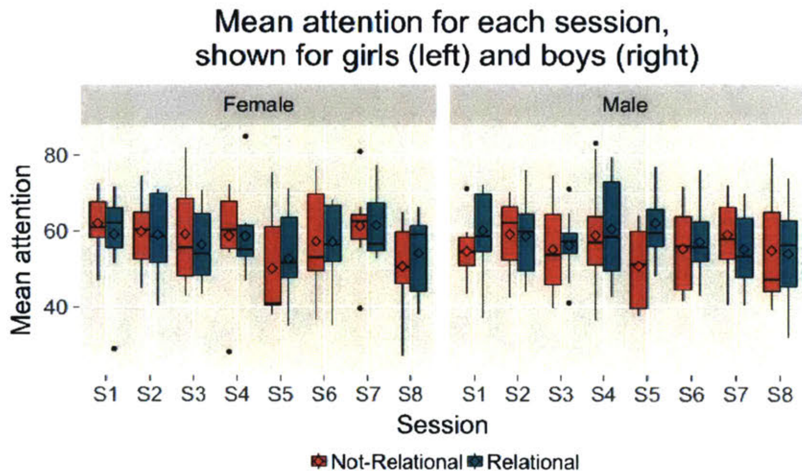


(b) Joy.

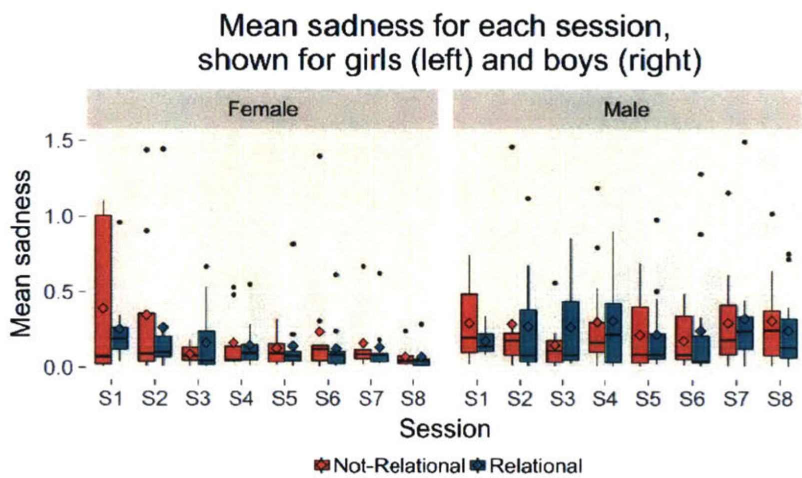


(c) Relaxation.

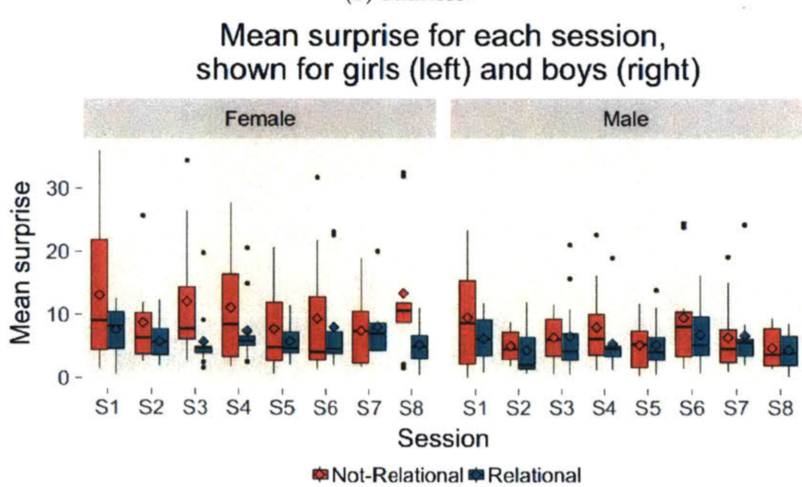
Figure 89: Children's overall affect for each session by condition and gender. Values can range from 0 (no expression present) to 100 (expression fully present), except valence, which can range from -100 to 100.



(a) Attention.

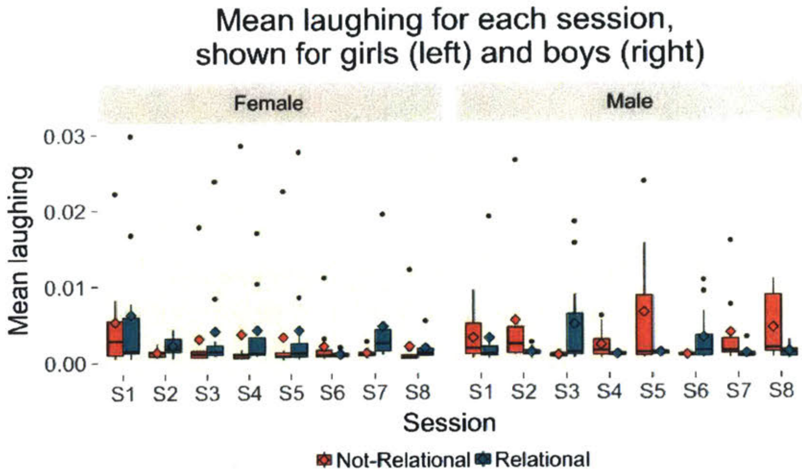


(b) Sadness.

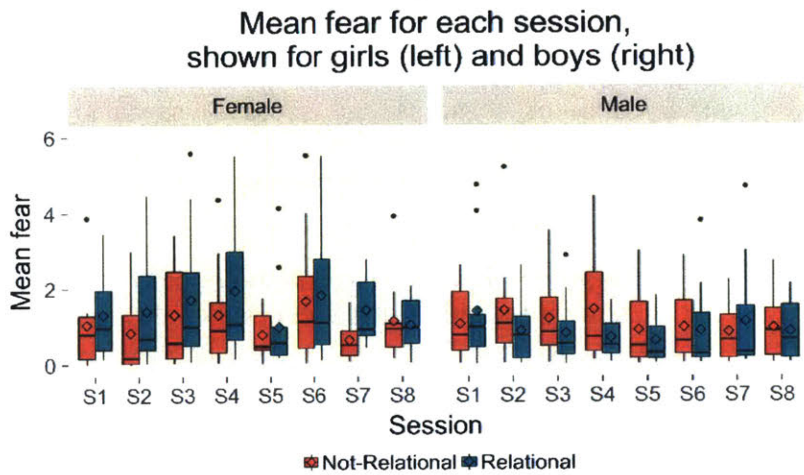


(c) Surprise.

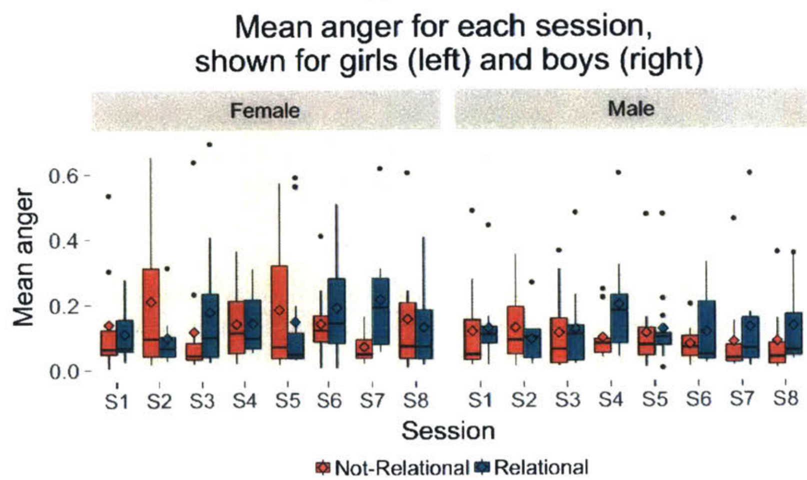
Figure 90: Children's overall affect for each session by condition and gender.Values can range from 0 (no expression present) to 100 (expression fully present), except valence, which can range from -100 to 100.



(a) Laughter.

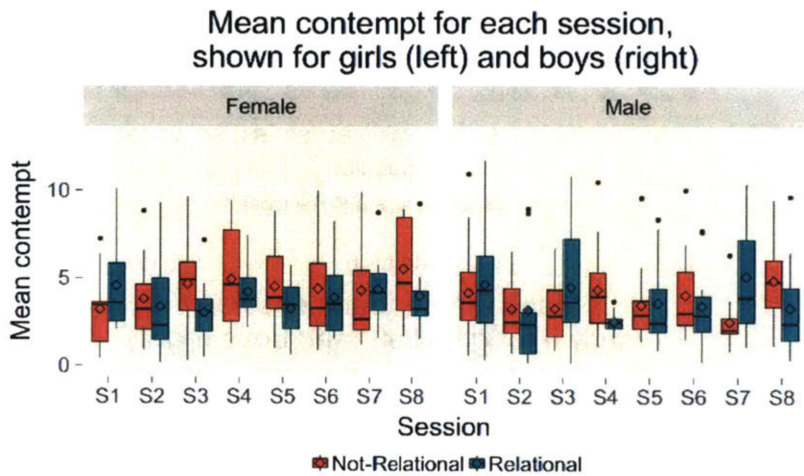


(b) Fear.

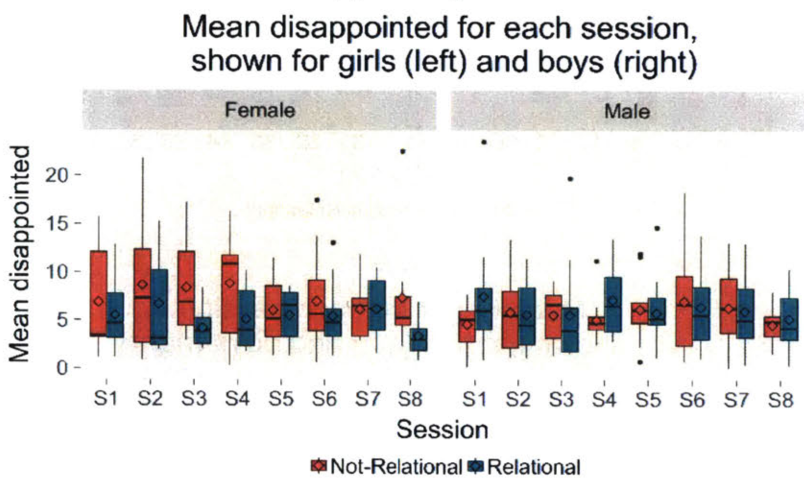


(c) Anger.

Figure 91: Children's overall affect for each session by condition and gender.Values can range from 0 (no expression present) to 100 (expression fully present), except valence, which can range from -100 to 100.

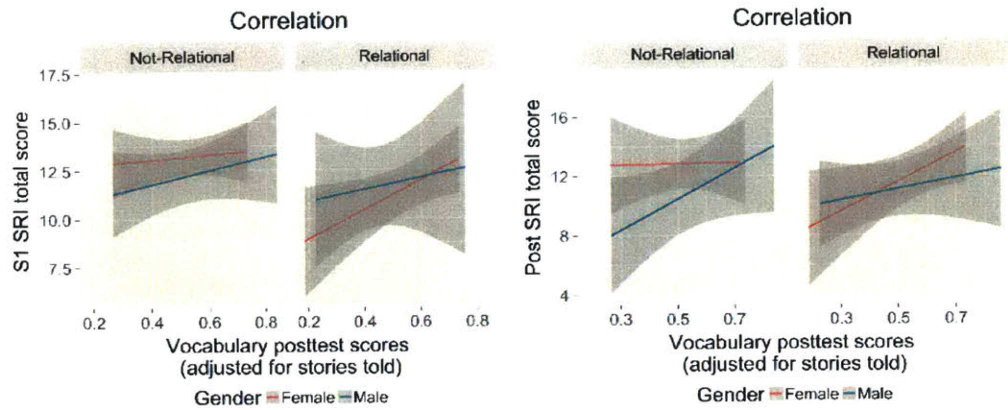


(a) Contempt.



(b) Disappointed.

Figure 92: Children's overall affect for each session by condition and gender. Values can range from 0 (no expression present) to 100 (expression fully present), except valence, which can range from -100 to 100.



(a) S1 test.

(b) Posttest.

Figure 93: Children's Social Relational Interview scores correlated with their vocabulary scores more strongly for girls in the *Relational* condition and boys in the *Not Relational* condition.

effect of time on children's joy, $F(7,265) = 6.28$, $p < 0.001$; and a significant interaction of time with condition and gender, $F(7,265) = 2.08$, $p = 0.046$ (Figure 89b). I observed a significant main effect of time on children's relaxation, $F(7,265) = 4.26$, $p < 0.001$; as well as a trend toward an interaction of time with condition and gender (Figure 89c). Girls tended to be more relaxed in the *Relational* condition than in the *Not Relational* condition. This follows the same pattern as children's joy and smiling.

Like before, I observed a significant main effect of time on children's engagement, $F(7,265) = 3.35$, $p = 0.002$. (Figure 88b). I observed a significant main effect of time on children's sadness, $F(7,265) = 3.39$, $p = 0.002$; as well as a significant interaction of time with condition and gender, $F(7,265) = 2.75$, $p = 0.009$ (Figure 90b). Children showed more sadness in Session 1 than in all other sessions, and this was driven primarily by girls in the *Not Relational* condition and boys in the *Relational* condition.

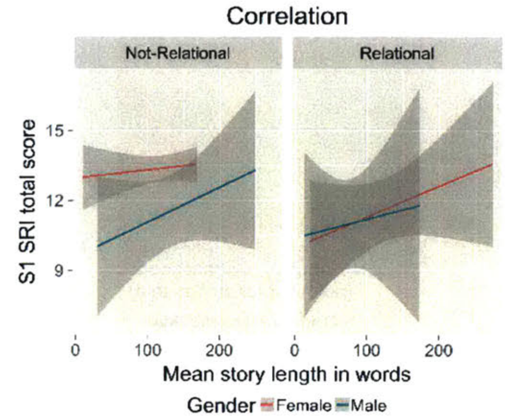
Like before, children's attention significantly varied by time, $F(7,266) = 3.44$, $p = 0.002$ (Figure 90a). Children's expressions of surprise varied by time, like before, $F(7,264) = 2.83$, $p = 0.007$ (Figure 90c). Again, there were no significant differences in children's expressions of laughter (Figure 91a), fear (Figure 91b), anger (Figure 91c), contempt (Figure 92a), or disappointment (Figure 92b).

10.3.3.11 *Relationship, Mirroring, and Learning*

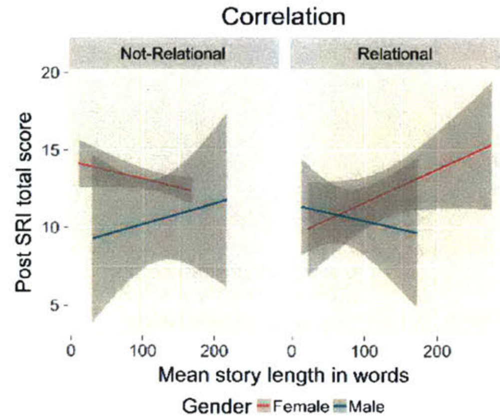
Many of the correlations between key relational measures used and various learning and mirroring behaviors displayed by children differed by gender. There were often stronger correlations for girls in the *Relational* condition and boys in the *Not Relational* condition, as shown in Table 19.

Girls in the *Relational* condition and boys in the *Not Relational* condition showed stronger correlations between their S1 S (SRI) scores and their vocabulary scores. They also showed stronger correlations with their S8 SRI scores (Figure 94). A similar pattern held for children's S1 and posttest SRI scores with their emulation of the robot during storytelling, their mean story length, and their use of keywords.

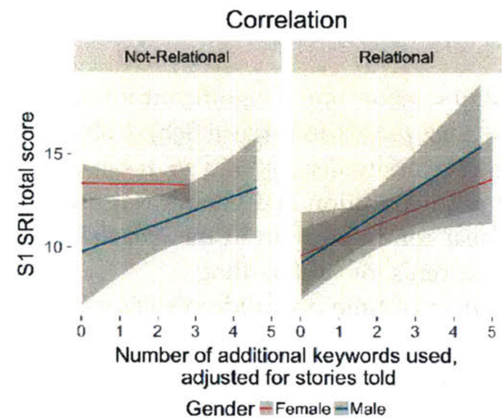
Girls in the *Relational* condition who placed the Tega robot near the human in the Picture Sorting Task were more likely than other children to also treat the robot socially



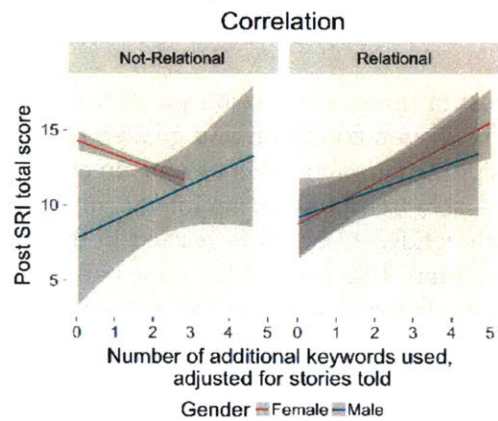
(a) S1 test.



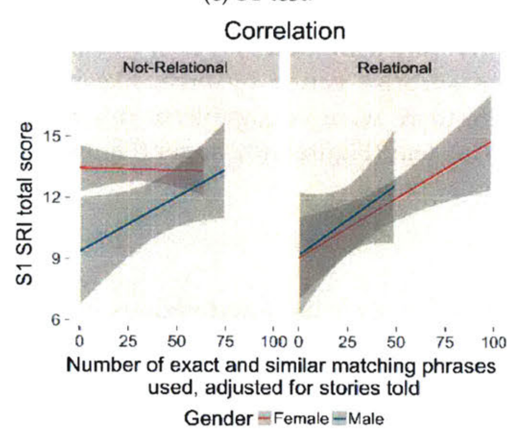
(b) Posttest.



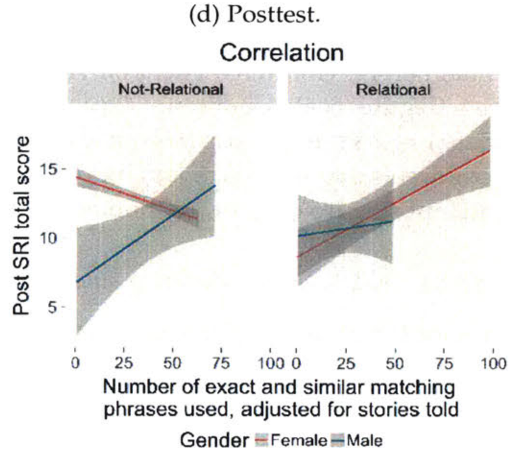
(c) S1 test.



(d) Posttest.



(e) S1 test.

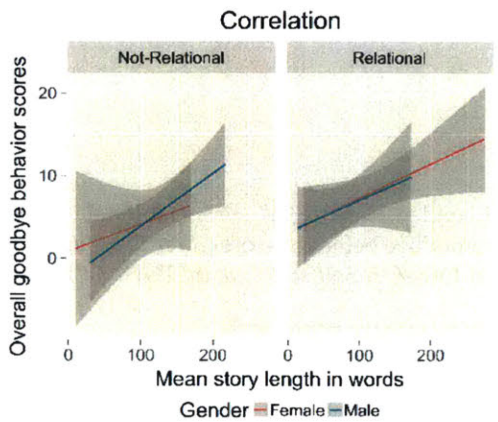


(f) Posttest.

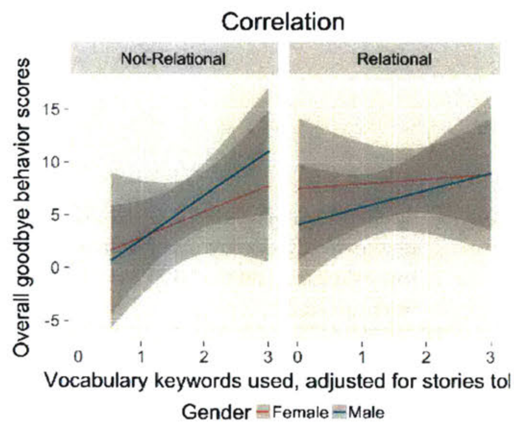
Figure 94: Correlations between children's Social Relational Interview (SRI) scores and their mean story length, use of keywords, and use of exact and similar matching phrases were generally stronger for girls in the *Relational* condition and boys in the *Not Relational* condition.

Table 19: Spearman correlations by condition and gender. Here, "SRI" = "Social Relational Interview", "Goodbye" refers to children's goodbye behavior scores, "APT" = "Anomalous Picture Task", "IOS" = "Inclusion of Other in Self task", and "PST" = "Picture Sorting Task"

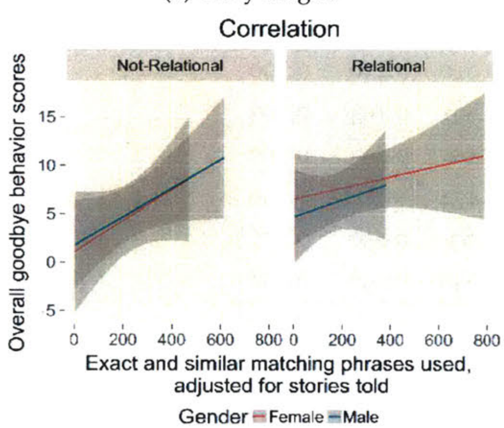
Measures		Girls		Boys	
		RR	NR	RR	NR
SRI S1	Vocabulary	0.573	0.185	0.405	0.424
SRI Post	Vocabulary	0.631	0.123	0.272	0.548
SRI S1	Phrase matching	0.743	-0.010	0.416	0.390
SRI Post	Phrase matching	0.729	-0.891	0.194	0.593
SRI S1	Story length	0.592	0.362	0.193	0.269
SRI Post	Story length	0.628	-0.474	-0.221	0.150
SRI S1	Keyword use	0.521	0.083	0.850	0.177
SRI Post	Keyword use	0.592	-0.846	0.390	0.488
Goodbye	Phrase matching	0.148	0.740	0.372	0.635
Goodbye	Story length	0.547	0.276	0.472	0.783
Goodbye	Keyword use	0.137	0.677	-0.259	0.401
Goodbye	PST S2	-0.549	0.139	0.066	0.576
Goodbye	PST Post	-0.472	-0.442	-0.469	-0.311
APT S1	Phrase matching	0.252	0.170	0.224	0.522
APT S8	Phrase matching	0.163	0.396	-0.132	0.039
APT S1	Story length	0.691	-0.283	0.295	0.490
APT S8	Story length	0.123	-0.086	0.122	0.233
APT S1	Keyword use	-0.077	-0.202	0.817	0.502
APT S8	Keyword use	0.315	0.010	0.656	0.371
IOS S1	Stories told	0.649	0.343	0.220	0.309
Negotiation	Vocabulary	0.116	0.185	0.688	0.318



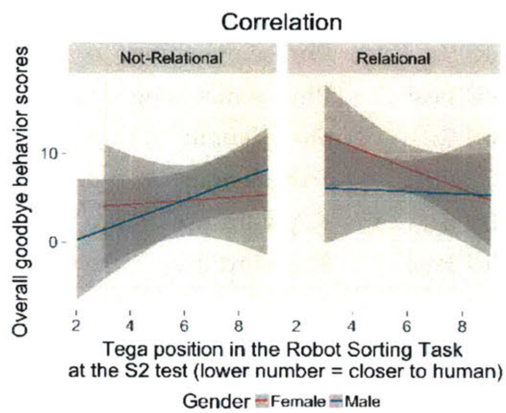
(a) Story length.



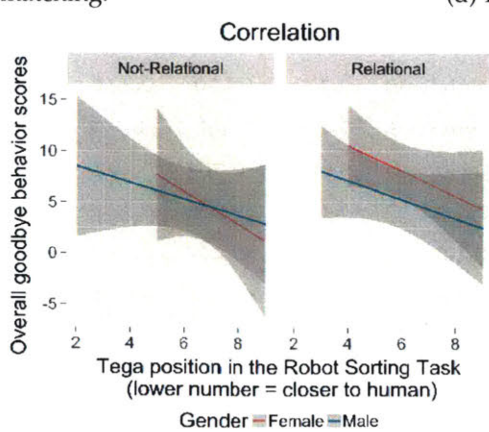
(b) Keywords.



(c) Phrase matching.

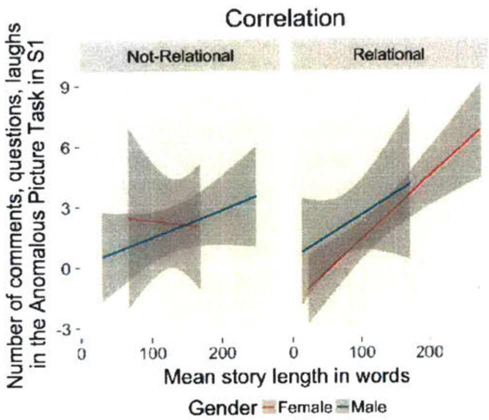


(d) PST S2.

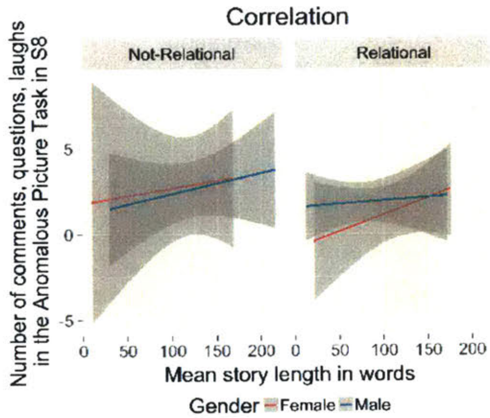


(e) PST Posttest.

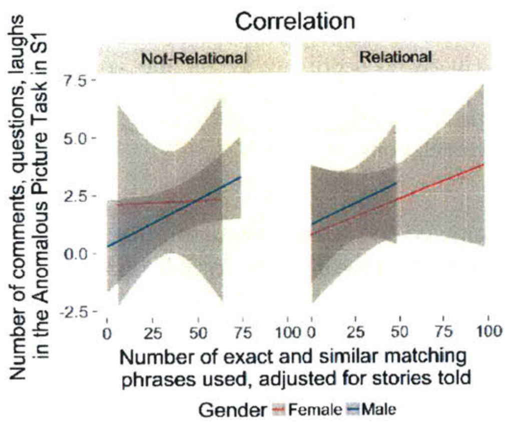
Figure 95: Children who treated the robot more socially when saying goodbye told longer stories, used more keywords, emulated the robot's phrases more during storytelling, and placed the Tega robot closer to the human in the Picture Sorting Task (PST).



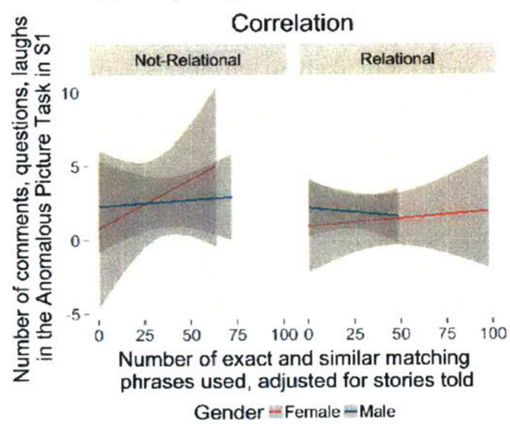
(a) Story length with S1 APT.



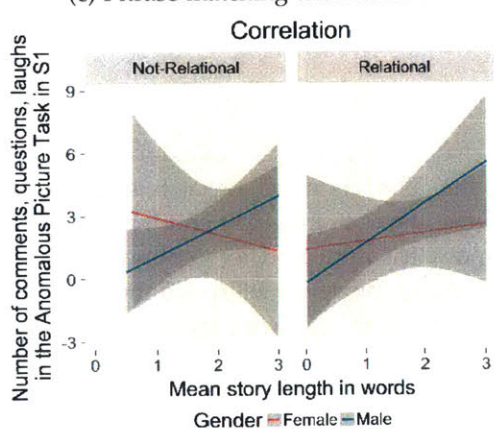
(b) Story length with S8 APT.



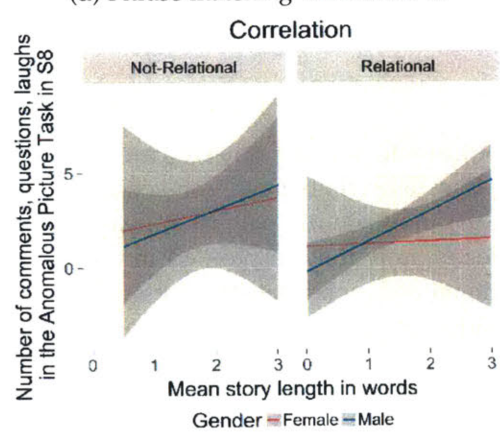
(c) Phrase matching with S1 APT.



(d) Phrase matching with S8 APT.

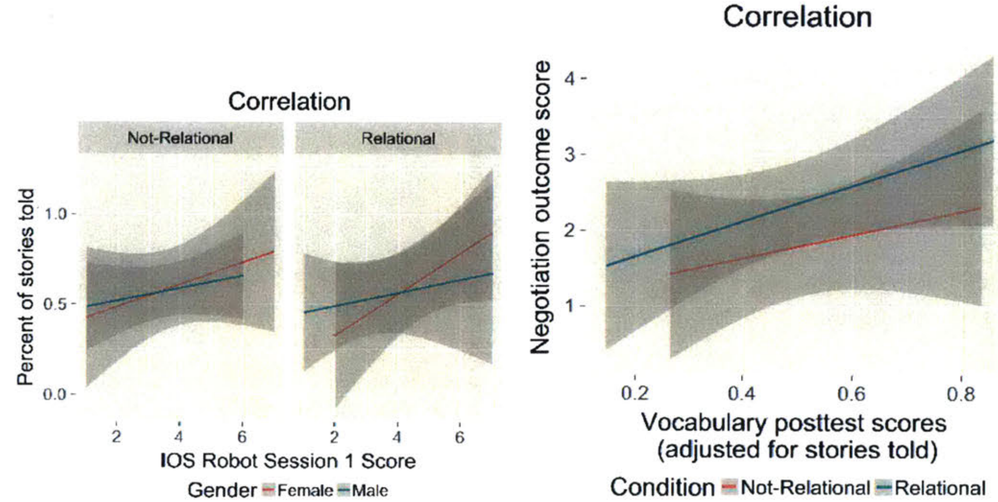


(e) Keyword use with S1 APT.



(f) Keyword use with S8 APT.

Figure 96: There were differences by condition in the correlation between children's social behavior during the Anomalous Picture Task (APT) and children's story length, phrase matching, and use of keywords.



(a) Children who rated the robot as closer in S1 on the Inclusion of Other in Self task were more likely to tell stories to the robot.

(b) Children who were more likely to seek equitable negotiation outcomes were more likely to score highly on the vocabulary posttest.

Figure 97: Other correlations.

when saying goodbye (Figure 95). They also showed the strongest correlation between their closeness ratings of the robot in S1 on the Inclusion of Other in Self task and the number of stories they told (Figure 97a).

Boys in the *Relational* condition showed the strongest correlations between their social behavior during the Anomalous Picture Task with the robot and their emulation of the robot, story length, and use of keywords (Figures 96). They also were more likely than other children to show a relation between their negotiation behavior and vocabulary scores (Figure 97b).

10.4 DISCUSSION

Contrary to my hypotheses, the differences I observed between genders were not as clear as girls being more social or rating the robot as more social than boys. Instead, an intriguing pattern appeared throughout the data in which children's gender affected how they responded in each condition. Girls appeared to respond to the relational and non-relational robots as expected; boys, however, followed the opposite pattern. Boys responded more positively to the robot in the *Not Relational* condition than in the *Relational* condition. Although this gender-condition interaction pattern sometimes appeared as a trend without statistical significance (likely due to the sample size), it appeared frequently enough to be of significant interest regardless. Below, I discuss the main findings with respect to this overarching pattern and then discuss the implications of these findings.

10.4.1 *Relationships and Behavior*

Children's ratings of their closeness to the robot, their perception of the robot as a social, relational other, and their actions reflecting relationship and rapport were frequently affected by children's gender. Thus, although the robot's relational behaviors appeared to have some of the expected effects toward increased feelings of closeness, rapport, and relationship, as discussed in Chapter 9, nearly all effects were moderated by children's gender.

Girls in the *Relational* condition, and frequently, boys in the *Not Relational* condition, tended to rate the robot as a greater social-relational agent, liked the robot more, felt closer to it, and showed more positive affect during the sessions. They often disclosed more information, and emulated more of the robot's language, and reported feeling more okay making mistakes in front of the robot. Boys in the *Relational* condition, on the other hand, showed the opposite pattern. Their ratings of the robot's social-relational nature decreased over time. They liked the robot less, felt less close to it, and showed less positive affect during the sessions. They used fewer words, disclosed less, and mirrored the robot the least. Girls in the *Not Relational* condition often followed this same pattern, though not to as great an extent.

The robot's relational behaviors appeared to affect children's social behavior with the robot. For example, during the Anomalous Picture Task, girls in the *Relational* condition looked at the robot less than other children did (and looked more at the pictures) and boys in the *Relational* condition looked longest at the robot (with less time looking at the pictures). Since children looked at the human less than they looked at the robot, and looked at the pictures longer with the human, this suggests these girls were responding to the robot more similarly to how they respond to a human, while the boys were responding differently. A similar pattern was seen in prior work with regards to a robot that was introduced as a social agent versus as a machine (Kory-Westlund et al., 2016a).

Children's affect differed most prominently in the earlier sessions, though the gender-condition interaction patterns were apparent throughout. Girls' affect with the relational robot and boys' affect with the non-relational robot were as expected, with these children showing more positive emotions, smiles, and laughter. This result jibes with prior work suggesting that laughter is social, and that people laugh more with familiar others (Manson et al., 2013; Provine, 2001, 2012). It seems likely that the boys in the *Relational* condition liked the robot less and felt less positively about it. Although their engagement and attention did not differ significantly from the other children, they showed fewer positive emotions during the interaction.

However, these differences I observed in children's emotion expression may also reflect gender differences in emotion expression. A recent meta-analysis found that girls show positive emotions and internalizing emotions (such as sadness, sympathy) more often than boys, while boys showed externalizing emotions (such as anger) more often (Chaplin and Aldao, 2013). These differences, however, were small and moderated by context and task. Other work suggests that girls tend to smile more during social interactions, perhaps especially so with unfamiliar others (Benenson, 2014). Perhaps as they got more familiar with the robot, the need to maintain a positive facade receded. Timothy Bickmore has explored making relational agents act more formally during early sessions and then get more familiar later on (Bickmore and Picard, 2005; Bickmore, Schulman, and Yin, 2010). The robot in this study did act somewhat less familiar and shy in the first session, so perhaps the change in the robot's behavior as well as children's familiarity with the robot affected their emotions.

One interesting result was that the gender interaction in the correlations was frequently less prominent for behaviors—such as whether children said goodbye or their behavior during the Anomalous Picture Task—than for interviews, such as the Social Relational Interview (Section 10.3.3.11). This suggests that some children *said* the robot was relational, but their *behavior* did not match up exactly with what they said. This was especially true of girls in the *Not Relational* condition, who frequently had high Social Relational Interview scores but did not treat the robot as as much of a social agent. In fact, they showed several negative correlations between their ratings of the robot's relational nature at the posttest and their emulation and story length. Girls in the *Not Relational* condition were also less likely to say goodbye, especially in Sessions 2 and 3. Perhaps these girls responded more negatively because of a mismatch in their expectations and in how the robot actually behaved. These girls may have had higher expectations of the robot than boys, which were not met, which significantly affected their view of the robot as a social other. They may have wanted to believe that the robot was a relational agent, and even reported that it was, but during interaction, reacted in the moment to the robot's less-relational behavior. Their experience of expectations not being met—such as the robot not necessarily remembering them or acting as relationally as they expected it would—could have led to more negative affect, less mirroring, telling shorter stories, and less social behavior such as saying goodbye.

One result that did not follow the pattern was children's responses to the SAQ. Even boys in the *Relational* condition reported becoming more accepting of the robot and of other kids via the SAQ, which suggests that exposure to the relational robot with its purported hearing disability—even though these boys tended to feel less rapport with it and less close to it—positively affected their judgment about robots and kids with disabilities.

Taken together, the results suggest that there was something about the relational robot that led boys in the *Relational* condition to feel less rapport, less engagement, and form less close a relationship with it, which in turn led these boys to tell shorter stories, use fewer keywords, and mirror the robot less. On the other hand, girls responded to the relational robot with greater rapport, greater engagement, and formed a closer relationship than the girls in the *Not Relational* condition. Why might we have seen these differences in relationship and rapport?

10.4.2 *Why the Gender-Condition Interaction?*

These gender-condition interaction patterns fit with current literature on girls' versus boys' approaches to social relationships, described briefly at the beginning of this chapter (Benenson, 2014; Benenson et al., 2018; Buhrmester and Furman, 1987; Gleason and Hohmann, 2006; Walker, Irving, and Berthelsen, 2002). Although individual children may not follow these gendered behavior patterns, the generalizations about girls' versus boys' approaches to social relationships are useful in attempting to understand the patterns seen in the data here.

The relational robot in this study explicitly shared information about itself, discussed its relationship to the child, and spent time on conversation that was not necessarily part of a game or activity. These actions are more typical of girls than of boys. Thus, these actions could have led children to think of the robot as more of a girl. The non-relational robot performed fewer of these actions. In prior work, we have seen girls give longer descriptions of their friends than boys, suggesting that girls may be more interested in describing their friendships and sharing information about their relationships (Kory-Westlund et al., 2018).

Because young boys are generally more interested in playing with groups of other boys than with girls, especially individual girls (Benenson, 2014), this could have led boys to be less interested in playing with the robot—especially with the relational robot. Several additional features of the interaction could have also contributed to this pattern. The mere fact that the robot interaction was one-on-one could have led girls to be more interested than boys, since girls tend to play one-on-one more often than boys do. Girls also tend to be more interested in activities with adults, especially women, than are boys (who are far more interested in playing with their peer group) (Benenson, 2014). The experimenters in this study were all adult women; this could have also attracted girls to the activity more than boys.

In addition, boys' play tends to involve action, battle, and defeating enemies; girls' play tends to involve nurturing activities (Benenson, 2014). None of the robot's Retell stories were action-adventure-battle stories; instead, they were about characters facing circumstantial adversity. *Baby Bird's First Nest* told the story of a young bird who fell out of her nest, built her own on the ground, and with the help of a frog, eventually made it home. *Baby Duck's New Friend* followed a young duck on an adventure downriver, where he learns he can fly and flies home by himself. *Henry's Happy Birthday* was about a boy whose expectations about his birthday did not line up with reality. *Possum and the Peeper* told the story of a possum who, with the help of several other animals, investigates an odd noise near his home. A few of the robot's Create stories involved battles or enemies, but since this corpus was large and each child only heard a fraction of the total possible stories, it was unlikely that the boys heard more than one of these stories. Thus, the content of the story activity was perhaps better suited to girls' tastes than to boys'. It could be worth exploring this in more depth later to see whether boys appeared to like particular stories better than girls. They generally said they liked all of the robot's stories, but because the robot was the one asking, children may have felt social pressure to respond positively and not hurt the robot's feelings. Looking more closely at the affect data in relation to the content of the stories may reveal new insights into children's story preferences.

The robot's behavior and speech may also have come across as more female, simply because the writer of the robot's speech (myself) was female. Prior work has found evidence that women and men may write and speak with different styles, enough so to be able to automatically categorize texts as being written by men or women with up to 80% accuracy (Argamon et al., 2003; Benenson, 2014; Koppel, Argamon, and Shimoni, 2002; Mulac, Bradac, and Gibbons, 2001). The voice for the robot was female, and shifted higher in pitch to appear younger. At young ages, the pitch of boys' and girls' speech does not differ greatly (Baker et al., 2008; Bennett, 1983; Gelfer and Denor, 2014; Hacki and Heitmüller, 1999; Weinberg and Zlatin, 1970). However, female and male voices may have noticeably different additional acoustic features (Klatt and Klatt, 1990), which children could have picked up on. An open question is whether I might see the same gender pattern in the results if the writer of the robot's speech and the voice of the robot were both male.

Another behavior divided by gender was children's responses in the Negotiation Task. Girls were more likely to allow the robot's choice or suggest a compromise. This is likely because girls and boys tend to approach conflict differently, with girls

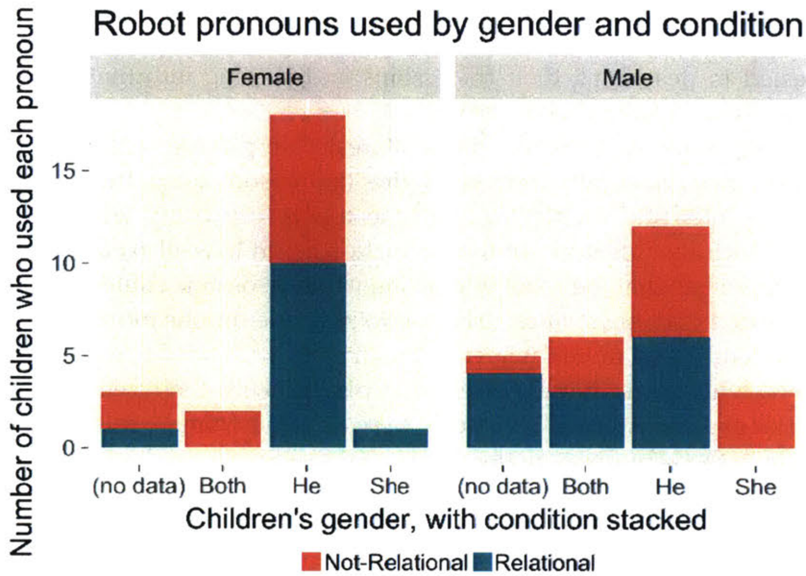


Figure 98: Children's pronoun use when talking about the robot by condition and gender.

more likely to be egalitarian and more likely to try to avoid conflict (Benenson, 2014; Benenson et al., 2018; Walker, Irving, and Berthelsen, 2002). The robot's own responses to the Negotiation Task involved an initial disagreement about which picture to tell stories about, but ultimate acquiescence to the child's choice—which was perhaps a more "girly" response.

10.4.3 *Robot Gender*

I did not explicitly ask children whether they thought the robot was a boy or a girl. Thus, we cannot know for sure whether children's perceptions of the robot's gender affected on their behavior. However, there were numerous questions where children explained their responses and described the robot, often referring to it with female or male pronouns (or both). I coded whether each child used pronouns to refer to the robot at any point, and if so, which pronouns they used. Figure 98 shows the breakdown of pronoun use by condition and gender. Girls most frequently used "he/his" to refer to the robot. Boys were more mixed. Six boys used both "he/his" and "she/hers" at different points; this makes it unclear whether they thought the robot was a boy or a girl, and were accidentally using the wrong pronoun some of the time. There were 8 children for whom I had no data (these children did not use pronouns when talking about the robot), including 4 boys in the *Relational* condition. Anecdotally, multiple children were not sure about the robot's gender and asked me or the other experimenters whether the robot was a boy or a girl. I overheard discussions amongst several of the children who participated in the study regarding the robot's gender, with one girl very insistent that it was a girl, and another insistent that it was a boy. Given all this, it is difficult to draw conclusions about the effects of robot gender perception on children's behavior.

In a prior longitudinal study with 17 children, when I explicitly asked children if they thought the robot was a boy or girl, I found that 75% of children assigned a similar fluffy robot (a green DragonBot) a gender that matched their own (Kory and Breazeal, 2014). For the mismatches, several girls thought the robot was a boy; only one boy thought the robot was a girl. Although that robot played a similar storytelling game to the robot in the present study, used the same voice, and included some similar dialogue, various behaviors and the morphology and color were different, which could easily have led to a different perception of the robot's gender. In particular, green is generally perceived as more gender-neutral in American culture than both red and blue (the colors of the robot in the present study).

Gender differences have been seen in other child-robot and child-agent education studies. In my earlier work, we found that girls rated the robot as more social and relational than boys did (Kory-Westlund et al., 2018). Kennedy, Baxter, and Belpaeme (2015) found that girls 7–8 years old improved more than boys of the same ages on a math learning task when a robot was present, and girls with a robot improved more than girls without a robot. It is unclear from that study what may have caused these differences. The authors hypothesized that the robot's social behavior distracted children, that the girls may have viewed the robot as less social, and thus, that they were less affected by the robot's socialness being distracting. However, the present study provides evidence for the opposite story: girls may have responded better to the robot's social behavior than boys, and aspects of rapport and liking were related to their increased learning.

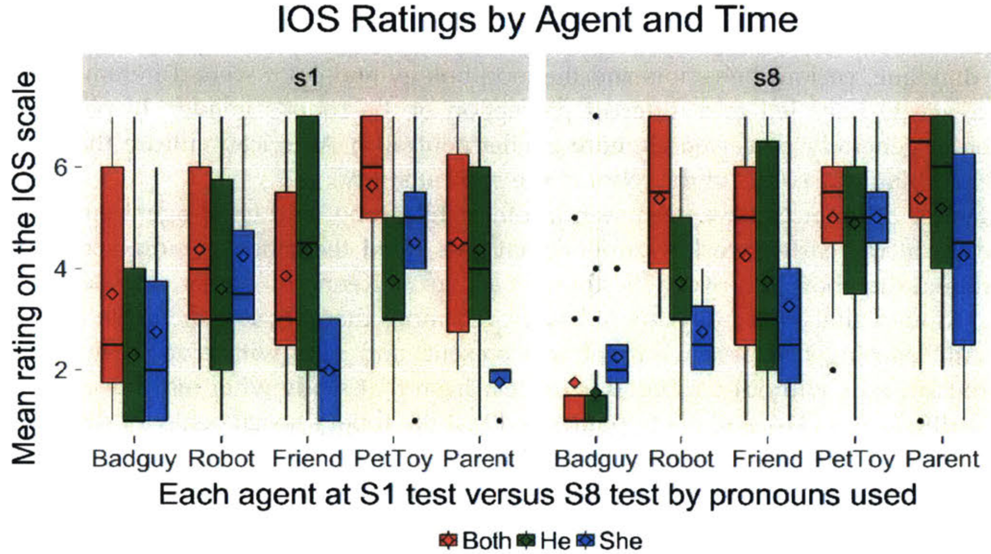
Burleson and Picard (2007) developed an affective virtual agent that helped children aged 11–13 years-old with a Towers of Hanoi problem-solving task. The agent used nonverbal mirroring and affective support interventions, just one or the other, or neither. They found that girls and boys frequently responded in opposite ways to the agent. For example, girls who worked with the agent that used both mirroring and support showed less frustration and more flow during the task than girls who worked with an agent that used just one or the other or neither; boys, on the other hand, showed more frustration. This pattern is in line with what I observed in this study. Other work has also found that girls engaged more with a virtual learning companion when it was embedded in a narrative activity than boys did, while boys reported higher mental demand when the agent was in the narrative as opposed to an agent that provided equivalent task support (Pezzullo et al., 2017).

10.4.3.1 *Results by Pronouns*

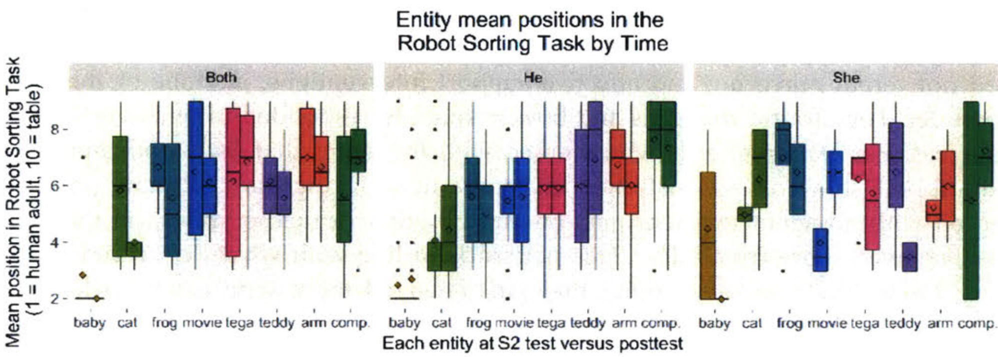
To learn more about whether and how children's perception of the robot's gender impacted their learning and relationship with the robot, I divided children by the pronouns they used to refer to the robot ("he," "she," or both). Children who used "he" and "she" consistently as opposed to both pronouns at different times could be considered to be less confused about the nature of the robot.

I analyzed children's overall vocabulary test scores, keyword use, emulation of the robot, Social Relational Interview sum scores, Picture Sorting Task placements, Inclusion of Other in Self ratings of the robot, reported liking of the robot, and overall acceptance score on the SAQ. I performed mixed ANOVAs, like before, substituting pronouns (between: he, she, both) for gender in the analyses. Because these analyses were post-hoc, I corrected for multiple comparisons with the Bonferroni correction, and only considered results significant when $p < 0.005$.

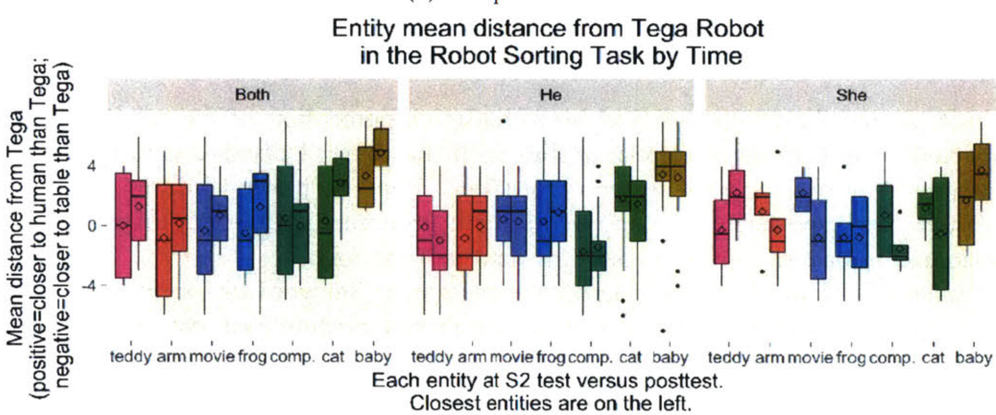
There were no differences regarding vocabulary test scores, children's use of keywords, or in children's emulation of the robot.



(a) Inclusion of Other in Self.

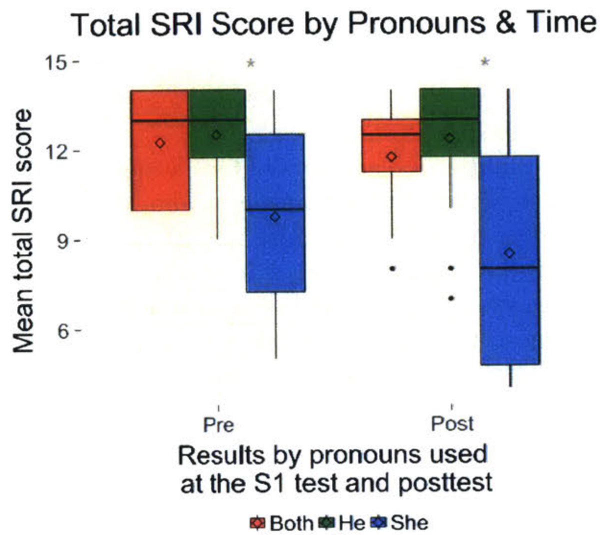


(b) PST positions.

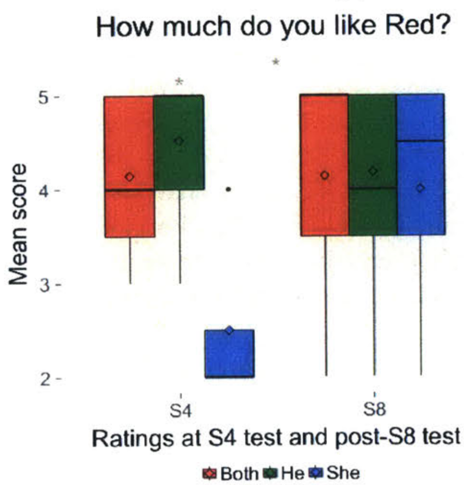


(c) PST distances.

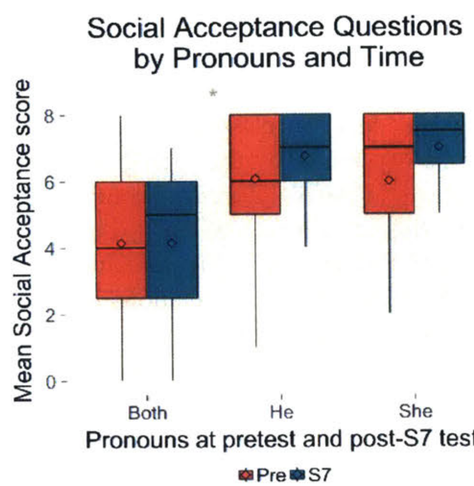
Figure 99: The pronouns children used to refer to the robot appeared to be related to some of their behavior.



(a) Social Relational Interview.



(b) Liking of the robot.



(c) Acceptance.

Figure 100: The pronouns children used to refer to the robot appeared to be related to some of their behavior.

There was a trend for children's sum Social Relational Interview scores to differ by pronoun use, $F(2,34) = 5.58$, $p = 0.008$. Children who called the robot "he" had higher scores than children who called the robot "she" (Figure 100a).

There were no significant differences regarding children's reported closeness on the Inclusion of Other in Self task. There was a slight trend, however, for children to say they felt closer to the robot if they used both pronouns, especially at the posttest, while children who used "she" tended to feel less close (Figure 99a). Similarly, children who used "she" liked the robot less initially, but this difference disappeared by S8, $F(2,35) = 5.97$, $p = 0.006$ (Figure 100b).

Children who used both pronouns were somewhat less accepting in the SAQ, $F(2,33) = 5.22$, $p = 0.011$ (Figure 100c).

There was a trend for children's placement of the entities in the Picture Sorting Task to differ by pronoun use, with trend for an interaction of pronouns with entity and time, $F(14,544) = 1.92$, $p = 0.023$. There was also a trend for an interaction of pronouns with entity and time for the distances relative to the Tega robot, $F(12,471) = 1.96$, $p = 0.026$ (Figures 99b and 99c). These interaction suggested that children who called the robot "he" or "she" placed the Tega robot closer to the human adult at the posttest than children who used both pronouns. There were multiple differences after S1 that evened out at the posttest, e.g., children who used "she" placed the robot arm and baby farther from the human, but placed the movie robot closer to the human than children who used "he" or "both." Children who used "she" placed the teddy bear far closer to the human at the posttest than other children. At S1, but not at the posttest, children who used "she" placed the human baby much closer to the Tega robot than other children.

These results showed some trends suggesting that children's use of different pronouns to refer to the robot may be related to their liking of and relationship with the robot, but no results were significant. Children who used "she" appeared to like the robot less, feel less close to it, and rate it as less social and relational. However, this may be explained by the fact that most of the children who called the robot "she" were boys in the *Not Relational* condition, who—based on the earlier analyses—tended to like the robot less, feel less close to it, and rate it as less social and relational. The SAQ and Picture Sorting Task results hint that children who used both pronouns may have had less strong associations between the robot and a human, but are not very conclusive on the matter.

10.4.4 *Learning and Relationships*

Various qualities of children's stories (e.g., length, unique words, target word use) varied as a result of both condition and gender. First, girls were more likely than boys to use the target vocabulary in their stories. This pattern was driven by the behavior of children in the *Relational* condition: Girls in the *Relational* condition and boys in the *Not Relational* condition were more likely to use the keywords in their stories than girls in the *Not Relational* condition and boys in the *Relational* condition. In addition, boys in the *Relational* condition told shorter stories than all other children and used fewer unique words in their stories. Some of these gender patterns were seen in the earlier dataset I analyzed (Section 9.1.1), but to a much lesser extent, likely because the robot in that prior study did not display different relational behaviors between conditions.

These differences are likely due to the connection I observed between children's learning and the relationship and rapport they had with the robot. As reported earlier (Section 10.3.3.11), there were multiple correlations between children's scores on the relationship assessments and children's vocabulary test scores, mirroring of the robot's stories, story length, and use of keywords in their stories. Most of these correlations were stronger for girls in the *Relational* condition and boys in the *Not Relational* condition. Given that these children also tended to report feeling closer and treated the robot as a greater relational other, this provides support for my hypotheses in Chapter 9 regarding links between children's friendships and their peer learning. It is also in line with similar correlations found in my earlier work (Sections 5.6 and 9.1.1) and other recent work linking learning to rapport (Sinha and Cassell, 2015a,b)

Another piece of evidence that children's relationships impacted learning was the correlation between children's negotiation behaviors and their vocabulary posttest scores. This correlation was more evident for boys than for girls. As discussed earlier, girls and boys may approach conflict differently; this perhaps led to girls in both conditions being more likely to seek an equitable solution, while boys who did so only did so as a reflection of their relationship or rapport with the robot.

11

GENERAL DISCUSSION

11.1 SOCIAL BEHAVIOR, RELATIONSHIPS, ENGAGEMENT, AND LEARNING

Through the nine studies discussed in earlier chapters, I used a deep dive into children's language learning with social robots to explore different aspects of relational AI, especially regarding the development of relationships and the impact of social and relational behaviors on children's engagement, emotion, and learning. I explored whether children can build relationships with relational robots, developed new assessments for understanding children's relationships, and examined what features of a relational robot can encourage or hinder the development of a relationship over time, as well as mechanisms by which a robot can influence children's engagement, emotion, and peer learning.

The data and results from these studies provide evidence for links between interpersonal social behavior, relational behavior, engagement, relationships, rapport, and learning. Some of these links are, by now, well-supported, through both my data and related work. Some links are tentative and need further investigation. Figure 101 summarizes some of the primary connections, which I discuss in more detail below, alongside other observations about children's interactions during my studies.

11.1.1 *Positive Affect*

In human-human interpersonal interaction, positive affect appears to be built through a combination of experiences shared together, social behaviors, and rapport (e.g., Boothby, Clark, and Bargh, 2014; Dijksterhuis, 2005; Dijksterhuis and Bargh, 2001; Tickle-Degnen and Rosenthal, 1990), and is also influenced by a wide variety of other factors that may or may not have anything to do with one's interlocutor (e.g., the weather). Much of the research on positive affect in social interaction has examined the effects of feeling positive emotions on how one interacts socially, such as effects on generosity, negotiation, and social problem-solving (Isen, 1987, 2002), rather than on which social or relational behaviors generate positive affect.

With regards to human-robot interaction, my research provides evidence that social and relational behaviors such as entrainment and backstory can promote positive affect (Chapters 8 and 9). In addition, some work has shown that personalization by the robot can promote positive affect (e.g., Gordon et al., 2016). However, more work is needed to establish these links more solidly.

11.1.2 *Rapport*

As discussed in Section 8.1.1, social contingency, mimicry and entrainment, and responsiveness are considered signals of rapport and can contribute to building rapport in both human-human relationships (Chartrand and Baaren, 2009; Dijksterhuis, 2005; Dijksterhuis and Bargh, 2001; Lakin et al., 2003; Rotenberg et al., 2003; Semin and Cacioppo, 2008; Tickle-Degnen and Rosenthal, 1990; Wiltermuth and Heath, 2009),

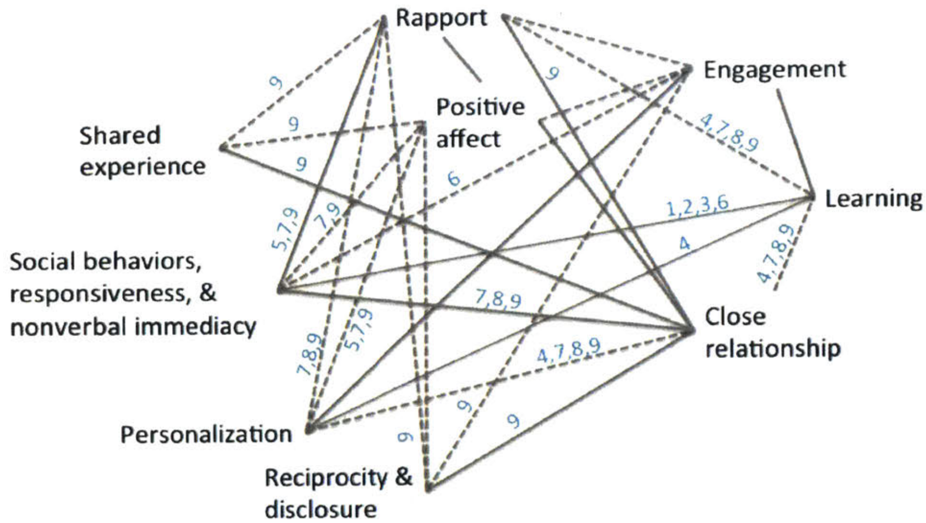


Figure 101: Connections between social behavior, relational behavior, engagement, rapport, relationships, and learning. Solid lines indicate established links, while dashed lines indicate that further investigation is needed. Numbers indicate the studies from this dissertation that provided evidence for the links.

and in human-agent relationships (e.g., Chapter 8, Bell, Gustafson, and Heldner, 2003; Breazeal, 2002; Gordon et al., 2016; Levitan et al., 2016; Suzuki and Katagiri, 2007).

It is not as clear whether other features of relationships, such as shared experiences or disclosure, may also contribute to the development of rapport, but there are tentative links based on my data (especially regarding children's language mirroring with the robot) and some prior work (e.g., Chapter 9, Berger, 2001).

11.1.3 *Engagement*

In human-robot and human-agent interaction, personalization has led to increased engagement (e.g., Bickmore, Schulman, and Yin, 2010; Coninx et al., 2016; Kanda et al., 2007; Leite et al., 2012b, 2014; Serholt and Barendregt, 2016). We have also seen that social behaviors, such as expressivity, can increase engagement (e.g., Kory-Westlund et al., 2017b).

There is some evidence that relational behaviors, such as disclosure and backstory, may increase engagement (e.g., Bickmore, Pfeifer, and Schulman, 2011; Gockley et al., 2005), but it is not entirely a clear story. In Bickmore, Pfeifer, and Schulman (2011), the relational agent tested performed not only self-disclosure, but also other nonverbal and social behaviors, so it is unclear which aspects of its relational behavior contributed to increased engagement. In both the Entrainment/Backstory study (Chapter 8) and in the Relational study (Chapter 9), I observed no difference in children's engagement by condition as measured via facial expressions. Perhaps in these studies, the robot and activities were interesting enough to generate engagement regardless of condition.

With regards to the effects of positive affect and rapport on engagement, there is some research showing that in learning activities, positive affect may contribute to engagement and negative affect (such as frustration or boredom) has been associated with disengagement (e.g., Baker, Rodrigo, and Xolocotzin, 2007; Baker et al., 2010;

D'Mello and Graesser, 2012; Reschly et al., 2008). In addition, Ireland and Henderson (2014) found that increased language style matching—a measure of overlap in speaking style that has been associated both with rapport and relationships (Babcock, Ta, and Ickes, 2014; Ireland et al., 2011; Niederhoffer and Pennebaker, 2002; Pennebaker, Mehl, and Niederhoffer, 2003; Tausczik and Pennebaker, 2010)—is also associated with increased social engagement.

11.1.4 *Close Relationships*

Factors that promote close relationships include shared experience, responsiveness and contingent social behaviors (e.g., entrainment), rapport, positive affect, and reciprocity (e.g., Berscheid and Reis, 1998; Buhrmester and Furman, 1987; Csikszentmihalyi and Halton, 1981; Davis, 1982; Hartup et al., 1988; Rubin, Bukowski, and Parker, 1998). These were discussed earlier as prominent features of human relationships (Section 5.1) and children's friendships (Section 5.2), and as necessary features of relational AI (Section 6.2.1). There is strong evidence in the human-human literature for these connections. My data provides initial evidence that these connections are also present for human-robot relationships.

One link that is less well established is whether personalization affects the development of close relationships. In human relationships, we do not talk about personalization so much as shared experience, and how people change in response to their relationship with each other, such as converging toward similar emotional reactions or preferences (e.g., Anderson, Keltner, and John, 2003; Bove, Sobal, and Rauschenbach, 2003). In human-robot research, most of the work on personalization has focused on measuring outcomes such as learning or engagement, not relationships (discussed in detail in Section 4.2). My data suggests that personalization of both curriculum and behavior may contribute to children's development of relationships with a robot and their perception of the robot as a social-relational agent (Chapter 8 and Sections 5.6, 9.1.1, and 9.2), but more work is needed.

11.1.5 *Learning*

There is strong evidence that personalization, social behaviors, and engagement all contribute to learning. Past research in both human-human and human-agent learning has linked engagement in learning activities to learning (e.g., Baker et al., 2010; Bickmore, Pfeifer, and Schulman, 2011; Carini, Kuh, and Klein, 2006; Craig et al., 2004; Csikszentmihalyi, 1990; D'Mello and Graesser, 2011; D'Mello and Graesser, 2012; D'Mello, Dieterle, and Duckworth, 2017; Pardos et al., 2014). We have also seen in human-robot and human-agent interactions with both children and adults that personalization of curriculum and behavior has led to increased learning gains (e.g., D'Mello et al., 2012; Gordon and Breazeal, 2015; Kory and Breazeal, 2014; Leyzberg, Spaulding, and Scassellati, 2014; Park et al., 2019; Ramachandran and Scassellati, 2015; Scassellati et al., 2018a; Thrun et al., 1999). One open question, however, is whether personalization leads to increased learning gains because it increases engagement and flow, as opposed to increasing learning directly.

With regards to social behaviors, research with humans has shown that social behaviors such as nonverbal immediacy is linked to increased learning gains (Christophel, 1990; Mehrabian, 1968; Witt, Wheeless, and Allen, 2004). With robots, we have seen that use of appropriate social cues, social contingency, nonverbal immediacy, vocal entrainment, and expressivity have led to increased learning and trust in the robot as an informant (e.g., Breazeal et al., 2016b; Kennedy, Baxter, and Belpaeme, 2017; Kory-Westlund et al., 2017a,b; Lubold, 2017; Lubold et al., 2018).

Regarding the impact of relationships and rapport on learning, we are seeing an increasing amount of evidence suggesting that they both can contribute positively to learning outcomes. As discussed in Section 5.4, prior work has shown, e.g., that the social bonds between children and teachers can predict learner performance (Wentzel, 1997), that children may learn math concepts from media characters more effectively when they have stronger parasocial relationships (Gola et al., 2013; Richards and Calvert, 2017), and that rapport led to improved learning in peer tutoring scenarios (Sinha and Cassell, 2015a,b). In my work, I also observed correlations between various measures of children's relationships, peer mirroring, and learning (Sections 5.6, 9.1.1, and 9.4.3.11).

11.1.6 *Individual Differences*

The studies about children's trust, relationships, and engagement with robots bring up intriguing questions about who children are comfortable with and how that affects their behavior. We have seen, e.g., that children selectively choose to learn from and endorse particular informants based on their trust (e.g., Breazeal et al., 2016b; Corriveau et al., 2009; Harris, 2007, 2012), and in many of these cases, it is children's familiarity with an interlocutor or their interlocutor's use of appropriate, contingent social cues that affect their level of trust. But there are other factors that influence who children are comfortable with, which in turn may affect their engagement, rapport, and development of a relationship. Gender is one, as we saw in the Relational AI study (Chapter 10). Children's own personality, such as their shyness, can affect their development of rapport, as can an adult experimenter's frequency of smiling (Rotenberg et al., 2003).

Thus, when we consider how to design technological agents to interact with children, and which behaviors those agents should use in order to engender trust and comfort, the answer is not straightforward. While some behaviors, such as rapport-building entrainment and mimicry or use of appropriate social cues, may be quite helpful, as we saw from the diagram above (Figure 101), children also respond to many other features of an agent—its gender, its embodiment, its morphology and appearance, its exuberance and personality, and more. Children are not necessarily drawn to the people or agents who are most like them, so we cannot rely on measuring the child's personality, exuberance, or behavior and merely match that and have it "work". We saw this to an extent in both the Entrainment/Backstory study and the Relational AI study (Chapters 8 and 9). Not all children reported liking the robot or a close relationship with it.

11.1.7 *Novelty*

One common question in human-robot and child-robot interaction research is *when does novelty wear off?* It is not disputed that novelty has some kind of an effect on initially increasing engagement, positive affect, and excitement about a robot interaction (and thus is a push for more longitudinal work to "get past" the novelty effect (e.g., Baxter et al., 2016)), but there is no consensus on what amount of exposure is needed to say that novelty has worn off, nor consensus on exactly what novelty *is* or how to measure it. It is also worth questioning whether novelty ever does completely "wear off." In human-human relationships, there seems to always be the potential for some novelty at some level. While interactions may reach some kind of steady state with less novelty, there is potential, e.g., for one's spouse of thirty years to still cause surprises.

So far, in longitudinal HRI studies, researchers have reported that novelty has worn off after 1–2 weeks, as assessed by increased boredom (e.g., Salter, Dautenhahn, and Bockhorst, 2004) or decreases in interaction time (e.g., Gockley et al., 2005; Kanda et al., 2004). That is, after some amount of time, the pattern of interaction changed. But because there are no consistent ways of measuring novelty, nor for determining whether observed interaction patterns were in fact a result of novelty wearing off versus, e.g., merely being boring after doing the activity a couple times, we do not know whether the effects reported in these studies are actually related to novelty.

During my longitudinal studies, in which children interacted with a robot approximately once a week for 10–20 minutes, I frequently observed increased engagement or positive emotion in earlier sessions, which appeared to decrease or flatten out over time. For example, in the Relational AI study (Chapter 9), many children (particularly in the relational condition, but with some variation by gender as well) showed more positive affect in the first three sessions than in later sessions. Should I attribute this to novelty? It is unclear. For example, because these studies have occurred during a school semester, some increases and decreases in engagement and affect may be more related to external events, such as increases in interest in the robot after a break from school (such as a spring break), or decreases near the end of a semester, when there are many fun "end of year" activities occurring in classrooms. In addition, we have seen some decreased interest in late spring, which coincides with when classrooms begin to spend more time outdoors, and children are more interested in participating in classroom activities that allow them time to be outside as opposed to in a small room with the robot.

To deeply understand the impact of novelty, we need to first clearly define novelty. Is novelty different than unfamiliarity, and if so, how? What are we contrasting novelty with—familiarization? Habituation? Boredom? How does novelty relate to engagement? For example, we could define the novelty effect in HRI as engagement that is due to newness rather than due to intrinsic qualities of a thing (e.g., a technology, a robot, a game) being engaging or fun. Then the question is, when does boredom or engagement overtake engagement-from-novelty? However, novelty is not intrinsically associated with either positive or negative valence and could lead an individual into a curious/interested state, or a state of threat/risk (Gillebaart, 2012). There are likely many individual differences in preference for seeking novelty, e.g., children may prefer novel toys and pictures over familiar ones, which may promote development and acquisition of new concepts (Gillebaart, 2012). Boredom-prone people may be more focused on novel experiences and may find them more interesting. Given this, defining novelty in terms of engagement may not make sense.

Leite, Martinho, and Paiva (2013) equated novelty with familiarization and habituation, suggesting that novelty has worn off when familiarization or habituation with the robot is stable, i.e., when a person does not react as much to it and starts preferring novel behaviors. They suggest using gaze or looking time to determine habituation. Using gaze, as well as other behavioral measures, is a reasonable suggestion. Some research suggests that children look less at a familiar peer, and look longer at an unfamiliar peer; they also may play more cooperatively with a familiar peer, and show more behaviors such as seeking attention, asking questions, and showing affection (Doyle, Connolly, and Rivest, 1980; McCornack, 1982).

In the Relational AI study, we examined children's gaze patterns during the APT (Section 9.4.3.5). If the robot was perceived as more novel than the human, and more novel at the pretest than at the posttest, we would expect to see a decrease in children's amount of looking time between the robot and the human, and from the pretest to the posttest. However, this was not what I observed. Children did spend more time looking at the robot than at the human experimenter at both times, but they also looked at the robot more at the posttest than at the pretest. This indicates that the relationship between novelty and children's gaze patterns is not as simple as decreasing novelty leading to decreased gaze. It may be that children were looking more at the robot as an attention-seeking behavior, which may also be related to greater familiarization, or perhaps they were looking longer because they knew it was their last session with the robot. Either way, it seems that more work is needed to understand how children's gaze patterns relate to their perception of novelty.

One useful framework for examining novelty may be Novelty Categorization Theory (NCT). NCT suggests that appraisal of events as novel relates to categorization, in that an event is novel if it does not fit in any existing categories one has (Förster, 2009; Förster, Marguc, and Gillebaart, 2010). Novel events are processed in a more global processing style that uses broader, more inclusive mental categories in order to assimilate the novel information and integrate it into existing mental categories or knowledge structures. Assimilated knowledge becomes more familiar and likeable. We could use this framework to evaluate people's categorizations of robots. For example, there are multiple tasks that measure global versus local processing on which people have performed differently when the task is framed as novel versus as familiar, such as the Gestalt Completion Task (discussed in Gillebaart, 2012). Perhaps we could administer these kinds of tasks either during or following a robot interaction to learn whether people are using a more global or local processing style, and thus, whether they are perceiving the activity with the robot as more or less novel.

11.2 CHILDREN'S EXPERIENCE OF RELATIONAL AI

Social robots, when explicitly designed as social agents, seem to be perceived as social-relational agents by children. The majority of children have reported that the robots they interacted with would feel sad, would try to help, wanted companionship, and had genuine feelings. They talked about the robots similarly to how they talked about their friends. They reported feeling as close to the robots as to their pets, friends, and parents, though depending on the exact behaviors displayed the robot, they might also report feeling closer to their best friend or to their parent than to the robot. When compared to other entities, the robot was placed in the middle—not as human as babies or cats, not as machine as a mechanical robot arm or a computer. Children responded socially to the robots as they do to humans, including mirroring their behaviors.

Designing social robots as social, relational agents enables them to close the interaction loop. Children respond to them as social agents; the robot can respond in kind, leading to more engaging, real, and reciprocal interactions. Social robots with relational AI are not a "mere technology": they become more than a playful object (Ackermann, 2005), a transitional object (Winnicott, 1953), an imaginary friend (Gleason, 1997; Gleason and Hohmann, 2006), a transactional system (e.g., persona-AI systems such as Alexa and Google Home), or even a relational object (Turkle et al., 2006a,b). They are not merely for projecting onto; they are for being with. Children treat social-relational robots as others with whom they can form long-term, social-emotional relationships. These robots share some features with pets, toys, computers, artifacts, and friends, but not exactly the same features as any of these. This social-relational nature is a unique aspect of this technology that provides new opportunities for interaction, intervention, and innovation.

The research discussed in this thesis provides evidence that when social robots are designed as social-relational agents—with expressivity, social contingency, nonverbal immediacy, and relational behaviors—children react to them as social-relational agents. The robot's social contingency and nonverbal immediacy contribute to children's reaction to the overall interaction rhythms, since they and the robot are both more in-tune with the ongoing social interaction. The robot's expressivity, animation, emotion, dynamicity, variation, and apparent agency contribute to children's reaction to it as a social agent with psychological and cognitive capabilities. Children treat the robot as a social being because it appears to be one—it seems to have an awareness of the child as its interlocutor, so children treat it as an interlocutor in turn.

Is reacting in the moment to a robot as a social-relational being the same as believing that the robot is, in fact, its own self with a social-relational life of its own? That is, when children step back from the interaction with the robot to reflect on it, becoming third-person observers of their own behavior and of the robot, how do they construe the robot? Prior work from a neuroscience perspective has shown that people act and think differently when being third-person observers than when being in second-person interaction with something (Schilbach et al., 2013). The data from my interviews and questionnaires suggests that upon reflection, children report that the robot is a social-relational agent. They are not only interacting with it as if it is social and relational, they believe it is on a deep enough level to report that belief to the experimenters.

Sherry Turkle has written a great deal about children's conceptualizations of computers and computerized toys (Turkle, 1985; Turkle, 2007; Turkle et al., 2006a,b). Early on, in the 1980's, she discussed how computers are evocative objects (Turkle, 1985; Turkle, 2005), in that they have marginal status—they are between other things. They raise metaphysical questions about infinity, self-reference, paradox, animism, and what it means to be alive. Today's robots and especially social robots continue to be evocative in this way, perhaps most especially with regards to what it means for something to be alive, animate, and intelligent.

However, Turkle emphasizes the way that evocative objects—particularly computers—both fascinate and disturb children. While I have observed children in fascination, few are outwardly disturbed. Instead, they show interest. They explore. They play. They test the limits of the system, and they play within its rules. Perhaps this is due to the kinds of systems I have observed children play with and the kinds of things those systems are capable of doing. Or, perhaps, it is due to the fact that children are now encountering so many different evocative objects, or encountering them earlier, so that the robot is not the first, newest, or most disturbing.

One of Turkle's examples is children who played with a version of a Speak and Spell that, when put in a certain mode, could not be turned off or switched to a different mode due to a bug in its programming. Children kept returning to that bug and were outwardly curious, concerned, and frightened. The bug contradicted their expectations of the toy as machine that could be turned on and off at will. The robotic systems and activities I have placed with children have not been as open-ended; they did not necessarily allow children to return to whatever mode or activity they want in their exploration and play. Children did ask questions, test the robot's reactive abilities, and see what they could do. For example, in the Relational AI study, children were prompted to wake up the robot to play with them by saying "ice cream!" One girl asked if she could try saying "hamburger!" next time to see if the robot would do something different. However, many children—especially in their first encounter with the robot—were more likely to watch, wait, follow as the robot led their interaction and their conversation.

Unlike the early computer games and robotic systems Turkle studied in the 1980s, today's social robots have even more "minds of their own." Current AI is making the rules of social robots complex enough that—like with people—the rules are not easily deciphered. As time goes on, this complexity will surely increase and the rules even less easy to predict and discover. Social robots are entering the social-relational world of people in a way that other technologies have not, with some of the messiness of social relationships that entails. But social robots do not have all the messiness of human relationships, at least not yet, and in many ways they are more like pets. They follow their own script, and cannot always be negotiated with. They are not as complex as humans in their sets of beliefs, desires, and needs—indeed, we could argue they do not actually "believe" or "desire" anything. They can be programmed to have the illusion of some of this complexity. For example, they could be programmed to be more likely to acquiesce to the child's creative vision during play than another child would be—such as always ultimately agreeing to do the child's choice when negotiating on which story scene to tell stories about—or they could be programmed to stick to their own choices, or to be as unaffected by the child's speech as the cat is. No doubt the illusion will grow more complex over time as new algorithms for processing sensory input and determining behavior are developed.

Turkle discussed how emotions were one differentiating factor between alive entities and computers. This may not be true anymore, since agents can be and are designed to display social and emotional characteristics. They can use sensory input to know something about the user's own emotional state. Today's robots are also more likely to direct activities. They are not merely responsive smart toys, they are reactive and proactive. This is a huge difference in the technology itself and in how it is likely to be perceived. The social-relational robots I have tested are agents more than they are machines. While they are programmed to follow particular scripts and react in particular ways to sensor readings, this is apparent only to the programmers and designers, not to the children who interact with them. To children, they take on a magical quality, a lifelike quality, an interactive quality. The appearance of the robots—fluffy, colorful, cheerful, with many of the mechanical and electrical components hidden—boosts children's recognition of the robot as agent for interacting rather than machine that can be controlled.

The experimenters' behaviors in the studies I have done bolster this perception. Because the experimenter generally introduces the robot as a social agent—e.g., giving it a name, talking to it, telling children they can converse with it—children may be, in part, following the lead of the knowledgable adult in the room. Turkle discussed how one criteria children use to determine an entity's aliveness is ethical discourse. How we treat other entities morally—e.g., is it okay to step on it, can we kill it, must we treat it gently—inform children's understanding of their status as alive or not. Bugs can be stepped on and killed, but parents are likely to scold children for stepping on other children or on their pets. The behavior of adults and experimenters around the robots that children interact with may be partly informing children's opinions. We tell children to be gentle with the robots. We ask them not to hit the robots. It's not okay to deliberately turn the robot off or crash it during an experiment. These behaviors may be telling children that the robot is a moral agent.

One study by Williams et al. (2018) has directly examined whether a voice-based smart toy (a talking doll) can influence children's social conformity and moral judgments. They observed 40 children aged 4–10 years old interact with the doll, with a human, or with no agent during two key tasks: a social conformity task in which children judged whether certain moral statements were okay or not okay (such as whether it was okay to tease another child), and a disobedience task, in which the experimenter showed children a box, asked them not to look in the box, and then left the room for a few moments. They found that the doll, like a human, could influence children's moral judgments, but that it did not affect children's disobedience. This work suggests that children did trust the opinions of the smart doll and did react to its moral suggestions—treating it as an agent with moral opinions.

Current culture does not dissuade the view that robots can be moral agents and friends. While there are plenty of American movies featuring killer robots and machines trying to take over the world, there are also an increasing number of popular movies featuring friendly, lifelike robotic agents or animated non-human sidekicks (who may or may not be robots). There are many similar agents featured in comics, books, and video games. Prominent film examples that I have found many 4–7-year-olds familiar with include R2D2, C3Po, and BB8 in *Star Wars*, Baymax in *Big Hero 6*, and WALL-E and Eve in WALL-E. Our research robots reflect the likeness of these robots and sidekicks—their animacy, their familiarity, their relatability.

Granted, there is a difference between the kinds of interactive, social research robots children may encounter in our lab or field studies, and the kinds of robotic toys and technologies available on the market for children to play with at home. While this difference is rapidly shrinking, the fact remains that in a controlled experimental study, there is more space for the illusion of life. Because the interaction is not open-ended and will not continue indefinitely, greater effort can be placed on creating an interactive agent that appears responsive, reactive, proactive, and lifelike—that is, within the constraints of that particular experimental interactive scenario.

When we move out of the lab into the world of commercial robots, home assistants, and AI-enabled smart toys, we see that children do explore a variety of questions and behaviors with them in an effort to understand them. For example, Druga et al. (2017) observed 26 children interact with a variety of agents, including an Amazon Alexa, Google Home, Anki's Cozmo robot, and the Julie Chatbot. They observed younger children asking questions about various agents as a person (such as its age and favorite color) and older children trying to understand how the agents worked, such as whether it had a phone inside it, what actions it could take, and what it knew. They also tested the limits of the devices, such as checking whether the agents could see as well as hear, or repeating questions to see if the agents would give a different answer. The children often said the agents were as intelligent as they were and that the agents were friendly and trustworthy. It did not appear, however, that children were at all confused about whether the agents were people—they did treat them socially, but they also appeared to understand that the agents had limits and had access to different knowledge or capabilities than people do. Sherry Turkle described the same kinds of explorations of computerized toys that we are seeing now with smart toys and AI assistants: exploring what the toy or agent is, how it is made, what it is capable of, what its limits are, and very often, treating it socially as an agent. I would expect that now, because so many of our devices are designed to interact in social ways (e.g., through voice interaction), children's interactions and explorations are leaning even more into the social realm.

In an attempt to understand how children perceived the intelligence of these agents in comparison to how they viewed human and animal intelligence, Druga et al. (2018) performed another study where they asked children and their parents to assess the intelligence of a mouse and of a small robot and of themselves during a maze-solving task. They found that most parents and children thought the robot was smarter than the mouse, and about half of participants said both agents were smarter than themselves. Importantly, older children often mirrored their parents' mental models about the intelligence of both agents. Parents' knowledge of agents—such as robots and mice—can heavily influence children's perceptions of these agents. This is interesting in light of the generational differences in the kinds and amounts of technology available to today's children versus their parents. With smart phones, tablets, computers, robots, AI-enabled home assistants and smart toys all becoming ubiquitous for many of today's children, we might expect that the understanding they have of computational devices is very different than that of their parents, especially given how socially-oriented many of these devices are. But computers have always been magical black boxes, to some extent at least. Perhaps parents' mental models of past computers can be transferred to current devices without losing much in terms of understanding how those devices work. It is also intriguing to wonder how one's lived experience with technology might change or update one's mental models of these devices, since with experience, one comes to recognize the capabilities and limits of a system. Druga, Williams, and colleagues (Druga et al., 2017, 2018; Williams et al., 2018) have pondered in their work how we might develop new activities and experiences that help children and their parents understand what these new technologies are capable of, how they work, and what they are, arguing that helping people learn how technology and AI function is critical to enabling them to live with technology in informed, ethical ways.

These observations and questions about children's and parents' experiences with current robots and smart devices shine light on the ethical questions that pervade the development of any new technology. Every new technology raises concerns about its use and misuse; its capacity for deception and social manipulation; its ability to promote emotional attachment and reliance; its authenticity; its transparency; its relation to privacy, security, and safety; its embodiment of existential threat to human specialness and uniqueness. Social robots are not fundamentally different than other technologies in this regard, though they may be unique in that they raise all of these concerns at once.

One question Sherry Turkle's work has raised is whether children know what robots are, and whether they ought to be informed—by an authority figure or by the robot itself—about a robot's true nature and its capabilities. This question reminds me of the movie Robot and Frank, in which the humanoid helper-robot explicitly and frequently reminds the elderly Frank (who is coming to rely on the robot) that, "I'm just a robot, Frank." But what does "just a robot" mean? Is it supposed to mean that robot is somehow just a machine, programmed, following scripts, doing what it is told? Social robots are more than mere machines in how we approach them and interact with them. For example, the mere fact that children frequently say goodbye to robots—when they do not say goodbye to toasters or iPads—lifts robots into a more peopleesque category. Perhaps "just a robot" is supposed to remind us that robots are not people, that they do not have minds of their own, that they have significant limitations and do not really understand anything? Is saying "just a robot" acknowledging the Chinese Room problem (Searle, 1980) behind today's robots? That is, the robot has the appearance of life and understanding without actually understanding. Due to the nature of its algorithms, it is reactive, responsive, proactive, even inventive. But it does not understand; it is not conscious; it has no "soul." Is this the "truth" we want children to understand about today's robots? The truth that we think they don't already understand?

I think children who interact with social robots have a sense of that truth. They use anthropomorphic language to talk about the robots, but they place them farther away from human adults, human babies, and cats, and closer to frogs, teddy bears, computers, and tables. They talk about their friends differently than robots. Robots are not their friends; they are their robot-friends; they are a new kind of relationship. Children know robots occupy an in-between space. They know robots have marginal status. And unlike adults of a generation who did not grow up living with pervasive smart devices and smart toys, I think children are okay with that. It's just a robot, Frank. Children are growing up believing that things other than humans can and do have intelligence, as Sherry Turkle predicted. The new ontological space that social robots occupy is just that: a new category, in between the others, erasing the line that used to divide a binary and sticking something new in the middle. I think children are not confused about the new category; in fact the opposite. They understand and need the new category because there are an increasing number of things in the world now that fit in it.

Asking whether children know what robots are assumes that children are confused about robots. From my observations, children are not confused. As I have said, they rightly think that robots occupy an in-between space with some qualities of living beings and some qualities of artifacts and some qualities of machines. That said, they may not know how robots work or what makes a robot function how it does. Turkle talked about how children in the 1980's often fixated on their toys' batteries as explanation for how the toys worked. Children would talk about how the batteries were food; they talked about the importance of the batteries. In the Relational AI study at part of an empathy assessment one morning, the robot talked about how it was sad because it could not play as much because it had to wait for its battery to charge. Several children expressed confusion: "Wait, you have batteries?" "Why do you have batteries?" The batteries are no longer a focal point—instead, they are invisible. Children are focusing on different parts of the robot, such as its smartphone face. Children understand that smartphones can do a lot; when they see that the robot has a smartphone component, they likely attribute some of the intelligence of the phone—which they are familiar with—to the robot. Turkle discussed how the smart toys of the 1980's were opaque. Children could not understand them using physics, so they turned to psychological explanations. Now, children also explain the robots in terms of other smart devices they have encountered. The opacity is still there, but it is different.

With early computational toys, children could learn to program and use that understanding of programming to understand how their toys might function. In our current world with its proliferation of smart devices, children need to learn about computers, robots, programming, and AI. Learning how a computer works, how a smartphone works, and how to teach an AI algorithm how to solve a problem or play a game removes a little bit of the "magic" behind their function. But I don't think that teaching children what makes social robots tick removes the magic and immediacy of interaction, as it is happening, in the moment. As one example, even the undergraduate and graduate students in our lab—all of whom have a hand in designing, programming, or testing social robots!—talk about them anthropomorphically, easily fall into conversation when we are testing new dialogue trees or interactive scenarios, happily make silly faces just to explore what animations the robot will play as it mirrors them, and generally interact socially when the robot engages them socially.

This anecdote says something profound about how powerful the social is for humans. We are social beings. We have evidence that *being with*—being present and immediate—strongly affects human psychology (Guerin, 1986; Henderson et al., 2006; Johnson et al., 2009; Kory-Westlund, Breazeal, and Ostrowski, in review; Li, 2015; Trope and Liberman, 2010). Turkle (1985) wrote about how early computers drew children in, leading them to deeply engage, becoming experts above and beyond their teachers in early programming languages and video games. But the power of social robots is different than the power of immersive games or new technological tools for creating. The power of social robots arguably comes, in large part, from their social presence and social power. This is what we have been uncovering throughout recent studies on children's engagement and learning with social robots: that social robots are unlike any other existing technology. Humans react to social agents very differently than they react to machines or even to voice-only agents, virtual agents, or distant embodied agents. Social robots have a power that we can and should leverage to support human flourishing.

12

RELATIONAL AI: DESIGN IMPLICATIONS

The studies I performed provide new and intriguing insights into children's behavior with social robots in both short-term and longitudinal learning interactions. I found evidence that not only was the robot's social and relational behavior indeed related to children's learning, affect, and imitation of the robot, but also that children's gender significantly impacted how they responded to the robot, their construal of the robot as a social-relational other, and the relationship they formed with it.

12.1 RELATIONAL BEHAVIOR AND LEARNING

The connection between the robot's relational behavior and children's learning and imitation is important for the future development of social robots for education. My work has provided evidence that the relationships children form do matter for their peer learning, in particular, their language learning. Playing with a social-relational peer-like robot afforded opportunities for children to learn the way they learn from their human peers. While there are numerous theories regarding how peer learning occurs (discussed in Section 3.1), there is less research exploring what modulates children's peer learning. Through multiple studies (Section 5.5, Section 5.6, Chapter 8, and Chapter 9), I showed that children's rapport, liking, and relationship will significantly impact their learning and language emulation. The robot's expressive, social, and relational behaviors all contributed to children's relationship and rapport.

Below, I discuss several different factors that impacted children's learning, engagement, relationship, and behavior that are important to consider when designing new relational robots, but especially when designing robots to help and support children during language learning activities.

12.2 MAINTAINING ENGAGEMENT OVER TIME

The Relational AI study (Chapter 9) showed that a social robot with relational AI could quite easily maintain children's attention and engagement in learning activities over many weeks. I think the following five features were particularly important, both in this study and those prior: (1) change, (2) shared experience, (3) backstory, (4) play and creativity, and (5) design as a social agent.

12.2.1 *Change*

In the studies reported earlier as well as in my earlier work, we have seen that features such as variation in dialogue and activity content, and personalization of various aspects of the interaction have increased children's engagement during repeated encounters with a similar robotic learning companion (Gordon et al., 2016; Kory, 2014). These features are important aspects of a relational robot's behavior. Change and variation, in particular, are necessary to maintain and build relationship, as has been seen in prior work (Bickmore, Schulman, and Yin, 2010; Kidd and Breazeal, 2008; Lee et al., 2012b). Change can include both "scripted" change—such as variation in how the agent speaks or acts, and backstory that is revealed over time—as well as "unscripted" change, such as personalizing different aspects of the interaction in response to the user's behavior.

With regards to personalization, the gender-related results discussed earlier are evidence that personalization of relational behaviors in social-relational robots should go beyond whether or not the robot entrains, whether or not it discloses information, whether or not it mirrors a child's affect or posture, and so forth. Instead, for the robot to engage children fully and to develop a relationship with individual children that benefits and supports them as individuals, the robot should be tailored to many different individual attributes of children, including aspects of gender, attitudes toward play and social conflict, and the content of play activities. Since gender is usually easily observable in the first interaction a robot has with a child, perhaps the robot could to start out using behaviors and personality traits that are generally favored by children of that gender, and then personalize further depending on the individual child's reaction. Knowing a child's gender is not a substitute for knowing an individual child, but generalizations about girls versus boys could lead us to design robots that in general may work better for individual children.

That said, the goal in developing a social robot for education may not be to provide the children with more of the same, more of what the child is comfortable with, more of what the child responds to best. Yes, the child may love stories about dinosaurs and cars, battles and defeating enemies, but they may need to hear stories about other things in order to grow, learn, and develop. Children need to be exposed to different ideas, different viewpoints, different personalities with whom they must connect and resolve conflicts. While a child may respond best to a particular personality or style, perhaps using that all the time limits the child's growth and does not invite them out of their comfort zone. This also should be explored in future robot design.

12.2.2 *Shared Experience*

Children have responded positively to the robot referencing their shared narrative—e.g., using their name, talking about stories told together, and mentioning facts learned about the child such as their favorite color. Children noticed when the robot used their name (Section 9.4.1.4). Anecdotally, they were visibly delighted when the robot shared a picture of their favorite animal. Prior work has also shown that including some kind of memory system in the robot that can be used to track and reference prior interactions with the user can be beneficial for engagement and positive affect (e.g., Kasap and Magnenat-Thalmann, 2010; Leite, Pereira, and Lehman, 2017). Including shared experience can be considered part of change and personalization. It can contribute to the sense that the robot "knows you" and help build a relationship.

12.2.3 *Backstory*

How an agent is introduced, the stories told about it, and the story told by it all influence human perception of the agent and their behavior with it (Darling, Nandy, and Breazeal, 2015; Klapper et al., 2014; Kory-Westlund et al., 2016a; Stenzel et al., 2012). In the Entrainment/Backstory study (Chapter 8), I found that telling children about some of the robot's limitations led them to be more accepting of the robot. Backstory can also be used to add interest in variation to dialogue to help maintain interest and engagement over time, e.g., as was done with the robot receptionist (Gockley et al., 2005). Thus, I recommend using the robot's story to help shape users' expectations about the robot through sharing the robot's history, capabilities, and limitations. The story can be used to establish the robot's character, in the same way we learn about other people through conversation and disclosure. This story can be told by people who lead interactions with the robot, such as experimenters, as well as by the robot itself during conversation.

12.2.4 *Play and Creativity*

Making interactions playful with space for children to express curiosity, explore, and create is important. While I think the Relational AI study's activities (Chapter 9) did not go far enough toward enabling playfulness and creativity, there still was some room for children to create and explore. Edith Ackermann has written about how toys that are "incredibles" can keep children coming back to play more—that is, toys that enable flexibility in play, opportunity for exploration, creativity, and variety (Ackermann, 2005). The framing of the robot as wanting to play and learn, with its hearing and listening difficulties, seemed to help children understand and deal with the robot's limitations.

For future language learning and storytelling projects, before choosing one storytelling activity or the other—and before using different activities in different sessions—I recommend asking first what the goal of the storytelling activity is. If the goal is primarily vocabulary learning, then adding word definitions throughout the stories and using the Retell format to encourage remembering the story and using those words in the same context in which they were learned would likely work well. I suggest this because in the Relational AI study (Chapter 9), in which the storytelling were based on the activities in prior studies (all of which had only ever included one kind of storytelling activity, rather than two), it was interesting—although somewhat expected—to find that children performed less imitation of the robot's story during the Create A Story task than during the Retell task. Children mimicked more when explicitly asked to do so. While the Retell activity also led to longer stories and greater use of key vocabulary words, this may have been because some of the Create A Story stories were from a corpus of stories other children had told, and thus, were themselves fairly short, and may or may not have had any key words in them. The robot also did not explicitly define or ask about the vocabulary words presented in the Create stories, which could have led to less use of relevant keywords.

If the goal is also to promote creativity and learn language in a play context, the Create format may work better. Perhaps it would be beneficial to add in some explicit teaching about word definitions to improve the vocabulary outcomes. Children's stories tended to be less similar to the robot's stories when they were invited to make up their own stories. That said, as I saw in earlier work, children frequently borrowed elements from the robot's stories when creating their own, as they do when playing with other children (Kory, 2014), and they did emulate the robot more over time (Section 5.6).

The framing of the activity also matters. In my earlier work, the storytelling activity was framed as play. The robot did not lead a long conversation before saying, "Want to play a story game?" In this study, on the other hand, the whole interaction was more formal. It was framed as practice for the robot around conversation and stories.

The robot asked to tell a story, not play a story game. It was certainly more structured than the prior storytelling activity; it included explicit conversation; it was less play-focused and more conversation-focused. Even the positioning of the tablet to the side of the robot as opposed to embedded in a play table (a "magic table") changed the focus and playfulness of the interaction; it removed the tablet-as-mediator component that appeared critical to encourage play. As a result, I think children were less focused on play during the robot interaction in this study, and that was a failing. Their stories during the Create activity were shorter than they were in the play-focused prior study. This framing of the interaction as play versus as an activity with the robot changed children's approach to it, and very likely decreased their creativity and engagement in the activity as play. Prior work has suggested that encountering an activity as play can increase creativity and engagement (Gray, 2013).

12.2.5 *Design as a Social Agent*

Social design includes all aspects of the robot's social behavior and communicative abilities—including whether and how it speaks, how it moves, its nonverbal behavior, and its social contingency. Through my studies, I have seen that all these social behaviors contribute to children's treatment of the robot as a social agent, and in turn, to their trust, relationship, and learning. Designing a robot from the ground up with social interaction in mind means considering how to make the robot's facial expressions, movement, gaze, dialogue, and other behaviors understandable to humans. It enables the robot to be responsive, expressive, and social.

The design of social robots as social agents may be more important for robots that interact with children than those that interact with adults. In a recent study of in-home use of a social home robot and a voice-only home assistant, Singh (2018) found that children were more drawn to the entertainment and social capabilities of the agents, while adults were more interested in the agents' functionality and usefulness.

12.3 DIALOGUE

One key aspect of social robots for language learning is dialogue. I have found in my studies, particularly the Relational AI study (Chapter 9), that the robots have managed to communicate and maintain engagement reasonably well, despite the dialogue being scripted and the available responses being limited—the interactions were fairly linear.

Part of the success of the robots' dialogue likely comes from my experience writing dialogue for social robotic learning companions that play story and language games with young children over numerous studies. Having observed hundreds of children interact with such robots, I know how the robot might phrase utterances to be understandable and interesting to this age group. I can anticipate the kinds of responses children might have to different questions the robot might ask, and include custom responses for the robot. This experience, in turn, comes in part from receiving feedback from experts in child development, education, and from children themselves. For designers who are writing dialogue for social robots, I suggest that getting input from others who have experience with the population in question, expertise in how that population responds to the technology in question, as well as feedback from the population itself through user studies or focus groups, is invaluable when creating new systems and interventions.

One point of difficulty in developing conversational abilities for a robot is ASR. There were often semi-awkward pauses in the conversation as the robot "listened" for children's responses. This happened most often if children were shy and did not speak, or if children were very quick to respond and spoke over the robot's last few words (such that the ASR was not listening yet and did not hear them). The backstory was important for explaining the robot's difficulty with hearing; it helped children be understanding. However, children sometimes did get annoyed or frustrated when they had to repeat themselves. Improving ASR would improve the fluency of the interaction dramatically. It would also help to adapt how long and when the robot "listened" for children's responses to individual children's speaking styles. For example, the robot could pick up on the fact that a particular child always spoke very quickly, and start listening sooner.

12.4 RELATIONSHIPS AND GENDER

It is important to realize that the robot's relational behaviors made a greater difference for girls than for boys (Chapter 10). Existing work with embodied conversational agents has explored how to engage under-represented groups, such as girls in STEM (Karacora et al., 2012; Kim and Lim, 2013; Krämer et al., 2016) and African American students in computing degrees (Gosha, 2013; Gosha and Middlebrook, 2017), as well as how to use animated characters to understand bullying in children (Woods et al., 2007) and how to modify the ethnic identity of an agent to improve engagement (Cassell et al., 2009; Iacobelli and Cassell, 2007). Some of this work has found that affiliation, rapport, and relationship impact girls' perception of the agents and their performance more than for boys. For example, Kim and Lim (2013) found that boys treated the agent as more a tool for learning and were fairly detached from it; girls, on the other hand, treated the agent as more of a friend or companion, using the agent's name and personal pronouns. Other work has suggested that the realism of the agent mattered more for males (Baylor and Kim, 2004); it seems likely that boys and girls will respond better to different agents with different behaviors—such as agents with different gender-stereotyped behaviors. Some of the work with embodied conversational agents has found that the agent's gender is important, but what that gender ought to be varies. For example, in some work, interacting with an opposite gender agent who displayed rapport enhanced math performance (Karacora et al., 2012; Krämer et al., 2016); in work on understanding bullying with younger children, same-gender agents led to better results (Woods et al., 2007). Age, task, and other aspects of context may impact children's preferences and behavior.

Even in the early days of computers, Turkle (1985) found that girls and boys related to the computers they were programming differently. Girls psychologized the computers more; they talked about them in more relational terms and imagined themselves as part of the system instead of being objective observers. This early work aligns with the patterns we are seeing now: girls frequently react more strongly to the psychological, social, and relational attributes of technological systems.

The effects children's gender had on their interactions suggest that we should pay more attention to the robot's gender and gendered behavior than previously thought. While there has been a growing amount of work in the last decade around personalization of social robots for individual children (discussed in Section 4.2), relatively few studies have explored personalization of personality, the content of curricula beyond aspects such as level (e.g., syntactic level, word difficulty), or the style of speech and behavior as relates to gendered expressions of emotion and ideas. It may be that trying to create a gender-neutral robot is not the right approach. Neutral may be perceived differently by boys and by girls. It may be that the robot needs to display much more strong, stereotypically masculine behaviors in order to come across as male (e.g., stereotypical behaviors of a five-year-old boys, such as making fart jokes). If the robot did come across as a strong male model, we might find that boys would like it more and imitate it more. There is some evidence that girls and women are culturally "allowed" to display a wider range of behaviors and still be considered women than are boys, who are subject to stronger cultural rules about what "counts" as appropriate male behavior (Benenson, 2014).

If neutral is not actually perceived as neutral by children, then attempts to make gender-neutral robots may be failing spectacularly. In all of the studies I have performed, I have rarely seen children not ascribe gender to the robot. There may be a biological basis for assuming that other social-relational agents have gender, since in the world we live in and throughout our ecological history, other agents have had gender (whether male, female, or a culturally acceptable third gender or in-between). It may be that, growing up in our current culture with our cultural history, children do not see "gender neutral" for robots—at least, perhaps not without an explicit backstory where the robot explicitly shares information about its gender (or lack thereof).

Providing a backstory about a robot's lack of gender, or nonbinary or gender neutral status may help change children's construal of the robot, but it also may not be sufficient to change how children react to it or perceive its behavior in relation to their own, given the strong stereotypes about what counts as male or female behavior (Benenson, 2014). This would certainly be worth exploring in future work. Can robots truly be gender neutral? Other technologies may be presented as neutral—computers, machines, and so forth. But even robot vacuum cleaners and robot dogs are assigned gender (Forlizzi, 2007; Hendriks et al., 2011; Kahn, Friedman, and Hagman, 2002; Sung et al., 2007; Weiss, Wurhofer, and Tscheligi, 2009). When a technology becomes relational, people seem to use the same social-relational mechanisms to understand and relate to it that they use when dealing with other people—including assuming gender. When designing new robots, explicitly considering gender and how the robot's behaviors will be perceived in relation to gender will be critical.

12.5 RELATIONAL ROBOT DESIGN

As discussed above, there are many factors that can be incorporated into the design of a relational robot, and numerous challenges in developing different relational features. In the relational robot tested in Chapter 9, I included a wide variety of relational AI features—including social behaviors such as speech entrainment, exuberance entrainment, affect mirroring, backchanneling, and some posture and gaze responsiveness; referencing shared experiences; relationship management actions; disclosure; and personalization of story level and story content. Below, I discuss which of these behaviors I think worked better than others, contributing more directly to the success of the robot, while which did not do as much as I thought they might, and what we might do to improve the robot's design, especially with regard to children's gender.

To start, the design of the relational robot worked well insofar as the robot was construed as a social agent. Because of the large number of prior studies in which we developed and tested different social behaviors, such as the robot's emotive, animated movements, its expressive voice, and its understandable dialogue, it was relatively straightforward to implement the same kinds of social behaviors on the relational robot. Children interacted with the relational robot socially. That said, there was definitely room for more personalized social behavior, in particular, in response to children's emotions. For example, the robot could do more to react to children's emotions during storytelling, e.g., reacting with a positive "Ooh!" versus a negative "Oh no!" to events in a child's story (this is easy to do when teleoperating the robot, as was done in Study 4 4.3, but much more difficult for an autonomous robot). Children often looked to the robot in conversation or storytelling for a reaction, but the robot did not always react (or react in the expected way), which could have led to increased disappointment with the robot over time. As another example, children sometimes reacted with frustration or annoyance when the robot did not hear them or did not understand them. These moments were opportunities for the robot to react, e.g., by saying "Sorry I didn't hear, I'm trying!" or by making a sad face—something to acknowledge and soothe children's frustration rather than ignoring it.

We may also need to design the robot to use different emotional reactions for different children. Because prior work suggests that there are gender differences in emotion expression (e.g., Chaplin and Aldao, 2013), the way the robot displayed emotions could have influenced children's perception of its gender. Boys may be more engaged by the robot if it acted more like a typical boy. One way to affect children's perception of the robot's gender could be with the robot's emotion displays—for example, the robot could react emotionally in a more boy-like way (e.g., showing more externalizing emotions such as anger) versus a more girl-like way (e.g., showing more positive and internalizing emotions, such as sadness and sympathy).

There were numerous other issues with the ASR and the robot's classifications of children's exuberance, as discussed earlier in Section 9.5.5. Children who spoke too quickly or too quietly for the ASR to recognize their speech were likely affected most by the ASR's limitations. The robot could do more to adapt its listening skills to individual children's communication styles—such as how long the robot's ASR should "listen" before determining that the child said something or not. Some children take longer to start speaking than others. Because speech and exuberance entrainment were two of the primary ways that the robot mirrored children, this could have contributed to decreased relationship formation, or even a perception of *less* rapport if the robot mirrored poorly.

As discussed above (Section 12.3), outside of problems with ASR, the robot's dialogue content appeared to flow reasonably well. There were a few places, such as when prompting for the story retell, that the robot could have used fewer pauses or provided clearer instructions, but overall, the dialogue had sufficient variation and sufficient responses to different things children might say to keep the interaction going.

That said, I think the content of the dialogue worked better for girls than for boys, as mentioned earlier in the sections about the robot's gender (Sections 10.4 and 12.4). This is interesting given that in earlier studies (e.g., Chapter 8, Section 5.6, Breazeal et al., 2016b; Gordon et al., 2016; Kory-Westlund et al., 2015a, 2017a,b; Kory, 2014), we had not seen strong gender differences in how children reacted to the robot's social behaviors or its dialogue. It was the addition of more relational behaviors in Study 9 (Chapter 9) that appeared to lead to gender differences. Several of the relational behaviors were part of the dialogue, such as the robot's disclosures, explicit discussion of its relationship with the child, and references to shared experiences. I think that if the relational conversation was phrased differently, boys may be more amenable to it—

e.g., the robot could still disclose information, but perhaps could frame the disclosure as a way of comparing its' skillsets to the boy's (since boys are often more interested in comparisons of skills relative to their peers) instead of as sharing personal information about itself (which is a more feminine framing). In addition, changing the emotional content of the robot's dialogue could also help, since as mentioned above, there may be gender differences in emotion expression. The robot could, for example, react with anger to an unfavorable negotiation scenario (a more boy-like response) instead of with sadness (a more girl-like response). It would also be worth examining whether having a man record the robot's voice as opposed to a woman would affect children's construal of the robot's gender (even if in both cases the voices were shifted up in pitch to be more child-like).

I do not think the relational robot's personalization of story content did that much for the interaction. The content personalization was based only on cosine similarity of the robot's stories to the child's stories, which is a fairly gross measure of linguistic similarity. I think content personalization would do better to match based on content words (both from within children's stories and from probing questions by the robot about things such as the child's favorite animal), perhaps in addition to overall linguistic similarity. Doing better at personalizing activity content and having a wider range of possible content available would help engage different children (e.g., boys versus girls).

The story level personalization was likely helpful for choosing appropriate stories with respect to syntactic level and language complexity, but the story leveling was limited in that it was set at the beginning of the study based on children's initial vocabulary test results and did not change thereafter. A recent study by Park et al. (2019) performed more dynamic story leveling, adjusting the personalized curriculum for the child over the course of the study, and this more dynamic approach appeared to be fairly effective with regards to engagement and learning. I would recommend using that kind of dynamic approach.

Of the other aspects of the robot's personalization to individual children, I think the robot's references to shared experiences were particularly helpful in engaging children. As mentioned earlier, children noticed when the robot used their names and were often visibly delighted when the robot remembered facts about them, like their favorite color and animal. Adding in more of this kind of shared narrative to the interaction would likely increase engagement and perception of relationship. For example, the robot could choose story content that includes the child's favorite animal. Some of this kind of personalization is difficult because it requires a higher degree of "understanding" of what kind of information can or should be referenced in different contexts. The references I had the relational robot use were chosen to be relatively straightforward: mentioning which story was told last time, mentioning whether the child had said they liked the story, showing a picture of the child's favorite animal. These pieces of information were relatively easy to collect using simple questions (e.g., "What's your favorite animal?") or tracking interaction content (e.g., a record of what stories have been told). While it may be relatively easy to add in additional simple questions (e.g., asking about other activities the child likes or things the child has done) or uses of this straightforward information (e.g., in choosing story content), referencing more arbitrary, less structured information (such as what children share through stories or disclosures) would be more difficult.

I think the relational robot could do more with its backstory. Sharing information about the robot's hearing/listening limitations was helpful in setting children's expectations about the interaction. Sharing additional information about the robot's other technical limitations—such as its difficulty correctly identifying emotions or its inability to answer arbitrary questions—would probably increase children's understanding of how the robot works as well as their acceptance of its other limitations. As mentioned earlier, the backstory could also be used to discuss the robot's gender. In general, I think it will important to use the robot's backstory to help children understand what relational technologies are capable of and to help them frame their relationships with relational technologies appropriately, and the relational robot I designed did not go far enough in this regard. Mismatches between children's expectations about the relational robot and its actual behavior could have led to increased disappointment, less rapport, or less strong a relationship over time.

Finally, as talked about above in Section 12.2.4, the framing of the activities with the robot as "practice" versus "play" could have affected children's interest in interacting with the robot. The relational robot did a lot with conversation and story, and framed some of these activities as hearing/listening practice, which some children may have found boring—after all, children want to play! Using more creative, playful activities could help solve this—or in the least, framing the same activities as play instead of as practice might help (a factor of both backstory and dialogue content).

12.6 ETHICAL IMPLICATIONS

My studies have raised multiple ethical questions. One important question involves children's disclosure. During the self-disclosure conversations, children were frequently willing to provide the robot with personal information, such as details about their families and friends, where they lived, how they got to school, preferences for what they liked and didn't like, and activities they were good or bad at. Many children provided this information even after interacting with the robot just once, or just a few times. Ethically speaking, should we be concerned? These data show that social robots have the potential to collect children's personal information. However, it is worth remembering that children are frequently willing to provide similar information to relatively unfamiliar humans as well. The bigger concerns with robots are not that children provide information, but instead, whether this information is stored privately and securely, and who has access to the information later. The next chapter will discuss these ethical questions further.

13

ETHICS OF RELATIONAL AI

The dream of relational AI is to help and support people in being people—to help people flourish, to augment and support human relationships, to enable people to be happier, healthier, more educated, and more able to lead the lives they want to live. Because of the distinct and novel power of relational AI as a social-relational technology, it has the potential to engage people in innovative ways across many domains. But with any new technology come ethical concerns about its appropriate use and potential for misuse. As mentioned earlier, social robots and relational AI may be unique in that they tend to raise many different ethical concerns—most of which are also encountered in other technologies and domains—all at once. Many of the ethical concerns are most contentious with children. All of these concerns are pressing given that social and relational technology is swiftly entering the market. People now live with a wide range of smart devices, AI-enabled digital assistants, personal home robots, and smart toys. Children are growing up with these technologies front and center.

Below, I discuss several of the primary ethical concerns about relational technologies, what we can learn from my data that can be applied to these issues, and implications for the ethical design of relational AI. The final section of this chapter (Section 13.6) summarizes the suggested ethical design guidelines.

13.1 SOCIAL BONDS AND AUTHENTICITY

One main concern about relational technology is that it will replace the social bonds people have or would have had with other people, a concern frequently voiced by Sherry Turkle (Turkle, 2007, 2017). One part of this concern pertains to deception—i.e., whether relational technologies are deceptive in their display of relationship, emotions, and empathy, causing people to think, act, and believe that they have emotional and relational capabilities that they do not "really" have (Coeckelbergh, Fourth 2012; Picard and Klein, 2002; Turkle, 2007).

Questions about deception and authenticity are, at the heart, about the effects of deception on people—i.e., deception is a problem because it causes harm to people. (An alternative framing is, of course, that deception or lying is inherently wrong and a morally unacceptable behavior. Which framing one prefers has more to do with whether one cares more about moral agency—i.e., what actions a moral agent can take—or about moral patiency—i.e., what can cause moral agents to suffer—than anything about deception per se.) One possible harmful effect relates to human attachment to and reliance on relational technology. Will we come to depend on it too much, when we should not, to our social detriment (e.g., Turkle, 2007)?

Coeckelbergh (Fourth 2012) presented a series of arguments about whether emotional robots are deceptive. He suggested that what robot ethicists really mean when arguing about emotional deception is either (1) that the robots intend to deceive; (2) that the emotions robots have are not real; or (3) that the robots pretend to be a kind of entity they are not. He argues that in the first case, it is not the robot that intends to deceive but the robot designer, and that designers often intend to deceive (e.g., in creating movies with talking animals, or in designing experiences at Disneyland where children get to meet favorite characters). No one is fooled that these characters are "real", though, or if they are, this is generally considered an acceptable kind of deception. In the case of robots, the question is whether people are fooled—and then, whether this is necessarily a bad thing. We have seen that many children do report that robots are real, social-relational entities who can think and feel. But more data are needed to determine what that means. Do children believe robots have emotions that are the same as human emotions? Is it the same for a robot to feel happy as for a child to feel happy? Is it bad for children to want robots to be happy? Maybe through feeling empathy for a robot, children can learn to feel empathy for other children, too. This leads into Coeckelbergh's second point.

Coeckelbergh questions what it means to have "real" or "authentic" emotions in the first place. Sherry Turkle, for example, has argued that social robots are inauthentic: they may provoke emotional attachment, trust, caring, and empathy that is not deserved because the relationship and the feelings are not reciprocal—i.e., not real, or deceptive (Turkle, 2007). But what does it mean for emotions to be authentic and for relationships to be reciprocal? Is it that the emotions originate in a biological entity (i.e., real = natural), or that the emotions express some kind of inner self or inner life? In the latter case, robots could perhaps have some kind of inner life or "virtual inner self." They could be built such that they have artificial emotions, and could for all intents and purposes have functional emotions (Arbib and Fellous, 2004; Fellous, 2004; Parisi and Petrosino, 2010). However, most current robots do not have artificial/functional emotions.

The question of reciprocality brings in Coeckelbergh's third point, i.e., whether robots are pretending to be something they are not or have something they do not. Are robots designed to show emotions, e.g., as part of attempting to respond appropriately to human emotions, pretending to be human-like when they should not? This question—like many discussions in robot ethics—assumes that robots are competing with humans, competing with our existing human-human relationships. But should we expect that relationships with social robots will be like our existing human-human relationships? Should we fear whether robots will replace our human-human relationships?

Reciprocality in relationships is the assumption that there is something symmetrical about the relationship. Both parties of the relationship add something of value to the relationship (such as social support, affection, and companionship), perhaps in equal measure, such that both parties get something from the relationship that makes maintaining the relationship worthwhile. Claiming that human-robot relationships are not reciprocal suggests that robots cannot add value to a human's life, or at least, that they cannot add value in an "authentic" way. However, many human-human relationships are not perfectly reciprocal. We can also find analogies in other domains for other kinds of relationships humans enter into that are not entirely reciprocal. Children deeply attach to many non-human entities, including pets, security blankets, pacifiers, toys, stuffed animals, and fictional characters (Melson, 1990; Passman and Halonen, 1979; Weiss, Wurhofer, and Tscheligi, 2009). Children's relationships with these entities are not reciprocal; some are partially imaginary—and yet we recognize that these relationships can add value. Children may frequently tell secrets to their teddy bears. They gain comfort and offer nurturance. With pets, one may clearly see that a child's pet dog "loves" the child, but not in a human-like way at all. People have deep conversations with chatbots and virtual therapists (Bickmore, Gruber, and

Picard, 2005; Bobicz and Richard, 2003; Pontier and Siddiqui, 2008); often, people consider these agents less judgmental than humans (Bickmore, Gruber, and Picard, 2005; Gratch et al., 2007; Lucas et al., 2014; Utami, Bickmore, and Kruger, 2017).

Reciprocality in equal measure is not a requirement of a relationship. Human-robot relationships may simply be one more different kind of relationship. People are capable of having many different relationships simultaneously. Children's robot-friends may be just one new different kind of relation.

Coeckelbergh also pointed out that even humans often pretend to have certain emotions when a display of others would be inconvenient or harmful, and asks whether we should always assume that others have good intentions when it comes to emotional communication. Is it necessarily a problem that a robot that acts as if it feels concern actually does not, from the point of view of subjectively experiencing a state similar to what a human would feel when feeling concern? Is it okay if the robot functionally experiences concern—i.e., that its displayed emotions reflect something about the inner workings of the robot's architecture, perceptions, state, goals, values, and so forth, even if these are not necessarily "human-like" and may not arise in the same way that human emotions do? Or if the robot uses displays of emotion in order to communicate with humans in a way that humans can, generally, easily understand?

I do not have the answers to all these questions, but I do have data from nearly 350 children. I would argue that to really understand whether children are at risk of developing unhealthy emotional attachments to robots (regardless of whether those attachments are similar to other attachments children form), as well as what kinds of opportunities costs there are to children (i.e., is the child missing other opportunities that might be better for the child by playing with the robot, e.g., spending less time with parents or human playmates when more social needs are met by the robot), we need to deeply understand the kinds of relationships children are actually forming. Speculation about the goods and bads of authenticity, attachment, and deception are not enough. Instead of merely asking whether the relationships people form with social robots and relational AI are harmful and positing possible harms (such as losing trust or feeling betrayed upon discovering that a robot is deceptive), we need data to understand what effects these relationships with relational AI actually have. The research I have presented in this thesis is a beginning, though a great deal more research is needed.

For example, we can acknowledge that, based on the data, children do understand that robots are not humans and that their robot-friends are not like their human friends (and tend to have significantly more limitations). Children use a robot's social cues as a cue to its trustworthiness, like they do with people, but they don't treat robots exactly like they do people. Throughout my work, we have seen that children's gaze patterns, descriptions of the robot versus their best friends, their ratings of closeness to the robot versus other entities, their attributions of different psychological and cognitive properties, and much more differ significantly. As discussed earlier (Section 11.2), having these data showing that children are fairly confident about the nature of robots as "betwixt and between" as opposed to confused about what category robots inhabit is important in developing data-driven theories about children's behavior.

We also need to understand the kinds of behaviors robots can use to promote, or inhibit, engagement, trust, emotional attachment, and dependence, and then implement these behaviors judiciously to promote the kinds of robot relationships that we think may be appropriate. Some research, including work presented in this thesis, has begun looking into these kinds of behaviors. For example, using appropriate, contingent social cues can increase trust (Section 3.5 Desteno et al., 2012; Hancock et al., 2011; Lee et al., 2013). Increasing a robot's use of social and relational behaviors such as shared experience, entrainment, and personalization can increase engagement (more discussed in Chapter 12).

When determining appropriate use of these behaviors, the same problems apply to social robots and relational AI as to any other technology. We can have all the ethical debates in the world about what we should and should not do, but it is the designers and developers of these technologies who need to have some moral understanding of what is appropriate and what is not, some ethical guidelines they follow, in order for the technology to be designed ethically. This is partly an issue of education, and is not specific to relational AI.

With regards to mitigating risks of emotional over-attachment or dependence, in a workshop on robot ethics in 2015 (Riek et al., 2015), we jokingly suggested adding "warning labels" to commercial social robots. Like the nutrition labels on food, these labels could let users know the ingredients of the robot: did it have high potential for emotional attachment? Did it use affect recognition technologies for responding in emotional ways to the user? More realistically, and perhaps more effectively, we can use the robot's own behaviors to support the development of positive, healthy relationships. After some amount of time, the robot could insist that the child find a human playmate instead, and turn itself off. The story the robot tells about itself could include reminders that it is just a robot—it has limitations and cannot respond to all of the child's questions, wants, or needs. When the child discloses information that the robot detects may be sensitive, the robot could suggest talking to a trusted friend or adult. These suggestions are in line with what Tim Bickmore has termed meta-relational communication (Bickmore, 2003): be clear about agent's capabilities and limitations and what expectations the user should have about the agent. Use backstory and framing to help communicate this and reinforce this.

The impetus for developing healthy relationships with relational AI is not only on the robot and robot designer, however. Children are also learning how to develop healthy relationships with other children, with their pets, and with many other entities, frequently with the guidance of teachers, parents, older siblings, and other caregivers. Keeping children's caregivers in the loop and encouraging dialogue between children and caregivers about the relational technology in children's lives can help them learn what kinds of behaviors are appropriate.

One important result I found was that social robots as they are now are not all-powerful at drawing children's attention. For example, during the Relational AI study (Chapter 9), for the first session or two, children were very excited about having their turns with the robot. They immediately dropped what they were doing in the classroom to come with the experimenter to see the robot. Later on, however, sometimes children did not want their turns. Sometimes, they wanted to stay in the class and do whatever they were doing there with their friends, and see the robot later. The robot was not magical. The novelty of the robot wore off—children liked playing with it, but it was competing for their attention with everything else the children generally did at school, and some children liked other activities better.

The affect patterns we saw in the Relational AI study (Section 9.4.3.10) mirror this anecdote. Many children showed more positive valence in earlier sessions than in later sessions, but this varied by gender and condition. In different studies and in different experimental conditions, we have seen children's engagement and attentiveness either decline, increase, or stay approximately the same over the course of a single session with a robot or even over the course of 7 or 8 sessions. The particular robot design—such as the actions the robot takes, its appearance, its personalization, and its social and relational behaviors—impacts children's interaction and relationship more than any intrinsic quality of social robots. For example, in Study 5 (Section 4.4), children's positive affect increased over 7 sessions when the robot used personalized affective feedback, but decreased when it did not. In Study 6 (Section 5.5), children's engagement increased when the robot was more expressive. In Study 7 (Chapter 8), children's attention increased from the first half to second half of the study in the *Entrainment* condition, but decreased in the *No Entrainment* condition.

If we want to design social robots that children want to continue interacting with over time (e.g., for education or health applications), we need to pay careful attention to the qualities that appear to increase their engagement, attention, interest, and relationship. Based on the research so far (discussed further in Chapter 12), these qualities include change and variation over time, space for play and creativity, personalization to the child, nonverbal immediacy, and appropriate relational behaviors, which will likely need to be tuned to a variety of individual differences, such as gender and personality.

13.2 SOCIAL MANIPULATION

A second ethical concern pertains to social manipulation and persuasion. Technologies often mediate and implicitly shape human interaction with and perception of the world, by encouraging or inviting some forms of actions while discouraging or inhibiting others (Verbeek, 2006). Some HRI research has focused on creating robots for behavior change—e.g., to help someone with particular weight loss goals to stay on track (Kidd and Breazeal, 2008). This kind of change is usually considered acceptable: it is a "positive" behavior change, with the goal of helping people achieve what they want to achieve. When used for "good," then persuasion, in a robot or in a human, is often seen as a positive attribute—it gets us to the end we want. If used for "bad," it is another story altogether. For example, robots that are used to provide the elderly with shopping assistance may be seen as beneficial (Iwamura et al., 2011), but those that target potential customers may raise some eyebrows (Kanda et al., 2008). The IEEE guidelines for Ethically Aligned Design address some of these issues [1].

There is also another aspect, which is that some human behaviors may seem unacceptable when placed in another entity. People are socially manipulative and persuasive all the time with each other—this is part of how social interaction works. However, while it may be acceptable for the car salesman to use conversation and rapport tactics to up-sell expensive features for a new car, a robot that does the same thing could be considered alarming. This relates to the existential threat robots pose to us, which I discuss further below in Section 13.5.

The question here is whether being social manipulative or being persuasive is acceptable for technology, and if so, to what degree. Verbeek (2006) argued that persuasion is not an intrinsic property of any technology, but comes from both the designer and the user. He argued that we should assess potentially persuasive technologies on three fronts: (1) whether the intended persuasions are morally justifiable, e.g., that

1 `https://standards.ieee.org/industry-connections/ec/autonomous-systems.html`, retrieved February 7, 2019.

they do not cause harm, and promote beneficence or justice; (2) that the methods of persuasion used are morally acceptable, e.g., that they respect human autonomy; and (3) that the outcomes or consequences of persuasion are morally justifiable. The biggest problem, here, however, is that people are not going to agree on what is considered morally acceptable or morally justifiable. People's opinions depend on their particular moral and ethical philosophies, which vary widely.

We can look to other domains for inspiration on how to handle these ethical quandaries. Marketing and advertising are two domains that frequently raise similar questions about social manipulation and which face similar problems regarding lack of consensus about what is ethical behavior (Drumwright and Murphy, 2009). Some marketing agencies have adopted codes of ethics promoting transparency, honesty of relationships, opinions, and identity—i.e., promoting the idea that they should make sure consumers know when they are being advertised to. Relational AI could follow this example of promoting transparency and honesty, e.g., using backstory or dialogue to explain to users what it is capable of and what its goals are for others' behavior. Informing users can be helpful, but there are additional questions we can raise about how much consumers can really trust the designers of technology, who is held responsible for the behaviors of users, how we can be sure that a technology will not have undesired effects as well as whatever is advertised, and whether being persuasive is in itself an ethical thing to do.

These questions are difficult to answer. Perhaps one response is to involve philosophers and ethicists more directly in the design of future relational technology. We need designers of technology to be aware of the ethical and moral issues involved in the things they are creating, and to attempt whenever possible to create technology that supports and affirms people in becoming who they want to be—that supports human flourishing.

One final point about transparency pertains to the use of technology in research studies. In research studies, participants (children and adults alike) should be fully debriefed at the end of the study regarding the technology's capabilities, such as whether the robot was in fact teleoperated, whether any aspects were human-controlled, and what sensing, modeling, and reacting capabilities the technology has. In our studies, we have always fully debriefed parents or adults, and generally only debriefed children when they expressed interest. We could go farther in the future toward ensuring that the robot's capabilities are fully explained to children. For example, children are generally made aware that the robot uses a camera to see them and a microphone to hear them, but we may not explain at the end of the study how the robot uses that information in behavior selection, unless the child or parent specifically asks about it. This kind of additional explanation could be considered part of helping AI education reach a broader audience.

13.3 ADDICTION AND HOLDING POWER

In her earlier work, Sherry Turkle discussed the holding power of simulated worlds and video games, which was in large part related to their interactivity, challenge, and rule-based structure (Turkle, 1985). The early games she discussed, such as Space Invaders and Asteroids, are much simpler than the far more immersive, interactive, and intense video games of today, but the risks she discussed around addiction to games and the games' holding power are real. Current video games and internet games are often addictive by design, using the same techniques as, e.g., gambling, to provide intermittent rewards that influence a player to want to keep playing and keep getting that reward (Kuss and Griffiths, 2012). Internet gaming addictions are similar on a neuronal and biochemical level to other addictions, such as cigarettes. Cigarettes, according to a 2014 report from the US Surgeon General (ASPA, 2014), are also addictive by design—the tobacco is treated such that the nicotine released can more quickly cross the cell membranes in the lungs. The question is, should we create things that are addictive, with strong holding power? The answer rests on one's personal philosophy: should we respect freedom of choice (e.g., to buy and play addictive games)? Or should we protect people from themselves (e.g, so they do not unintentionally fall into debt because they bought too many in-game items)?

Like with most new technologies, some people question whether social robots will also be addicting. But robots are not video games. Robots are embodied in the world. As social agents, they are not a whole interactive world the way a game is. Robots have the potential to lead the user into an imaginary world, which could have gameplay and reward systems, but only so far, because on their own, robots are still situated in the real world. They can only go so far into virtual realms. This is an intriguing difference between embodied, physical, present agents and virtual or telepresent agents. Virtual agents can accompany users through 2D and 3D virtual worlds and virtual realities, as in video games. They can be characters in interactive, enticing, immersive settings. Robots, however, may not have as much holding power because of their physical presence. The story becomes more interesting (though in a good or bad way remains to be seen) when we start to discuss new form factors and ways of migrating an agent between different embodiments—as discussed in (Kory-Westlund, Breazeal, and Ostrowski, in review)—since this could allow a robot to connect users to more powerful and alluring virtual realms. All that said, designing a robot for addiction is different than designing a robot for empathy. Designers of social robots should consider what goals they are designing towards—and hopefully, design with empathy and human flourishing in mind, rather than addiction.

Robots can allow for more open-ended play than video games as well, particularly more so than the early video games Turkle analyzed in her work. While often still rule-based and with limitations on content and activities, robots can provide space for creativity and exploration in a way that video games cannot. This was the idea of my early storytelling activity (Kory and Breazeal, 2014), based on earlier work with SAM the Castlemate (Ryokai, Vaucelle, and Cassell, 2003), where the robot provided a context for storytelling and the child was invited to make up their own stories. This was a limited scenario, but could easily be made more open-ended, such as allowing the child to choose the story scenes or create their own, or equipping the robot with narrative generative capabilities, so that it could co-create new stories instead of relying on an existing story corpus.

13.4 PRIVACY, SECURITY, AND SAFETY

Another set of concerns are about safety, privacy, security. Physical safety is the easiest to address, and seems the smallest issue for social robots, since we already have adopted the principal of "safety by design" in manufacturing artifacts for our personal use (though physical safety is certainly a concern for military robots, especially regarding autonomy—see (Arkin, 2009; Lin, Bekey, and Abney, 2008)). Emotional safety was discussed earlier, in the considerations of deception and emotional attachment. However, the capability of social robots to monitor and surveil beyond the capacities of human sensing (e.g., through the use of infrared or ultrasonic sensors, or at ranges or distances unavailable to humans on their own) is concerning. Ryan Calo describes three areas of privacy that we should be concerned about: direct surveillance—i.e., robots that magnify the human capacity to observe; increased access—e.g., new access to historically protected spaces, like inside homes; and social meaning—e.g., people may act differently as a result of feeling observed and evaluated (Calo, 2010). These issues are not unique to social robots or relational AI—they arise with many current technologies, such as laptops and devices in the internet of things (Arnold, 2010; Goldman, 2015).

The first question to ask with regards to relational AI for children is whether children have a right to privacy, and if so, from whom. Children seem to think privacy is as important as adults do, but their conceptions and expectations about privacy may be very different than adults (Shmueli and Blecher-Prigat, 2011). However, Shmueli and Blecher-Prigat (2011) argue that while it is generally assumed that parents will protect children and are assumed to act in children's best interests, children's privacy tends to be undervalued in family situations. There is unprecedented surveillance of young people—e.g., automatic monitoring of texting, geolocation tools that allow parents to track their children via their cell phones, recording of website and browser history for parental review, blacklisting sites (Arnold, 2010). Most surveillance of children is used specifically when parents are not present or are busy (e.g., blacklisting websites or geolocators in phones). While the technology itself may be neutral, this use of it is not. Parents likely see themselves as protecting their children against potential threats—cyberbullying, hacking, harmful web content, so forth—and to some extent they are justified. The need here is for compromise between granting children their privacy, and protecting them and providing safety.

However, children greatly value their privacy—their physical integrity, and their freedom to be unobserved during play or other activities, communicative privacy if they have a diary, being able to speak without close supervision, and so forth. Even Piaget acknowledged that autonomy is very important for children's psychological wellbeing (Arnold, 2010). As much as possible, we should err on the side of treating children as people. Although they are young, have less capacity to regulate their data, enter informed relationships, and may not be fully independent yet, they are still people and should be respected as such. Furthermore, children may resent parental "snooping." Privacy is about dignity, respect, and trust. Surveillance is used in place of trust; it allows observance and control, but not trust (Shmueli and Blecher-Prigat, 2011). Surveilling a child denies that child trust—without being trusted themselves, will they learn to trust? Perhaps one ought to encourage children to seek out guidance when they are uncertain or perceive threats, rather than overestimating the dangers and imprisoning children.

A bigger question, however, is how any information collected may be misused. Vendors who offer surveillance services, companies that sell social robots that use these data either to track children or just to personalize the robot to the child or to the family, anyone who collects or stores data, could mine this data, could sell it, or could lose it to a third party. Surveillance can be misused, whether it is surveillance by a social robot, surveillance by parents, or by another party.

With regards to social robots and relational AI, the sheer amount of data that could be collected about a user—about a child—is enormous. Much of it is highly personal data; some of it may even be considered medical data (e.g., one's heart rate variability or other physiological data as measured with a webcam (Kwon, Kim, and Park, 2012;

Poh, McDuff, and Picard, 2011), which in the robot may be used to determine if the user is cognitively stressed (McDuff, Gontarek, and Picard, 2014), so that it can provide an appropriately supportive response). What data are allowable for a robot or AI to collect? Who should have access to the data? How can the data be protected?

The data that are allowable should be determined by three factors: first, what data are needed for the robot to fulfill its tasks; second, only data that can be sufficiently protected; and third, only data that are acceptable to the users of the robot. Similarly, the individuals who should have full access to any data collected should also be determined by several factors: first, if it is necessary for someone to view the data for the robot to fulfill its tasks; second, if an individual about whom the data has been collected wishes to view it, or wishes it shared. This is necessarily in the direction of placing privacy and control in the hands of users; users should as much as possible be in control of the data collected about them. Here are a few examples to illustrate these contextual and situational factors.

If the robot is a medical device that is helping doctors to learn what is medically wrong with a child in order to help treat that child, it may be acceptable to collect physiological data for the child's health records, and for that data to be HIPAA-protected (Drew Simshaw, 2015). Furthermore, if the robot is used to diagnose mental disorders, then it may have a mandatory reporting policy about anything it learns from the child. The older the child, the more they should be informed about what data are collected and why. An intriguing and difficult question arises if, at some point in the future, a robot was better at sensing and helping doctors diagnose certain medical conditions than any other method available. Would the child then be allowed to maintain privacy and refuse data collection if their parent decided it was best for their health? It seems unlikely. Another question pertains to children's trust in the robot. If they are unaware that the robot will report back to a medical professional, they may share secrets with the robot that they would not otherwise have shared. This presents a dilemma (some dilemmas described in (Bethel, Stevenson, and Scassellati, 2011; Bhakta et al., 2014; Kory-Westlund and Breazeal, 2015a)). Should children always be told up front that a robot may "tell" on them? What if the benefits from not telling—such as a medical professional gaining valuable information from the child about his or her mental and emotional state—outweigh the costs of that deception? It may be that after one such "betrayal," the child may refuse to tell the robot anything again.

If, however, the robot's role was as a learning companion, situated in a classroom or home, then perhaps its requisite data to collect is less; it may need to record right and wrong answers to report back to the teacher, and it may need to use video and audio data in the moment in order to respond appropriately, but it may not need to record anything else. That said, when nine teachers in a Boston-area preschool were asked whether they thought a social robot in their classroom should collect data about children, nearly all agreed that it should, and that it should report data back to them about the children it interacted with, not just regarding their academic performance but much more, such as their academic skill development, attention, participation levels, and social skills such as eye contact and verbal responses (Kory-Westlund et al., 2016b). They also generally thought that preschool children should not have access to the data the robot collected, though one teacher suggested that children's access should be dependent on grade level and that older children should be told more. Teachers were split over whether children should be told that their teachers could access the robot's monitoring.

Other intriguing questions arise around the case of long-term interactions, since relational AI necessarily considers long-term relationships. In any long-term relationship, we expect that our interaction partner will learn and change over time. We expect our partner to remember us, to remember shared experiences, and to change as a result of those shared experiences. If a robot is to be situated as a long-term companion, in a healthcare setting, in a classroom, or elsewhere, it will need some kind of memory. This entails data collection about very personal and personally identifiable information.

Also related to long-term interaction, particularly in the home, is another aspect of privacy as defined by (Calo, 2010): social meaning. People react to the sense of being observed and evaluated. We treat social robots as we do humans, both psychologically and neurologically (Meltzoff et al., 2010; Stenzel et al., 2012). The presence of social robots in homes may reduce time for solitude, interiority, and self-reflection. But this time is important—we require solitude in some quantity; we require moments "off stage" to be ourselves (Csikszentmihalyi, 1997). Perhaps we need to limit the amount of surveillance and time spent with social robots not simply because these data could be misused, but for our own wellbeing.

As has been related, there are many problems and ethical concerns surrounding privacy and security for relational AI. How can we solve them?

There are several levels here: (1) what data can companies collect about people; (2) how do companies store and transmit the data they collect; and (3) who else has access to the collected data. Regarding the first (what data can be collected), we need security and privacy by design in products, as well as the existing safety by design. We may need regulations to make this happen. For example, we might regulate surveillance, and what data can be collected about, e.g., consumers by social store robots, or we might amend the Electronic Communications Privacy Act to require warrants to access any data from inside homes (Calo, 2010). Roboticists could adopt an ethical code similar to codes that professionals in other fields follow that emphasizes privacy, accuracy, intellectual property, and access (Calo, 2010; Riek and Don Howard, 2014).

Regarding the second level (how data are stored and transmitted), foremost, we need accountability from the companies and other legal entities who hold protected data, such as the companies that produce social robots or companies that provide necessary services that social robots utilize such as speech recognition tools. Federal regulation is needed. This could be similar to HIPAA-protected data regulations. If a company has shown negligence when data is lost, they might have to pay a large fine—this might motivate them to actually protect their customers' data. Cases such as recent Target and Home Depot credit card hacks show that many companies are still using outdated operating systems and have few mechanisms in place for detecting intrusion into their systems or auditing access to their servers (*Home Depot Hit By Same Malware as Target* 2014; *Home Depot's Suspected Breach Looks Just Like the Target Hack* 2014), and furthermore, these were not the only large hacks that year (*The Big Data Breaches of 2014* 2015). Standards need to be imposed that apply not only to how data are stored, but also how data are transmitted, involving secure, encrypted connections that match to industry standard security practices. That said, even encrypted data is not entirely secure. Since current networks and systems were not designed from the ground up for security, observers can gain metadata about messages transmitted even if they cannot access the contents of the encrypted messages. Knowing to whom data goes, when it was sent, and how much data was sent is still very valuable information, and there is no current solution to this.

Further regulations we may need include mandatory disclosure of data loss within a tight timeframe, so that consumers and users will be aware when their data has been compromised. We also need intrusion detection and auditing mechanisms for servers and data centers, so that when data breaches do occur—and they will—companies can actually figure out what happened using data forensics. Finally, there should be some threshold for small businesses for following all these guidelines (e.g., if they have fewer than a thousand users), so as not to discourage small businesses; in addition, small businesses tend to carry much less risk, and are less often targets.

Given that data breaches will occur, and now involving children's data in addition to adults' data, how do we mitigate the risk? Imposing standards regarding security, encryption, data forensics, and so forth as discussed above will all be a good start. We may also need to consider whether some technology ought *not* to share all the data it obtains about users with the company that creates it—that is, whether some data used, e.g., for the personalization of behavior in response to a user ought to be kept privately within the technology and not transmitted back to the parent company. This may make targeted attacks on particular companies less likely to reveal highly personal data, since the company would not directly have access to that data.

Another point here is that users and consumers should be aware of what data are being collected about them. Reynolds and Picard (2004) found that participants felt their privacy was more respected when a web service using affect detection provided an ethical contract than when it was not. This may also be due to the level of transparency; they may have felt more trusted, rather than deceived through the omission of a contract.

For the third level (who has access to the data), there are two big concerns. One is advertising. When data are sufficiently protected, how can advertising work? Currently, many companies sell personal data to advertising and marketing companies; some of it is likely in the category of data that should be more protected than it currently is. To some extent, this is valuable for both the advertisers and the consumers—it is more useful all around for people to see relevant ads. Perhaps we can still get sufficiently targeted ads without giving access to all our personal data. Perhaps we could have "opt-in" options (though then, the option may become part of every EULA that we do not read and simply click "agree" to so our software will install). Personally identifiable information can be removed before it is sold, but sometimes identities can still be reconstructed (de Montjoye et al., 2013, 2015).

The other big concern is government. Right now, regardless of how well a company or organization protects its data, the data can be forcibly and secretly obtained by the government through National Security Letters and similar warrantless searches (which are unconstitutional due to their violation of people's privacy and the freedom of speech of the entity holding the data records) (*National Security Letters*). Imagine the consequences if the government became hostile to certain sub-populations within the country, and used its power to obtain all of their data. Furthermore, given the lack of unbreakable encryption standards, the government may also then be able to hijack smart devices, including those that interact with children, in order to obtain additional data or to interact directly—e.g., in order to question children about their beliefs or their parents' beliefs, provide disinformation about others, or persuade children to inform on others. While perhaps an extreme case, it serves to point out that data needs legal protection from the government. The political climate in the United States, and in other places around the world, needs to change, such that politicians realize the need for unbreakable encryption into which the US government does not get a secret back door (if they can break in, others could, too). The government does not need nor should it have access to our data or our children's data.

Furthermore, if the US government currently thinks it should have access to the data held by a multi-national company, what might happen if another country where that company operates enacts a similar statue like the National Security Letters? Then the company might be left in the unfortunate position of providing data about US citizens to a foreign government, and about that foreign government's citizens back to the US government—and neither government would likely tolerate this. This further serves to point out that our current legal protections and regulations need to be amended to deal with current technological concerns, including all those raised by social robots and relational AI (Calo, 2010).

These are just a few steps we could take toward ensuring the privacy and security of user data. No doubt there are many others, and no doubt it will take some time before these measures are enacted.

13.5 EXISTENTIAL THREAT

Some of the fears people have about social robots are existential. We do not want to be replaced—we want to hold on to our unique, special place in the universe, the sole beings that are human. Social robots encroach on our uniqueness as the only beings we know of that could use language, that could use our sophisticated social cues and have our level of social intelligence (though many other animals use social signals and have social groups, e.g., (Moore, 2013; Tomasello et al., 1997)), that could have our level of creativity, and so forth. Some social robots are designed to look like us (Nishio, Ishiguro, and Hagita, 2007). We have created sculptures, animations, puppets, and other artistic and life-like representations of humans before, but never in a form that could move and talk in a way that is as much like us as these android robots could be (note that right now, their movement and likeness, while pretty good, is still lacking).

I see a parallel here between this existential threat robots pose and to an observation Sherry Turkle made about children learning to program the computers of the 1980's (Turkle, 1985). She wrote that some people found young children who manipulated machines, manipulating the abstract and symbolic before they could properly hold a pencil and write, disconcerting. It was counter to the image they held of children as innocent, juvenile, unable to read or write until a certain age. Computers were a cultural symbol of a "loss of direct contact with other people, the construction of a private world, a flight from real things to their representations" (p. 95). In the same way, social robots now pose a threat to our image of ourselves as special.

However, existential threats are not new. There was resistance to the Copernican revolution, when it was revealed that the solar system is in fact heliocentric, not geocentric. There was resistance to the theory of evolution, and the idea that humans were not specially designed and placed on the Earth—we, like all other creatures, had simply evolved. We discovered that chimpanzees use tools—so we are not the only tool-users on Earth. And so forth. Throughout history, there are cases where some scientific or philosophical revelation has pointed out that humans are not quite as special and unique as we thought we were. Perhaps social robots are just one data point in this trend.

In addition, many of the ethical questions about a technology that looks like us or acts like us, but is not us, have been previously explored in nearly a century's worth of science fiction. If social robots truly are new in their similarity to humans in appearance, language, and intelligence, we can at least look to science fiction writers to see some possible problems, consequences, and solutions (e.g., Asimov, 1942, 1954, 2004).

13.6 ETHICAL DESIGN GUIDELINES

In summary, I make the following recommendations regarding the ethical design of relational AI and related technology:

- Design responsibly. Involve philosophers and ethicists, who have specific training in relevant ethical and moral frameworks and applications, in the design of new technology. Design with empathy and human flourishing in mind rather than strictly for addiction or profit. In particular, with children, consider technology that enables open-ended play and provides space for creativity and exploration.

- Be informed by data as well as theory. An increasing number of research studies are exploring questions highly relevant to the ethical design of relational AI, such as questions about engagement, trust, and attachment. We need to use the data from both human-human studies and human-agent studies to learn how people actually form relationships, develop trust, and interact with relational agents, and use these data to inform future design. Chapter 12 discussed this further.

- Involve all stakeholders. In particular, when designing relational AI for children, involve children's caregivers. Children look to their caregivers for information about technology and model their behavior and attitudes when interacting with technology.

- Be transparent and honest. Inform users about what a technology can do and what it will do. Use the technology's packaging, introduction, framing, and backstory to share information and set user expectations appropriately about the technology, its capabilities, and its limitations. Verify that users actually understood the technology's capabilities and limitations.

- Implement security and privacy by design as well as safety by design. Collect only data that are needed for agent to fulfill its tasks, only data that can be sufficiently protected, and only data that are acceptable to users. Be transparent about what set of data are collected, how data are stored and transmitted, and how data are used.

14 CONCLUSION

14.1 CONTRIBUTIONS

14.1.1 *Technical*

The primary technical contribution of this thesis is the development of a relational AI system. As part of the Entrainment/Backstory and Relational AI studies (Chapters 8 and 9), I developed an autonomous social robot with relational AI that played storytelling games, adapted the level and content of the stories presented to children, selected dialogue based on each child's exuberance level (using features such the child's speaking rate and rate of answering questions), mirrored affect and some aspects of posture, performed appropriate gaze and lookat behaviors, and entrained its speech to match the child's, among other behaviors. Importantly, I developed a speech entrainment module that listens for an audio stream containing speech, morphs speech audio files such that they match the speaking rate and pitch of the heard speech, and streams these files to the robot for playback.

As part of the autonomous robot system, I created a web-based experimenter interface to facilitate study session administration. This interface allowed the experimenter to select the appropriate participant ID number before starting the robot, which was used to tag all data and load appropriate configuration files; it also allowed the experimenter to pause or stop the interaction with the robot on demand. The interface enabled experimenter input during several key assessments during the interaction to ensure appropriate robot behavior (such as the Negotiation Task and the Extra Picture activity). For the long-term interactions in the Relational AI study (Chapter 9), I compiled a new story corpus comprising stories that children had told the robot during prior work 4.3, and used this in accompaniment with two other existing story corpuses, I created numerous scripts for long-term interaction with the robot, which including varied dialogue content for all sections of the interactions, and varied tasks and activities. I recorded many new utterances for the Tega robot.

Many of the digital materials I created are available online. The open-source software is available at:

- Audio Entrainer: The audio entrainment module used in my work. `https://github.com/mitmedialab/rr_audio_entrainer/` under a GNU General Public License v3.0.
- RR messages: A ROS package with messages for the relational AI projects. `https://github.com/mitmedialab/rr_msgs` under an MIT License.
- Opal: A Unity tablet app used in multiple studies that can, e.g., display storybooks. It is ROS-enabled to allow for remote control. `https://github.com/mitmedialab/SAR-opal-base` under an MIT License.
- Opal messages: A ROS package with messages for talking to the Opal app. `https://github.com/mitmedialab/opal_msgs` under an MIT License.

- Text analysis tools: Scripts for determining exact and similar matching phrases. https://github.com/mitmedialab/text_analyses_tools under an MIT License.

- RR tools: Various tools and scripts for analysis and processing for the relational AI projects. https://github.com/mitmedialab/rr_tools under an MIT License.

- ASR: Uses the Google Cloud API to process audio and send results over ROS. https://github.com/mitmedialab/asr_google_cloud under an MIT License.

- Tega Teleop: A python ROS node for teleoperating the Tega robot. https://github.com/mitmedialab/tega_teleop under an MIT License.

Other study procedures, materials, and assessments are available on figshare at:

- Relationship assessments—the Social-Relational Interview, Narrative Description, Inclusion of Other in Self task, and Self-disclosure task (Kory-Westlund et al., 2018): 10.6084/m9.figshare.5419102.

- Additional relationship assessments—the Picture Sorting Task and Social Acceptance Questionnaire: 10.6084/m9.figshare.7575911.

- Study 7: Entrainment/Backstory procedures and assessments: 10.6084/m9.figshare.7175273.

- Study 9: Relational AI procedures and assessments: 10.6084/m9.figshare.7627289.

14.1.2 *Empirical*

This work contributes new data for understanding children's conceptualizations of social robots and the kinds of relationships children form with social robots over time. It builds on a growing body of literature examining social robots as companions, peers, and tutors for young children. I investigated several questions regarding (1) how social robots can use social-relational capabilities to engage children as peers, thus affording opportunities for peer modeling and learning; (2) how children's construal of and relationships with social robots develop and change over time; and (3) how children's relationships and rapport with social robots affect their long-term engagement and learning.

I collected a unique dataset about children's relationships with a social robot over time (Chapter 9), which enabled me to look beyond whether children liked the robot or not or whether they learned new words or not; after all, most children report liking robots, especially if they interact only once. The results showed that we can use relational AI to enable a robot to act in a more social and relational manner, and that the robot's relational behavior can impact children's perception of the robot, their relationship with it, and their learning. The main findings include:

- Children in the *Relational* condition reported that the robot was a more human-like, social, relational agent and responded to it in more social and relational ways. They often showed more positive affect, disclosed more information over time, and reported becoming more accepting of both the robot and other children with disabilities.

- Children in the *Relational* condition showed stronger correlations between their scores on the relationships assessments and their learning and behavior, such as their vocabulary posttest scores, emulation of the robot's language during storytelling, and use of target vocabulary words.

- Regardless of condition, children who rated the robot as a more social and relational agent were more likely to treat it as such, as well as showing more learning.

- Children's behavior showed that they thought of the robot and their relationship with it differently than their relationships with their parents, friends, and pets. They appeared to understand that the robot was an "in between"entity that had some properties of both alive, animate beings and inanimate machines.

These results provide evidence for links between children's imitation of the robot during storytelling, their affect and valence, and their construal of the robot as a social-relational other. A large part of the power of social robots seems to come from their social presence and social power. This work builds on prior research linking children's peer learning to rapport and relationships (Gola et al., 2013; Richards and Calvert, 2017; Sinha and Cassell, 2015a,b).

In addition, children's behavior depended on both the robot's behavior and their own personalities and inclinations. Girls and boys imitated, interacted, and responded differently to a robotic agent with social-relational capabilities and to one without. These gender differences are reflected in multiple prior studies (Baylor and Kim, 2004; Burleson and Picard, 2007; Kennedy, Baxter, and Belpaeme, 2015; Kim and Lim, 2013; Pezzullo et al., 2017), suggesting that we ought to pay greater attention to children's gender and individual differences when creating new technologies to engage and support them.

Through this work, I found that even a fairly scripted interaction without complex planning or machine learning capabilities was able to engage children over many sessions and help them learn and grow. Complex real-time AI, dialogue planning, and so forth is not always needed for successful, engaging interactions, as Bickmore (2003) previously suggested.

The studies reported in this thesis also demonstrate the potential of social robots as a tool for psychologists and social psychologists to use to study human relationships and human interactions. Robots afford more fine-grained control of social behaviors than we can achieve with human actors, thus enabling us to test, e.g., the impact of particular verbal or non-verbal features on engagement, trust, and learning (Desteno et al., 2012; Kory-Westlund et al., 2017b) (also Chapter 8).

These studies also provided important data that can inform the design of new relational technology. I provided design recommendations (Chapter 12), and discussed how we can address several important ethical concerns regarding relational technology (Chapter 13). I also present a model showing how social behavior, relational behavior, engagement, rapport, relationships, and learning are linked (Section 11.1).

Finally, as part of this work, I developed and tested numerous new assessments for understanding and measuring children's relationships with social robots and other entities. They are available online from the links provided above in Section 14.1.1. These assessments can enable researchers to better understand the relationships children form with social and relational technologies.

14.1.3 *Theoretical*

The data I collected can inform our theories about how children construe and relate to social-relational technology. As discussed in Chapter 11, children's understanding of current social technology shares some similarities with earlier observations about children's interactions with early computers and computerized toys (Kahn et al., 2006; Turkle, 1985; Turkle, 2005). Today's social robots and other social-relational agents are far more complex than the agents studied by Turkle and Kahn, but children do not seem to be more confused or disturbed by them—in fact the opposite. Children today seem more accepting of social technology as interactive agents that are in-between the usual dualistic categories of alive/animate beings and inanimate artifacts, perhaps because they are growing up with these agents present in their everyday lives.

Children's interactions with social robots are driven by the immediacy of the interaction. They are in the moment, responding socially and naturally to an agent that engages them socially, and do not need to reflect on a meta-level about what that agent "really is" during interaction. However, even when children step back and think about what robots are, they still report that robots are social-relational beings. Children's opinions seem to be shaped in part by the others around them, perhaps more so than children who interacted with early computers, since technology is far more ubiquitous today. Adults, parents, experimenters, teachers, and other children influence children's mental models about robot cognition and intelligence, and affect their view of robots as social and moral entities.

Because of the complexity and increasing social-relational capabilities of current technology, there are an increasing number of ethical questions involved in the design and deployment of these technologies. I provided a discussion of several critical ethical issues and some suggestions for how we might address these issues in future designs. Many of these ethical issues are not new, but social robots with relational AI appear to be special in that they raise many different ethical concerns all at once.

The data I collected also inform our understanding of how children learn language and how children interface with their peers. In developmental psychology, when investigating children's language learning, speakers have generally been treated as equal—i.e., one speaker is generally treated as as good as another. There has been less interest in examining different characteristics of individual speakers, so long as there is a quality speaker providing decent language input. However, based on my research and related work examining the impact of children's trust, judgments of credibility, rapport, and engagement on their language learning, I hypothesize that not all speakers are treated equally by children. Children will not imitate the language of peers or adults equally, and will not consider all of them to be equally trustworthy informants. The studies I report here provide some evidence for this hypothesis. For example, children tended to score more highly on vocabulary posttests and use more target vocabulary words in their stories when they thought of the robot as a greater social-relational other—such as rating the robot as closer to themselves, rating the robot as more social, and using more social behaviors such as saying goodbye to the robot. Children's perception of an agent's social-relational qualities and the agent's nonverbal immediacy seem to impact their engagement and learning.

14.2 FUTURE WORK AND OPPORTUNITIES

As discussed in Section 9.5, relational AI provides new opportunities for engaging children in social learning activities, and in particular, engaging girls. Future studies could explore what kinds of relational behaviors are more likely to engage boys, as well as how different behaviors influence children's perception of the robot's gender. We should ask children what they like most and least about the robot to gain a better understanding of what boys versus girls thought about the robot. We could explore whether the robot can provide a backstory about its gender (or lack thereof) that will encourage children to treat the robot one way or another, and whether the robot's other behaviors can "override" its backstory with respect to how children construe it as a gendered agent.

The Relational AI study examined differences between a relational robot and a non-relational robot. One important next step will be to examine the contributions that different kinds of relational behaviors make to child-robot relationships, and to children's engagement and learning. Relational behaviors could be divided up in many different ways, for example, by timescale (e.g., behaviors developing in the present on shorter timescales in matters of seconds or minutes, versus behaviors that develop over longer times, such as days, weeks, or years) or by modality (e.g., verbal vs. nonverbal cues, linguistic vs. non-linguistic). It will be important to assess, for example, whether the socially contingent aspects of being relational—such as audio entrainment and affect mirroring, which both happen close to the present as opposed to over a longer span of time—are sufficient to promote learning and engagement, or whether we need "more stuff." We need to determine how necessary memory is, and then what kinds of memory are needed—for example, memory of facts, events, emotions, and personal preferences or other details. Memory, in particular, has ethical implications regarding privacy and data storage. We could also examine whether, e.g., providing the robot with a "true" back story (e.g., the robot was built in the lab, visits schools, and so forth) versus a fictional back story (e.g., robot is from outer space, here to visit earth and learn) has different effects on the relationship. How does the robot's backstory impact the robot's credibility, or the perception of the robot as a character?

A great deal of work could be done to increase and improve the relational behaviors used by relational AI. Because we have evidence that children's rapport and relationships are affected by the robot's behaviors, and also differ as a result of children's individual differences, we could build classifiers to learn what kinds of behaviors by children—such as gaze patterns, affect, entrainment, and dialogue patterns—predict closer relationships with the robot. We could use the dataset from the Relational AI study (Chapter 9), which includes videos, audio, as well as ground truth from the various relationship assessments, as a start. Using this kind of classifier, we could develop relational AI that can autonomously categorize, adapt, and adjust its relationship state during longitudinal interactions with children.

Long-term interactions with children raise many interesting questions. For example, the robot may ask children questions about their knowledge or preferences. Sometimes children's preferences are stable for long periods of time, and sometimes they change frequently. What knowledge and which preferences are stable, and for how long? At what point in time should the robot probe again to gain more information or reassess what it knows about the child? How should it go about reassessing? We we can also learn more about what keeps people in relationships. What are the contributions of change, affiliation, liking, desire for understanding, sharing and hearing stories, and so forth? At what point do relationships move from novelty to familiarity and habituation, and how does novelty continue to play a role in relationship continuation (e.g., in the form of change and new stories shared from when people are apart)?

These kinds of question show that we may need to measure more aspects of relationships. The relationship assessments developed, tested, and used here measure some aspects of children's relationships; there may be some overlap in the constructs measured and some features of relationships that were not sufficiently captured. We could draw on work on mother-child dialogue, which found variations in the connectedness displayed in conversation (Ensor and Hughes, 2008). Connectedness refers to the heart of the conversation: how much things each person says are picked up by the other. We could examine relationships between a child's temperament and how they interact, such as prosody, expressivity, or willingness to respond verbally to questions. Performing more detailed interviews, like those performed by Turkle with children in the early 1980's (Turkle, 1985), could provide new insight into how children's relationships with technology form and progress.

Children are frequently accompanied by others in educational contexts, such as friends, siblings, parents, and teachers. They learn from all these others. It will be important to explore how group interactions affect children's learning and development of a relationship and rapport with social robots. It will also be important to assess transference. E.g., could playing language and storytelling games with robots motivate children to be more excited about language learning in other contexts as well? How well might other skills children learn or practice with robots, such as social skills, transfer to interaction with other people? Answering these questions will help us learn how to integrate robots into existing educational contexts, such as homes and schools.

Fish and Pinkerman (Fish and Pinkerman, 2003) discussed how high-SES parents tend to repeat and expand on what children say, following the child's interest in a sensitive and responsive way and using more affirmations, while lower-SES parents tended to repeat their own utterances and use more commands and prohibitions. This observation suggests two possible interventions. First, the robot may be more effective if it acted more like a high-SES parent—that is, affirming the child's natural curiosity, letting the child drive the learning, and providing additional information when the child shows interest. Second, the importance of maternal interaction and parental support for children's language learning cannot be understated. Future work should examine how the robot can bring parents into the interaction and perhaps teach or model contingent responsiveness and other beneficial behaviors. For example, one project has already demonstrated that a virtual character in a digital storybook can model dialogic question asking for parents, thus prompting parents to ask more questions when reading with their children themselves (Boteanu et al., 2016). The issue of how to help lower-SES parents better support in their children's early language development is receiving a lot of attention, with interventions being developed and tested by organizations such as the Thirty Millon Words Center [1] and the Habla Conmigo Academy[2], among others. In addition, the Laboratory for Social Machines at the MIT Media Lab is working on the Playful Words project[3], which focuses on bringing chil-

1 http://tmwcenter.uchicago.edu/, retrieved Jan 21, 2019.

2 https://www.hablaconmigoacademy.org/, retrieved Jan 21, 2019.

3 http://playfulwords.org/, retrieved Jan 21, 2019.

dren together with peers, teachers, and parents in machine-guided literacy activities on mobile devices.

One question that arose during Study 6 (Kory-Westlund et al., 2017b) (Also see Section 5.5) was whether the robot's expressivity led children to actually encode information differently, or whether they were simply more apt to emulate the robot when it was expressive. People may also encode information differently in relation to the shared reality constructed with another person (Echterhoff, Higgins, and Levine, 2009; Harris, 2017). Future work could probe this to better understand the effects the robot's relational behaviors have on learning.

Relational AI has great potential to support children in learning skills beyond language. Preschool teachers we have interviewed have repeatedly said that they would be very interested in robots that could help children learn and practice the social-emotional skills that are frequently critical to preschool curricula, such as sharing and turn-taking (Kory-Westlund et al., 2016b). Other work has demonstrated the potential of robots as tools for supporting older children in learning about emotional intelligence (e.g., Leite et al., 2015), as well as children with ASD in learning and practice social skills, such as joint attention, turn-taking, sequencing, and emotion understanding (Kim and Lim, 2013; Scassellati et al., 2018a). Relational AI could be used to support these kinds of interventions. Social robots have also been used as pediatric companions to support children in hospitals (Jeong et al., 2018). Since healthcare, like education, is frequently a long-term endeavor that can benefit from strong relationships, relational AI could also be used in these domains.

Because of the power social and relational interaction has for humans, relational AI has the potential to engage and empower not only children across many domains, but also other populations—older children, adults, and the elderly. We can and should use relational AI to help all people flourish, to augment and support human relationships, and to enable people to be happier, healthier, more educated, and more able to lead the lives they want to live.

ASPA (2014). *The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General, 2014*. URL: http://www.surgeongeneral.gov/library/reports/50-years-of-progress/index.html (visited on 08/25/2015).

Ackermann, Edith K. (2002). "Language Games, Digital Writing, Emerging Literacies: Enhancing Kids' Natural Gifts as Narrators and Notators." In: *Proceedings of the 3rd Hellenic Conference on Information and Communication Technologies in Education*. Ed. by A. Dimitracopoulou. Vol. 1. University of Aegean, Rhodes, Greece., pp. 31–38.

— (2005). "Playthings That Do Things: A Young Kid's "Incredibles"!" In: *Proceedings of the 2005 Conference on Interaction Design and Children*. IDC '05. New York, NY, USA: ACM, pp. 1–8. ISBN: 978-1-59593-096-5. DOI: 10.1145/1109540.1109541.

Ahmad, Muneeb Imtiaz, Omar Mubin, and Joanne Orlando (2016). "Children Views' on Social Robot's Adaptations in Education." In: *Proceedings of the 28th Australian Conference on Computer-Human Interaction*. OzCHI '16. New York, NY, USA: ACM, pp. 145–149. ISBN: 978-1-4503-4618-4. DOI: 10.1145/3010915.3010977.

Alemi, M., A. Meghdari, and M. Ghazisaedy (2015). "The Impact of Social Robotics on L2 Learners' Anxiety and Attitude in English Vocabulary Acquisition." In: *International Journal of Social Robotics*, pp. 1–13. ISSN: 1875-4791, 1875-4805. DOI: 10.1007/s12369-015-0286-y.

Alemi, Minoo, Ali Meghdari, and Maryam Ghazisaedy (2014). "Employing Humanoid Robots for Teaching English Language in Iranian Junior High-Schools." In: *International Journal of Humanoid Robotics* 11.03. ISSN: 0219-8436. DOI: 10.1142/S0219843614500224.

Anderson, Cameron, Dacher Keltner, and Oliver P. John (2003). "Emotional Convergence between People over Time." In: *Journal of Personality and Social Psychology* 84.5, pp. 1054–1068. ISSN: 1939-1315 0022-3514. DOI: 10.1037/0022-3514.84.5.1054.

Anderson, James N., Nancie Davidson, Hazel Morton, and Mervyn A. Jack (2008). "Language Learning with Interactive Virtual Agent Scenarios and Speech Recognition: Lessons Learned." In: *Computer Animation and Virtual Worlds* 19.5, pp. 605–619. ISSN: 1546-427X. DOI: 10.1002/cav.265.

Anderson, Sarah E., Christina D. Economos, and Aviva Must (Oct. 2008). "Active Play and Screen Time in US Children Aged 4 to 11 Years in Relation to Sociodemographic and Weight Status Characteristics: A Nationally Representative Cross-Sectional Analysis." In: *BMC Public Health* 8.1, p. 366. ISSN: 1471-2458. DOI: 10.1186/1471-2458-8-366.

André, Elisabeth (2011). "Design and Evaluation of Embodied Conversational Agents for Educational and Advisory Software." In: *Gaming and Simulations: Concepts, Methodologies, Tools and Applications*. IGI Global, pp. 668–686.

Arbib, Michael A. and Jean-Marc Fellous (2004). "Emotions: From Brain to Robot." In: *Trends in cognitive sciences* 8.12, pp. 554–561.

Argamon, Shlomo, Moshe Koppel, Jonathan Fine, and Anat Rachel Shimoni (2003). "Gender, Genre, and Writing Style in Formal Written Texts." In: *TEXT-THE HAGUE THEN AMSTERDAM THEN BERLIN-* 23.3, pp. 321–346.

Arkin, Ronald C. (2009). "Ethical Robots in Warfare." In: *Technology and Society Magazine, IEEE* 28.1, pp. 30–33.

Arnold, Bruce (2010). "Digital Handcuffs or Electronic Nannies: Children, Privacy and Emerging Surveillance Technologies." In:

Aron, Arthur, Elaine N. Aron, and Danny Smollan (1992). "Inclusion of Other in the Self Scale and the Structure of Interpersonal Closeness." In: *Journal of Personality and Social Psychology* 63.4, pp. 596–612. ISSN: 1939-1315 0022-3514. DOI: 10.1037/0022-3514.63.4.596.

Aron, Arthur, Elaine N. Aron, Michael Tudor, and Greg Nelson (1991). "Close Relationships as Including Other in the Self." In: *Journal of Personality and Social Psychology* 60.2, pp. 241–253. ISSN: 1939-1315 0022-3514. DOI: 10.1037/0022-3514.60.2.241.

Asimov, Isaac (1942). "Runaround." In: *Astounding Science Fiction* 29.1, pp. 94–103.

— (1954). *Caves of Steel*. New York, NY: Random House Publishing Group. ISBN: 978-0-307-79241-9.

— (2004). *I, Robot*. Vol. 1. Spectra.

Azevedo, Roger, Ronald S. Landis, Reza Feyzi-Behnagh, Melissa Duffy, Gregory Trevors, Jason M. Harley, François Bouchet, Jonathan Burlison, Michelle Taub, Nicole Pacampara, Mohamed Yeasin, A. K. M. Mahbubur Rahman, M. Iftekhar Tanveer, and Gahangir Hossain (2012). "The Effectiveness of Pedagogical Agents' Prompting and Feedback in Facilitating Co-Adapted Learning with MetaTutor." In: *Intelligent Tutoring Systems*. Ed. by Stefano A. Cerri, William J. Clancey, Giorgos Papadourakis, and Kitty Panourgia. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 212–221. ISBN: 978-3-642-30950-2.

Babcock, Meghan J., Vivian P. Ta, and William Ickes (Jan. 2014). "Latent Semantic Similarity and Language Style Matching in Initial Dyadic Interactions." In: *Journal of Language and Social Psychology* 33.1, pp. 78–88. ISSN: 0261-927X. DOI: 10.1177/0261927X13499331.

Bailenson, J, K Swinth, C Hoyt, S Persky, A Dimov, and J Blascovich (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." In: *Presence* 14.4, pp. 379–393. ISSN: 1054-7460. DOI: 10.1162/105474605774785235.

Bailenson, Jeremy N., Nick Yee, Jim Blascovich, Andrew C. Beall, Nicole Lundblad, and Michael Jin (2008). "The Use of Immersive Virtual Reality in the Learning Sciences: Digital Transformations of Teachers, Students, and Social Context." In: *The Journal of the Learning Sciences* 17.1, pp. 102–141.

Baker, Ryan S. J. d., Ma. Mercedes T. Rodrigo, and Ulises E. Xolocotzin (2007). "The Dynamics of Affective Transitions in Simulation Problem-Solving Environments." In: *Affective Computing and Intelligent Interaction*. Ed. by Ana C. R. Paiva, Rui Prada, and Rosalind W. Picard. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 666–677. ISBN: 978-3-540-74889-2.

Baker, Ryan SJd, Sidney K. D'Mello, Ma Mercedes T. Rodrigo, and Arthur C. Graesser (2010). "Better to Be Frustrated than Bored: The Incidence, Persistence, and Impact of Learners' Cognitive-Affective States during Interactions with Three Different Computer-Based Learning Environments." In: *International Journal of Human-Computer Studies* 68.4, pp. 223–241.

Baker, Susan, Barbara Weinrich, Megan Bevington, Katharine Schroth, and Emily Schroeder (June 2008). "The Effect of Task Type on Fundamental Frequency in Children." In: *International Journal of Pediatric Otorhinolaryngology* 72.6, pp. 885–889. ISSN: 0165-5876. DOI: 10.1016/j.ijporl.2008.02.019.

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

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

Bar, Moshe, Maital Neta, and Heather Linz (2006). "Very First Impressions." In: *Emotion* 6.2, pp. 269–278. ISSN: 1931-1516 1528-3542. DOI: 10.1037/1528-3542.6.2.269.

Barakova, Emilia I., Prina Bajracharya, Marije Willemsen, Tino Lourens, and Bibi Huskens (Dec. 2015). "Long-Term LEGO Therapy with Humanoid Robot for Children with ASD." In: *Expert Systems* 32.6, pp. 698–709. ISSN: 1468-0394. DOI: 10.1111/exsy.12098.

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

Bartneck, Christoph (2008). "Who like Androids More: Japanese or US Americans?" In: *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium On*. IEEE, pp. 553–557.

Batliner, A., S. Steidl, and E. Noth (2011). "Associating Children's Non-Verbal and Verbal Behaviour: Body Movements, Emotions, and Laughter in a Human-Robot Interaction." In: IEEE, pp. 5828–5831. ISBN: 978-1-4577-0538-0. DOI: 10.1109/ICASSP.2011.5947686.

Baxter, Paul, Rachel Wood, and Tony Belpaeme (2012). "A Touchscreen-Based 'Sandtray' to Facilitate, Mediate and Contextualise Human-Robot Social Interaction." In: *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*. HRI '12. New York, NY, USA: ACM, pp. 105–106. ISBN: 978-1-4503-1063-5. DOI: 10.1145/2157689.2157707.

Baxter, Paul, James Kennedy, Emmanuel Senft, Severin Lemaignan, and Tony Belpaeme (2016). "From Characterising Three Years of HRI to Methodology and Reporting Recommendations." In: *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*. HRI '16. Piscataway, NJ, USA: IEEE Press, pp. 391–398. ISBN: 978-1-4673-8370-7.

Baylor, Amy L. and Yanghee Kim (2004). "Pedagogical Agent Design: The Impact of Agent Realism, Gender, Ethnicity, and Instructional Role." In: *Intelligent Tutoring Systems*. Ed. by James C. Lester, Rosa Maria Vicari, and Fábio Paraguaçu. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 592–603. ISBN: 978-3-540-30139-4.

Baylor, Amy, Jeeheon Ryu, and E. Shen (2003). "The Effects of Pedagogical Agent Voice and Animation on Learning, Motivation and Perceived Persona." In: *World Conference on Educational Multimedia, Hypermedia and Telecommunications*. Vol. 2003, pp. 452–458.

Bell, Linda, Joakim Gustafson, and Mattias Heldner (2003). "Prosodic Adaptation in Human-Computer Interaction." In: *Proceedings of ICPHS*. Vol. 3, pp. 833–836.

Belpaeme, Tony, James Kennedy, Aditi Ramachandran, Brian Scassellati, and Fumihide Tanaka (Aug. 2018). "Social Robots for Education: A Review." In: *Science Robotics* 3.21, eaat5954. ISSN: 2470-9476. DOI: 10.1126/scirobotics.aat5954.

Benenson, Joyce F. (Jan. 2014). *Warriors and Worriers: The Survival of the Sexes*. Oxford University Press. ISBN: 978-0-19-997224-1.

Benenson, Joyce F., Ariel Durosky, Jennifer Nguyen, Alexandra Crawford, Evelyne Gauthier, and Éloise Dubé (2018). "Gender Differences in Egalitarian Behavior and Attitudes in Early Childhood." In: *Developmental Science* 0.0, e12750. ISSN: 1467-7687. DOI: 10.1111/desc.12750.

Bennett, Suzanne (Mar. 1983). "A 3-Year Longitudinal Study of School-Aged Children's Fundamental Frequencies." In: *Journal of Speech, Language, and Hearing Research* 26.1, pp. 137–141. ISSN: 1092-4388. DOI: 10.1044/jshr.2601.137.

Berger, Leigh (2001). "Inside Out: Narrative Autoethnography as a Path Toward Rapport." In: *Qualitative Inquiry* 7.4, pp. 504–518. ISSN: 1077-8004. DOI: 10.1177/107780040100700407.

Bergmann, Kirsten and Manuela Macedonia (2013). "A Virtual Agent as Vocabulary Trainer: Iconic Gestures Help to Improve Learners' Memory Performance." In: *Intelligent Virtual Agents*. Ed. by Ruth Aylett, Brigitte Krenn, Catherine Pelachaud, and Hiroshi Shimodaira. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 139–148. ISBN: 978-3-642-40415-3.

Bers, Marina Umaschi, Edith Ackermann, Justine Cassell, Beth Donegan, Joseph Gonzalez-Heydrich, David Ray DeMaso, Carol Strohecker, Sarah Lualdi, Dennis Bromley, and Judith Karlin (1998). "Interactive Storytelling Environments: Coping with Cardiac Illness at Boston's Children's Hospital." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM Press/Addison-Wesley Publishing Co., pp. 603–610.

Berscheid, Ellen and Harry T. Reis (1998). "Attraction and Close Relationships." In: *The Handbook of Social Psychology, Vols. 1-2, 4th Ed.* Ed. by D. T. Gilbert, S. T. Fiske, and G. Lindzey. New York, NY, US: McGraw-Hill, pp. 193–281. ISBN: 978-0-19-521376-8.

Bethel, C.L., M.R. Stevenson, and B. Scassellati (Oct. 2011). "Secret-Sharing: Interactions between a Child, Robot, and Adult." In: *2011 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 2489–2494. DOI: 10.1109/ICSMC.2011.6084051.

Bhakta, Roy, Maggi Savin-Baden, Gemma Tombs, Roy Bhakta, Maggi Savin-Baden, and Gemma Tombs (2014). "Sharing Secrets with Robots?" In: *World Conference on Educational Multimedia, Hypermedia and Telecommunications*. Vol. 2014, pp. 2295–2301. ISBN: 978-1-939797-08-7.

Bickmore, Timothy W. (2003). "Relational Agents : Effecting Change through Human-Computer Relationships." Thesis. Massachusetts Institute of Technology.

Bickmore, Timothy W. and Rosalind W. Picard (June 2005). "Establishing and Maintaining Long-Term Human-Computer Relationships." In: *ACM Trans. Comput.-Hum. Interact.* 12.2, pp. 293–327. ISSN: 1073-0516. DOI: 10.1145/1067860.1067867.

Bickmore, Timothy W., Dina Utami, Robin Matsuyama, and Michael K. Paasche-Orlow (2016). "Improving Access to Online Health Information With Conversational Agents: A Randomized Controlled Experiment." In: *Journal of Medical Internet Research* 18.1, e1. DOI: 10.2196/jmir.5239.

Bickmore, Timothy and Justine Cassell (2001). "Relational Agents: A Model and Implementation of Building User Trust." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '01. New York, NY, USA: ACM, pp. 396–403. ISBN: 978-1-58113-327-1. DOI: 10.1145/365024.365304.

— (2005). "Social Dialongue with Embodied Conversational Agents." In: *Advances in Natural Multimodal Dialogue Systems*. Ed. by Jan C. J. van Kuppevelt, Laila Dybkjær, and Niels Ole Bernsen. Text, Speech and Language Technology 30. Springer Netherlands, pp. 23–54. ISBN: 978-1-4020-3932-4 978-1-4020-3933-1. DOI: 10.1007/1-4020-3933-6_2.

Bickmore, Timothy, Amanda Gruber, and Rosalind Picard (2005). "Establishing the Computer–Patient Working Alliance in Automated Health Behavior Change Interventions." In: *Patient Education and Counseling* 59.1, pp. 21–30. ISSN: 0738-3991. DOI: 10.1016/j.pec.2004.09.008.

Bickmore, Timothy, Laura Pfeifer, and Daniel Schulman (2011). "Relational Agents Improve Engagement and Learning in Science Museum Visitors." In: *Intelligent Virtual Agents*. Ed. by Hannes Högni Vilhjálmsson, Stefan Kopp, Stacy Marsella, and Kristinn R. Thórisson. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 55–67. ISBN: 978-3-642-23974-8.

Bickmore, Timothy, Daniel Schulman, and Langxuan Yin (2010). "Maintaining Engagement in Long-Term Interventions with Relational Agents." In: *Applied Artificial Intelligence* 24.6, pp. 648–666.

Bickmore, Timothy, Ha Trinh, Reza Asadi, and Stefan Olafsson (2018). "Safety First: Conversational Agents for Health Care." In: *Studies in Conversational UX Design*. Ed. by Robert J. Moore, Margaret H. Szymanski, Raphael Arar, and Guang-Jie Ren. Human–Computer Interaction Series. Cham: Springer International Publishing, pp. 33–57. ISBN: 978-3-319-95579-7. DOI: 10.1007/978-3-319-95579-7_3.

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

Biocca, F, C Harms, and J Burgoon (2003). "Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria." In: *Presence* 12.5, pp. 456–480. ISSN: 1054-7460. DOI: 10.1162/105474603322761270.

Bloom, Lois (1974). "Talking, Understanding, and Thinking: Developmental Relationship between Receptive and Expressive Language." In: *Language Perspectives, Acquisition, Retardation and Intervention*, pp. 285–311. DOI: 10.7916/D88S4VMC.

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

Bobicz, K. P. and D. C. S. Richard (2003). "The Virtual Therapist: Behavior Therapy in a Digital Age." In: *Behavior Therapist* 26.3, pp. 265–269.

Boothby, Erica J., Margaret S. Clark, and John A. Bargh (Dec. 2014). "Shared Experiences Are Amplified." In: *Psychological Science* 25.12, pp. 2209–2216. ISSN: 1467-9280. DOI: 10.1177/0956797614551162.

Borrie, Stephanie A. and Julie M. Liss (June 2014). "Rhythm as a Coordinating Device: Entrainment With Disordered Speech." In: *Journal of Speech, Language, and Hearing Research* 57.3, pp. 815–824. ISSN: 1092-4388. DOI: 10.1044/2014_JSLHR-S-13-0149.

Boteanu, Adrian, Sonia Chernova, David Nunez, and Cynthia Breazeal (2016). "Fostering Parent–Child Dialog through Automated Discussion Suggestions." In: *User modeling and user-adapted interaction* 26.5, pp. 393–423. ISSN: 1573-1391.

Bove, Caron F, Jeffery Sobal, and Barbara S Rauschenbach (Feb. 2003). "Food Choices among Newly Married Couples: Convergence, Conflict, Individualism, and Projects." In: *Appetite* 40.1, pp. 25–41. ISSN: 0195-6663. DOI: 10.1016/S0195-6663(02)00147-2.

Breazeal, Cynthia L. (2004). *Designing Sociable Robots*. MIT Press. ISBN: 978-0-262-52431-5.

Breazeal, Cynthia (Jan. 2002). "Regulation and Entrainment in Human-Robot Interaction." In: *The International Journal of Robotics Research* 21.10-11, pp. 883–902. ISSN: 0278-3649, 1741-3176. DOI: 10.1177/0278364902021010096.

Breazeal, Cynthia, Kerstin Dautenhahn, and Takayuki Kanda (2016). "Social Robotics." In: *Springer Handbook of Robotics*. Ed. by Bruno Siciliano and Oussama Khatib. Springer International Publishing, pp. 1935–1972. ISBN: 978-3-319-32550-7 978-3-319-32552-1. DOI: 10.1007/978-3-319-32552-1_72.

Breazeal, Cynthia, Atsuo Takanishi, and Tetsunori Kobayashi (2008). "Social Robots That Interact with People." In: *Springer Handbook of Robotics*. Ed. by Bruno Siciliano and Oussama Khatib. Springer Berlin Heidelberg, pp. 1349–1369. ISBN: 978-3-540-23957-4 978-3-540-30301-5. DOI: 10.1007/978-3-540-30301-5_59.

Breazeal, Cynthia, Robin Morris, Stephanie Gottwald, Tinsley Galyean, and Maryanne Wolf (2016a). "Mobile Devices for Early Literacy Intervention and Research with Global Reach." In: *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*. ACM, pp. 11–20.

Breazeal, Cynthia, Paul L. Harris, David DeSteno, Jacqueline M. Kory-Westlund, Leah Dickens, and Sooyeon Jeong (2016b). "Young Children Treat Robots as Informants." In: *Topics in Cognitive Science*, pp. 1–11. ISSN: 1756-8765. DOI: 10.1111/tops.12192.

Brehm, Sharon S. (1992). *Intimate Relationships, 2nd Ed.* Vol. xviii. The McGraw-Hill series in social psychology. New York, NY, England: Mcgraw-Hill Book Company. ISBN: 978-0-07-007443-9.

Bretherton, Inge (1992). "The Origins of Attachment Theory: John Bowlby and Mary Ainsworth." In: *Developmental Psychology* 28.5, pp. 759–775. ISSN: 1939-0599(Electronic);0012-1649(Print). DOI: 10.1037/0012-1649.28.5.759.

Brunick, Kaitlin L., Marisa M. Putnam, Lauren E. McGarry, Melissa N. Richards, and Sandra L. Calvert (Apr. 2016). "Children's Future Parasocial Relationships with Media Characters: The Age of Intelligent Characters." In: *Journal of Children and Media* 10.2, pp. 181–190. ISSN: 1748-2798. DOI: 10.1080/17482798.2015.1127839.

Buhrmester, Duane and Wyndol Furman (1987). "The Development of Companionship and Intimacy." In: *Child Development* 58.4, pp. 1101–1113. ISSN: 0009-3920. DOI: 10.2307/1130550.

Burgoon, Judee K. and Jerold L. Hale (1984). "The Fundamental Topoi of Relational Communication." In: *Communication Monographs* 51.3, pp. 193–214. ISSN: 1479-5787 0363-7751. DOI: 10.1080/03637758409390195.

Burleson, Winslow and R. W. Picard (2004). "Affective Agents: Sustaining Motivation to Learn through Failure and a State of Stuck." In: *Social and Emotional Intelligence in Learning Environments Workshop In Conjunction with the 7th International Conference on Intelligent Tutoring Systems, Maceio-Alagoas, Brasil*.

Burleson, Winslow and Rosalind W. Picard (2007). "Gender-Specific Approaches to Developing Emotionally Intelligent Learning Companions." In: *Intelligent Systems, IEEE* 22.4, pp. 62–69.

Cafaro, Angelo, Hannes Högni Vilhjálmsson, and Timothy Bickmore (Aug. 2016). "First Impressions in Human–Agent Virtual Encounters." In: *ACM Trans. Comput.-Hum. Interact.* 23.4, 24:1–24:40. ISSN: 1073-0516. DOI: 10.1145/2940325.

Cairney, Trevor H. and Lynne Munsie (Feb. 1995). "Parent Participation in Literacy Learning." In: *The Reading Teacher* 48.5, pp. 392–403. ISSN: 0034-0561.

Calo, Ryan (2010). "Robots and Privacy." In: *Robot Ethics: The Ethical and Social Implications of Robotics*. Ed. by Patrick Lin, George Bekey, and Keith Abney. Cambridge, MA: MIT Press, pp. 187–201.

Calvert, Sandra L. (Jan. 2017). "Chapter 5 - Parasocial Relationships With Media Characters: Imaginary Companions for Young Children's Social and Cognitive Development." In: *Cognitive Development in Digital Contexts*. Ed. by Fran C. Blumberg and Patricia J. Brooks. San

Diego: Academic Press, pp. 93–117. ISBN: 978-0-12-809481-5. DOI: 10.1016/B978-0-12-809481-5.00005-5.

Carini, Robert M., George D. Kuh, and Stephen P. Klein (2006). "Student Engagement and Student Learning: Testing the Linkages*." In: *Research in Higher Education* 47.1, pp. 1–32.

Carney, Dana R., C. Randall Colvin, and Judith A. Hall (Oct. 2007). "A Thin Slice Perspective on the Accuracy of First Impressions." In: *Journal of Research in Personality* 41.5, pp. 1054–1072. ISSN: 0092-6566. DOI: 10.1016/j.jrp.2007.01.004.

Cassell, Justine (Jan. 2004). "Towards a Model of Technology and Literacy Development: Story Listening Systems." In: *Journal of Applied Developmental Psychology* 25.1, pp. 75–105. ISSN: 0193-3973. DOI: 10.1016/j.appdev.2003.11.003.

Cassell, Justine and Timothy Bickmore (Dec. 2000). "External Manifestations of Trustworthiness in the Interface." In: *Commun. ACM* 43.12, pp. 50–56. ISSN: 0001-0782. DOI: 10.1145/355112.355123.

— (Feb. 2003). "Negotiated Collusion: Modeling Social Language and Its Relationship Effects in Intelligent Agents." In: *User Modeling and User-Adapted Interaction* 13.1-2, pp. 89–132. ISSN: 0924-1868, 1573-1391. DOI: 10.1023/A:1024026532471.

Cassell, Justine and Kimiko Ryokai (2001). "Making Space for Voice: Technologies to Support Childrenâ€™s Fantasy and Storytelling." In: *Personal and ubiquitous computing* 5.3, pp. 169–190.

Cassell, Justine, Andrea Tartaro, Yolanda Rankin, Vani Oza, Candice Tse, and Frances Searle (2007). "Virtual Peers for Literacy Learning." In: *EDUCATIONAL TECHNOLOGY-SADDLE BROOK THEN ENGLEWOOD CLIFFS NJ-* 47.1, p. 39.

Cassell, Justine, Kathleen Geraghty, Berto Gonzalez, and John Borland (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, USA: ACM, pp. 135–142. ISBN: 978-1-60558-772-1. DOI: 10.1145/1647314.1647338.

Chambers, Christine T. and Charlotte Johnston (Jan. 2002). "Developmental Differences in Children's Use of Rating Scales." In: *Journal of Pediatric Psychology* 27.1, pp. 27–36. ISSN: 0146-8693. DOI: 10.1093/jpepsy/27.1.27.

Chandra, Shruti, Raul Paradeda, Hang Yin, Pierre Dillenbourg, Rui Prada, and Ana Paiva (2018). "Do Children Perceive Whether a Robotic Peer Is Learning or Not?" In: *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. HRI '18. New York, NY, USA: ACM, pp. 41–49. ISBN: 978-1-4503-4953-6. DOI: 10.1145/3171221.3171274.

Chang, Angela and Cynthia Breazeal (2011). "TinkRBook: Shared Reading Interfaces for Storytelling." In: *Proceedings of the 10th International Conference on Interaction Design and Children*. IDC '11. New York, NY, USA: ACM, pp. 145–148. ISBN: 978-1-4503-0751-2. DOI: 10.1145/1999030.1999047.

Chang, Angela, Cynthia Breazeal, Fardad Faridi, Tom Roberts, Glorianna Davenport, Henry Lieberman, and Nick Montfort (2012). "Textual Tinkerability: Encouraging Storytelling Behaviors to Foster Emergent Literacy." In: *CHI '12 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '12. New York, NY, USA: ACM, pp. 505–520. ISBN: 978-1-4503-1016-1. DOI: 10.1145/2212776.2212826.

Chang, Chih-Wei, Jih-Hsien Lee, Po-Yao Chao, Chin-Yeh Wang, and Gwo-Dong Chen (2010). "Exploring the Possibility of Using Humanoid Robots as Instructional Tools for Teaching a Second Language in Primary School." In: *Journal of Educational Technology & Society* 13.2, pp. 13–24. ISSN: 1176-3647.

Chaplin, Tara M. and Amelia Aldao (2013). "Gender Differences in Emotion Expression in Children: A Meta-Analytic Review." In: *Psychological bulletin* 139.4, p. 735.

Chartrand, Tanya L. and Rick van Baaren (2009). "Human Mimicry." In: *Advances in experimental social psychology* 41, pp. 219–274.

Chaspari, Theodora and Jill Fain Lehman (2016). "An Acoustic Analysis of Child-Child and Child-Robot Interactions for Understanding Engagement during Speech-Controlled Computer Games." In: *INTERSPEECH*, pp. 595–599.

Chen, Huili (2018). "Adaptive Role Switching in Socially Interactive Agents for Children's Language Learning." Master's Thesis. Cambridge, MA: Massachusetts Institute of Technology.

Chiong, C., Ree, J., Takeuchi, L., and Erickson, I. (2012). "Print Books vs. e-Books: Comparing Parent-Child Co-Reading on Print, Basic, and Enhanced e-Book Platforms." In: New York: The Joan Ganz Cooney Center.

Chiong, C. and C. Shuler (2010). *Learning: Is There an App for That? Investigations of Young Children's Usage and Learning with Mobile Devices and Apps. Investigations of Young*. New York. Retrieved from http://www-tc. pbskids. org/read/files/cooney_learning_apps. pdf.

Chisholm, Kim and Janet Strayer (Apr. 1995). "Verbal and Facial Measures of Children's Emotion and Empathy." In: *Journal of Experimental Child Psychology* 59.2, pp. 299–316. ISSN: 0022-0965. DOI: 10.1006/jecp.1995.1013.

Christophel, Diane M. (Oct. 1990). "The Relationships among Teacher Immediacy Behaviors, Student Motivation, and Learning." In: *Communication Education* 39.4, pp. 323–340. ISSN: 0363-4523. DOI: 10.1080/03634529009378813.

Clark, Richard E. (Jan. 1983). "Reconsidering Research on Learning from Media." In: *Review of Educational Research* 53.4, pp. 445–459. ISSN: 0034-6543, 1935-1046. DOI: 10.3102/00346543053004445.

Coeckelbergh, M. (Fourth 2012). "Are Emotional Robots Deceptive?" In: *IEEE Transactions on Affective Computing* 3.4, pp. 388–393. ISSN: 1949-3045. DOI: 10.1109/T-AFFC.2011.29.

Cole, Ron, Angela Halpern, Lorraine Ramig, Sarel Van Vuuren, Nattawut Ngampatipatpong, and Jie Yan (2007). "A Virtual Speech Therapist for Individuals with Parkinson's Disease." In: *EDUCATIONAL TECHNOLOGY-SADDLE BROOK THEN ENGLEWOOD CLIFFS NJ-* 47.1, p. 51.

Coninx, Alexandre et al. (2016). "Towards Long-Term Social Child-Robot Interaction: Using Multi-Activity Switching to Engage Young Users." In: *Journal of Human-Robot Interaction* 5.1, pp. 32–67. ISSN: Coninx, Alexandre and Baxter, Paul and Oleari, Elettra and Bellini, Sara and Bierman, Bert and Henkemans, Olivier Blanson and Canamero, Lola and Cosi, Piero and Enescu, Valentin and Espinoza, Raquel Ros and Hiolle, Antoine and Humbert, Remi and Kiefer, Bernd and Kruijff-korbayova, Ivana and Looije, Rosemarijn and Mosconi, Marco and Neerincx, Mark and Paci, Giulio and Patsis, Georgios and Pozzi, Clara and Sacchitelli, Francesca and Sahli, Hichem and Sanna, Alberto and Sommavilla, Giacomo and Tesser, Fabio and Demiris, Yiannis and Belpaeme, Tony (2016) Towards long-term social child-robot interaction: using multi-activity switching to engage young users. Journal of Human-Robot Interaction, 5 (1). pp. 32-67. ISSN 2163-0364.

Connell, Sabrina L., Alexis R. Lauricella, and Ellen Wartella (Jan. 2015). "Parental Co-Use of Media Technology with Their Young Children in the USA." In: *Journal of Children and Media* 9.1, pp. 5–21. ISSN: 1748-2798. DOI: 10.1080/17482798.2015.997440.

Coolahan, Kathleen, John Fantuzzo, Julia Mendez, and Paul McDermott (2000). "Preschool Peer Interactions and Readiness to Learn: Relationships between Classroom Peer Play and Learning Behaviors and Conduct." In: *Journal of Educational Psychology* 92.3, pp. 458–465. ISSN: 1939-2176 0022-0663. DOI: 10.1037/0022-0663.92.3.458.

Corriveau, Kathleen H., Paul L. Harris, Elizabeth Meins, Charles Fernyhough, Bronia Arnott, Lorna Elliott, Beth Liddle, Alexandra Hearn, Lucia Vittorini, and Marc De Rosnay (2009). "Young Childrens' Trust in Their Mothers' Claims: Longitudinal Links with Attachment Security in Infancy." In: *Child development* 80.3, pp. 750–761.

Craig, Scotty, Arthur Graesser, Jeremiah Sullins, and Barry Gholson (2004). "Affect and Learning: An Exploratory Look into the Role of Affect in Learning with AutoTutor." In: *Journal of Educational Media* 29.3, pp. 241–250. ISSN: 1358-1651. DOI: 10.1080/1358165042000283101.

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. New York, NY: Harper and row. ISBN: 978-0-06-187672-1.

Csikszentmihalyi, Mihaly (1997). *Finding Flow: The Psychology of Engagement with Everyday Life*. Basic Books.

Csikszentmihalyi, Mihaly and Eugene Halton (Oct. 1981). *The Meaning of Things: Domestic Symbols and the Self*. Cambridge University Press. ISBN: 978-0-521-28774-6.

Curenton, Stephanie M., Michelle Jones Craig, and Nadia Flanigan (2008). "Use of Decontextualized Talk across Story Contexts: How Oral Storytelling and Emergent Reading Can Scaffold Children's Development." In: *Early education and development* 19.1, pp. 161–187.

D'Mello, S. K. and A. C. Graesser (2011). "The Half-Life of Cognitive-Affective States during Complex Learning." In: *Cognition and Emotion* 25.7, pp. 1299–1308.

D'Mello, S. K. and A. Graesser (2012). "Dynamics of Affective States during Complex Learning." In: *Learning and Instruction* 22, pp. 145–157.

D'Mello, Sidney, Ed Dieterle, and Angela Duckworth (2017). "Advanced, Analytic, Automated (AAA) Measurement of Engagement During Learning." In: *Educational psychologist* 52.2, pp. 104–123. ISSN: 0046-1520. DOI: 10.1080/00461520.2017.1281747.

D'Mello, Sidney, Andrew Olney, Claire Williams, and Patrick Hays (2012). "Gaze Tutor: A Gaze-Reactive Intelligent Tutoring System." In: *International Journal of Human-Computer Studies* 70.5, pp. 377–398.

Darling, K., P. Nandy, and C. Breazeal (Aug. 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)*, pp. 770–775. DOI: 10.1109/ROMAN.2015.7333675.

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

De Lisi, Richard and Susan L. Golbeck (1999). "Implications of Piagetian Theory for Peer Learning." In: *Cognitive Perspectives on Peer Learning*. Ed. by A. M. O and A. King. The Rutgers Invitational Symposium On Education Series. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, pp. 3–37. ISBN: 978-0-8058-2447-6 978-0-8058-2448-3.

DeLay, Dawn, Laura D. Hanish, Carol Lynn Martin, and Richard A. Fabes (2016). "Peer Effects on Head Start Children's Preschool Competency." In: *Developmental Psychology* 52.1, pp. 58–70. ISSN: 1939-0599 0012-1649. DOI: 10.1037/dev0000066.

Deshmukh, Amol, Aidan Jones, Srinivasan Janarthanam, Helen Hastie, Tiago Ribeiro, Ruth Aylett, Ana Paiva, Ginevra Castellano, Mary Ellen Foster, Lee J. Corrigan, and others (2015). "An Empathic Robotic Tutor in a Map Application." In: *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, pp. 1923–1924.

Desteno, David, Cynthia Breazeal, Robert H. Frank, David Pizarro, Jolie Baumann, Leah Dickens, and Jin Joo Lee (Dec. 2012). "Detecting the Trustworthiness of Novel Partners in Economic Exchange." In: *Psychological Science* 23.12, pp. 1549–1556. ISSN: 1467-9280. DOI: 10.1177/0956797612448793.

Dijksterhuis, Ap (2005). "Why We Are Social Animals: The High Road to Imitation as Social Glue." In: *Perspectives on imitation: From neuroscience to social science* 2, pp. 207–220.

Dijksterhuis, Ap and John A. Bargh (2001). "The Perception-Behavior Expressway: Automatic Effects of Social Perception on Social Behavior." In: *Advances in experimental social psychology* 33, pp. 1–40.

D'mello, Sidney and Art Graesser (2012). "AutoTutor and Affective AutoTutor: Learning by Talking with Cognitively and Emotionally Intelligent Computers That Talk Back." In: *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2.4, p. 23.

Doyle, Anna-Beth, Jennifer Connolly, and Louis-Paul Rivest (1980). "The Effect of Playmate Familiarity on the Social Interactions of Young Children." In: *Child Development*, pp. 217–223.

Drew Simshaw (2015). "Data Collection, Privacy, and the Role of Regulatory Agencies in Healthcare Robotics." In: *Workshop on The Emerging Policy and Ethics of Human-Robot Interaction, Held in Conjunction with the 10th ACM/IEEE International Conference on Human-Robot Interaction*.

Druga, Stefania, Randi Williams, Cynthia Breazeal, and Mitchel Resnick (2017). ""Hey Google Is It OK If I Eat You?": Initial Explorations in Child-Agent Interaction." In: *Proceedings of the 2017 Conference on Interaction Design and Children*. IDC '17. New York, NY, USA: ACM, pp. 595–600. ISBN: 978-1-4503-4921-5. DOI: 10.1145/3078072.3084330.

Druga, Stefania, Randi Williams, Hae Won Park, and Cynthia Breazeal (2018). "How Smart Are the Smart Toys?: Children and Parents' Agent Interaction and Intelligence Attribution." In: *Proceedings of the 17th ACM Conference on Interaction Design and Children*. IDC '18. New York, NY, USA: ACM, pp. 231–240. ISBN: 978-1-4503-5152-2. DOI: 10.1145/3202185.3202741.

Drumwright, Minette E. and Patrick E. Murphy (2009). "The Current State of Advertising Ethics: Industry and Academic Perspectives." In: *Journal of Advertising* 38.1, pp. 83–108.

Duck, Steve (1991). *Understanding Relationships*. Vol. vii. New York, NY, US: Guilford Press. ISBN: 978-0-89862-758-9 978-0-89862-470-0.

Duranti, Alessandro and Charles Goodwin (1992). *Rethinking Context: Language as an Interactive Phenomenon*. 11. Cambridge University Press.

Echterhoff, Gerald, E. Tory Higgins, and John M. Levine (2009). "Shared Reality: Experiencing Commonality with Others' Inner States about the World." In: *Perspectives on Psychological Science* 4.5, pp. 496–521. ISSN: 1745-6916. DOI: 10.1111/j.1745-6924.2009.01161.x.

Elias, Maurice J. (2006). "The Connection between Academic and Social-Emotional Learning." In: *The educator's guide to emotional intelligence and academic achievement*, pp. 4–14.

Engel, Susan (1995). *The Stories Children Tell: Making Sense of the Narratives of Childhood*. Macmillan.

Ensor, Rosie and Claire Hughes (2008). "Content or Connectedness? Mother–Child Talk and Early Social Understanding." In: *Child Development* 79.1, pp. 201–216. ISSN: 1467-8624. DOI: 10.1111/j.1467-8624.2007.01120.x.

Favazza, Paddy C. and Samuel L. Odom (July 1996). "Use of the Acceptance Scale to Measure Attitudes of Kindergarten-Age Children." In: *Journal of Early Intervention* 20.3, pp. 232–248. ISSN: 1053-8151. DOI: 10.1177/105381519602000307.

Favazza 1954, Paddy C., Leslie Phillipsen, and Poonam Kumar (2000). "Measuring and Promoting Acceptance of Young Children with Disabilities." In: *Exceptional Children* 66.4, pp. 491–508. ISSN: 00144029.

Feil-Seifer, D. and M.J. Mataric (2011). "Socially Assistive Robotics." In: *IEEE Robotics Automation Magazine* 18.1, pp. 24–31. ISSN: 1070-9932. DOI: 10.1109/MRA.2010.940150.

Fellous, Jean-Marc (2004). "From Human Emotions to Robot Emotions." In: *Architectures for Modeling Emotion: Cross-Disciplinary Foundations, American Association for Artificial Intelligence*, pp. 39–46.

Finkelstein, Samantha, Stefan Scherer, Amy Ogan, Louis-Phillipe Morency, and Justine Cassell (2012). "Investigating the Influence of Virtual Peers as Dialect Models on Students' Prosodic Inventory." In:

Fish, Margaret and Brenda Pinkerman (2003). "Language Skills in Low-SES Rural Appalachian Children: Normative Development and Individual Differences, Infancy to Preschool." In: *Journal of applied developmental psychology* 23.5, pp. 539–565.

Flewitt, Rosie, David Messer, and Natalia Kucirkova (Sept. 2015). "New Directions for Early Literacy in a Digital Age: The iPad." In: *Journal of Early Childhood Literacy* 15.3, pp. 289–310. ISSN: 1468-7984. DOI: 10.1177/1468798414533560.

Fogg, B. J. and Hsiang Tseng (1999). "The Elements of Computer Credibility." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '99. New York, NY, USA: ACM, pp. 80–87. ISBN: 978-0-201-48559-2. DOI: 10.1145/302979.303001.

Forlizzi, Jodi (2007). "How Robotic Products Become Social Products: An Ethnographic Study of Cleaning in the Home." In: *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. HRI '07. New York, NY, USA: ACM, pp. 129–136. ISBN: 978-1-59593-617-2. DOI: 10.1145/1228716.1228734.

Förster, Jens (2009). "Cognitive Consequences of Novelty and Familiarity: How Mere Exposure Influences Level of Construal." In: *Journal of Experimental Social Psychology* 45.2, pp. 444–447. ISSN: 0022-1031. DOI: 10.1016/j.jesp.2008.10.011.

Förster, Jens, Janina Marguc, and Marleen Gillebaart (2010). "Novelty Categorization Theory." In: *Social and Personality Psychology Compass* 4.9, pp. 736–755. ISSN: 1751-9004. DOI: 10.1111/j.1751-9004.2010.00289.x.

Fraley, Barbara and Arthur Aron (Mar. 2004). "The Effect of a Shared Humorous Experience on Closeness in Initial Encounters." In: *Personal Relationships* 11.1, pp. 61–78. ISSN: 1475-6811. DOI: 10.1111/j.1475-6811.2004.00071.x.

Freed, Natalie Anne (2012). ""This Is the Fluffy Robot That Only Speaks French": Language Use between Preschoolers, Their Families, and a Social Robot While Sharing Virtual Toys." Master's Thesis. Cambridge, MA: Massachusetts Institute of Technology.

Fuchs, Douglas, Lynn S. Fuchs, Patricia G. Mathes, and Deborah C. Simmons (Jan. 1997). "Peer-Assisted Learning Strategies: Making Classrooms More Responsive to Diversity." In: *American Educational Research Journal* 34.1, pp. 174–206. ISSN: 0002-8312. DOI: 10.3102/00028312034001174.

Furman, Wyndol (1996). "The Measurement of Friendship Perceptions: Conceptual and Methodological Issues." In: *The company they keep: Friendship in childhood and adolescence*, pp. 41–65.

Gächter, Simon, Chris Starmer, and Fabio Tufano (June 2015). "Measuring the Closeness of Relationships: A Comprehensive Evaluation of the 'Inclusion of the Other in the Self' Scale." In: *PLOS ONE* 10.6, e0129478. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0129478.

Gallagher, Tanya M. (Oct. 1993). "Language Skill and the Development of Social Competence in School-Age Children." In: *Language, Speech & Hearing Services in Schools* 24.4, p. 199. ISSN: 01611461.

Gaudiello, Ilaria, Sébastien Lefort, and Elisabetta Zibetti (2015). "The Ontological and Functional Status of Robots: How Firm Our Representations Are?" In: *Computers in Human Behavior* 50, pp. 259–273. ISSN: 0747-5632. DOI: 10.1016/j.chb.2015.03.060.

Gelfer, Marylou Pausewang and Sara L. Denor (Aug. 2014). "Speaking Fundamental Frequency and Individual Variability in Caucasian and African American School-Age Children." In: *American Journal of Speech-Language Pathology* 23.3, pp. 395–406. ISSN: 1058-0360. DOI: 10.1044/2014_AJSLP-13-0016.

Gillebaart, M. (2012). *Something Old, Something New: When People Favor Novelty over Familiarity and How Novelty Affects Creative Processes.* ISBN: 978-94-6191-330-2.

Gleason, Tracy R. (Apr. 1997). "Invisible Friends and Personified Objects: Qualitative Differences in Relationships with Imaginary Companions." In:

— (Nov. 2002). "Social Provisions of Real and Imaginary Relationships in Early Childhood." In: *Developmental Psychology* 38.6, pp. 979–992. ISSN: 0012-1649. DOI: http://dx.doi.org.libproxy.mit.edu/10.1037/0012-1649.38.6.979.

Gleason, Tracy R. and Lisa M. Hohmann (Feb. 2006). "Concepts of Real and Imaginary Friendships in Early Childhood." In: *Social Development* 15.1, p. 128. ISSN: 0961205X.

Gockley, R., A. Bruce, J. Forlizzi, M. Michalowski, A. Mundell, S. Rosenthal, B. Sellner, R. Simmons, K. Snipes, A.C. Schultz, and Jue Wang (2005). "Designing Robots for Long-Term Social Interaction." In: *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2005. (IROS 2005)*, pp. 1338–1343. DOI: 10.1109/IROS.2005.1545303.

Gola, Alice Ann Howard, Melissa N. Richards, Alexis R. Lauricella, and Sandra L. Calvert (2013). "Building Meaningful Parasocial Relationships Between Toddlers and Media Characters to Teach Early Mathematical Skills." In: *Media Psychology* 16.4, pp. 390–411. ISSN: 1521-3269. DOI: 10.1080/15213269.2013.783774.

Goldman, David (Feb. 2015). *Your Samsung TV Is Eavesdropping on Your Private Conversations.* URL: http://money.cnn.com/2015/02/09/technology/security/samsung-smart-tv-privacy/index.html (visited on 08/25/2015).

Golinkoff, Roberta Michnick, Kathryn A. Hirsh-Pasek, and Dorothy G. Singer (2006). "Play= Learning: A Challenge for Parents and Educators." In: *Play= Learning: How Play Motivates and Enhances Children's Cognitive and Social-Emotional Growth: How Play Motivates and Enhances Children's Cognitive and Social-Emotional Growth*, p. 1.

Goodwin, Kristy and Kate Highfield (2012). "iTouch and iLearn: An Examination of "Educational" Apps." In: *Early Education and Technology for Children Conference*, pp. 14–16.

Gordon, Goren and Cynthia Breazeal (2015). "Bayesian Active Learning-Based Robot Tutor for Children's Word-Reading Skills." In: *Proceedings of the 29th AAAI Conference on Artificial Intelligence.*

Gordon, Goren, Cynthia Breazeal, and Susan Engel (2015a). "Can Children Catch Curiosity from a Social Robot?" In: *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. HRI '15. New York, NY, USA: ACM, pp. 91–98. ISBN: 978-1-4503-2883-8. DOI: 10.1145/2696454.2696469.

— (2015b). "Can Children Catch Curiosity from a Social Robot?" In: *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, pp. 91–98.

Gordon, Goren, Samuel Spaulding, Jacqueline Kory-Westlund, Jin Joo Lee, Luke Plummer, Marayna Martinez, Madhurima Das, and Cynthia Breazeal (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.

Gosha, Kinnis (2013). "The Application Of Embodied Conversational Agents For Mentoring African American STEM Doctoral Students." In: *All Dissertations*.

Gosha, Kinnis and Kamal Middlebrook (2017). "Broadening Participation Research Project: Exploring Computing Careers Through a Virtual Career Exploration Fair Using Embodied Conversational Agents (Abstract Only)." In: *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*. SIGCSE '17. New York, NY, USA: ACM, pp. 708–708. ISBN: 978-1-4503-4698-6. DOI: 10.1145/3017680.3022425.

Gottwald, Stephanie, Robin Morris, Maryanne Wolf, and Tinsley Galyean (Dec. 2017). "Bringing the Bottom Billion into Basic Literacy: How We Can and Why We Must." In: *New Directions for Child and Adolescent Development* 2017.158, pp. 93–104. ISSN: 1534-8687. DOI: 10.1002/cad.20225.

Grammer, K., K. B. Kruck, and M. S. Magnusson (1998). "The Courtship Dance: Patterns of Nonverbal Synchronization in Opposite-Sex Encounters." In: *Journal of Nonverbal behavior* 22.1, pp. 3–29.

Gratch, Jonathan, Ning Wang, Jillian Gerten, Edward Fast, and Robin Duffy (Sept. 2007). "Creating Rapport with Virtual Agents." In: *Intelligent Virtual Agents*. Springer, Berlin, Heidelberg, pp. 125–138. DOI: 10.1007/978-3-540-74997-4_12.

Gray, Peter (Mar. 2013). *Free to Learn: Why Unleashing the Instinct to Play Will Make Our Children Happier, More Self-Reliant, and Better Students for Life*. Basic Books. ISBN: 978-0-465-03791-9.

Griffin, Terri M., Lowry Hemphill, Linda Camp, and Dennis Palmer Wolf (June 2004). "Oral Discourse in the Preschool Years and Later Literacy Skills." In: *First Language* 24.2, pp. 123–147. ISSN: 0142-7237. DOI: 10.1177/0142723704042369.

Guerin, Bernard (Jan. 1986). "Mere Presence Effects in Humans: A Review." In: *Journal of Experimental Social Psychology* 22.1, pp. 38–77. ISSN: 0022-1031. DOI: 10.1016/0022-1031(86)90040-5.

Guernsey, Lisa, Michael Levine, Cynthia Chiong, and Maggie Severns (2012). "Pioneering Literacy in the Digital Wild West: Empowering Parents and Educators." In: *Washington, DC: Campaign for Grade-Level Reading*.

Hacki, T. and S. Heitmüller (Oct. 1999). "Development of the Child's Voice: Premutation, Mutation." In: *International Journal of Pediatric Otorhinolaryngology* 49, Supplement 1, S141–S144. ISSN: 0165-5876. DOI: 10.1016/S0165-5876(99)00150-0.

Hancock, Peter A., Deborah R. Billings, Kristin E. Schaefer, Jessie Y. C. Chen, Ewart J. de Visser, and Raja Parasuraman (Oct. 2011). "A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction." In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 53.5, pp. 517–527. ISSN: 0018-7208, 1547-8181. DOI: 10.1177/0018720811417254.

Hargrave, Anne C and Monique Sénéchal (2000). "A Book Reading Intervention with Preschool Children Who Have Limited Vocabularies: The Benefits of Regular Reading and Dialogic Reading." In: *Early Childhood Research Quarterly* 15.1, pp. 75–90. ISSN: 0885-2006. DOI: 10.1016/S0885-2006(99)00038-1.

Harris, Justin, Roberta Golinkoff, and Kathy Hirsh-Pasek (2011). "Lessons from the Crib for the Classroom: How Children Really Learn Vocabulary." In: *Handbook of Early Literacy Researchedited*, pp. 49–65.

Harris, Paul L. (2007). "Trust." In: *Developmental science* 10, pp. 135–138.

— (2012). *Trusting What You're Told: How Children Learn from Others*. Harvard University Press.

Harris, Paul L. (2017). "Tell, Ask, Repair: Early Responding to Discordant Reality." In: *Motivation Science* 3.3, pp. 275–286. ISSN: 2333-8113. DOI: 10.1037/mot0000075.

Hart, Betty and Todd R. Risley (1995). *Meaningful Differences in the Everyday Experience of Young American Children.* ERIC.

Harter, Susan and Robin Pike (1984). "The Pictorial Scale of Perceived Competence and Social Acceptance for Young Children." In: *Child Development* 55.6, pp. 1969–1982. ISSN: 0009-3920. DOI: 10.2307/1129772.

Hartshorne, Joshua K., Joshua B. Tenenbaum, and Steven Pinker (Aug. 2018). "A Critical Period for Second Language Acquisition: Evidence from 2/3 Million English Speakers." In: *Cognition* 177, pp. 263–277. ISSN: 0010-0277. DOI: 10.1016/j.cognition.2018.04.007.

Hartup, Willard W., Brett Laursen, Mark I. Stewart, and Amy Eastenson (1988). "Conflict and the Friendship Relations of Young Children." In: *Child Development* 59.6, pp. 1590–1600. ISSN: 0009-3920. DOI: 10.2307/1130673.

Haviland, Jeannette M. and Mary Lelwica (1987). "The Induced Affect Response: 10-Week-Old Infants' Responses to Three Emotion Expressions." In: *Developmental Psychology* 23.1, pp. 97–104. ISSN: 1939-0599(Electronic);0012-1649(Print). DOI: 10.1037/0012-1649.23.1.97.

Henderson, Marlone D., Kentaro Fujita, Yaacov Trope, and Nira Liberman (2006). "Transcending the "Here": The Effect of Spatial Distance on Social Judgment." In: *Journal of Personality and Social Psychology* 91.5, pp. 845–856. ISSN: 1939-1315(Electronic);0022-3514(Print). DOI: 10.1037/0022-3514.91.5.845.

Hendriks, Bram, Bernt Meerbeek, Stella Boess, Steffen Pauws, and Marieke Sonneveld (Apr. 2011). "Robot Vacuum Cleaner Personality and Behavior." In: *International Journal of Social Robotics* 3.2, pp. 187–195. ISSN: 1875-4805. DOI: 10.1007/s12369-010-0084-5.

Hirsh-Pasek, Kathy, Jennifer M. Zosh, Roberta Michnick Golinkoff, James H. Gray, Michael B. Robb, and Jordy Kaufman (May 2015). "Putting Education in "Educational" Apps: Lessons From the Science of Learning." In: *Psychological Science in the Public Interest* 16.1, pp. 3–34. ISSN: 1529-1006. DOI: 10.1177/1529100615569721.

Hoff, Erika (Mar. 2006). "How Social Contexts Support and Shape Language Development." In: *Developmental Review* 26.1, pp. 55–88. ISSN: 0273-2297. DOI: 10.1016/j.dr.2005.11.002.

Hoffman, Diane M. (2009). "Reflecting on Social Emotional Learning: A Critical Perspective on Trends in the United States." In: *Review of Educational Research* 79.2, pp. 533–556.

Home Depot Hit By Same Malware as Target (2014). URL: https://krebsonsecurity.com/2014/09/home-depot-hit-by-same-malware-as-target/ (visited on 08/25/2015).

Home Depot's Suspected Breach Looks Just Like the Target Hack (2014). URL: http://www.bloomberg.com/bw/articles/2014-09-02/home-depots-credit-card-breach-looks-just-like-the-target-hack (visited on 08/25/2015).

Hood, Deanna, Séverin Lemaignan, and Pierre Dillenbourg (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, USA: ACM, pp. 83–90. ISBN: 978-1-4503-2883-8. DOI: 10.1145/2696454.2696479.

Hoque, M. and Rosalind W. Picard (2011). "Acted vs. Natural Frustration and Delight: Many People Smile in Natural Frustration." In: *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference On*. IEEE, pp. 354–359.

Hoque, Mohammed Ehsan, Daniel J. McDuff, and Rosalind W. Picard (2012). "Exploring Temporal Patterns in Classifying Frustrated and Delighted Smiles." In: *IEEE Transactions on Affective Computing*, pp. 323–334.

Horvath, Adam O. | Greenberg (1989). "Development and Validation of the Working Alliance Inventory." In: *Journal of Counseling Psychology* 36.2, pp. 223–33.

Huttenlocher, Janellen, Marina Vasilyeva, and Priya Shimpi (Feb. 2004). "Syntactic Priming in Young Children." In: *Journal of Memory and Language* 50.2, pp. 182–195. ISSN: 0749-596X. DOI: 10.1016/j.jml.2003.09.003.

Huttenlocher, Janellen, Wendy Haight, Anthony Bryk, Michael Seltzer, and Thomas Lyons (1991). "Early Vocabulary Growth: Relation to Language Input and Gender." In: *Devel-*

opmental Psychology 27.2, pp. 236–248. ISSN: 1939-0599 0012-1649. DOI: 10.1037/0012-1649.27.2.236.

Huttenlocher, Janellen, Marina Vasilyeva, Elina Cymerman, and Susan Levine (Nov. 2002). "Language Input and Child Syntax." In: *Cognitive Psychology* 45.3, pp. 337–374. ISSN: 0010-0285. DOI: 10.1016/S0010-0285(02)00500-5.

Huttenlocher, Janellen, Heidi Waterfall, Marina Vasilyeva, Jack Vevea, and Larry V. Hedges (Dec. 2010). "Sources of Variability in Children's Language Growth." In: *Cognitive Psychology* 61.4, pp. 343–365. ISSN: 0010-0285. DOI: 10.1016/j.cogpsych.2010.08.002.

Hwang, Inseok, Chungkuk Yoo, Chanyou Hwang, Dongsun Yim, Youngki Lee, Chulhong Min, John Kim, and Junehwa Song (2014). "TalkBetter: Family-Driven Mobile Intervention Care for Children with Language Delay." In: *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, pp. 1283–1296.

Hyun, Eunja, Hyunmin Yoon, and Sooryun Son (2010). "Relationships Between User Experiences and Children's Perceptions of the Education Robot." In: *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction*. HRI '10. Piscataway, NJ, USA: IEEE Press, pp. 199–200. ISBN: 978-1-4244-4893-7.

Iacobelli, Francisco and Justine Cassell (2007). "Ethnic Identity and Engagement in Embodied Conversational Agents." In: *Intelligent Virtual Agents*. Ed. by Catherine Pelachaud, Jean-Claude Martin, Elisabeth André, Gérard Chollet, Kostas Karpouzis, and Danielle Pelé. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 57–63. ISBN: 978-3-540-74997-4.

Ingram, David (1974). "The Relationship between Comprehension and Production." In: *Language Perspectives: Acquisition, Retardation, and Intervention*. Baltimore, MD, US: University Park Press, pp. xv, 670–xv, 670. ISBN: 978-0-8391-0687-6.

Ireland, Molly E. and Marlone D. Henderson (Feb. 2014). "Language Style Matching, Engagement, and Impasse in Negotiations." In: *Negotiation and Conflict Management Research* 7.1, pp. 1–16. ISSN: 1750-4716. DOI: 10.1111/ncmr.12025.

Ireland, Molly E., Richard B. Slatcher, Paul W. Eastwick, Lauren E. Scissors, Eli J. Finkel, and James W. Pennebaker (Jan. 2011). "Language Style Matching Predicts Relationship Initiation and Stability." In: *Psychological Science* 22.1, pp. 39–44. ISSN: 0956-7976. DOI: 10.1177/0956797610392928.

Isen, Alice M. (1987). "Positive Affect, Cognitive Processes, and Social Behavior." In: *Advances in Experimental Social Psychology*. Ed. by Leonard Berkowitz. Vol. 20. Academic Press, pp. 203–253. DOI: 10.1016/S0065-2601(08)60415-3.

— (2002). "A Role for Neuropsychology in Understanding the Facilitating Influence of Positive Affect on Social Behavior and Cognitive Processes." In: *Handbook of positive psychology*, pp. 528–540.

Iwamura, Yamato, Masahiro Shiomi, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita (2011). "Do Elderly People Prefer a Conversational Humanoid as a Shopping Assistant Partner in Supermarkets?" In: *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference On*. IEEE, pp. 449–457.

Jeong, Sooyeon, Cynthia Breazeal, Deirdre Logan, and Peter Weinstock (2018). "Huggable: The Impact of Embodiment on Promoting Socio-Emotional Interactions for Young Pediatric Inpatients." In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. New York, NY, USA: ACM, 495:1–495:13. ISBN: 978-1-4503-5620-6. DOI: 10.1145/3173574.3174069.

Johal, Wafa, Paul Vogt, James Kennedy, Mirjam de Haas, Ana Paiva, Ginevra Castellano, Sandra Okita, Fumihide Tanaka, Tony Belpaeme, and Pierre Dillenbourg (2017). "Workshop on Robots for Learning: R4L." In: *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. HRI '17. New York, NY, USA: ACM, pp. 423–424. ISBN: 978-1-4503-4885-0. DOI: 10.1145/3029798.3029801.

Johal, Wafa, James Kennedy, Vicky Charisi, Hae Won Park, Ginevra Castellano, and Pierre Dillenbourg (2018). "Robots for Learning - R4L: Inclusive Learning." In: *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. HRI '18. New York, NY, USA: ACM, pp. 397–398. ISBN: 978-1-4503-5615-2. DOI: 10.1145/3173386.3173567.

Johnson, Amy Janan, Michel M. Haigh, Elizabeth A. Craig, and Jennifer a. H. Becker (Dec. 2009). "Relational Closeness: Comparing Undergraduate College Students' Geographically Close and Long-Distance Friendships." In: *Personal Relationships* 16.4, pp. 631–646. ISSN: 1475-6811. DOI: 10.1111/j.1475-6811.2009.01243.x.

Johnson, Jacqueline S and Elissa L Newport (Jan. 1989). "Critical Period Effects in Second Language Learning: The Influence of Maturational State on the Acquisition of English as a Second Language." In: *Cognitive Psychology* 21.1, pp. 60–99. ISSN: 0010-0285. DOI: 10.1016/0010-0285(89)90003-0.

Johnson, W. Lewis and James C. Lester (Mar. 2016). "Face-to-Face Interaction with Pedagogical Agents, Twenty Years Later." In: *International Journal of Artificial Intelligence in Education* 26.1, pp. 25–36. ISSN: 1560-4306. DOI: 10.1007/s40593-015-0065-9.

— (2018). "Pedagogical Agents: Back to the Future." In: *AI Magazine* 39.2.

Jong, Nivja H. de and Ton Wempe (May 2009). "Praat Script to Detect Syllable Nuclei and Measure Speech Rate Automatically." In: *Behavior Research Methods* 41.2, pp. 385–390. ISSN: 1554-351X, 1554-3528. DOI: 10.3758/BRM.41.2.385.

Jonsson, Ing-Marie, Clifford Nass, Helen Harris, and Leila Takayama (2005). "Matching In-Car Voice with Driver State: Impact on Attitude and Driving Performance." In: *Driving Assessment 2005: 3rd International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle DesignNissan Technical Center - North AmericaUniversity of Iowa, Iowa CityHonda R & D Americas, IncorporatedToyota Technical Center, U.S.A.Federal Motor Carrier Safety AdministrationNational Highway Traffic Safety Administration5DT, Inc.DriveSafety, Inc.Human Factors and Ergonomics SocietySeeing MachinesTransportation Research BoardUniversity of CalgaryUniversity of LeedsUniversity of Minnesota, MinneapolisUniversity of Michigan Transportation Research InstituteVirginia Polytechnic Institute and State University, Blacksburg*. ISBN: 978-0-87414-151-1.

Judge, Sharon, Kimberly Floyd, and Tara Jeffs (Jan. 2015). "Using Mobile Media Devices and Apps to Promote Young Children's Learning." In: *Young Children and Families in the Information Age*. Ed. by Kelly L. Heider and Mary Renck Jalongo. Educating the Young Child 10. Springer Netherlands, pp. 117–131. ISBN: 978-94-017-9183-0 978-94-017-9184-7.

Justice, Laura M., Yaacov Petscher, Christopher Schatschneider, and Andrew Mashburn (Nov. 2011). "Peer Effects in Preschool Classrooms: Is Children's Language Growth Associated With Their Classmates' Skills?" In: *Child Development* 82.6, pp. 1768–1777. ISSN: 1467-8624. DOI: 10.1111/j.1467-8624.2011.01665.x.

Kaestle, Carl F., Anne Campbell, Jeremy D. Finn, and Sylvia T. Johnson (2001). *Adult Literacy and Education in America*.

Kahn, Peter H., Batya Friedman, and Jennifer Hagman (2002). "I Care about Him as a Pal: Conceptions of Robotic Pets in Online Aibo Discussion Forums." In: *CHI'02 Extended Abstracts on Human Factors in Computing Systems*. ACM, pp. 632–633.

Kahn, Peter H., Batya Friedman, Deanne R. Perez-Granados, and Nathan G. Freier (2006). "Robotic Pets in the Lives of Preschool Children." In: *Interaction Studies* 7.3, pp. 405–436.

Kahn, Peter H., Aimee L. Reichert, Heather E. Gary, Takayuki Kanda, Hiroshi Ishiguro, Solace Shen, Jolina H. Ruckert, and Brian Gill (2011). "The New Ontological Category Hypothesis in Human-Robot Interaction." In: *Proceedings of the 6th International Conference on Human-Robot Interaction*. ACM, pp. 159–160.

Kahn, Peter H., Takayuki Kanda, Hiroshi Ishiguro, Nathan G. Freier, Rachel L. Severson, Brian T. Gill, Jolina H. Ruckert, and Solace Shen (2012). ""Robovie, You'll Have to Go into the Closet Now": Children's Social and Moral Relationships with a Humanoid Robot." In: *Developmental psychology* 48.2, p. 303.

Kanda, T., M. Shimada, and S. Koizumi (2012). "Children Learning with a Social Robot." In: *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 351–358. DOI: 10.1145/2157689.2157809.

Kanda, T., R. Sato, N. Saiwaki, and H. Ishiguro (Oct. 2007). "A Two-Month Field Trial in an Elementary School for Long-Term Human #x2013;Robot Interaction." In: *IEEE Transactions on Robotics* 23.5, pp. 962–971. ISSN: 1552-3098. DOI: 10.1109/TRO.2007.904904.

Kanda, Takayuki, Takayuki Hirano, Daniel Eaton, and Hiroshi Ishiguro (2004). "Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial." In: *Hum.-Comput. Interact.* 19.1, pp. 61–84. ISSN: 0737-0024. DOI: 10.1207/s15327051hci1901\&2_4.

Kanda, Takayuki, Dylan F. Glas, Masahiro Shiomi, Hiroshi Ishiguro, and Norihiro Hagita (2008). "Who Will Be the Customer?: A Social Robot That Anticipates People's Behavior from Their Trajectories." In: *Proceedings of the 10th International Conference on Ubiquitous Computing*. ACM, pp. 380–389.

Kanero, Junko, Vasfiye Geçkin, Cansu Oranç, Ezgi Mamus, Aylin C. Küntay, and Tilbe Göksun (Sept. 2018). "Social Robots for Early Language Learning: Current Evidence and Future Directions." In: *Child Development Perspectives* 12.3, pp. 146–151. ISSN: 1750-8606. DOI: 10.1111/cdep.12277.

Karacora, Bilge, Morteza Dehghani, Nicole Kramer-Mertens, and Jonathan Gratch (2012). "The Influence of Virtual Agents' Gender and Rapport on Enhancing Math Performance." In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 34.

Kasap, Zerrin and Nadia Magnenat-Thalmann (2010). "Towards Episodic Memory-Based Long-Term Affective Interaction with a Human-like Robot." In: *RO-MAN, 2010 IEEE*. IEEE, pp. 452–457.

— (2012). "Building Long-Term Relationships with Virtual and Robotic Characters: The Role of Remembering." In: *The Visual Computer* 28.1, pp. 87–97.

Kelley, Harold H, Ellen Berscheid, Andrew Christensen, J Harvey, T Huston, G Levinger, E McClintock, L Peplau, and D Peterson (1983). *Close Relationships*. OCLC: 902175384. New York: Freeman. ISBN: 978-0-7167-1443-9.

Kennedy, J., P. Baxter, E. Senft, and T. Belpaeme (Mar. 2016a). "Heart vs Hard Drive: Children Learn More from a Human Tutor than a Social Robot." In: *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 451–452. DOI: 10.1109/HRI.2016.7451801.

Kennedy, James, Paul Baxter, and Tony Belpaeme (2015). "The Robot Who Tried Too Hard: Social Behaviour of a Robot Tutor Can Negatively Affect Child Learning." In: *Proc. HRI*. Vol. 15.

— (Jan. 2017). "Nonverbal Immediacy as a Characterisation of Social Behaviour for Human–Robot Interaction." In: *International Journal of Social Robotics* 9.1, pp. 109–128. ISSN: 1875-4791, 1875-4805. DOI: 10.1007/s12369-016-0378-3.

Kennedy, James, Paul Baxter, Emmanuel Senft, and Tony Belpaeme (2016b). "Social Robot Tutoring for Child Second Language Learning." In: *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 231–238.

Kennedy, James, Séverin Lemaignan, Caroline Montassier, Pauline Lavalade, Bahar Irfan, Fotios Papadopoulos, Emmanuel Senft, and Tony Belpaeme (2017). "Child Speech Recognition in Human-Robot Interaction: Evaluations and Recommendations." In: *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. HRI '17. New York, NY, USA: ACM, pp. 82–90. ISBN: 978-1-4503-4336-7. DOI: 10.1145/2909824.3020229.

Kidd, Cory D. and Cynthia Breazeal (2008). "Robots at Home: Understanding Long-Term Human-Robot Interaction." In: *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference On*. IEEE, pp. 3230–3235.

Kim, Elizabeth S., Lauren D. Berkovits, Emily P. Bernier, Dan Leyzberg, Frederick Shic, Rhea Paul, and Brian Scassellati (May 2013). "Social Robots as Embedded Reinforcers of Social Behavior in Children with Autism." In: *Journal of Autism and Developmental Disorders* 43.5, pp. 1038–1049. ISSN: 0162-3257, 1573-3432. DOI: 10.1007/s10803-012-1645-2.

Kim, Yanghee and Amy L. Baylor (Nov. 2006). "A Social-Cognitive Framework for Pedagogical Agents as Learning Companions." In: *Educational Technology Research and Development* 54.6, pp. 569–596. ISSN: 1042-1629, 1556-6501. DOI: 10.1007/s11423-006-0637-3.

Kim, Yanghee and Jae Hoon Lim (2013). "Gendered Socialization with an Embodied Agent: Creating a Social and Affable Mathematics Learning Environment for Middle-Grade Females." In: *Journal of educational psychology* 105.4, p. 1164.

Klapper, André, Richard Ramsey, Daniël Wigboldus, and Emily S. Cross (2014). "The Control of Automatic Imitation Based on Bottom–Up and Top–Down Cues to Animacy: Insights from Brain and Behavior." In: *Journal of Cognitive Neuroscience* 26.11, pp. 2503–2513. ISSN: 0898-929X. DOI: 10.1162/jocn_a_00651.

Klatt, Dennis H. and Laura C. Klatt (Feb. 1990). "Analysis, Synthesis, and Perception of Voice Quality Variations among Female and Male Talkers." In: *The Journal of the Acoustical Society of America* 87.2, pp. 820–857. ISSN: 0001-4966. DOI: 10.1121/1.398894.

Knox, W. Bradley, Samuel Spaulding, and Cynthia Breazeal (2016). "Learning from the Wizard: Programming Social Interaction Through Teleoperated Demonstrations (Extended Abstract)." In: *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*. AAMAS '16. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, pp. 1309–1310. ISBN: 978-1-4503-4239-1.

Koppel, Moshe, Shlomo Argamon, and Anat Rachel Shimoni (Nov. 2002). "Automatically Categorizing Written Texts by Author Gender." In: *Literary and Linguistic Computing* 17.4, pp. 401–412. ISSN: 0268-1145. DOI: 10.1093/llc/17.4.401.

Kory-Westlund, Jacqueline M. and Cynthia Breazeal (2015a). "Deception, Secrets, Children, and Robots: What's Acceptable?" In: *Workshop on The Emerging Policy and Ethics of Human-Robot Interaction, Held in Conjunction with the 10th ACM/IEEE International Conference on Human-Robot Interaction*.

— (2015b). "The Interplay of Robot Language Level with Children's Language Learning During Storytelling." In: *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*. HRI'15 Extended Abstracts. New York, NY, USA: ACM, pp. 65–66. ISBN: 978-1-4503-3318-4. DOI: 10.1145/2701973.2701989.

Kory-Westlund, Jacqueline M., Cynthia Breazeal, and Anastasia K. Ostrowski (in review). "Physical or Virtual Robots—Which Is Better? A Survey of Embodiment and Presence in Human-Robot Interaction." In:

Kory-Westlund, Jacqueline M., Leah Dickens, Sooyeon Jeong, Paul Harris, David DeSteno, and Cynthia Breazeal (2015a). "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, The Netherlands.

Kory-Westlund, Jacqueline M., Goren Gordon, Sam Spaulding, Jin Joo Lee, Luke Plummer, Marayna Martinez, Madhurima Das, and Cynthia Breazeal (2015b). "Learning a Second Language with a Socially Assistive Robot." In: *New Friends: The 1st International Conference on Social Robots in Therapy and Education*. Almere, The Netherlands.

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

Kory-Westlund, Jacqueline M., Goren Gordon, Samuel Spaulding, Jin Joo Lee, Luke Plummer, Marayna Martinez, Madhurima Das, and Cynthia Breazeal (2016b). "Lessons From Teachers on Performing HRI Studies with Young Children in Schools." In: *The Eleventh ACM/IEEE International Conference on Human Robot Interation*. HRI '16. Piscataway, NJ, USA: IEEE Press, pp. 383–390. ISBN: 978-1-4673-8370-7.

Kory-Westlund, Jacqueline M., Jin Joo Lee, Luke Plummer, Fardad Faridi, Jesse Gray, Matt Berlin, Harald Quintus-Bosz, Robert Hartmann, Mike Hess, Stacy Dyer, dos Santos Kris, Aðalgeirsson, Sigurður Örn, Gordon, Goren, Spaulding, Samuel, Martinez, Marayna, Das, Madhurima, Archie, Maryam, Jeong, Sooyeon, and Breazeal, Cynthia (2016c). "Tega: A Social Robot." In: *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. DOI: 10.1109/HRI.2016.7451856.

Kory-Westlund, Jacqueline M., Leah Dickens, Sooyeon Jeong, Paul L. Harris, David DeSteno, and Cynthia L. Breazeal (2017a). "Children Use Non-Verbal Cues to Learn New Words from Robots as Well as People." In: *International Journal of Child-Computer Interaction*. ISSN: 2212-8689. DOI: 10.1016/j.ijcci.2017.04.001.

Kory-Westlund, Jacqueline M., Sooyeon Jeong, Hae Won Park, Samuel Ronfard, Aradhana Adhikari, Paul Lansley Harris, David DeSteno, and Cynthia Breazeal (2017b). "Flat versus Expressive Storytelling: Young Children's Learning and Retention of a Social Robot's Narrative." In: *Frontiers in Human Neuroscience* 11. ISSN: 1662-5161. DOI: 10.3389/fnhum.2017.00295.

Kory-Westlund, Jacqueline M., Hae Won Park, Randi Williams, and Cynthia Breazeal (2017c). "Measuring Children's Long-Term Relationships with Social Robots." In: *Workshop on Perception and Interaction dynamics in Child-Robot Interaction, held in conjunction with the Robotics: Science and Systems XIII.*

Kory-Westlund, Jacqueline M., Hae Won Park, Randi Williams, and Cynthia Breazeal (2018). "Measuring Young Children's Long-Term Relationships with Social Robots." In: *Proceedings of the 17th ACM Conference on Interaction Design and Children*. ACM, pp. 207–218.

Kory, Jacqueline M., Sooyeon Jeong, and Cynthia L. Breazeal (2013). "Robotic Learning Companions for Early Language Development." In: *In J. Epps, F. Chen, S. Oviatt, & K. Mase (Eds.), Proceedings of the 15th ACM on International Conference on Multimodal Interaction*. New York, NY: ACM: ACM, pp. 71–72.

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

Kory, Jacqueline and Cynthia Breazeal (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*, pp. 643–648. DOI: 10.1109/ROMAN.2014.6926325.

Kozima, Hideki, Marek P. Michalowski, and Cocoro Nakagawa (2009). "Keepon." In: *International Journal of Social Robotics* 1.1, pp. 3–18.

Krämer, Nicole C., Bilge Karacora, Gale Lucas, Morteza Dehghani, Gina Rüther, and Jonathan Gratch (Aug. 2016). "Closing the Gender Gap in STEM with Friendly Male Instructors? On the Effects of Rapport Behavior and Gender of a Virtual Agent in an Instructional Interaction." In: *Computers & Education* 99, pp. 1–13. ISSN: 0360-1315. DOI: 10.1016/j.compedu.2016.04.002.

Kruijff-Korbayová, Ivana, Elettra Oleari, Anahita Bagherzadhalimi, Francesca Sacchitelli, Bernd Kiefer, Stefania Racioppa, Clara Pozzi, and Alberto Sanna (Oct. 2015). "Young Users' Perception of a Social Robot Displaying Familiarity and Eliciting Disclosure." In: *Social Robotics*. Springer, Cham, pp. 380–389. DOI: 10.1007/978-3-319-25554-5_38.

Kuhl, Patricia K. (2007). "Is Speech Learning "Gated" by the Social Brain?" In: *Developmental science* 10.1, pp. 110–120.

— (Sept. 2011). "Social Mechanisms in Early Language Acquisition: Understanding Integrated Brain Systems Supporting Language." In: *The Oxford Handbook of Social Neuroscience*. Ed. by Jean Decety and John T. Cacioppo.

Kuhl, Patricia K., Feng-Ming Tsao, and Huei-Mei Liu (2003). "Foreign-Language Experience in Infancy: Effects of Short-Term Exposure and Social Interaction on Phonetic Learning." In: *Proceedings of the National Academy of Sciences* 100.15, pp. 9096–9101.

Kuss, Daria and Mark Griffiths (Apr. 2012). "Internet Gaming Addiction: A Systematic Review of Empirical Research." In: *International Journal of Mental Health & Addiction* 10.2, pp. 278–296. ISSN: 15571874. DOI: 10.1007/s11469-011-9318-5.

Kutner, Mark, Elizabeth Greenberg, and Justin Baer (2006). *A First Look at the Literacy of America's Adults in the 21st Century. NCES 2006-470*. ED Pubs, P.

Kwon, Sungjun, Hyunseok Kim, and Kwang Suk Park (2012). "Validation of Heart Rate Extraction Using Video Imaging on a Built-in Camera System of a Smartphone." In: *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. IEEE, pp. 2174–2177.

Ladd, Gary W., Becky J. Kochenderfer, and Cynthia C. Coleman (June 1996). "Friendship Quality as a Predictor of Young Children's Early School Adjustment." In: *Child Development* 67.3, pp. 1103–1118. ISSN: 1467-8624. DOI: 10.1111/j.1467-8624.1996.tb01785.x.

Lakey, Brian and Sheldon Cohen (2000). "Social Support Theory and Measurement." In: *Social Support Measurement and Intervention: A Guide for Health and Social Scientists*. Ed. by S. Cohen, L. G. Underwood, and B. H. Gottlieb. New York, NY, US: Oxford University Press, pp. 29–52. ISBN: 978-0-19-512670-9.

Lakin, Jessica L., Valerie E. Jefferis, Clara Michelle Cheng, and Tanya L. Chartrand (2003). "The Chameleon Effect as Social Glue: Evidence for the Evolutionary Significance of Nonconscious Mimicry." In: *Journal of Nonverbal Behavior* 27.3, pp. 145–162.

Lasseter, John (1987). "Principles of Traditional Animation Applied to 3D Computer Animation." In: *ACM Siggraph Computer Graphics*. Vol. 21. ACM, pp. 35–44.

Lee, Hee Rin, JaYoung Sung, Selma Šabanović, and Joenghye Han (2012a). "Cultural Design of Domestic Robots: A Study of User Expectations in Korea and the United States." In: *RO-MAN, 2012 IEEE*. IEEE, pp. 803–808.

Lee, Jin Joo, W. Bradley Knox, Jolie B. Wormwood, Cynthia Breazeal, and David DeSteno (Dec. 2013). "Computationally Modeling Interpersonal Trust." In: *Frontiers in Psychology* 4. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2013.00893.

Lee, Min Kyung, Jodi Forlizzi, Sara Kiesler, Paul Rybski, John Antanitis, and Sarun Savetsila (2012b). "Personalization in HRI: A Longitudinal Field Experiment." In: *Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference On*. IEEE, pp. 319–326.

Lee, Sungjin, Hyungjong Noh, Jonghoon Lee, Kyusong Lee, Gary Geunbae Lee, Seongdae Sagong, and Munsang Kim (2011). "On the Effectiveness of Robot-Assisted Language Learning." In: *ReCALL* 23.01, pp. 25–58.

Leite, I., C. Martinho, A. Pereira, and A. Paiva (Sept. 2009). "As Time Goes by: Long-Term Evaluation of Social Presence in Robotic Companions." In: *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 669–674. DOI: 10.1109/ROMAN.2009.5326256.

Leite, Iolanda, Carlos Martinho, and Ana Paiva (Apr. 2013). "Social Robots for Long-Term Interaction: A Survey." In: *International Journal of Social Robotics* 5.2, pp. 291–308. ISSN: 1875-4791, 1875-4805. DOI: 10.1007/s12369-013-0178-y.

Leite, Iolanda, André Pereira, and Jill Fain Lehman (2017). "Persistent Memory in Repeated Child-Robot Conversations." In: *Proceedings of the 2017 Conference on Interaction Design and Children*. IDC '17. New York, NY, USA: ACM, pp. 238–247. ISBN: 978-1-4503-4921-5. DOI: 10.1145/3078072.3079728.

Leite, Iolanda, Ginevra Castellano, AndrÃ© Pereira, Carlos Martinho, and Ana Paiva (2012a). "Long-Term Interactions with Empathic Robots: Evaluating Perceived Support in Children." In: Social Robotics. Springer, pp. 298–307.

Leite, Iolanda, AndrÃ© Pereira, Ginevra Castellano, Samuel Mascarenhas, Carlos Martinho, and Ana Paiva (2012b). "Modelling Empathy in Social Robotic Companions." In: Advances in User Modeling. Springer, pp. 135–147.

Leite, Iolanda, Ginevra Castellano, André Pereira, Carlos Martinho, and Ana Paiva (Aug. 2014). "Empathic Robots for Long-Term Interaction." In: *International Journal of Social Robotics* 6.3, pp. 329–341. ISSN: 1875-4791, 1875-4805. DOI: 10.1007/s12369-014-0227-1.

Leite, Iolanda, Marissa McCoy, Monika Lohani, Daniel Ullman, Nicole Salomons, Charlene Stokes, Susan Rivers, and Brian Scassellati (2015). "Emotional Storytelling in the Classroom: Individual versus Group Interaction between Children and Robots." In: *Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI'15). ACM (to Appear).*

Lester, James C., Sharolyn A. Converse, Susan E. Kahler, S. Todd Barlow, Brian A. Stone, and Ravinder S. Bhogal (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, USA: ACM, pp. 359–366. ISBN: 978-0-89791-802-2. DOI: 10.1145/258549.258797.

Levitan, Rivka, Stefan Benus, Ramiro H. Gálvez, Agustín Gravano, Florencia Savoretti, Marian Trnka, Andreas Weise, and Julia Hirschberg (2016). "Implementing Acoustic-Prosodic Entrainment in a Conversational Avatar." In: *INTERSPEECH*, pp. 1166–1170.

Leyzberg, Daniel, Samuel Spaulding, and Brian Scassellati (2014). "Personalizing Robot Tutors to Individuals' Learning Differences." In: *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, pp. 423–430.

Li, Jamy (May 2015). "The Benefit of Being Physically Present: A Survey of Experimental Works Comparing Copresent Robots, Telepresent Robots and Virtual Agents." In: *International*

Journal of Human-Computer Studies 77, pp. 23–37. ISSN: 1071-5819. DOI: 10.1016/j.ijhcs.2015.01.001.

Liew, Jeffrey (June 2012). "Effortful Control, Executive Functions, and Education: Bringing Self-Regulatory and Social-Emotional Competencies to the Table." In: *Child Development Perspectives* 6.2, pp. 105–111. ISSN: 1750-8606. DOI: 10.1111/j.1750-8606.2011.00196.x.

Lin, Patrick, George Bekey, and Keith Abney (2008). *Autonomous Military Robotics: Risk, Ethics, and Design*. Tech. rep. DTIC Document.

Lin, Tzu-Jung, Laura M. Justice, Narmada Paul, and Andrew J. Mashburn (2016). "Peer Interaction in Rural Preschool Classrooms: Contributions of Children's Learning-Related Behaviors, Language and Literacy Skills, and Problem Behaviors." In: *Early Childhood Research Quarterly* 37, pp. 106–117. ISSN: 0885-2006. DOI: 10.1016/j.ecresq.2016.04.001.

Lubold, N., H. Pon-Barry, and E. Walker (Dec. 2015). "Naturalness and Rapport in a Pitch Adaptive Learning Companion." In: *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, pp. 103–110. DOI: 10.1109/ASRU.2015.7404781.

Lubold, Nichola (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, USA: ACM, pp. 297–300. ISBN: 978-1-4503-4656-6. DOI: 10.1145/3027063.3027132.

Lubold, Nichola, Erin Walker, and Heather Pon-Barry (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, USA: IEEE Press, pp. 255–262. ISBN: 978-1-4673-8370-7.

Lubold, Nichola, Erin Walker, Heather Pon-Barry, and Amy Ogan (2018). "Automated Pitch Convergence Improves Learning in a Social, Teachable Robot for Middle School Mathematics." In: *Artificial Intelligence in Education*. Ed. by Carolyn Penstein Rosé, Roberto Martínez-Maldonado, H. Ulrich Hoppe, Rose Luckin, Manolis Mavrikis, Kaska Porayska-Pomsta, Bruce McLaren, and Benedict du Boulay. Lecture Notes in Computer Science. Springer International Publishing, pp. 282–296. ISBN: 978-3-319-93843-1.

Lucas, Gale M., Jonathan Gratch, Aisha King, and Louis-Philippe Morency (2014). "It's Only a Computer: Virtual Humans Increase Willingness to Disclose." In: *Computers in Human Behavior* 37, pp. 94–100. ISSN: 0747-5632. DOI: 10.1016/j.chb.2014.04.043.

Manson, Joseph H., Gregory A. Bryant, Matthew M. Gervais, and Michelle A. Kline (Nov. 2013). "Convergence of Speech Rate in Conversation Predicts Cooperation." In: *Evolution and Human Behavior* 34.6, pp. 419–426. ISSN: 1090-5138. DOI: 10.1016/j.evolhumbehav.2013.08.001.

Mashburn, Andrew J., Laura M. Justice, Jason T. Downer, and Robert C. Pianta (May 2009). "Peer Effects on Children's Language Achievement During Pre-Kindergarten." In: *Child Development* 80.3, pp. 686–702. ISSN: 1467-8624. DOI: 10.1111/j.1467-8624.2009.01291.x.

Mathes, Patricia G., Jill K. Howard, Shelley H. Allen, and Douglas Fuchs (Jan. 1998). "Peer-Assisted Learning Strategies for First-Grade Readers: Responding to the Needs of Diverse Learners." In: *Reading Research Quarterly* 33.1, pp. 62–94. ISSN: 1936-2722. DOI: 10.1598/RRQ.33.1.4.

Matsuzoe, Shizuko and Fumihide Tanaka (2012). "How Smartly Should Robots Behave?: Comparative Investigation on the Learning Ability of a Care-Receiving Robot." In: *RO-MAN, 2012 IEEE*. IEEE, pp. 339–344.

McClelland, Megan M and Frederick J Morrison (2003). "The Emergence of Learning-Related Social Skills in Preschool Children." In: *Early Childhood Research Quarterly* 18.2, pp. 206–224. ISSN: 0885-2006. DOI: 10.1016/S0885-2006(03)00026-7.

McClure, Catherine, Maureen Cunningham, Sheana Bull, Stephen Berman, and Mandy A. Allison (Nov. 2018). "Using Mobile Health to Promote Early Language Development: A Narrative Review." In: *Academic Pediatrics* 18.8, pp. 850–854. ISSN: 1876-2859. DOI: 10.1016/j.acap.2018.07.010.

McCornack, Barbara L. (Dec. 1982). "Effects of Peer Familiarity on Play Behavior in Preschool Children." In: *Journal of Genetic Psychology* 141.2, p. 225. ISSN: 00221325.

McDuff, Daniel, Sarah Gontarek, and Rosalind Picard (2014). "Remote Measurement of Cognitive Stress via Heart Rate Variability." In: *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*. IEEE, pp. 2957–2960.

McDuff, Daniel, Abdelrahman Mahmoud, Mohammad Mavadati, May Amr, Jay Turcot, and Rana el Kaliouby (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, USA: ACM, pp. 3723–3726. ISBN: 978-1-4503-4082-3. DOI: 10.1145/2851581.2890247.

Mehrabian, Albert (Jan. 1968). "Some Referents and Measures of Nonverbal Behavior." In: *Behavior Research Methods & Instrumentation* 1.6, pp. 203–207. ISSN: 1554-3528. DOI: 10.3758/BF03208096.

Melson, Gail F. (June 1990). "Studying Children's Attachment to Their Pets: A Conceptual and Methodological Review." In: *Anthrozoös* 4.2, pp. 91–99. ISSN: 0892-7936. DOI: 10.2752/089279391787057297.

Melson, Gail F., Jr. Kahn Peter H., Alan Beck, and Batya Friedman (Sept. 2009). "Robotic Pets in Human Lives: Implications for the Human–Animal Bond and for Human Relationships with Personified Technologies." In: *Journal of Social Issues* 65.3, pp. 545–567. ISSN: 1540-4560. DOI: 10.1111/j.1540-4560.2009.01613.x.

Meltzoff, Andrew N., Patricia K. Kuhl, Javier Movellan, and Terrence J. Sejnowski (2009). "Foundations for a New Science of Learning." In: *Science* 325.5938, pp. 284–288.

Meltzoff, Andrew N., Rechele Brooks, Aaron P. Shon, and Rajesh PN Rao (2010). "Social Robots Are Psychological Agents for Infants: A Test of Gaze Following." In: *Neural Networks* 23.8, pp. 966–972.

Mendelson, Morton J. and Frances E. Aboud (1999). "Measuring Friendship Quality in Late Adolescents and Young Adults: McGill Friendship Questionnaires." In: *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement* 31.2, p. 130.

Moore, Richard (Nov. 2013). "Social Learning and Teaching in Chimpanzees." In: *Biology & Philosophy* 28.6, pp. 879–901. ISSN: 01693867. DOI: 10.1007/s10539-013-9394-y.

Movellan, Javier, Micah Eckhardt, Marjo Virnes, and Angelica Rodriguez (2009). "Sociable Robot Improves Toddler Vocabulary Skills." In: *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*. ACM, pp. 307–308.

Mulac, Anthony, James J. Bradac, and Pamela Gibbons (Jan. 2001). "Empirical Support for the Gender-as-Culture HypothesisAn Intercultural Analysis of Male/Female Language Differences." In: *Human Communication Research* 27.1, pp. 121–152. ISSN: 0360-3989. DOI: 10.1111/j.1468-2958.2001.tb00778.x.

Naigles, Letitia R. and Lara Mayeux (2001). "Television as Incidental Language Teacher." In: *Handbook of children and the media*, pp. 135–152.

Nalin, M., I. Baroni, I. Kruijff-Korbayová, L. Cañamero, M. Lewis, A. Beck, H. Cuayáhuitl, and A. Sanna (2012). "Children's Adaptation in Multi-Session Interaction with a Humanoid Robot." In: *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, pp. 351–357. DOI: 10.1109/ROMAN.2012.6343778.

National Security Letters. URL: https://www.eff.org/issues/national-security-letters (visited on 08/25/2015).

Neuman, Susan B., Ashley M. Pinkham, and Tanya Kaefer (2013). "Building Word and World Knowledge in the Early Years." In: *International Handbook of Research on Children's Literacy, Learning, and Culture*. Ed. by Kathy Hall, Teresa Cremin, Barbara Comber, and Luis C. Moll. John Wiley & Sons, Ltd, pp. 199–214. ISBN: 978-1-118-32334-2.

Neuman, Susan B. and Kathy Roskos (1992). "Literacy Objects as Cultural Tools: Effects on Children's Literacy Behaviors in Play." In: *Reading Research Quarterly* 27.3, pp. 203–225. ISSN: 0034-0553. DOI: 10.2307/747792.

Neumann, Michelle M. (Jan. 2018). "Using Tablets and Apps to Enhance Emergent Literacy Skills in Young Children." In: *Early Childhood Research Quarterly* 42, pp. 239–246. ISSN: 0885-2006. DOI: 10.1016/j.ecresq.2017.10.006.

Newcomb, Andrew F. and Catherine L. Bagwell (1995). "Children's Friendship Relations: A Meta-Analytic Review." In: *Psychological Bulletin* 117.2, pp. 306–347. ISSN: 1939-1455 0033-2909. DOI: 10.1037/0033-2909.117.2.306.

Niederhoffer, Kate G. and James W. Pennebaker (Dec. 2002). "Linguistic Style Matching in Social Interaction." In: *Journal of Language and Social Psychology* 21.4, pp. 337–360. ISSN: 0261-927X. DOI: 10.1177/026192702237953.

Nishio, Shuichi, Hiroshi Ishiguro, and Norihiro Hagita (2007). *Geminoid: Teleoperated Android of an Existing Person*. INTECH Open Access Publisher.

Nishiwaki, Yusaku, Masugi Furukawa, Soshi Yoshikawa, Nihan Karatas, and Michio Okada (2017). "iBones: A Weak Robot to Construct Mutual Coordination for Handing Out Tissue Packs." In: *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. HRI '17. New York, NY, USA: ACM, pp. 413–413. ISBN: 978-1-4503-4885-0. DOI: 10.1145/3029798.3036654.

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

Nowicki, Stephen and Marshall P. Duke (1974). "A Preschool and Primary Internal-External Control Scale." In: *Developmental Psychology* 10.6, pp. 874–880. ISSN: 1939-0599 0012-1649. DOI: 10.1037/h0037253.

Nuñez, David S. (David Scott) (2015). "GlobalLit : A Platform for Collecting, Analyzing, and Reacting to Children's Usage Data on Tablet Computers." Thesis. Massachusetts Institute of Technology.

Ong, Ethel Chua Joy, Danielle Grace Consignado, Sabrina Jane Ong, and Zhayne Chong Soriano (2017). "Building a Semantic Ontology for Virtual Peers in Narrative-Based Environments." In: *Trends in Artificial Intelligence: PRICAI 2016 Workshops*. Lecture Notes in Computer Science. Springer International Publishing, pp. 65–76. ISBN: 978-3-319-60675-0.

Paez, Mariela M., Patton O. Tabors, and Lisa M. Lopez (2007). "Dual Language and Literacy Development of Spanish-Speaking Preschool Children." In: *Journal of applied developmental psychology* 28.2, pp. 85–102.

Pahl, Kristine M. and Paula M. Barrett (July 2007). "The Development of Social–Emotional Competence in Preschool-Aged Children: An Introduction to the Fun FRIENDS Program." In: *Journal of Psychologists and Counsellors in Schools* 17.1, pp. 81–90. ISSN: 1839-2520, 1037-2911. DOI: 10.1375/ajgc.17.1.81.

Palestra, Giuseppe, Giovanna Varni, Mohamed Chetouani, and Floriana Esposito (2016). "A Multimodal and Multilevel System for Robotics Treatment of Autism in Children." In: *Proceedings of the International Workshop on Social Learning and Multimodal Interaction for Designing Artificial Agents*. DAA '16. New York, NY, USA: ACM, 3:1–3:6. ISBN: 978-1-4503-4560-6. DOI: 10.1145/3005338.3005341.

Pardos, Zach A., Ryan S. J. D. Baker, Maria San Pedro, Sujith M. Gowda, and Supreeth M. Gowda (2014). "Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes." In: *Journal of Learning Analytics* 1.1, pp. 107–128. ISSN: 1929-7750. DOI: 10.18608/jla.2014.11.6.

Parisi, Domenico and Giancarlo Petrosino (2010). "Robots That Have Emotions." In: *Adaptive Behavior* 18.6, pp. 453–469.

Park, H. W., R. A. Coogle, and A. Howard (May 2014). "Using a Shared Tablet Workspace for Interactive Demonstrations during Human-Robot Learning Scenarios." In: *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2713–2719. DOI: 10.1109/ICRA.2014.6907248.

Park, H. W. and A. M. Howard (May 2015). "Retrieving Experience: Interactive Instance-Based Learning Methods for Building Robot Companions." In: *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6140–6145. DOI: 10.1109/ICRA.2015.7140061.

Park, H. W. and A. Howard (Mar. 2013). "Providing Tablets as Collaborative-Task Workspace for Human-Robot Interaction." In: *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 207–208. DOI: 10.1109/HRI.2013.6483574.

Park, H. W., M. Gelsomini, J. J. Lee, T. Zhu, and C. Breazeal (May 2017a). "Backchannel Opportunity Prediction for Social Robot Listeners." In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2308–2314. DOI: 10.1109/ICRA.2017.7989266.

Park, Hae Won, Rinat Rosenberg-Kima, Maor Rosenberg, Goren Gordon, and Cynthia Breazeal (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, USA: ACM, pp. 137–145. ISBN: 978-1-4503-4336-7. DOI: 10.1145/2909824.3020213.

Park, Hae Won, Mirko Gelsomini, Jin Joo Lee, and Cynthia Breazeal (2017c). "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, USA: ACM, pp. 100–108. ISBN: 978-1-4503-4336-7. DOI: 10.1145/2909824.3020245.

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

Park, Sanghoon (2018). "Virtual Pedagogical Agents for English Language Teaching and Learning." In: *The TESOL Encyclopedia of English Language Teaching*. American Cancer Society, pp. 1–9. ISBN: 978-1-118-78423-5. DOI: 10.1002/9781118784235.eelt0912.

Parker, Jeffrey G., Kenneth H. Rubin, Stephen A. Erath, Julie C. Wojslawowicz, and Allison A. Buskirk (2015). "Peer Relationships, Child Development, and Adjustment: A Developmental Psychopathology Perspective." In: *Developmental Psychopathology*. John Wiley & Sons, Ltd, pp. 419–493. ISBN: 978-0-470-93938-3. DOI: 10.1002/9780470939383.ch12.

Passman, Richard H. and Jane S. Halonen (June 1979). "A Developmental Survey of Young Children's Attachments to Inanimate Objects." In: *The Journal of Genetic Psychology* 134.2, pp. 165–178. ISSN: 0022-1325. DOI: 10.1080/00221325.1979.10534051.

Pennebaker, James W., Matthias R. Mehl, and Kate G. Niederhoffer (2003). "Psychological Aspects of Natural Language Use: Our Words, Our Selves." In: *Annual Review of Psychology* 54.1, pp. 547–577. DOI: 10.1146/annurev.psych.54.101601.145041.

Perkins, Suzanne C., Eric D. Finegood, and James E. Swain (2013). "Poverty and Language Development: Roles of Parenting and Stress." In: *Innovations in Clinical Neuroscience* 10.4, pp. 10–19. ISSN: 2158-8333.

Pezzullo, Lydia G., Joseph B. Wiggins, Megan H. Frankosky, Wookhee Min, Kristy Elizabeth Boyer, Bradford W. Mott, Eric N. Wiebe, and James C. Lester (2017). ""Thanks Alisha, Keep in Touch": Gender Effects and Engagement with Virtual Learning Companions." In: *Artificial Intelligence in Education*. Ed. by Elisabeth André, Ryan Baker, Xiangen Hu, Ma. Mercedes T. Rodrigo, and Benedict du Boulay. Lecture Notes in Computer Science. Springer International Publishing, pp. 299–310. ISBN: 978-3-319-61425-0.

Philippot, Pierre, Robert S. Feldman, and Erik J. Coats (1999). *The Social Context of Nonverbal Behavior*. Cambridge University Press.

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

Picard, Rosalind W. and Jonathan Klein (2002). "Computers That Recognise and Respond to User Emotion: Theoretical and Practical Implications." In: *Interacting with Computers* 14.2, pp. 141–169. ISSN: 0953-5438. DOI: 10.1016/S0953-5438(01)00055-8.

Poh, Ming-Zher, Daniel J. McDuff, and Rosalind W. Picard (2011). "Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam." In: *Biomedical Engineering, IEEE Transactions on* 58.1, pp. 7–11.

Pontier, Matthijs and Ghazanfar F. Siddiqui (2008). "A Virtual Therapist That Responds Empathically to Your Answers." In: *Intelligent Virtual Agents*. Springer, pp. 417–425.

Porzel, Robert, Annika Scheffler, and Rainer Malaka (2006). "How Entrainment Increases Dialogical Effectiveness." In: *Proceedings of the IUI*. Vol. 6. Citeseer, pp. 35–42.

Provine, Robert R. (Dec. 2001). *Laughter: A Scientific Investigation*. Penguin. ISBN: 978-1-101-65925-0.

— (2012). *Curious Behavior: Yawning, Laughing, Hiccupping, and Beyond*. Harvard University Press.

Pütten, Astrid M. Rosenthal-von der, Carolin Straßmann, and Nicole C. Krämer (Sept. 2016). "Robots or Agents – Neither Helps You More or Less During Second Language Acquisition." In: *Intelligent Virtual Agents*. Lecture Notes in Computer Science. Springer, Cham, pp. 256–268. ISBN: 978-3-319-47664-3 978-3-319-47665-0. DOI: 10.1007/978-3-319-47665-0_23.

Radesky, Jenny S. and Dimitri A. Christakis (Oct. 2016). "Increased Screen Time: Implications for Early Childhood Development and Behavior." In: *Pediatric Clinics* 63.5, pp. 827–839. ISSN: 0031-3955, 1557-8240. DOI: 10.1016/j.pcl.2016.06.006.

Radesky, Jenny S., Jayna Schumacher, and Barry Zuckerman (Jan. 2015). "Mobile and Interactive Media Use by Young Children: The Good, the Bad, and the Unknown." In: *Pediatrics* 135.1, pp. 1–3. ISSN: 0031-4005, 1098-4275. DOI: 10.1542/peds.2014-2251.

Ramachandran, Aditi and Brian Scassellati (2015). "Developing Adaptive Social Robot Tutors for Children." In: *2015 AAAI Fall Symposium Series*.

Read, Janet C. (Apr. 2008). "Validating the Fun Toolkit: An Instrument for Measuring Children's Opinions of Technology." In: *Cognition, Technology & Work* 10.2, pp. 119–128. ISSN: 1435-5558, 1435-5566. DOI: 10.1007/s10111-007-0069-9.

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

Receptiviti API (2017). URL: http://receptiviti2.wpengine.com/.

Reeves, Byron and Clifford Nass (1996). *How People Treat Computers, Television, and New Media like Real People and Places*. CSLI Publications and Cambridge university press.

Reitter, David, Frank Keller, and Johanna D. Moore (2011). "A Computational Cognitive Model of Syntactic Priming." In: *Cognitive science* 35.4, pp. 587–637. ISSN: 0364-0213.

Reschly, Amy L., E. Scott Huebner, James J. Appleton, and Susan Antaramian (2008). "Engagement as Flourishing: The Contribution of Positive Emotions and Coping to Adolescents' Engagement at School and with Learning." In: *Psychology in the Schools* 45.5, pp. 419–431. ISSN: 1520-6807. DOI: 10.1002/pits.20306.

Reynolds, Carson and Rosalind Picard (2004). "Affective Sensors, Privacy, and Ethical Contracts." In: *CHI'04 Extended Abstracts on Human Factors in Computing Systems*. ACM, pp. 1103–1106.

Richards, Melissa N. and Sandra L. Calvert (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*. Ed. by Rachel Barr and Deborah Nichols Linebarger. Cham: Springer International Publishing, pp. 141–163. ISBN: 978-3-319-45102-2. DOI: 10.1007/978-3-319-45102-2_9.

Ricks, Daniel J. and Mark B. Colton (2010). "Trends and Considerations in Robot-Assisted Autism Therapy." In: *Robotics and Automation (ICRA), 2010 IEEE International Conference On*. IEEE, pp. 4354–4359.

Riek, Laurel D. and Don Howard (2014). "A Code of Ethics for the Human-Robot Interaction Profession." In: *WeRobot 2014 Conference, University of Miami*.

Riek, Laurel D., Woodrow Hartzog, Don Howard, A. Moon, and Ryan Calo (2015). "The Emerging Policy and Ethics of Human Robot Interaction." In: *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*. ACM, pp. 247–248.

Rintjema, Emmy, Rianne van den Berghe, Anne Kessels, Jan de Wit, and Paul Vogt (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, USA: ACM, pp. 219–220. ISBN: 978-1-4503-5615-2. DOI: 10.1145/3173386.3177059.

Robins, B., K. Dautenhahn, R. Te Boekhorst, and A. Billard (2005). "Robotic Assistants in Therapy and Education of Children with Autism: Can a Small Humanoid Robot Help Encourage Social Interaction Skills?" In: *Universal Access in the Information Society* 4.2, pp. 105–120.

Romeo, Rachel R., Julia A. Leonard, Sydney T. Robinson, Martin R. West, Allyson P. Mackey, Meredith L. Rowe, and John D. E. Gabrieli (May 2018a). "Beyond the 30-Million-Word Gap: Children's Conversational Exposure Is Associated With Language-Related Brain Function." In: *Psychological Science* 29.5, pp. 700–710. ISSN: 0956-7976. DOI: 10.1177/0956797617742725.

Romeo, Rachel R., Joshua Segaran, Julia A. Leonard, Sydney T. Robinson, Martin R. West, Allyson P. Mackey, Anastasia Yendiki, Meredith L. Rowe, and John D. E. Gabrieli (Sept. 2018b). "Language Exposure Relates to Structural Neural Connectivity in Childhood." In: *Journal of Neuroscience* 38.36, pp. 7870–7877. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.0484-18.2018.

Ros, Raquel, Elettra Oleari, Clara Pozzi, Francesca Sacchitelli, Daniele Baranzini, Anahita Bagherzadhalimi, Alberto Sanna, and Yiannis Demiris (2016). "A Motivational Approach to Support Healthy Habits in Long-Term Child–Robot Interaction." In: *International Journal of Social Robotics* 8.5, pp. 599–617. ISSN: 1875-4791, 1875-4805. DOI: 10.1007/s12369-016-0356-9.

Rotenberg, Ken J. (Sept. 1995). "Development of Children's Restrictive Disclosure to Friends." In: *The Journal of Genetic Psychology* 156.3, pp. 279–292. ISSN: 0022-1325. DOI: 10.1080/00221325.1995.9914823.

Rotenberg, Ken J. and Luanne Mann (1986). "The Development of the Norm of the Reciprocity of Self-Disclosure and Its Function in Children's Attraction to Peers." In: *Child Development* 57.6, pp. 1349–1357. ISSN: 0009-3920. DOI: 10.2307/1130414.

Rotenberg, Ken J., Nancy Eisenberg, Christine Cumming, Ashlee Smith, Mike Singh, and Elizabeth Terlicher (Jan. 2003). "The Contribution of Adults' Nonverbal Cues and Children's Shyness to the Development of Rapport between Adults and Preschool Children." In: *International Journal of Behavioral Development* 27.1, pp. 21–30. ISSN: 0165-0254. DOI: 10.1080/01650250143000571.

Rubin, Kenneth H., William M. Bukowski, and Jeffrey G. Parker (1998). "Peer Interactions, Relationships, and Groups." In: *Handbook of Child Psychology*. John Wiley & Sons, Inc. ISBN: 978-0-470-14765-8. DOI: 10.1002/9780470147658.chpsy0310.

Ryokai, Kimiko, Cati Vaucelle, and Justine Cassell (2003). "Virtual Peers as Partners in Storytelling and Literacy Learning." In: *Journal of Computer Assisted Learning* 19.2, pp. 195–208.

Sadoughi, Najmeh, André Pereira, Rishub Jain, Iolanda Leite, and Jill Fain Lehman (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*. ACM, pp. 91–99.

Sage, Kara D. and Dare Baldwin (2010). "Social Gating and Pedagogy: Mechanisms for Learning and Implications for Robotics." In: *Neural Networks* 23.8, pp. 1091–1098.

Salter, T., K. Dautenhahn, and R. Bockhorst (Sept. 2004). "Robots Moving out of the Laboratory - Detecting Interaction Levels and Human Contact in Noisy School Environments." In: *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No.04TH8759)*, pp. 563–568. DOI: 10.1109/ROMAN.2004.1374822.

Scassellati, Brian, Henny Admoni, and Maja Mataric (2012). "Robots for Use in Autism Research." In: *Annual review of biomedical engineering* 14, pp. 275–294.

Scassellati, Brian, Laura Boccanfuso, Chien-Ming Huang, Marilena Mademtzi, Meiying Qin, Nicole Salomons, Pamela Ventola, and Frederick Shic (Aug. 2018a). "Improving Social Skills in Children with ASD Using a Long-Term, in-Home Social Robot." In: *Science Robotics* 3.21, eaat7544. ISSN: 2470-9476. DOI: 10.1126/scirobotics.aat7544.

Scassellati, Brian, Jake Brawer, Katherine Tsui, Setareh Nasihati Gilani, Melissa Malzkuhn, Barbara Manini, Adam Stone, Geo Kartheiser, Arcangelo Merla, Ari Shapiro, David Traum, and Laura-Ann Petitto (2018b). "Teaching Language to Deaf Infants with a Robot and a Virtual Human." In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. New York, NY, USA: ACM, 553:1–553:13. ISBN: 978-1-4503-5620-6. DOI: 10.1145/3173574.3174127.

Schechter, Carlota and Beth Bye (2007). "Preliminary Evidence for the Impact of Mixed-Income Preschools on Low-Income Children's Language Growth." In: *Early Childhood Research Quarterly* 22.1, pp. 137–146. ISSN: 0885-2006. DOI: 10.1016/j.ecresq.2006.11.005.

Schilbach, Leonhard, Bert Timmermans, Vasudevi Reddy, Alan Costall, Gary Bente, Tobias Schlicht, and Kai Vogeley (Aug. 2013). "Toward a Second-Person Neuroscience." In: *Behavioral and Brain Sciences* 36.4, pp. 393–414. ISSN: 0140-525X, 1469-1825. DOI: 10.1017/S0140525X12000660.

Schwab, Jessica F. and Casey Lew-Williams (2016). "Language Learning, Socioeconomic Status, and Child-Directed Speech." In: *Wiley interdisciplinary reviews. Cognitive science* 7.4, pp. 264–275. ISSN: 1939-5078. DOI: 10.1002/wcs.1393.

Schwienhorst, Klaus (2002a). "The State of VR: A Meta-Analysis of Virtual Reality Tools in Second Language Acquisition." In: *Computer Assisted Language Learning* 15.3, pp. 221–239. ISSN: 0958-8221. DOI: 10.1076/call.15.3.221.8186.

— (2002b). "Why Virtual, Why Environments? Implementing Virtual Reality Concepts in Computer-Assisted Language Learning." In: *Simulation & Gaming* 33.2, pp. 196–209. ISSN: 1046-8781. DOI: 10.1177/1046878102332008.

Searle, John R. (1980). "Minds, Brains, and Programs." In: *Behavioral and Brain Sciences* 3.3, pp. 417–424. ISSN: 1469-1825, 0140-525X. DOI: 10.1017/S0140525X00005756.

Selwyn, Neil (Jan. 2004). "Reconsidering Political and Popular Understandings of the Digital Divide." In: *New Media & Society* 6.3, pp. 341–362. ISSN: 1461-4448, 1461-7315. DOI: 10.1177/1461444804042519.

Semin, Gün R. and John T. Cacioppo (2008). "Grounding Social Cognition: Synchronization, Coordination, and Co-Regulation." In:

Sénéchal, Monique (Feb. 1997). "The Differential Effect of Storybook Reading on Preschoolers' Acquisition of Expressive and Receptive Vocabulary." In: *Journal of Child Language* 24.1, pp. 123–138. ISSN: 1469-7602, 0305-0009.

Senechal, Thibaud, Daniel McDuff, and Rana el Kaliouby (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, USA: IEEE Computer Society, pp. 10–18. ISBN: 978-1-4673-9711-7. DOI: 10.1109/ICCVW.2015.11.

Serholt, Sofia and Wolmet Barendregt (2016). "Robots Tutoring Children: Longitudinal Evaluation of Social Engagement in Child-Robot Interaction." In: *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*. NordiCHI '16. New York, NY, USA: ACM, 64:1–64:10. ISBN: 978-1-4503-4763-1. DOI: 10.1145/2971485.2971536.

Serholt, Sofia, Wolmet Barendregt, Iolanda Leite, Helen Hastie, Andrew Jones, Ana Paiva, Asimina Vasalou, and Ginevra Castellano (2014). "Teachers' Views on the Use of Empathic Robotic Tutors in the Classroom." In: *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium On*. IEEE, pp. 955–960.

Setapen, Adam Michael (2012). "Creating Robotic Characters for Long-Term Interaction." Master's Thesis. Cambridge, MA: MIT.

Severson, Rachel L. and Stephanie M. Carlson (2010). "Behaving as or Behaving as If? Children's Conceptions of Personified Robots and the Emergence of a New Ontological Category." In: *Neural Networks* 23.8, pp. 1099–1103.

Shmueli, Benjamin and Ayelet Blecher-Prigat (Jan. 2011). *Privacy for Children*. SSRN Scholarly Paper ID 1746540. Rochester, NY: Social Science Research Network.

Short, Elaine, Katelyn Swift-Spong, Jillian Greczek, Aditi Ramachandran, Alexandru Litoiu, Elena Corina Grigore, David Feil-Seifer, Samuel Shuster, Jin Joo Lee, Shaobo Huang, and others (2014). "How to Train Your Dragonbot: Socially Assistive Robots for Teaching Children about Nutrition through Play." In: *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium On*. IEEE, pp. 924–929.

Shuler, Carly, Zachary Levine, and Jinny Ree (2012). "iLearn II An Analysis of the Education Category of Apple's App Store." In:

Singer, Dorothy G., Roberta Michnick Golinkoff, and Kathy Hirsh-Pasek (2006). *Play= Learning: How Play Motivates and Enhances Children's Cognitive and Social-Emotional Growth*. Oxford University Press.

Singer, Dorothy G., Jerome L. Singer, Heidi D'Agnostino, and Raeka DeLong (2009). "Children's Pastimes and Play in Sixteen Nations: Is Free-Play Declining?" In: *American Journal of Play* 1.3, pp. 283–312. ISSN: 1938-0399.

Singh, Nikhita (2018). "Talking Machines: Democratizing the Design of Voice-Based Agents for the Home." Master's Thesis. Cambridge, MA, USA: Massachusetts Institute of Technology.

Sinha, Tanmay and Justine Cassell (June 2015a). "Fine-Grained Analyses of Interpersonal Processes and Their Effect on Learning." In: *Artificial Intelligence in Education*. Springer, Cham, pp. 781–785. DOI: 10.1007/978-3-319-19773-9_115.

— (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, USA: ACM, pp. 13–20. ISBN: 978-1-4503-3986-5. DOI: 10.1145/2823513.2823516.

Sisson, Susan B., Stephanie T. Broyles, Birgitta L. Baker, and Peter T. Katzmarzyk (2010). "Screen Time, Physical Activity, and Overweight in U.S. Youth: National Survey of Children's Health 2003." In: *Journal of Adolescent Health* 47.3, pp. 309–311. ISSN: 1054-139X. DOI: 10.1016/j.jadohealth.2010.02.016.

Smidl, Sarah Lynn (2006). "Portraits of Laughter in" Kid" Ergarten Children: The Giggles and Guffaws That Support Teaching, Learning, and Relationships." In:

Snow, Catherine E., Michelle V. Porche, Patton O. Tabors, and Stephanie R. Harris (2007). *Is Literacy Enough? Pathways to Academic Success for Adolescents*. ERIC.

Sorenson, David N. (Apr. 1989). "A Fundamental Frequency Investigation of Children Ages 6–10 Years Old." In: *Journal of Communication Disorders* 22.2, pp. 115–123. ISSN: 0021-9924. DOI: 10.1016/0021-9924(89)90028-2.

Spaulding, Samuel, Goren Gordon, and Cynthia Breazeal (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, pp. 864–872. ISBN: 978-1-4503-4239-1.

Spencer-Oatey, Helen (1996). "Reconsidering Power and Distance." In: *Journal of Pragmatics* 26.1, pp. 1–24. ISSN: 0378-2166. DOI: 10.1016/0378-2166(95)00047-X.

Stenzel, Anna, Eris Chinellato, Maria A. Tirado Bou, Angel P. del Pobil, Markus Lappe, and Roman Liepelt (2012). "When Humanoid Robots Become Human-like Interaction Partners: Corepresentation of Robotic Actions." In: *Journal of Experimental Psychology: Human Perception and Performance* 38.5, p. 1073.

Stone, Sandra J. and James F. Christie (1996). "Collaborative Literacy Learning During Sociodramatic Play in a Multiage (K-2) Primary Classroom." In: *Journal of Research in Childhood Education* 10.2, pp. 123–133. ISSN: 0256-8543. DOI: 10.1080/02568549609594895.

Sung, Ja-Young, Lan Guo, Rebecca E. Grinter, and Henrik I. Christensen (2007). ""My Roomba Is Rambo": Intimate Home Appliances." In: *UbiComp 2007: Ubiquitous Computing*. Ed. by John Krumm, Gregory D. Abowd, Aruna Seneviratne, and Thomas Strang. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 145–162. ISBN: 978-3-540-74853-3.

Suzuki, Noriko and Yasuhiro Katagiri (2007). "Prosodic Alignment in Human-Computer Interaction." In: *Connection Science* 19.2, pp. 131–141.

Takeuchi, Lori, Reed Stevens, and others (2011). "The New Coviewing: Designing for Learning through Joint Media Engagement." In: *New York, NY: The Joan Ganz Cooney Center at Sesame Workshop*.

Tanaka, F. and T. Kimura (Sept. 2009). "The Use of Robots in Early Education: A Scenario Based on Ethical Consideration." In: *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 558–560. DOI: 10.1109/ROMAN.2009.5326227.

Tanaka, F., K. Isshiki, F. Takahashi, M. Uekusa, R. Sei, and K. Hayashi (2015). "Pepper Learns Together with Children: Development of an Educational Application." In: *2015 IEEE-RAS*

15th International Conference on Humanoid Robots (Humanoids), pp. 270–275. DOI: 10.1109/HUMANOIDS.2015.7363546.

Tanaka, Fumihide, Aaron Cicourel, and Javier R. Movellan (Nov. 2007). "Socialization between Toddlers and Robots at an Early Childhood Education Center." In: *Proceedings of the National Academy of Sciences* 104.46, pp. 17954–17958. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0707769104.

Tanaka, Fumihide and Shizuko Matsuzoe (2012). "Children Teach a Care-Receiving Robot to Promote Their Learning: Field Experiments in a Classroom for Vocabulary Learning." In: *Journal of Human-Robot Interaction* 1.1, pp. 78–95.

Tartaro, Andrea and Justine Cassell (2015). "Developing Reciprocity With Technology and Storytelling." In: *Children's Play, Pretense, and Story: Studies in Culture, Context, and Autism Spectrum Disorder*, p. 72.

Taub, Michelle, Seth A. Martin, Roger Azevedo, and Nicholas V. Mudrick (2016). "The Role of Pedagogical Agents on Learning: Issues and Trends." In: *Handbook of Research on 3-D Virtual Environments and Hypermedia for Ubiquitous Learning*, pp. 362–386. DOI: 10.4018/978-1-5225-0125-1.ch015.

Tausczik, Yla R. and James W. Pennebaker (Mar. 2010). "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods." In: *Journal of Language and Social Psychology* 29.1, pp. 24–54. ISSN: 0261-927X. DOI: 10.1177/0261927X09351676.

Taylor, Marjorie (2001). *Imaginary Companions and the Children Who Create Them*. Oxford University Press. ISBN: 978-0-19-514629-5.

— (2008). "Imaginary Companions: Pretending They Are Real but Knowing They Are Not." In: *American Journal of Play* 1.1, pp. 47–54. ISSN: ISSN-1938-0399.

Teale, William H. and Elizabeth Sulzby (1986). *Emergent Literacy: Writing and Reading. Writing Research: Multidisciplinary Inquiries into the Nature of Writing Series*. Ablex Publishing Corporation, 355 Chestnut St. ISBN: 978-0-89391-385-4.

The Big Data Breaches of 2014 (2015). URL: http://www.forbes.com/sites/moneybuilder/2015/01/13/the-big-data-breaches-of-2014/ (visited on 08/25/2015).

Thrun, Sebastian, Maren Bennewitz, Wolfram Burgard, Armin B. Cremers, Frank Dellaert, Dieter Fox, Dirk Hähnel, Charles Rosenberg, Nicholas Roy, Jamieson Schulte, and others (1999). "MINERVA: A Second-Generation Museum Tour-Guide Robot." In: *Robotics and Automation, 1999. Proceedings. 1999 IEEE International Conference On*. Vol. 3. IEEE.

Tickle-Degnen, Linda and Robert Rosenthal (Oct. 1990). "The Nature of Rapport and Its Nonverbal Correlates." In: *Psychological Inquiry* 1.4, pp. 285–293. ISSN: 1047-840X. DOI: 10.1207/s15327965pli0104_1.

Tomasello, Michael, Josep Call, Jennifer Warren, G. Thomas Frost, Malinda Carpenter, and Katherine Nagell (1997). "The Ontogeny of Chimpanzee Gestural Signals: A Comparison across Groups and Generations." In: *Evolution of communication* 1.2, pp. 223–259.

Topping, Keith J. (Dec. 2005). "Trends in Peer Learning." In: *Educational Psychology* 25.6, pp. 631–645. ISSN: 0144-3410.

Trope, Yaacov and Nira Liberman (2010). "Construal-Level Theory of Psychological Distance." In: *Psychological Review* 117.2, pp. 440–463. ISSN: 1939-1471(Electronic);0033-295X(Print). DOI: 10.1037/a0018963.

Tudge, Jonathan and Barbara Rogoff (1989). "Peer Influences on Cognitive Development: Piagetian and Vygotskian Perspectives." In: *Interaction in Human Development*. Ed. by M. H. Bornstein and J. S. Bruner. Crosscurrents in contemporary psychology. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc, pp. 17–40. ISBN: 978-0-8058-0035-7.

Turkle, S. (1985). "The Second Self: Computers and the Human Spirit." In:

Turkle, Sherry (Sept. 2005). *The Second Self: Computers and the Human Spirit*. MIT Press. ISBN: 978-0-262-25067-2.

— (Jan. 2007). "Authenticity in the Age of Digital Companions." In: *Interaction Studies* 8.3, pp. 501–517. DOI: 10.1075/is.8.3.11tur.

— (2017). *Alone Together: Why We Expect More from Technology and Less from Each Other*. Hachette UK.

Turkle, Sherry, Cynthia Breazeal, Olivia Dasté, and Brian Scassellati (2006a). "Encounters with Kismet and Cog: Children Respond to Relational Artifacts." In: *Digital Media: Transformations in Human Communication*, pp. 1–20.

Turkle, Sherry, Will Taggart, Cory D. Kidd, and Olivia Dasté (Dec. 2006b). "Relational Artifacts with Children and Elders: The Complexities of Cybercompanionship." In: *Connection Science* 18.4, pp. 347–361. ISSN: 0954-0091. DOI: 10.1080/09540090600868912.

Utami, D., T. W. Bickmore, and L. J. Kruger (2017). "A Robotic Couples Counselor for Promoting Positive Communication." In: *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 248–255. DOI: 10.1109/ROMAN.2017.8172310.

Vaala, Sarah, Anna Ly, and Michael H. Levine (2015). *Getting a Read on the App Stores: A Market Scan and Analysis of Children's Literacy Apps. Full Report*. Joan Ganz Cooney Center at Sesame Workshop.

Valdesolo, Piercarlo and David DeSteno (2011). "Synchrony and the Social Tuning of Compassion." In: *Emotion-APA* 11.2, p. 262.

Valdez-Menchaca, Marta C. and Grover J. Whitehurst (1992). "Accelerating Language Development through Picture Book Reading: A Systematic Extension to Mexican Day Care." In: *Developmental Psychology* 28.6, pp. 1106–1114. ISSN: 1939-0599 0012-1649. DOI: 10.1037/0012-1649.28.6.1106.

Vardoulakis, Laura Pfeifer, Lazlo Ring, Barbara Barry, Candace L. Sidner, and Timothy Bickmore (2012). "Designing Relational Agents as Long Term Social Companions for Older Adults." In: *SpringerLink*. Springer Berlin Heidelberg, pp. 289–302. DOI: 10.1007/978-3-642-33197-8_30.

Vázquez, C., L. Xia, T. Aikawa, and P. Maes (July 2018). "Words in Motion: Kinesthetic Language Learning in Virtual Reality." In: *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*, pp. 272–276. DOI: 10.1109/ICALT.2018.00069.

Veletsianos, George and Gregory S. Russell (2014). "Pedagogical Agents." In: *Handbook of Research on Educational Communications and Technology*. Ed. by J. Michael Spector, M. David Merrill, Jan Elen, and M. J. Bishop. New York, NY: Springer New York, pp. 759–769. ISBN: 978-1-4614-3185-5. DOI: 10.1007/978-1-4614-3185-5_61.

Verbeek, Peter-Paul (2006). "Persuasive Technology and Moral Responsibility Toward an Ethical Framework for Persuasive Technologies." In: *Persuasive* 6, pp. 1–15.

Vogt, Paul, Mirjam de Haas, Chiara de Jong, Peta Baxter, and Emiel Krahmer (2017). "Child-Robot Interactions for Second Language Tutoring to Preschool Children." In: *Frontiers in Human Neuroscience* 11. ISSN: 1662-5161. DOI: 10.3389/fnhum.2017.00073.

Vogt, Paul, Rianne van den Berghe, Mirjam de Haas, Laura Hoffmann, Junko Kanero, Ezgi Mamus, Jean-Marc Montanier, Ora Oudgenoeg-Paz, Daniel Hernandez Garcia, and Fotios Papadopoulos (2019). "Second Language Tutoring Using Social Robots. A Large-Scale Study." In: *IEEE/ACM Int. Conf. on Human-Robot Interaction (HRI 2019)*.

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

Walker, Sue, Kym Irving, and Donna Berthelsen (June 2002). "Gender Influences on Preschool Children's Social Problem-Solving Strategies." In: *The Journal of Genetic Psychology* 163.2, pp. 197–209. ISSN: 0022-1325. DOI: 10.1080/00221320209598677.

Wallbridge, Christopher D., Rianne van den Berghe, Daniel Hernández Garcia, Junko Kanero, Séverin Lemaignan, Charlotte Edmunds, and Tony Belpaeme (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, USA: ACM, pp. 54–60. ISBN: 978-1-4503-5953-5. DOI: 10.1145/3284432.3284433.

Weinberg, Bernd and Marsha Zlatin (June 1970). "Speaking Fundamental Frequency Characteristics of Five- and Six-Year-Old Children with Mongolism." In: *Journal of Speech, Language, and Hearing Research* 13.2, pp. 418–425. ISSN: 1092-4388. DOI: 10.1044/jshr.1302.418.

Weiss, Astrid, Daniela Wurhofer, and Manfred Tscheligi (June 2009). ""I Love This Dog"—Children's Emotional Attachment to the Robotic Dog AIBO." In: *International Journal of Social Robotics* 1.3, pp. 243–248. ISSN: 1875-4791, 1875-4805. DOI: 10.1007/s12369-009-0024-4.

Wentzel, Kathryn R. (1997). "Student Motivation in Middle School: The Role of Perceived Pedagogical Caring." In: *Journal of Educational Psychology* 89.3, pp. 411–19. ISSN: 0022-0663.

Whitehurst, G. J., Falco, F. L., Lonigan, C. J., Fischel, J. E., B. D. DeBaryshe, M. C. Valdez-Menchaca, and M. Caulfield (1988). "Accelerating Language Development through Picture Book Reading." In: *Developmental Psychology* 24.4, pp. 552–559. ISSN: 1939-0599 0012-1649. DOI: 10.1037/0012-1649.24.4.552.

Wik, Preben and Anna Hjalmarsson (2009). "Embodied Conversational Agents in Computer Assisted Language Learning." In: *Speech Communication*. Spoken Language Technology for Education 51.10, pp. 1024–1037. ISSN: 0167-6393. DOI: 10.1016/j.specom.2009.05.006.

Williams, Randi, Christian Vázquez Machado, Stefania Druga, Cynthia Breazeal, and Pattie Maes (2018). ""My Doll Says It's Ok": A Study of Children's Conformity to a Talking Doll." In: *Proceedings of the 17th ACM Conference on Interaction Design and Children*. IDC '18. New York, NY, USA: ACM, pp. 625–631. ISBN: 978-1-4503-5152-2. DOI: 10.1145/3202185.3210788.

Wiltermuth, Scott S. and Chip Heath (2009). "Synchrony and Cooperation." In: *Psychological Science* 20.1, pp. 1–5.

Winnicott, D. (1953). "Transitional Objects and Transitional Phenomena; a Study of the First Not-Me Possession." In: *The International Journal of Psychoanalysis* 34, pp. 89–97. ISSN: 1745-8315 0020-7578.

Wise, Barbara, Ron Cole, Sarel Van Vuuren, Scott Schwartz, Lynn Snyder, Nattawut Ngampatipatpong, Jariya Tuantranont, and Bryan Pellom (2005). "Learning to Read with a Virtual Tutor: Foundations to Literacy." In: *Interactive literacy education: Facilitating literacy environments through technology*, pp. 31–75.

Witt, Paul L., Lawrence R. Wheeless, and Mike Allen (June 2004). "A Meta-analytical Review of the Relationship between Teacher Immediacy and Student Learning." In: *Communication Monographs* 71.2, pp. 184–207. ISSN: 0363-7751. DOI: 10.1080/036452042000228054.

Wolf, M, S Gottwald, T Galyean, R Morris, and C Breazeal (2014). "The Reading Brain, Global Literacy, and the Eradication of Poverty." In: *Bread and Brain, Education and Poverty*. Scripta Varia 125: Vatican City.

Wolf, Maryanne, Stephanie Gottwald, Cynthia Breazeal, Tinsley Galyean, and Robin Morris (2017). ""I Hold Your Foot:" Lessons from the Reading Brain for Addressing the Challenge of Global Literacy." In: *Children and Sustainable Development: Ecological Education in a Globalized World*. Ed. by Antonio M. Battro, Pierre Léna, Marcelo Sánchez Sorondo, and Joachim von Braun. Cham: Springer International Publishing, pp. 225–238. ISBN: 978-3-319-47130-3. DOI: 10.1007/978-3-319-47130-3_17.

Woods, Sarah, Lynne Hall, Kerstin Dautenhahn, and Dieter Wolke (Jan. 2007). "Implications of Gender Differences for the Development of Animated Characters for the Study of Bullying Behavior." In: *Computers in Human Behavior* 23.1, pp. 770–786. ISSN: 0747-5632. DOI: 10.1016/j.chb.2004.11.018.

You, Zhen-Jia, Chi-Yuh Shen, Chih-Wei Chang, Baw-Jhiune Liu, and Gwo-Dong Chen (2006). "A Robot as a Teaching Assistant in an English Class." In: *Advanced Learning Technologies, 2006. Sixth International Conference On*. IEEE, pp. 87–91.

Youngblade, Lise M., Kathryn A. Park, and Jay Belsky (Dec. 1993). "Measurement of Young Children's Close Friendship: A Comparison of Two Independent Assessment Systems and Their Associations with Attachment Security." In: *International Journal of Behavioral Development* 16.4, pp. 563–587. ISSN: 0165-0254. DOI: 10.1177/016502549301600403.

Zhao, Ran, Tanmay Sinha, Alan W. Black, and Justine Cassell (2016). "Socially-Aware Virtual Agents: Automatically Assessing Dyadic Rapport from Temporal Patterns of Behavior." In: *International Conference on Intelligent Virtual Agents*. Springer International Publishing, pp. 218–233.

Zheng, Dongping and Kristi Newgarden (2011). "Rethinking Language Learning: Virtual Worlds as a Catalyst for Change." In:

Zheng, Dongping, Michael F. Young, Manuela Maria Wagner, and Robert A. Brewer (2009). "Negotiation for Action: English Language Learning in Game-Based Virtual Worlds." In:

The Modern Language Journal 93.4, pp. 489–511. ISSN: 1540-4781. DOI: 10.1111/j.1540-4781.2009.00927.x.

de Montjoye, Yves-Alexandre, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel (2013). "Unique in the Crowd: The Privacy Bounds of Human Mobility." In: *Scientific reports* 3.

de Montjoye, Yves-Alexandre, Laura Radaelli, Vivek Kumar Singh, and others (2015). "Unique in the Shopping Mall: On the Reidentifiability of Credit Card Metadata." In: *Science* 347.6221, pp. 536–539.

van Dijk, Jan A. G. M. (Aug. 2006). "Digital Divide Research, Achievements and Shortcomings." In: *Poetics*. The digital divide in the twenty-first century 34.4–5, pp. 221–235. ISSN: 0304-422X. DOI: 10.1016/j.poetic.2006.05.004.

van Maris, Anouk, Hagen Lehmann, Lorenzo Natale, and Beata Grzyb (2017). "The Influence of a Robot's Embodiment on Trust: A Longitudinal Study." In: *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. HRI '17. New York, NY, USA: ACM, pp. 313–314. ISBN: 978-1-4503-4885-0. DOI: 10.1145/3029798.3038435.