Relational AI: Creating long-term interpersonal interaction and shared experiences with social robots

by

Jacqueline M. Kory Westlund

Dissertation Proposal for the degree of Doctor of Philosophy in Media Arts and Sciences at the Massachusetts Institute of Technology

Spring 2016

Dissertation Advisor: _____________________________ Date: _____________________

Dr. Cynthia Breazeal
Associate Professor of Media Arts and Sciences
Personal Robots Group
MIT Media Lab

Committee Member: _____________________________ Date: _____________________

Dr. Paul Harris
Professor of Education
Harvard Graduate School of Education

Committee Member: _____________________________ Date: _____________________

Dr. Rosalind Picard
Professor of Media Arts and Sciences
Affective Computing Group
MIT Media Lab
Abstract

Literacy, language, and interpersonal skills are arguably the most important skills any child will learn, as they can greatly impact children's later educational and life success. However, many children do not receive sufficient support, instruction, or practice in developing these crucial skills. Many of the proposed technological interventions that address this need engage children passively and target older children, despite the fact that most early language and interpersonal interventions are most effective if they target young children aged 3-5 years using active, dialogic, social methods. In this thesis, we will examine the use of social robots as a technology to support preschool children in learning early literacy, language, and social-emotional skills. We hypothesize that a key aspect of why social robots can benefit children's learning is their nature as a relational technology—that is, a technology that can build long-term, social-emotional relationships with users. Thus, through a series of empirical child-robot interaction studies, this thesis will first establish the role of social robots as relational technologies. We will demonstrate their capabilities as learning companions for young children that afford opportunities for social engagement and reciprocal interaction, particularly peer-to-peer modeling. We will propose a framework by which we can understand how children conceptualize social robots as relational agents. Second, we introduce the term relational AI to refer to a subset of relational technologies that are also autonomous and change through time. In order to examine how relational AI can impact longitudinal child-robot learning interactions, we will develop a computational relational AI model. Through testing the model in a longitudinal study with a social robot, we will 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.
Introduction

Literacy, language, and interpersonal skills are arguably the most important skills any child will learn. These skills are critical for nearly all subsequent learning and can greatly impact children's later educational and life success (Fish & Pinkerman, 2003; Hart & Risley, 1995; Paez, Tabors, & Lopez, 2007; Snow, Porche, Tabors, & Harris, 2007). But it is not mere exposure to more vocabulary-building curricula that will help—prior research shows that children learn language best when they are active participants, using words to create and share meaning (Bloom, 2000; Duranti & Goodwin, 1992; Vygotsky, 1978). This kind of scenario is social and interactive by definition: a dialogic context where children are engaged as speakers and listeners for the purpose of communicating. Thus, it is no surprise that social cues have been shown to be critical for language development (Bloom, 2000; Corriveau et al., 2009; Harris, 2007, 2012; Meltzoff, Kuhl, Movellan, & Sejnowski, 2009; Sage & Baldwin, 2010), and furthermore, that a lack of social interaction with a partner may even impair language learning (Kuhl, 2007, 2011; Naigles & Mayeux, 2001). Because of the importance of social interaction, children's early interpersonal and social-emotional skills are also 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, Fantuzzo, Mendez, & McDermott, 2000; Gallagher, 1993).

However, not all children currently receive enough support, instruction, or practice in learning early literacy or interpersonal skills for numerous reasons, such as lack of resources, staff, or time; or the need for additional practice or therapy. Many interventions have been developed to help fill this gap, often using technology such as computers, tablets, iPads, and even robots to supplement children's early education in personalized ways. These technologies are already becoming a regular feature in children's lives. Many young children in the United States—as well as in other parts of the world—regularly consume digital content: playing games, reading eBooks, watching movies.

But there are several areas where these technologies could be improved. First, many of the interventions target older children, but that is not ideal. Based on the existing evidence, interventions will be most effective if they target young children aged 3-5 years for three reasons. First, fixing the 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 & Risley, 1995). Second, developmentally, language learning may proceed more rapidly in the early years (Johnson & Newport, 1989). Third, preschool children are also actively developing the social and emotional skills that promote their readiness to learn, facilitate positive peer interactions, and contribute to their academic and life success (Coolahan et al., 2000; Gallagher, 1993; McClelland & Morrison, 2003; Pahl & Barrett, 2007). Preschool is thus a critical age.

Second, many of the existing technological interventions engage children passively as watchers, listeners, and consumers—not as creators or actors. But as we have seen, the subjects that are so critical for children to learn at this young age—early literacy skills, language, social and emotional skills—are all rooted in interpersonal, social interaction. These children are learning how to be social agents. They are learning how to communicate, using words, emotions, and other nonverbal signals, through interaction with the social agents around them.

In this thesis, we hypothesize that when creating technology to support children in learning early literacy, language, and social-emotional skills, it will be technology that embodies the social and interpersonal—as well as employing the many other benefits of technology in general, such as accessibility, rapid customization and the easy addition of new content, and the potential to deploy at scale—that will afford the greatest benefits. It will be technology that targets preschool-age children.
and engages them in rich dialogic contexts as active participants. To this end, in this thesis, we will discuss relational technologies: technologies that are, inherently, interpersonal and social, reciprocal and relational, personalized through time. One such technology is social robots. Through a series of empirical child-robot interaction studies, we will establish the role of social robots as relational technologies. We will demonstrate their capabilities as learning companions for young children that afford opportunities for social engagement and reciprocal interaction, particularly peer-to-peer modeling. We will propose a framework by which we can understand how children conceptualize social robots as relational agents, different from all the other agents children encounter. Finally, through developing a computational model and testing it during a longitudinal study, we will show that social robots that utilize relational AI—personalizing to children through time to create and reference a shared relational narrative—are a new, powerful educational tool, unlike any other existing technology, that we can leverage to support children’s early education and development.

Related Work

Social robotics

Social robotics is a new research field, spanning less than two decades (C. Breazeal, Dautenhahn, & Kanda, 2016; C. L. Breazeal, 2004; C. Breazeal, Takanishi, & Kobayashi, 2008; Feil-Seifer & Mataric, 2011). While science fiction authors have for decades debated the benefits, risks, and ethical dilemmas surrounding technological artifacts endowed with varying degrees of human-like capabilities, the existence of any such technology in the real world is recent. As such, there is much to explore regarding how humans perceive, understand, and interface with these devices. Social robots are unique in that they combine traditional computers and machines with the embodied, situated world. They have physical bodies, share physical spaces with humans, and leverage human behaviors—such as speech, movement, and nonverbal signals—to communicate with us in more natural ways.

Social robots for learning

Various social robots have been developed and tested as tutors and learning companions for young children in the domain of literacy and language learning. These robots have played vocabulary learning games (C. Breazeal, Harris, et al., 2016; Chang, Lee, Chao, Wang, & Chen, 2010; Freed, 2012; Gordon et al., 2016; Kanda, Hirano, Eaton, & Ishiguro, 2004; Kennedy, Baxter, Senft, & Belpaeme, 2016; J. Kory Westlund, Gordon, et al., 2015; S. Lee et al., 2011; Movellan, Eckhardt, Virnes, & Rodriguez, 2009; Tanaka & Matsuzoe, 2012), led storytelling activities (Chang et al., 2010; Kory, 2014; Kory & Breazeal, 2014; J. Kory Westlund & Breazeal, 2015) and reading activities (Gordon & Breazeal, 2015), and helped children learn handwriting skills (Hood, Lemaignan, & Dillenbourg, 2015). The robots are most often situated as tutors or instructors (e.g., Chang et al., 2010; Kennedy et al., 2016; Leyzberg, Spaulding, & Scassellati, 2014), but some are situated as peers (e.g., Gordon et al., 2016; Kanda et al., 2004; Kory, 2014), or as younger peers whom the children have to help or teach (e.g., Hood et al., 2015; Tanaka & Matsuzoe, 2012). With regards to social and emotional learning, the majority of the research targets children with Autism Spectrum Disorder, using social robots to encourage social interaction (Kim et al., 2013; Ricks & Colton, 2010; Scassellati, Admoni, & Mataric, 2012); fewer studies have been performed with typically developing children (e.g., Leite et al., 2015).

This prior research shows that children will engage with, learn from, and respond socially to social robots in learning contexts. However, many of the existing 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: 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? What kind of relationship will children develop with the robot over time? A few studies have started addressing these questions (Gordon et al., 2016; Kory, 2014; Leite, Castellano, Pereira, Martinho, & Paiva, 2012; Movellan et al., 2009). Collectively, they suggest that personalization of the robotic system to the child over time—such as personalizing the content presented or the robot's affective feedback—will lead to greater engagement and improved learning outcomes. Related work in other learning contexts with older children and adults lends similar evidence regarding the benefits of personalization (D’mello & Graesser, 2012; Gordon & Breazeal, 2015; Kasap & Magnenat-Thalmann, 2012; Ramachandran & Scassellati, 2015; Thrun et al., 1999), as well as evidence that change in the robot's speech and behavior can help maintain user engagement and build a long-term relationship (T. Bickmore, Schulman, & Yin, 2010; Kidd & Breazeal, 2008; M. K. Lee et al., 2012).

Relationships with technology

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 & Reis, 1998; Csikszentmihalyi & 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 & 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 & Reis, 1998; T. Bickmore & Cassell, 2001; Burgoon & Hale, 1984; Cassell & Bickmore, 2000; Fogg & Tseng, 1999; Spencer-Oatey, 1996; Trope & Liberman, 2010), since 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 & 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 & Halonen, 1979).

Relational Agents, Relational Objects, and More

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., T. Bickmore et al., 2010), Bickmore uses the term relational agents more narrowly to refer exclusively to conversational virtual humans. 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 do not specifically try to establish or maintain relationships, even if users interact with them longitudinally, are excluded. For example, there are several persona-AI systems, such as Alexa, Google Assistant, and Google Home, that are transactional and non-relational.

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, Breazeal, Dasté, & Scassellati, 2006; Turkle, Taggart, Kidd, & Dasté, 2006). 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.

Finally, Winnicott explored the idea of transitional objects (Winnicott, 1953). In this case, the objects are seen as symbolic of the self, of an other, or of a relationship, but are not studied as things with which one is having a relationship, and thus are not considered relational.

Research Questions

Social Robots as a Relational Technology

First, given the existing evidence that social robots can be effective tutors and learning companions, why are they effective? What design features of the robots positively impact children's learning and attitudes through time? To understand how social robots can benefit children's learning, we examine children's interactions with the technology. For example, we have begun seeing children perform peer-to-peer modeling of curiosity, affect, and language with the robots (Gordon, Breazeal, & Engel, 2015; Kory, 2014; Kory Westlund et al., in review). As a result of these studies, we argue that a key aspect of why social robots benefit children's learning is their nature as a relational technology. Thus, this thesis will first explore how children perceive and conceptualize social robots in learning contexts, and how children relate to these robots through time.

Relational AI

The second subject this thesis will address the core nature of relational technologies. We introduce the term relational AI to refer to autonomous relational technologies, especially the underlying computational models, algorithms, and mechanisms by which they operate. We will discuss in detail which features are necessary and sufficient to qualify as relational AI. For example, if relationships are construed as necessarily longitudinal, then one criterion of relational AI is it deals with repeated encounters with users through time. Since existing through time necessitates change, a second criterion is that the relational AI must change over time in response to the relationship. Relational AI comprises a subset of relational technologies, since, e.g., some may have some but not all the features necessary for relationship establishment and maintenance; may be teleoperated; or may not change through time. In this thesis, we will examine how adding features of relational AI to a social robot will impact longitudinal child-robot learning interactions. How might relational AI impact children's learning, engagement, and relationships?

Research Plan

Completed work

Multiple studies contributing to this thesis have already been completed. These studies
primarily address the first set of research questions regarding the nature of robots as learning companions for young children, and children's construals of social robots as relational technologies. These 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 and the robot's expressivity impact children's engagement, learning, and judgments of the robot's credibility. Children apply social judgments to the robots, treat them as peers, and even display peer-to-peer modeling of the robot's affect and language. These studies also suggest that children view social robots as something betwixt and between the dualistic categories of alive, animate beings and inanimate objects: not technology like iPads or computers, not machines like toasters and dishwashers, not biological like pets or plants, and certainly not human—something other. Below, these study and their contributions are summarized briefly.

**Study 1: Children will 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 (J. Kory Westlund, Dickens, et al., 2015). 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. This study provided evidence that children will learn from social robots, and will think of them as social partners.

**Study 2: Children pay attention 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 (J. Kory Westlund et al., in review). Children viewed two images of unfamiliar animals at once, and their interlocutor (human or robot) named one of the animals. 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.

**Study 3: The contingency of a robot's nonverbal social cues impacts whether children treat the robot as a credible informant**

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 (C. Breazeal, Harris, et al., 2016). We found that children treated both robots as interlocutors and as informants from whom they could seek information. 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.

**Study 4: Children will treat a social robot as a peer and show peer-to-peer modeling of language over time**

The next study examined the potential of a social robotic learning companion to support children's early language development over time (Kory, 2014; Kory & Breazeal, 2014; J. Kory Westlund & Breazeal, 2015). Seventeen children aged 4-6 played a storytelling game with a robot eight times over two months. Children took turns with the robot telling stories about characters in scenes depicted on a tablet. Target vocabulary words were embedded in the stories the robot told. We 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. Furthermore, these children also showed more peer-to-peer language modeling, using phrases and words from the robot's stories in their own stories. This study provided evidence that personalizing the content the robot presents to individual children will lead to greater learning gains, consistent with prior work on personalization. Children will engage a social robot as a peer over multiple encounters, and will demonstrate peer-to-peer language modeling as a result.

**Study 5: Personalizing the robot's affective feedback over time impacts children's own affect**

To further probe personalization and long-term child-robot interaction, we performed a study in which children played a second-language learning game with a social robot and with a virtual character on a tablet (Gordon et al., 2016; J. Kory Westlund, Gordon, et al., 2015). 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 provides 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.

**Study 6: Children show more peer-to-peer modeling of affect and language with an expressive social robot than with a less expressive robot**

Our next 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., in review). 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. 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 additional 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.

**Study 7: Relational AI: Personalization and reciprocity through time**

**Research Goals**

To address the second set of research questions regarding the core nature of relational AI, we will create a computational relational AI model that brings together many of the elements studied so far, and evaluate the model with a social robot. The model will incorporate key features of relational AI such as encounters through time, personalization, and social reciprocity to explore how children experience relational technologies, and how relational AI impacts children's learning, engagement, and relationships during child-robot learning interactions.

**Relational AI Model**

**Features of Relational AI**

As alluded to earlier, we will explore the impact of several key features of relationships on child-robot learning interactions. Foremost is the element of time. We take the stance that relationships are necessarily longitudinal. Thus, one primary criterion of relational AI is that it deals with repeated interactions with users through time.

Second, since existing through time necessitates change, the AI must change over time. Specifically, relational AI must 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 must be clearly seen to affect the 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, & John, 2003) or similar choices of food (Bove, Sobal, & Rauschenbach, 2003). However, 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, only a few studies so far have examined 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, Walker, & Pon-Barry, 2016; Ramachandran & Scassellati, 2015). 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.

Another aspect of time as experienced by humans is a sense of past, present, and future. Our relationships with others are partially based on a sense of shared experiences, i.e., activities we have done together in the past or are performing together now. During our interactions, we reference our shared narrative: we acknowledge our shared experiences 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 & Aron, 2004). Relational AI should create and reference a shared narrative as well.

Finally, since we want to model a positive relationship, we must also look at elements of successful human relationships. For example, one such feature is rapport (Berscheid & Reis, 1998), which is often indicated via behaviors such as entrainment/mirroring and social reciprocity (Davis, 1982; Dijksterhuis, 2005; Dijksterhuis & Bargh, 2001). This feature is more immediate, happening in the now. Relational AI should both affect and be reciprocally affected by the user.
Model Architecture

Relational AI operates on two timescales: the immediate, and the extended. The immediate includes behaviors exhibited in the present—such as rapport, entrainment, and mirroring—that reflect the current relationship in the now. The extended includes the past and future: behaviors that more deeply reflect the shared narrative and the relationship through time. Our model will include several immediate and extended behaviors (see Figure 1).

Immediate. The model will use input from cameras to detect children’s face position, head pose, and affect. A microphone will provide input for recognizing when the child is speaking as well as some basic speech. These inputs will be used to generate face tracking, affect mirroring, pause detection and backchanneling, face recognition, and shared narrative references.

Extended. The model will use input from cameras to detect children’s faces. Between sessions, children’s speech from the previous session will be transcribed and any posttests scored to determine task content and performance. These inputs will be used to generate face/person recognition behaviors, content personalization, and shared narrative referencing. The face/person recognition will allow the robot to personally recognize and greet children each session. This provides a sense of continuity and grounds the interaction in the shared experience of playing together multiple times.

The content personalization will primarily adjust the level or complexity of the task content. We will be using a story-based learning activity; thus, content personalization will adjust the robot's stories.

Figure 1: The model will use inputs from a camera, microphone, speech/transcripts, and the child’s performance (via posttests). These inputs will be used to generate robot behaviors, including face tracking, affect mirroring, pause detection and backchanneling, face recognition, and shared narrative references.
based on the child's performance to match the child's current ability. It may also adjust the language or content the robot uses in later stories based on how the child tells or retells stories, i.e., peer-to-peer modeling by the robot of the child's language, similar to how we have seen children exhibit this emulation of the robot's language in prior work. Finally, we could also adjust the content of the stories based on the child's preferences, e.g., telling a dinosaur story if the child especially likes dinosaurs. This could be positioned as the robot's preferences about story content changing to accommodate the child's preferences. Similarly, we could adapt the robot's emotional reactions to the stories based on the child's reactions (using the affect data), since people in close relationships may converge in their emotional responses to events (Anderson et al., 2003).

The shared narrative references will be used during “small talk” periods of the interaction session to refer to past session activities (e.g., mention past stories told: “Last time, we...”, “Remember, I told a story about...”), ask questions building on prior conversations (e.g., the model could associate children's affect with robot stories told, and later recall this: “Last week, I told a story about a dinosaur, but you thought it was kind of scary...”, “Since you liked the monkey story so much, I thought I'd tell another story about animals...”), and highlight future activities (e.g., “Next time, we'll...”) or predict future activities (e.g., based on children's affect during past robot stories, predict future affect: “Since you liked that story, I bet you'll love the story I have for next time!”). These references will help create a sense of continued shared experience and position the robot as an active participant in the interaction.

**Evaluation**

**Pilot Study**

A one-session pilot study will be conducted in the lab to test the immediate relational robot behaviors and the learning game. Approximately 30 participants will be recruited from the PRG participant mailing list. Each participant will play a learning game with a robot for 10-15 minutes. For half the participants, the robot will use the immediate relational behaviors; for the other half, the robot will not use these behaviors. We will use measures similar to those used in Study 6 (Kory Westlund et al., in review) to assess children's learning and engagement: pre and post learning tests, emotion tracking, and brief interviews. We expect that children who play with the relational robot will show more learning and deeper engagement.

**Longitudinal Study**

**Overview.** This study will evaluate the computational relational AI model, addressing the question of how a social robot that utilizes relational AI impacts children's engagement, learning, and relationship with the robot during a longitudinal learning scenario.

**Participants.** Approximately 80 typically-developing children aged 4-6 years will be recruited from Boston-area schools to participate in the study.

**Procedure.** Participants will interact with a robot (either Tega or Jibo) approximately once per week for 2-3 months. During each interaction session, the children will play a story-based learning game with the robot for about 10-15 minutes, based on the games used in (Kory, 2014; Kory Westlund et al., in review). In the game, the robot and child will both have opportunities to tell or retell stories to each other. Target vocabulary words will be embedded in the robot's stories. This game will provide opportunities for learning, engagement, and references to the child-robot shared narrative to occur.

**Conditions.** Children will be randomly assigned to play with either (1) a robot that uses the relational AI model to determine its immediate and extended relational behavior (Relational condition); (2) a robot that uses the relational AI model to determine only its immediate relational behaviors, but
does not use extended relational behaviors (Immediate condition); (3) a robot that uses the relational AI model to determine only its extended relational behaviors, but does not use immediate relational behaviors (Extended condition) or (4) a robot that does not use the relational AI model at all, but is just as expressive in terms of its immediate behavior (None condition) (see Figure 2).

**Measures.** We will record video and audio data of each session. We will measure children's learning via vocabulary posttests and analyzing the content of their story retells, following the methods used in (Kory Westlund, et al., in review). We will also use their story transcripts to assess peer-to-peer language modeling. Children's head pose, affect, attention, and engagement will be measured using Affdex from Affectiva, Inc ([http://affectiva.com/](http://affectiva.com/)). Children's affect will also be used to determine the extent of affect mirroring by the children of the robot. We will perform pre- and post-interviews to assess children's construal of the robot, perceptions of their relationship with it and its relationship with them, and self-reported engagement and liking of the robot. We will also code children's transcribed speech for references to the child-robot shared narrative (e.g., references to past activities the child and robot did together) and references to the robot as a social other (e.g., using second-person pronouns such as “we”).

We will have children perform the Anomalous Picture Task, used previously in (Kory Westlund, Martinez, Archie, Das, & Breazeal, 2016), with the robot and with an experimenter (as a control) before and after the full interaction. We will code children's gaze patterns during this task, since we have previously seen differences in children's gaze patterns as a result of their judgments about their partner as a social other.

We will also include several pre- and post-test measures to assess children's relationship with the robot, and their perception of it as trustworthy or credible.

**Hypotheses.** Based on the results of our prior studies and relevant literature, we expect that just being an expressive, contingent social agent (Immediate) isn't enough to maintain a successful relationship over time. The agent must also use extended relational behaviors to acknowledge the shared narrative, and change in response to shared experiences over time (Relational). However, using either immediate or extended relational behaviors will still do more than using none (None). Thus, we expect that interacting with the Relational robot will lead to increased rapport, liking, trust, attachment, perceived credibility, deeper engagement, deeper learning, social behaviors (e.g., peer-to-peer modeling), and a stronger relationship than in the Immediate and Extended conditions, but that the Immediate condition will still lead to increases of all these factors when compared to the None condition. The Extended condition will lead to less rapport, liking, and trust than the Immediate and None conditions due to the lack of rapport-building behaviors, which may also affect children's learning and the robot's perceived credibility, but will still lead to deeper engagement than the None condition due to the changes over time. The Relational robot will be construed as a more human-like, social agent than in the other conditions. The Immediate robot will be seen as more social than the Extended and None robots, while the Extended will be seen as having a continued existence in a human-like way more than the Immediate and None robots.

**Understanding children's conceptualization of social robots**

In addition to the data to be collected during the relational AI model evaluation study, we will also perform additional analyses of data collected during the previously completed studies in order to further explore how children conceptualize social robots as a relational technology. Existing data includes interview and questionnaire responses from children regarding their perception of social robots as being like people or iPads, and as having certain mental, perceptual, and biological qualities; data about gaze patterns during different tasks with the robots, which may reflect children's views of the robot as a social other; and data about children's peer-to-peer modeling with the robot.
Contributions

In this thesis, we will introduce and explore the concepts of relational technologies and relational AI. Through a series of empirical child-robot interaction studies, we will establish the role of social robots as relational technologies. We will demonstrate their capabilities as learning companions for young children that afford opportunities for social engagement and reciprocal interaction. We will introduce a novel theory regarding children's conceptualization of social robots as relational agents, and will discuss the potential impact, opportunities, and ethical dilemmas presented by these technologies. We will create and evaluate a novel computational relational AI model for use with a social robot in a longitudinal child-robot learning interaction. In the process, we will create reusable software infrastructure for coordinating the robot, learning game, and relational AI model. Finally, using the data collected throughout the series of studies, we will show that social robots that utilize relational AI are a new, powerful educational tool, unlike any other existing technology, that we can leverage to support children's early education and development.

Figure 2: With no relational AI, the robot provides content, but does not change in relation to the child's behavior. With immediate relational AI, the robot responds to the child's behavior in the present. With extended relational AI, the robot responds through time. With both immediate and extended relational AI, the robot reacts to both the child's behavior in the present and through time.
## Timeline

<table>
<thead>
<tr>
<th>January 2017</th>
<th>Final written draft of proposal to committee</th>
</tr>
</thead>
<tbody>
<tr>
<td>February 2017</td>
<td>Proposal signed by committee</td>
</tr>
<tr>
<td></td>
<td>Submit to MASCOM</td>
</tr>
<tr>
<td>March 2017</td>
<td>MASCOM proposal approval</td>
</tr>
<tr>
<td></td>
<td>Schedule proposal critique</td>
</tr>
<tr>
<td></td>
<td>Proposal critique</td>
</tr>
<tr>
<td>April 2017</td>
<td>Finalize evaluation metrics and assessments</td>
</tr>
<tr>
<td></td>
<td>IRB / COUHES for longitudinal study</td>
</tr>
<tr>
<td></td>
<td>Project development</td>
</tr>
<tr>
<td>May 2017</td>
<td>Contact schools about study</td>
</tr>
<tr>
<td></td>
<td>All game content for interaction finalized</td>
</tr>
<tr>
<td>June 2017</td>
<td>Project development</td>
</tr>
<tr>
<td></td>
<td>Pilot testing</td>
</tr>
<tr>
<td>July 2017</td>
<td>Project development</td>
</tr>
<tr>
<td></td>
<td>Pilot testing</td>
</tr>
<tr>
<td>August 2017</td>
<td>Project development</td>
</tr>
<tr>
<td></td>
<td>Pilot testing</td>
</tr>
<tr>
<td>September 2017</td>
<td>Begin data collection for longitudinal study</td>
</tr>
<tr>
<td></td>
<td>Begin writing: introduction, background, theory, methods</td>
</tr>
<tr>
<td>October 2017</td>
<td>Data collection continues</td>
</tr>
<tr>
<td></td>
<td>Continue writing</td>
</tr>
<tr>
<td>November 2017</td>
<td>Data collection continues</td>
</tr>
<tr>
<td></td>
<td>Continue writing</td>
</tr>
<tr>
<td>December 2017</td>
<td>Complete data collection</td>
</tr>
<tr>
<td></td>
<td>Continue writing</td>
</tr>
<tr>
<td></td>
<td>Data analysis</td>
</tr>
<tr>
<td>January 2018</td>
<td>Data analysis</td>
</tr>
<tr>
<td></td>
<td>Continue writing: results, discussion, conclusion</td>
</tr>
<tr>
<td>February 2018</td>
<td>Data analysis</td>
</tr>
<tr>
<td></td>
<td>Continue writing</td>
</tr>
<tr>
<td>March 2018</td>
<td>Dissertation Defense</td>
</tr>
<tr>
<td>April 2018</td>
<td>Final revisions and writing</td>
</tr>
<tr>
<td>May 2018</td>
<td>Turn in dissertation by May deadline!</td>
</tr>
</tbody>
</table>
Resources Required

This research will require the following equipment and other resources:

- 2-4 Jibo or Tega robots
- Supporting robot study equipment (e.g., data collection equipment, tablets or touchscreens)
- Participation from several local preschool or kindergarten classrooms
- Someone to help with data collection for the final study
Biographical Information

Author
Jacqueline M. Kory Westlund is a Ph.D. student in the Personal Robots Group at the MIT Media Lab under Dr. Cynthia Breazeal. She earned her Master’s of Media Arts and Sciences in 2014. Her research focuses on developing and evaluating social robotic learning companions to support young children's language learning and social and emotional development. She was awarded an NSF Graduate Research Fellowship in 2012-2015 and an MIT Media Lab Learning Innovation Fellowship in 2016-2017 to support her research at MIT. Her work has been featured in IEEE Spectrum, IEEE Automation Blog, New Scientist, and on CNN. Prior to coming to the MIT Media Lab, she spent a year researching human emotion and learning with Dr. Sidney D’Mello’s Emotive Computing group at the University of Memphis and the University of Notre Dame. She worked at NASA for two summers, developing autonomous vehicles and laser space robots. She holds a BA in cognitive science from Vassar College.

Thesis Advisor: Dr. Cynthia Breazeal

Dr. Cynthia Breazeal is an Associate Professor of Media Arts and Sciences at the Massachusetts Institute of Technology where she founded and directs the Personal Robots Group at the Media Lab. She is a pioneer of social robotics and Human Robot Interaction. Her research focuses on developing the principles, techniques, and technologies for personal robots that are socially intelligent, interact and communicate with people in human-centric terms, work with humans as peers, and learn from people as an apprentice. She has authored the book “Designing Sociable Robots”, has published over 100 peer-reviewed articles in journals and conferences on the topics of autonomous robotics, artificial intelligence, human robot interaction, and robot learning. She has won numerous awards, including the National Academy of Engineering’s Gilbreth Lecture Award, Technology Review’s TR35 Award, TIME magazine’s Best Inventions of 2008, an ONR Young Investigator Award, and numerous best paper and best technology inventions at top academic conferences. She received her S.B. (1989) in Electrical and Computer Engineering from the University of California, Santa Barbara. She did her graduate work at the MIT Artificial Intelligence Lab, and received her SM (1993) and ScD (2000) in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology.

Thesis Reader: Dr. Paul Harris

Dr. Paul Harris is the Victor S. Thomas Processor of Education at the Harvard Graduate School of Education, where he directs the Early Childhood Lab. He is interested in the early development of cognition, emotion and imagination. After studying psychology at the University of Sussex and the University of Oxford, he taught at the University of Lancaster, the Free University of Amsterdam, and the London School of Economics. In 1980, he moved to Oxford where he became a professor of developmental psychology and fellow of St John's College. In 2001, he migrated to Harvard University, where he currently studies how far children rely on their own first-hand observation or, alternatively, trust what other people tell them—especially when they try to understand a domain of knowledge in which first-hand observation is difficult. He has published over 100 peer-reviewed articles in journals and conferences, as well as seven books and edited volumes on children's understanding of mental states and emotion, the development of pretend play and imagination, and children's trust in what
others tell them. His latest book, *Trusting What You’re Told: How Children Learn from Others*, received the Eleanor Maccoby Book Award from the American Psychological Association and the Cognitive Development Society Book Award.

Dr. Harris has been a close collaborator of the Personal Robots Group for the past five years, serving as co-PI on grants with Dr. Cynthia Breazeal.

**Thesis Reader: Dr. Rosalind Picard**

Dr. Rosalind Picard is founder and director of the Affective Computing research group at the MIT Media Lab, co-director of the Lab's Advancing Wellbeing Initiative, and faculty chair of MIT's Mind+Hand+Heart Initiative. She has co-founded Empatica, Inc., which creates wearable sensors and analytics to improve health, and Affectiva, Inc., which delivers technology to help measure and communicate emotion. Picard has authored or co-authored more than 250 scientific articles and chapters spanning computer vision, pattern recognition, machine learning, human-computer interaction, wearable sensors, neurology, and affective computing. She is a recipient of several “best paper” prizes, holds patents including wearable and non-contact sensors, algorithms, and systems for sensing, recognizing, and responding respectfully to human affective information; and has been honored with dozens of distinguished and named lectureships and other international awards. In 2005, she was named a Fellow of the IEEE for contributions to image and video analysis and affective computing. CNN named her one of seven "Tech Superheroes to Watch in 2015." She has given over 100 invited keynote talks. Picard holds a bachelor's degree in electrical engineering with highest honors from the Georgia Institute of Technology, and master's and doctorate degrees, both in electrical engineering and computer science, from MIT.
References


Dijksterhuis, A. (2005). Why we are social animals: The high road to imitation as social glue. Perspectives on Imitation: From Neuroscience to Social Science, 2, 207–220.


Freed, N. A. (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). Massachusetts Institute of Technology, Cambridge, MA.


Turkle, S., Taggart, W., Kidd, C. D., & Dasté, O. (2006). Relational artifacts with children and elders:
https://doi.org/10.1080/09540090600868912