Producing Acoustic-Prosodic Entrainment

in a Robotic Learning Companion to Build Learner Rapport

by

Nichola Lubold

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved August 2018 by the
Graduate Supervisory Committee:

Erin Walker, Co-Chair
Heather Pon-Barry, Co-Chair
Kurt VanLehn
Diane Litman
Visar Berisha

ARIZONA STATE UNIVERSITY

December 2018

ABSTRACT

With advances in automatic speech recognition, spoken dialogue systems are assuming increasingly social roles. There is a growing need for these systems to be socially responsive, capable of building rapport with users. In human-human interactions, rapport is critical to patient-doctor communication, conflict resolution, educational interactions, and social engagement. Rapport between people promotes successful collaboration, motivation, and task success. Dialogue systems which can build rapport with their user may produce similar effects, personalizing interactions to create better outcomes.

This dissertation focuses on how dialogue systems can build rapport utilizing acoustic-prosodic entrainment. Acoustic-prosodic entrainment occurs when individuals adapt their acoustic-prosodic features of speech, such as tone of voice or loudness, to one another over the course of a conversation. Correlated with liking and task success, a dialogue system which entrains may enhance rapport. Entrainment, however, is very challenging to model. People entrain on different features in many ways and how to design entrainment to build rapport is unclear. The first goal of this dissertation is to explore how acoustic-prosodic entrainment can be modeled to build rapport.

Towards this goal, this work presents a series of studies comparing, evaluating, and iterating on the design of entrainment, motivated and informed by human-human dialogue. These models of entrainment are implemented in the dialogue system of a robotic learning companion. Learning companions are educational agents that engage students socially to increase motivation and facilitate learning. As a learning companion's ability to be socially responsive increases, so do vital learning outcomes. A second goal of this dissertation is to

i

explore the effects of entrainment on concrete outcomes such as learning in interactions with robotic learning companions.

This dissertation results in contributions both technical and theoretical. Technical contributions include a robust and modular dialogue system capable of producing prosodic entrainment and other socially-responsive behavior. One of the first systems of its kind, the results demonstrate that an entraining, social learning companion can positively build rapport and increase learning. This dissertation provides support for exploring phenomena like entrainment to enhance factors such as rapport and learning and provides a platform with which to explore these phenomena in future work.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Over the past few years, spoken dialogue systems have become increasingly ubiquitous. Individuals can order food, make appointments and send text messages via their voiced personal assistants; they speak with digital customer service agents (Barres et al., 2007), and many popular toys now feature spoken dialogue interaction (Matteo et al., 2009). As people engage in ever more complex conversations with these systems, there is an increasing need for systems that can socially engage the user or build **rapport** by promoting feelings of harmony and social connection. In human-human interactions, rapport has been shown to be important to the success of communication between patients and doctors (Moira et al., 1999), various forms of negotiation and conflict resolution (Drolet and Morris, 2000), educational interactions (Tsui, 1996; Frisby and Myers, 2008; Ogan et al., 2012), and caregiving (Burns 1984; Miller et al., 2016). When dialogue partners feel more rapport, they enjoy the interaction more, are more engaged, they are more motivated, and collaborate better. There is early evidence that systems that can build rapport with users can replicate similar effects as seen in human-human interactions. People who feel more rapport for an agent-partner tend to be more motivated, engaged, and have higher task-success (Szafir & Mutlu, 2012; Kang, Gratch, & Watts, 2009; Huang, Morency, & Gratch, 2011). A dialogue system that can build rapport has the potential to more effectively personalize interactions and optimize experiences.

Dialogue systems are convenient for building rapport and enhancing social engagement because they can unite two channels of rapport building behavior. With

1

dialogue systems, it is possible to influence rapport through both the *content* of what is said and by *how* it is said, through prosody. Influencing rapport through the *content* of a dialogue system has been shown to have promising effects; introducing social dialogue such as inclusive language, off-topic conversation, and praise enhances trust, motivation, and engagement (Bickmore and Cassell, 2001, Bickmore 2003, Kanda et al., 2007, Ogan et al., 2012, Gulz, Haake, and Silvervarg, 2011). Influencing rapport through the *prosody* of a dialogue system, such as patterns of stress and intonation, has been less explored. Prosody conveys important metacommunicative information in conversation; when speakers modulate their tone of voice, speak more quickly or softly, these changes reveal details about how the speaker is feeling and what they want the listener to know. Prosody makes the speech signal a rich modality for enhancing social behavior and building rapport.

One particular prosodic phenomenon of human-human conversation which has been linked to rapport and other social factors is **acoustic-prosodic entrainment**. Entrainment occurs when individuals adapt their prosodic features of speech, such as pitch or tone of voice, loudness, or speaking rate, to one another over the course of a conversation. According to the Communication Accommodation Theory (CAT), individuals accommodate or entrain to their partner to achieve social approval (Giles & Smith, 1979). This theory suggests an individual on the receiving end of a high level of prosodic adaptation is likely to feel a greater sense of self-esteem, satisfaction, and rapport for their partner than if they were a receiver of low adaptation. In support of this, entrainment in human-human conversation has been found to be related to social engagement factors, including trust (Benus et al., 2018), social bias (Benus, Levitan, & Hirshberg, 2012), positive and negative affect between married couples (Lee et al., 2010),

2

engagement (Gravano et al., 2014), learning (Ward & Litman, 2007; Thomason, Nguyen, & Litman, 2013), and task success (Borrie, Lubold, & Pon-Barry, 2015). It is possible that a system which can model the prosodic fluctuations of a human conversational partner and entrain prosodically has the potential to increase social engagement beyond a system which only takes advantage of manipulating content through, for example, social dialogue. An entraining dialogue system may enhance feelings of rapport, where greater rapport may facilitate higher task success, motivation, and satisfaction with the interaction.

Modeling acoustic-prosodic entrainment, however, is challenging. People can entrain on many features of speech (pitch, intensity, speaking rate, voice quality) in various manners (converging on a feature vs. matching on a turn-by-turn basis, globally vs. locally). While entrainment has been explored extensively in human-human interactions, it is unclear how entrainment can be designed in a human-agent interaction such that the agent's entrainment positively influences social engagement. It is also unclear how entrainment can be combined with the more well-known content-based approaches for building rapport. Therefore, I pose the following two research questions in this work:

RQ 1: How can acoustic-prosodic entrainment be modeled in a system to positively influence social responses?

RQ 2: How does automated entrainment influence rapport when combined with content-based approaches for building rapport?

To answer these two questions, I present an iterative exploration on designing and implementing acoustic-prosodic entrainment in a dialogue system. I evaluate the effects of entrainment with social dialogue, which has had prior success in building rapport in human-robot and human-agent interactions. The dialogue system is implemented in a

robotic learning companion which provides an excellent application for testing and evaluating the posed research questions. The learning companion is a form of educational agent that engages students socially to increase motivation, provide emotional support and facilitate learning. Based on the theory that learning is social (Vygotsky, 1979), learning companions rely on building rapport with learners to influence socio-motivational factors and increase learning. In this work, the learning companion is implemented as a teachable robotic agent. Learning companions as teachable agents involve learners in a "learning-by-teaching" experience, where learners teach the agent about a subject domain (Chou, Chan, and Lin, 2003). When teaching others, learners attend more to the problem, reflect on misconceptions when correcting their peers' errors, and elaborate on their knowledge as they construct explanations (Roscoe and Chi, 2007). Learning by teaching can improve domain knowledge (Kauchak and Eggen, 1993), self-efficacy (Frager and Stern, 1970), and peer attitudes (Griffin and Griffin, 1998).

The teachable robot platform presents a unique opportunity for exploring the design and outcomes of an entraining, social dialogue system. Teachable agent interactions are thought to benefit from increased social engagement. Student tutors who feel more invested or feel more rapport for their agent have been found to learn more (Leelawong and Biswas, 2008; Ogan et al., 2012). Human-human dyads who exhibit higher rapport tend to have greater success in peer tutoring interactions. However, it is unclear how rapport and learning interact in these scenarios. A meaningful implementation of entrainment within a teachable robot may provide insight into both the relationship between entrainment and rapport and the relationship between rapport and learning in human-human and human-agent interactions. This suggests two additional research questions for this work:

4

RQ 3: How does entrainment influence learning in a robotic learning companion?

RQ 4: What insights regarding human-human and human-agent interactions can we gain by manipulating social behavior in a robotic learning companion?

Three main contributions emerge from this work. First, through the iterative exploration of how entrainment can be designed and implemented, I present an entraining, social dialogue system which is both robust and modular, capable of being explored in future contexts. One of the first dialogue systems of its kind, the system incorporates entrainment alongside other rapport-building verbal behavior and demonstrates successful, positive effects on rapport and learning. Secondly, this thesis contributes knowledge towards an integrated theory of entrainment and rapport by exploring the effects of acoustic-prosodic entrainment on rapport and learning in interactions with a teachable robot. Entrainment occurs in many contexts and across many domains. Rapport is also equally important in many contexts and domains. The significant results of this dissertation provide support for exploring how entrainment can foster rapport in other human-agent interactions where social engagement is equally important. Similarly, the findings regarding the relationship between entrainment, rapport, and learning provides a foundation for exploring other similar rapport-building phenomena in learning interactions. Finally, as a part of exploring the effects of entrainment with a teachable robot, this dissertation provides an understanding of how defining characteristics such as gender influence responses; this understanding is an important step in the creation of personalized systems to enhance social factors.

This thesis is organized as follows. Part I provides an overview of human-human entrainment and my own work analyzing human-human interactions. The goal of this section was to provide the initial insight needed to identify potential models for

implementing entrainment in a dialogue system. Part II contains the bulk of this dissertation. I took an iterative approach to designing, implementing, and evaluating entrainment in a dialogue system. The introduction to Part II in Chapter 8 summarizes the overall approach consisting of six iterations. Each iteration and its results are described in the remaining chapters. Part III concludes this thesis with an overview of the most significant findings and directions for future work.

PART I:

FOUNDATIONS IN HUMAN-HUMAN ENTRAINMENT

CHAPTER 2

MOTIVATION AND RESEARCH GOALS FOR PART I

This thesis poses four research questions regarding acoustic-prosodic entrainment, including how it can be modeled in a system, the effects automated entrainment might have on feelings of rapport and learning, and the insights we might gain by manipulating entrainment in a human-agent interaction. Part I of this thesis examines existing work on human-human entrainment and presents new work on human-human entrainment to provide insight into the design of potential entrainment models and possible effects.

Prior work on human-human entrainment can provide a general direction for how to model entrainment as well as some insight into the potential positive effects on social responses (Gravano et al. 2014; Levitan et al. 2012). However, prior work does not specify how entrainment specifically relates to feelings of rapport. Individuals can entrain in many ways; it is an open question whether rapport is positively related to entrainment in all ways for all features. This information is important for identifying appropriate models of automated entrainment that might be able to build rapport.

In addition to how entrainment relates to rapport, it is unclear how entrainment combines with content-based dialogue approaches for building rapport. Dialogue systems present a distinct opportunity to manipulate both content and prosody to build rapport. In manipulating content and prosodic entrainment, it is possible the effects of entrainment might differ for different content. For example, entrainment might be more influential when individuals are engaged in social, rapport-building dialogue or it might be influential in learning-oriented task-based dialogue. Understanding if individuals entrain differently

when engaged in different kinds of dialogue can inform the design of automated entrainment and guide interpretation of responses to a social, entraining system.

Given the lack of clarity on how entrainment relates to rapport or differs for different dialogue behaviors, I pose the following two research questions for Part I: **(1) how are different forms of acoustic-prosodic entrainment related to rapport? (2) how does entrainment differ for social and knowledge building dialogue?** To answer these questions, I present an analysis of prior work on human-human entrainment followed by a description of a supplementary, exploratory analysis on a corpus of human-human conversational data collected from a set of dyads as they worked together on a set of math problems. This exploratory analysis provides additional insight into the questions regarding how entrainment relates to rapport and different dialogue behaviors.

Chapter 3 summarizes the prior work on human-human entrainment, Chapter 4 describes the dataset I used for the supplementary analysis. An analysis on entrainment and rapport is presented in Chapter 5. Chapter 6 looks at the relationship between prosodic entrainment, social dialogue, and learning dialogue content. Chapter 7 summarizes key take-aways and provides conclusions regarding how these results might guide the design, hypotheses, and interpretations of a social, entraining dialogue system for a robotic learning companions.

CHAPTER 3

RELATED WORK

3.1   HUMAN-HUMAN ENTRAINMENT

Entrainment is quite prevalent in human-human interactions and occurs on multiple dimensions in addition to prosody. People have been found to mimic and adapt their facial expressions, their body language, and the content of their speech in addition to their speaking style (Hess & Blairy, 2001; Lakin & Chartrand, 2003; Nenkova et al., 2008; Levitan & Willson, 2012). Two primary theories have been proposed to explain why entrainment appears to be so prevalent: The Communication Accommodation Theory and the Interactive Alignment Theory.

The Communication Accommodation Theory (CAT) proposes that individuals will either entrain or dis-entrain as a means of achieving solidarity or to dissociate themselves from their interaction partner. A socio-psychological theory explaining entrainment based on CAT argues that the phenomenon is driven by the need to achieve certain social effects and is based on the idea of similarity-attraction. The similarity-attraction theory posits that, "The more similar the attitudes and beliefs are to those of others, the more likely it is for them to be attracted to us." (Giles & Smith, 1979). Individuals use entrainment to obtain social approval from their interlocutor. This theory suggests that an individual on the receiving end of a high level of accommodation is likely to develop a greater sense of self-esteem and satisfaction and to feel more rapport for their speaking partner than if they were a receiver of low accommodation.

According to the Interactive Alignment Model (IAM), entrainment contributes to and facilitates the construction of shared mental representations over the course of a conversation through the alignment of situation models (Pickering & Garrod, 2013). For example, at the end of a successful conversation, dialogue partners have similar representations of the time and location of events, the main characters involved, etc. This situation model alignment occurs automatically, and as it occurs, dialogue partners align at many levels, including lexical, semantic, syntactic, and acoustic-prosodic. Alignment at one level leads inexorably to alignment at other levels and that alignment at one level is enhanced by greater alignment at other levels. This process happens without any form of explicit negotiation; to explain how this happens without individuals engaging in a discussion, Pickering and Garrod propose that speakers are primed by each other to utilize the same forms. This idea of interactive priming is a method for operationalizing entrainment and has been used as a measure for entrainment at lexical, syntactic, and acoustic-prosodic levels.

Regardless of why individuals entrain, analyses in human-human interaction have examined entrainment along two primary time-scales which are referred to as local and global. Local entrainment is measured on a turn-by-turn basis while global entrainment is measured across the course of a conversation, typically by comparing the beginning to the end. In the literature, there are five different types of local and global entrainment: synchrony, convergence, divergence, proximity, and priming. Synchrony is typically only measured locally; convergence, divergence, proximity, and priming are measured both locally and globally. Another form of entrainment known as compensation is discussed theoretically; however, no current literature appears to have explored it. Figure 2.1 depicts

11

the types of entrainment as they might be observed in a rough approximation of two speakers' signals. The definitions for each type are below.

- **Proximity**: The broadest, most general form of entrainment, proximity occurs when individuals match or mirror one another. It implies no direction or actual contingency. Generally, the distance between the two speakers' raw acoustic-prosodic features is indicative of how much the two speakers entrain by proximity.

- **Synchrony**: Exhibited when individuals exhibit similar rhythmic qualities and coordination of features. As defined by Burgoon, Stern, and Dillman (2007), simultaneous synchrony refers to when behaviors occur at the same time, while concatenous synchrony occurs when the behaviors are part of a sequential speaker-speaker pattern – concatenous synchrony is much more common in analyzing entrainment in dialogue.

- **Convergence**: One of the most frequent measures of entrainment, convergence refers to when individuals start with very different dialogue features and become increasingly similar over time. When convergence is present, the difference between two speakers' acoustic-prosodic features shrinks over time

- **Divergence**: The opposite of convergence, divergence, is the process of interaction whereby an individual adopts behaviors that are increasingly dissimilar from that of their partner.

- **Compensation**: Currently not measured in entrainment analyses, compensation occurs when individuals over-adapt to one another, going between extremes

- **Priming**: Not depicted in Figure 2.1, priming is defined as a process where the occurrence of a stimulus (the prime) influences the processing of a subsequent

stimulus (the target). Priming is typically used to measure entrainment on lexical features though it has been applied to acoustic-prosodic features.

In a review of 68 papers on acoustic-prosodic entrainment analyses of human-human data, 34 explored some form of local entrainment, 13 explored global entrainment, and 21 explored some form of both global and local entrainment. Proximity and convergence/divergence dominated the analyses, with 39% of papers analyzing some form of proximity and 59% analyzing some form of convergence / divergence. Most of these works contributed either a new approach for measuring entrainment or an understanding regarding how entrainment relates to communicative and social constructs such as communicative success, liking, and trust. Of those in the latter group, which contributed understanding to how entrainment is related to social constructs, many explored proximity
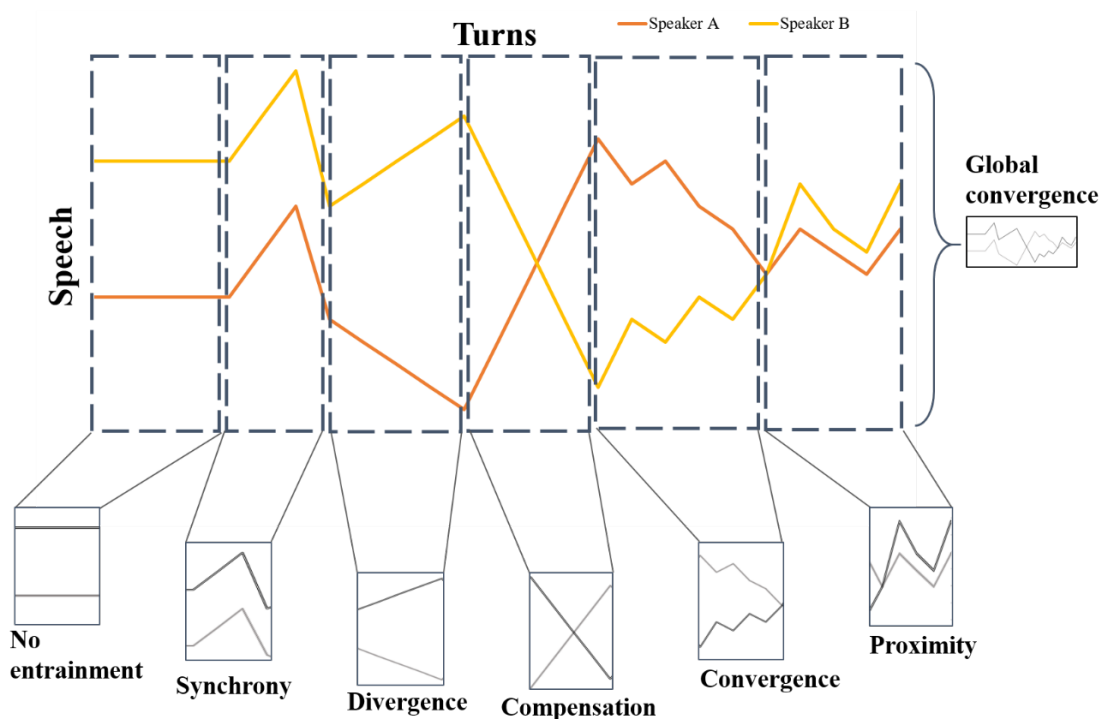
Figure 2.1. A Visual Depiction of Different Types of Entrainment

Global	**Proximity**	Pitch contour	↗	Engagement [Gravano et al. 2014]
		Positive words	↗	Trust [Scissors et al. 2009]
		Negative words	↘	Trust [Scissors et al. 2009]
		Pitch, intensity…	↙	Trying to be liked [Levitan et al. 2012]
		Pitch, intensity	↖	Conversational quality (ASD) [Bone et al 2013]
	Convergence	Pitch, intensity	—	Agreement [Bonin et al. 2013]
		Pitch	↗	Solidarity [Gregory et al. 1993]
		Pitch, intensity…	↗	Backchannel preceding cues [Levitan et al. 2011]
		Intensity	↗	Social desirability [Natale 1975]
		Speaking rate	↗	Rapport [Sinha & Cassell 2015]
		Boundary tone	↗	Common ground [Mushin 2003]
		Speaking rate	↘	Bias (vocal expectations) [Sidaras 2011]
Local	**Proximity**	Pitch, intensity…	↘	Latency [Levitan et al. 2015]
		Pitch, intensity…	↗	Backchannel [Levitan et al. 2015]
		Intensity	↗	Filled pauses [Benus et al. 2012; Benus 2009]
		Pitch, energy…	↗	Positive/negative affect [Lee et al 2010 / 2011]
		Pitch…	↗	Rapport [Lubold 2014]
		Pitch, intensity…	↗	Trying to be liked [Levitan 2012]
		Intensity, CF	↗	Favorable voting [Benus 2014]
		Phonetic, pitch	↗	Positive bias [Babel 2010; Babel 2012; Babel 2012]
		Overlaps	↗	Engagement [Kousidis 2009]
	Convergence	Pitch, intensity…	—	Backchannels [Lubold 2015]
		Pitch	↗	Grounding [Lubold 2015]
		Pitch, intensity…	↗	Transactive contributions [Gweon 2013]
		Speaking rate	↗	Liking [Schweitzer & Lewandowski 2013]
	Synchrony	Pitch, intensity…	—	Rapport [Lubold 2014]
		Pitch, intensity…	—	Agreement [Vaughn 2011]

Table 2.1. Relationship Between Entrainment and Rapport. Many types of social factors were considered. The ellipsis (…) indicates that there were additional acoustic-prosodic features explored.

and/or convergence and found positive relationships as shown in Table 2.1. The overwhelmingly positive findings may be due to non-reporting of non-significant results.

The positive results are supported by a popular theory of rapport originating from Tickle-Degnen and Rosenthal's (1990) work. This theory suggests that there should be a connection between feelings of rapport and evidence of entrainment. Tickle-Degnen and Rosenthal describe rapport as that moment or feeling when two people "click" or "have chemistry" and suggest that rapport consists of three components: positivity, mutual attention, and coordination. Positivity represents feelings of mutual friendliness and caring; mutual attention is a feeling that the other partner is involved, that there is an intense mutual interest in what the other is saying or doing. Coordination occurs when partners are 'in sync,' when behaviorally there is a high degree of coordination. Based on the definition of entrainment, it can be considered a form of coordination, on multiple levels. This suggests a connection between entrainment and rapport. Most analyses have targeted rapport-related attributes such as liking, engagement, and desirability.

Within human-human analyses, most approaches looking at acoustic-prosodic entrainment have utilized prosodic analysis. Prosodic analysis involves extracting individual features such as pitch, intensity, vocal quality, and speaking rate from differing levels of dialogue (i.e. from a turn, from within a turn, or from the first half of the dialogue). As can be seen from Table 2.1, most analyses have involved pitch and intensity.

In human-human tutoring dialogues, Ward and Litman (2007) found that students converge towards their human tutor on measures of pitch and intensity. Thomason and Litman (2013) explored whether there is any relationship between learning and entrainment with ITSPOKE, an intelligent tutoring system (Litman & Silliman 2004). Participant

entrainment towards the system on both pitch and intensity was significantly correlated to learning. Levitan and Hirschberg (2011) found that speaker intensity was the one feature on which individuals tended to be more similar to their partner then to themselves.

It is clear from the prior work that entrainment is connected to social factors and possibly to learning. I build on this prior work to explore how prosodic entrainment is directly related to measures of rapport and to dialogue behaviors indicative of rapport, such as social dialogue, and of learning, like grounding.

3.2   RAPPORT

Rapport has been defined and measured in several ways in prior work, including as self-reported rapport through questionnaires (i.e. "I felt a connection with the robot"), as perceptual rapport through third party perceptions where an individual observes and rates an interaction for rapport, and as behavioral rapport where the user's behaviors are used as assessment of their rapport (i.e. does the individual smile, do they use rapport-building language) (Pantic et al., 2007; De Carolis et al., 2015; Foster, Gaschler, and Guiliani, 2013; Bechade et al., 2015). For Part I, I utilize measures of self-reported rapport and perceptual rapport. In Part II, I use self-reported, behavioral, and perceptual measures.

As a self-reported measure, rapport has been assessed as general rapport related to feelings of connection and harmony (Gratch et al., 2007), as coordination, positivity, and attention (Sinha & Cassell, 2015; Tickle-Degnen and Rosenthal, 1990), and as social presence. For the work in Part I, I measure rapport as general rapport based on feelings of closeness. In Part II, I measure self-reported rapport as general rapport, as coordination, positivity, and attention, and as social presence. Social presence as a measure of rapport

is more common with agents and robots (Huang, Morency, & Gratch, 2011). Social presence has been described as the "level of awareness of the co-presence of another human, being, or intelligence" and as the "feeling that one has some level of access or insight into the other's intentional, cognitive, or affective states" (Biocca & Nowak, 2001). With robots and agents, social presence may be an important measure of rapport given the inherently remote properties typically associated with these technologies.

For perceptual rapport, prior work has looked at measuring rapport by evaluating "slices" of the audio and video (Ambady and Rosenthal, 1992). Third party observers than listen and watch the clips and answer questions about the degree of rapport they perceive. The size of the clips ranges; prior work typically looked at "thin-slices" of 30-seconds. This approach to measuring rapport has been used in prior work to rate the degree of rapport between peer tutors and tutees (Madaio, Ogan, & Casssell, 2016). I use a similar approach to measure perceptual rapport. In Part I, I use longer clips of approximately two-minutes while in Part II I use audio clips of 20-30 seconds.

To measure behavioral rapport, rapport theory suggests that linguistic behaviors which are indicative of politeness may provide insight into a user's feelings of rapport. Spenser-Oatey (2005) suggests an individual's use of politeness is an example of how individuals manage rapport. For example, if an individual praises their conversational partner, this may positively enhance their partner's feelings towards them. If an individual is rude to their conversational partner by calling them a name, this may introduce face-threat, hindering rapport. Bell, Arnold, and Haddock (2009) performed an analysis of linguistic politeness and interpretation of its meaning in peer tutoring scenarios. Based on the dialogue of two pairs of tutors and tutees, the authors analyzed different politeness

strategies on the part of the tutor based on verbal behaviors such as inclusive language, praise, and humor that were suited to the peer-tutoring domain. In first-time sessions, the tutors appeared to be reluctant to utilize positive politeness behaviors such as inclusive language and praise; over the course of multiple sessions, these behaviors increased and aligned to building rapport. Similar behaviors such as praise, inclusive language, name usage, and formal politeness have been found to be associated with positive rapport in other prior works (Ogan 2012a; Wheldall 1985). Exploring how students utilize similar linguistic strategies when tutoring a robotic learning companion may provide insight into their level of engagement in building a relationship with the companion and the rapport they feel for the robot. For this work, I measure behavioral rapport through their linguistic behaviors.

3.3   KNOWLEDGE BUILDING DIALOGUE

Dialogue can facilitate learning in several ways. One of the more popular learning behaviors analyzed in conversational dialogue is grounding. Grounding refers to the joint activity of speakers and listeners establishing common ground—a shared understanding of their mutual knowledge, beliefs, and assumptions. In learning interactions, effective conversational grounding has been shown to be a critical component of successful collaboration. Grounding with successful collaboration has been shown to facilitate learning (Traum & Dillenbourg, 1998; Traum 1999; Baker, 1999).

Assessing grounding can be challenging. Clark and Shaefer (1989) first introduced the idea of measuring common ground through coordinated contributions. Each contribution consists of two parts. In the first part, a speaker utters a contribution with an intended interpretation. Their dialogue partner responds, indicating that they heard the

utterance. They also indicate some interpretation of that utterance. How the listener interprets the speaker's utterance depends on the common ground established. If there is high mutual knowledge, the speaker's intended interpretation and the listener's actual interpretation will be more similar.   Misinterpretations result when knowledge coordination is incomplete. If a speaker is not satisfied with a listener's interpretation or the listener is uncertain about their own interpretation, additional acts are taken to expand or repair the initial utterance. As dialogue partners communicate, they come to some level of agreement on their interpretations, resulting in common ground.

The structural complexity of Clark and Shaefer's model is difficult to transition to computational systems, because there are multiple states to the process of grounding. A single contribution can consist of multiple turns between speakers as they work towards an agreed interpretation of an utterance, expanding and repairing as needed to establish what mutual knowledge. Several modifications and adaptations have been proposed for converting a grounding model to HCI with a focus on collaborative engagement (Traum 1999; Cahn 1999). One adaptation proposes identifying grounding "units" of discourse structure and that these "common ground units" (CGUs) are the basic unit of collaborative structure (Core et al., 1999). This adaptation focuses on the part of contributions which is directly related to adding or confirming mutual knowledge. Mushin, Stirling, Fletcher, and Wales utilized this adaptation to analyze the structure of common ground units in terms of prosody (Mushin 2003). I base the analysis of grounding on this work, identifying behaviors using the concept of grounding contributions.

3.2   SOCIAL DIALOGUE

Social dialogue has been explored as various forms in both human-human interactions and human-agent. Ogan and colleagues (2012) found that joking and teasing in conversation between peer tutors resulted in higher learning gains and Silvervarg, Gulz, and Sjödén (2010) found that students performed better when they evinced positive attitudes towards math. Bickmore and Cassell (2000) found that anecdotes and off-task, small talk had a social and trust-building effect in conversation with embodied conversational agents while Van Mulken, André, and Müller (1998) found that engaging in off-task conversation results in participants feeling more comfortable, relaxed, and at-ease with the task. Other works have also shown that off-task dialogue can promote trust and rapport, improving the learning environment (Bickmore, 2003). These findings suggest that a place to start in analyzing social dialogue between collaborators may be off-task versus on-task dialogue. It is possible that while on-task dialogue is primarily related to grounding mutual knowledge regarding the problem at hand, off-task dialogue may be more about creating common ground as two unique, independent individuals. This may result in nuanced differences when individuals entrain prosodically, if entrainment is related to rapport.

CHAPTER 4

DATA AND FEATURES

In this chapter, I describe the human-human dialogue corpus I collected to analyze the relationship between entrainment, rapport, and different dialogue behaviors indicative of rapport and learning. I also describe the acoustic-prosodic features I extracted to measure entrainment, how I measured dialogue behaviors pertinent to learning via grounding, and how I measured behaviors pertinent to rapport as social dialogue.

4.1   PROCEDURE

To explore these relationships, I collected a set of eight 30-40-minute dialogues from 16 undergraduate college students with basic knowledge of algebra and geometry. The students worked together in pairs as peers and were randomly assigned to their partners. I gave each student a tablet containing a version of the Formative Assessment with Computation Technologies (FACT) application (http://fact.engineering.asu.edu/). The students worked together face-to-face and the application encouraged collaborative interaction through the use of a shared workspace, as shown in Figure 4.1.

Figure 4.1. Students Collaborating with FACT Tablet Application

The application was designed to support and provide formative assessment for K-12 students solving mathematical problems. The mathematical problems available in the FACT application are part of the Mathematics Assessment Project. An example problem can be seen in Figure 4.2. The problems were designed with a goal to make knowledge and reasoning visible; the iterative refinement required to solve the problem was intended to generate conversation and drive collaboration as seen in the sample below:

A: Ohhh . . . negative. Wait, this doesn't help anything

B: Well it's just a bad equation because it's a fraction

A: I clearly can't do this

B: No, it's okay I can do it. So, y equals 10 minus x

The students did not receive any mathematics-based training before the experiment. They began with a 10-minute introductory exercise to ensure they were comfortable using the tablet-interface of the FACT application. In the body of the experiment, the pairs of

Figure 4.2. Screenshot of an Example MAP Problem from the FACT Application

students worked together to solve two math problems (grade level 9 and above) using the tablets. Sessions were of variable length (M = 26.0 minutes, SD = 6.25).

I recorded high-quality audio data using unidirectional microphones with separate audio channels for each speaker. I manually labeled dialogue turns in the following manner. I identified the beginning of a turn as anytime a participant introduced verbal articulation; the end of a turn was either when the participant ceased that articulation or concluded the overall utterance. Laughter and filled-pauses were included. Overlapping speech resulted in overlapping turns. Each student made on average 96 turns (SD = 57).

I then further segmented each turn into inter-pausal units or IPUs. An IPU is a pause-free unit of speech separated from any other speech by at least 50ms (Levitan & Hirschberg, 2011). Turns are composed of one or more IPUs. For example,

> B: No, it's okay I can do it. So, y equals 10 minus x... [silence for .33 seconds] … I mean negative x plus 10

is composed of two IPUs where the first IPU is "No it's okay I can do it. So, y equals 10 minus x" is the initial IPU of the turn followed by a pause greater than 50ms, in this case 0.33 seconds or 330ms, and the final IPU of the turn "I mean negative x plus 10." I extracted prosodic features from the level of the IPU and from the level of an utterance.

4.2   PROSODIC FEATURES

With a corpus of human-human dialogue, I was interested in exploring how prosodic entrainment relates to rapport and different dialogue behaviors. I chose to analyze entrainment across five of the most commonly assessed prosodic features: intensity, pitch (F0), jitter, speaking rate, and shimmer. Intensity was the normalized intensity, and pitch

was the fundamental frequency F0. Jitter was defined as the varying pitch and was calculated as pitch period length deviations; shimmer consisted of variations in loudness and was calculated as the amplitude deviations between pitch period lengths.

To extract these features, I used OpenSmile (Eyben, Wöllmer, & Schuller, 2010). For the speaking rate, I applied the approach from de Jong and Wempe, which automatically detects syllables and estimates speaking rate based on syllables per second (de Jong & Wempe, 2009). For each feature, I extracted several functionals, including the mean, maximum, and minimum. Table 4.1 describes these features and the functionals.

Feature	Description	Functionals
Pitch	F0: The fundamental frequency	mean maximum value max value position min value position standard deviation
Intensity	The normalized intensity	mean maximum value minimum value max value position min value position standard deviation
Voice Quality	Local Jitter: frame-to-frame jitter (pitch period length deviations)	mean maximum value max value position min value position standard deviation
	DDP Jitter: Differential frame-to-frame jitter (the 'jitter of the jitter')	mean maximum value max value position min value position standard deviation
	Shimmer: (amplitude deviations between pitch periods)	mean maximum value max value position min value position standard deviation
Speaking Rate	Measured in estimated syllables per second	N/A

Table 4.1. Acoustic-Prosodic Features and Their Functionals

When comparing speakers with different vocal tracts, I normalized features affected by the vocal tract so that they would lie in the same range. This was due primarily to differences in gender, so I normalized the female pitch mean and max by scaling them to lie in the same range as the male values; all other non-pitch features were raw. I describe the measures of entrainment using these features in Chapters 5 and 6.

4.3   RAPPORT MEASURES

In this section I describe how I measured rapport to evaluate the relationship between entrainment and rapport. To measure rapport, I looked at observational rapport and I validated the measure by comparing it to self-reported rapport obtained from five of the eight dyads. Prior to having annotators listen to the recorded audio, I manually selected four two-minute segments from each dialogue (32 segments in total). These segments optimized the amount of dialogue pertaining to the math problems and minimized the amount of silence. I manually annotated turn boundaries in each two-minute segment, defining a turn as a continuous speech utterance by a single speaker, including filled pauses and laughter (Traum & Heeman, 1997). I obtained a measure of perceptual rapport by having three annotators listen to only the audio of each of the 32 two-minute-long conversational segments. Since there were four segments per dyad, I randomized the order in which the annotators listened to the segments. For each segment, the three annotators responded to the following statement using a three-point Likert scale (Agree, Neutral, and Disagree): "There is a sense of closeness between Students A and B"

This question was adopted from the rapport scale statements developed by Gratch and colleagues (Gratch et al., 2007). I checked for inter-rater agreement using percent

agreement and Cohen's Kappa; the average pairwise percent agreement across all segments was 63.5% while the average pairwise Cohen's Kappa was 0.41. This was lower than I would have liked but was not entirely unexpected (Schuller et al., 2009; Acosta & Ward, 2011). However, given that the level of agreement between the annotators was lower, I validated the perceptual observations against the measures of self-reported rapport I collected from five of the eight dyads.

For the self-reported rapport, I posed to each of the participants' two questions with a similar connotation at the end of the session. The participants responded to the following statements, again using a three-point Likert scale (Agree, Neutral, Disagree):

"My partner created a sense of closeness between us'"

"I tried to create a sense of closeness between us"

There were two primary differences between the questions I posed to the annotators and those I posed to the participants. The first difference was the participants' self-reported responses were based upon the entire 30-40-minute session rather than a two-minute segment. The second difference was that while the annotators responded to a single, consolidated statement, each participant answered both questions listed above.

I validated the measures of perceptual rapport against these self-reported rapport responses. Both scales were on a Likert-scale from 1 to 3. I coded all responses as 0, 0.5, or 1, corresponding to "Disagree," "Neutral," and "Agree." I then aggregated the perceptual scores by calculating the average of the three observers' ratings. This resulted in a single value between 0 and 1 for each segment. For the self-reported rapport, I also found the average of both participants' responses to obtain a single value between 0 and 1.
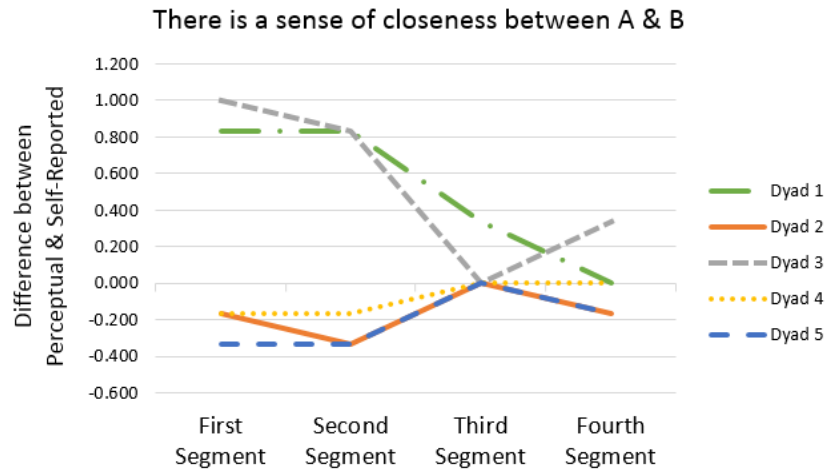
Figure 4.3. Difference in Perceived vs. Self-Reported Rapport for the Five Dyads

I compared the results from the perceptual annotators to the self-reported rapport scores for the five dyads for whom I had self-reported rapport scores. As I divided the dialogue of each dyad into four two-minute segments and the annotators provided perceptual observations for each segment, I examined the difference between the perceptual score of each segment and the overall self-reported score for that dyad. The segments were aligned temporally, in the order in which they occurred in the dialogue. Figure 4.3 depicts the results of this comparison for all five dyads.

Examining Figure 4.3, I found that for the first two segments, the perceptual observers were not aligned with the views of the participants. When I looked at the last two segments, the perceptual rapport scores began to reflect the self-reported rapport from the participants with increasing accuracy. This may be an attribute or reflection of the fact that participants' rapport responses were only collected at the end of the interaction. Perceptual observations from earlier in the interaction are not necessarily inaccurate measures, but they are potentially more representative of in-the-moment rapport rather than reflections of how participants felt at the end. To facilitate consistency, I chose a measure of rapport

27

using the perceptual scores from the two latter segments of each dyad. I use this measure to examine the relationship between entrainment and rapport in Chapter 5.

4.4   ASSESSING KNOWLEDGE BUILDING DIALOGUE

In addition to evaluating the relationship between entrainment and rapport, I was interested in how entrainment differs for dialogue behaviors which are thought to be important to learning. In this section, I discuss how I measured dialogue behaviors pertinent to grounding. Grounding is a critical component of successful collaborative interactions and is indicative of both cognitive and social factors in learning interactions.

For this analysis, I identified two common ground features based on common ground units, attempting to capture those parts of the dialogue related purely to mutual understanding (Nakatani & Traum, 1999). I identified (1) backchannels, and (2) "grounding" contributions as the two features to explore. Backchannels are defined as short non-disruptive segments of speech which a listener utters to let the speaker know they are listening. With backchannels, the purpose is not to take the floor but simply to indicate that the listener is keeping up and the speaker does not need to initiate a repair. The second feature, "grounding" contributions are contributions which are defined as specifically adding, confirming, or updating mutual knowledge. In grounding contributions, a turn is labeled as grounding if it is relevant to the preceding turn. I defined how it is relevant by analyzing whether it contains the response to a question or if there is a clear reference to repeated content from a previous turn by the other speaker. Examples of backchannels and "grounding" contributions are:

	Student 1:	So, I have 8x + 10y and I want to maximize that...
Backchannel	Student 2:	Right
	Student 1:	It looks like she can decorate 10 boomerangs or 20...?
Grounding Contribution	Student 2:	I think it is only just 10 boomerangs total

Grounded contributions and back channels were coded by two annotators and agreement was calculated with Cohen's kappa. Agreement for backchannels was 0.76 and the kappa for collaborative contributions was 0.72.

4.5   ASSESSING SOCIAL DIALOGUE

Finally, I am interested in assessing the relationship between entrainment and social dialogue. I broke the dialogues down into three categories: (1) problem-solving, (2) activity-related, and (3) social. In problem-solving dialogue, the students were actively working on the problem, discussing the solution, and in general, attempting to solve it. This was essentially on-task dialogue and grounding in on-task dialogue is likely to enable collaborators to co-construct knowledge. In social dialogue, I considered social discussion; everything outside the problem-solving task or the activity. Grounding in social dialogue

On-Task/ Off-Task	Topic	% of Corpus	Example(s)
On-task	Problem-Solving	79%	"Because if you add 7 to it, yeah"
Off-task	Activity Related	13%	"Yeah, see mine [screen] froze" "That was just a blank card [application tool]"
	Social	8%	"What are you studying?" "We don't need math!"

Table 4.2. On-Topic and Off-Topic Statistics in Corpus

may be likely to build rapport and improve the relationship between the collaborators. I added a third category, activity-related, to represent utterances where students were not problem-solving or communicating socially, but instead discussing the application itself. Table 4.2 details the topics and the distribution across the corpus. Two individual annotators coded for the topics as well and agreement was measured with Cohen's kappa; the overall average was 0.84.

CHAPTER 5

RAPPORT AND ENTRAINMENT IN COLLABORATIVE LEARNING

Using the corpus described in Chapter 4, I investigated whether the relationship between entrainment and rapport is observable and whether it differs for different forms of entrainment, looking for insight into **how different forms of acoustic-prosodic entrainment might be related to rapport.** The following section summarizes how entrainment was measured for the analysis in this chapter and the results and conclusions are given in Sections 5.2 and 5.3 respectively.

5.1   METHOD

I investigated the relationship between rapport and entrainment with three of the most popular measures of entrainment, proximity, convergence, and synchrony. Inspired by Levitan and Hirschberg (2011), I measured proximity, convergence, and synchrony locally, on a turn-by-turn basis.

Proximity is a measure of entrainment which looks at how closely the two speakers are to each other at a specific point in time as compared to the rest of the conversation. To determine proximity, I ran a paired samples t-test where each pair was composed of two differences. The first difference was the absolute difference between a speaker and their partner at an adjacent turn. The second difference was the absolute difference between a speaker and their partner at ten non-adjacent turns.

Convergence is the degree to which speakers become more similar over the course of the entire conversation. If convergence does not exist, the two speakers may grow further

apart over time (diverge). I calculated convergence by using Pearson's correlation in a two-tailed t-test on time and the absolute difference between a speaker and their partner.

Synchrony is the quality of interaction which occurs when speakers stay "in sync" as they converse. As they converse, they modulate their prosody in tandem. If two speakers are not in sync, this means there is no pattern in how they modulate their voices. To find synchrony, I computed Pearson's correlation coefficient with a two-tailed t-test on the speakers' feature values at adjacent turns.

The above measures of proximity, convergence, and synchrony are based on those suggested by Levitan and Hirschberg (2011). I follow them in considering the results for these measures to be statistically significant when $p < 0.01$ and the results with $p < 0.05$ to approach significance.

5.2   RESULTS

I first investigated whether acoustic-prosodic entrainment existed within the data. I then explored whether there was a relationship between acoustic-prosodic entrainment and perceptual rapport, reporting on features of acoustic-prosodic entrainment which correlated significantly with the rapport scores.

I observed evidence of significant entrainment as shown in Table 5.1. Looking at all three measures of entrainment, all three forms appeared to be significant. Overall participants entrained more on proximity than convergence or synchrony. This means that speakers were most likely to entrain by matching each other at adjacent turns. While synchrony and convergence were present across the corpus, the correlations were smaller.

32
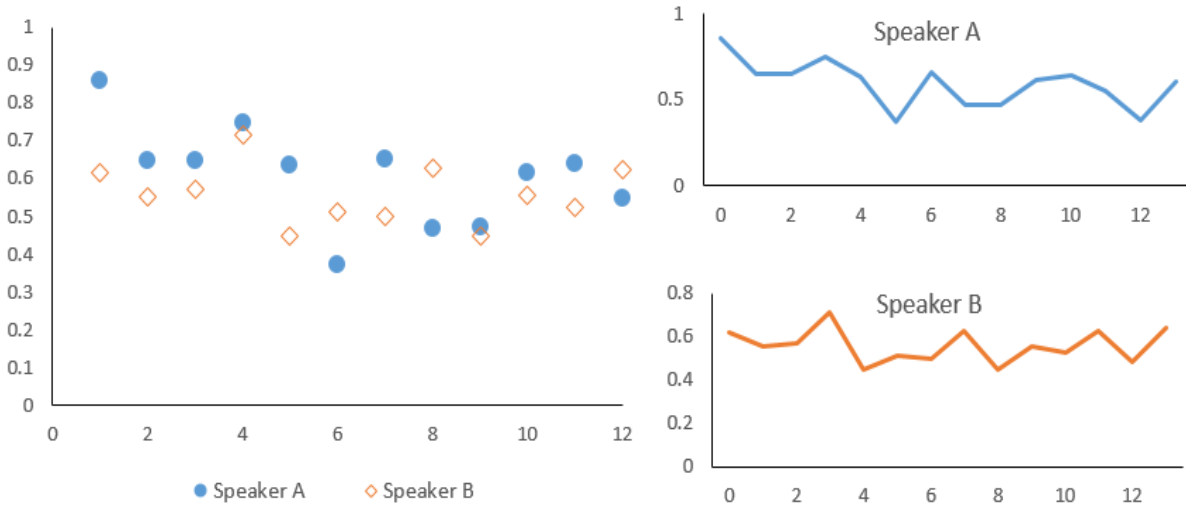
Figure 5.1. Entrainment Between Two Speakers from FACT Corpus

Figure 5.1 depicts each measure of entrainment with a two-minute sample with particularly high-levels of entrainment.

The results on proximity aligned with prior work; like Levitan and Hirschberg (2011), I found speakers matched each other most significantly in terms of intensity, suggesting that the speakers may be changing their normal behavior in intensity to conform to that of their partner. This also aligns with Coulston and colleagues' findings (Coulston, Oviatt, & Darves, 2002) that most children actively accommodated their loudness.

In contrast to Levitan and Hirschberg, I found significance in only a subset of the features I examined for synchrony and convergence, where as they found significance in every feature. Speakers exhibit synchrony when they adjust their speech in tandem with that of their partners. In the corpus, speakers entrained synchronously on intensity to a significant degree; however, the correlations were weak. For convergence, I found that only local jitter was significant when I examined it at the turn-level across the corpus.

33

	Feature	Functional	Paired t-test *t*
Proximity	Intensity	position max	2.29
		std dev	-2.84*
		max	-2.83*
	Pitch – F0	mean	-1.98

	Feature	Functional	Pearsons's Corr. *r*
Synchrony	Intensity	mean	.12*
		std dev	.11*
		max	.09
	Pitch – F0	mean	.08

	Feature	Functional	Pearsons's Corr. *r*
Convergence	Local Jitter	position max	-.09
		max	-.08

Table 5.1. Results of Measuring Proximity, Convergence, and Synchrony. Values shown are significant at $p < 0.05$; values marked with an * are significant at $p < 0.01$.

While I found significant acoustic-prosodic features for all three aspects of entrainment, I did not find the same level of entrainment as found by Levitan and Hirschberg. This could be due to several factors. One may be the differences in domain. Niederhoffer and Pennebaker found that entrainment was associated with the degree of engagement (Niederhoffer & Pennebaker, 2002). The Columbia Games corpus makes use of the gaming domain and may have been more likely to have higher levels of engagement.

While proximity may be the most significant form of entrainment when looking across the entire set of dialogues, it may not be the most significant form of entrainment for each dyad. In addition to looking at entrainment across the whole corpus, I explored entrainment within each individual dyad. As shown in Table 5.2, I found that not every dyad entrains in all three ways. Synchrony was the most common form of entrainment;

	Dyads							
	1	2	3	4	5	6	7	8
Proximity								
Intensity	✓	✓						✓
Pitch			✓		✓	✓		✓
Voice Quality	✓	✓			✓		✓	
Speaking Rate						✓		
Synchrony								
Intensity	✓	✓	✓			✓		✓
Pitch	✓		✓	✓	✓			
Voice Quality	✓		✓	✓	✓		✓	
Speaking Rate		✓						
Convergence								
Intensity		✓		✓			✓	
Pitch	✓							✓
Voice Quality	✓	✓	✓				✓	✓
Speaking Rate		✓						

Table 5.2. Evidence of Entrainment Within Individual Dyads. $p < 0.01$

every dyad entrained synchronously and for five out of the eight dyads, this was the most significant form. Proximity was a close second; seven out of eight of the dyads also entrained with proximity, matching each other on a turn-by-turn basis. The least common form of entrainment within each dyad was convergence, with only five of the eight dyads showing any signs of convergence, and it was also the least significant form of entrainment.

The acoustic-prosodic features which were important for each measure differed depending on the dyad and the measure. Intensity, pitch, and voice quality were distributed across the eight dyads. Speaking rate was entrained on the least. This could be due to the approach for measuring speaking rate, which looks at IPU as the unit of analysis, and the nature of the dialogues, where the IPUs were often shorter in duration.

Finally, I identified whether there was a relationship between entrainment and rapport by comparing the entrainment scores from the dialogues to the perceptual rapport

	Feature	Functional	Pearson's Corr. R
Proximity	Pitch – F0	max	.842*
		mean	.804
		std dev	.510
	Jitter – DDP	max	.644*
		std dev	.512
Synchrony	Pitch – F0	std dev	.568
	Jitter – Local	std dev	.741*
Convergence	Pitch – F0	std dev	.586
	Jitter – Local	std dev	.634*

Table 5.3. Significant Relationships Between Rapport and Entrainment. Values are significant at $p < 0.05$; values marked with an * are significant at $p < 0.01$

scores for the last two audio segments of each dyad. Based on the comparisons with self-reported rapport, the final two segments reflected a more accurate picture of the perceptual rapport. I compared entrainment and rapport with Pearson's correlation coefficient with a two-tailed t-test. The results are given in Table 5.3.

I found proximity had the most acoustic-prosodic indicators of rapport. Interestingly, proximal entrainment on intensity was not related to rapport despite being the most significant feature of proximal entrainment overall. While people entrain more on intensity in general, proximal entrainment on pitch may be more pertinent when looking for indicators of rapport. Figure 5.2 illustrates how speakers entrained differently on pitch mean when higher rapport was observed versus less rapport.

Synchrony where speakers change their behavior in sync, was positively correlated to rapport for two acoustic-prosodic features: the standard deviation of the pitch (F0) and the standard deviation of the local frame-to-frame jitter. Looking at convergence, I found these same two features. All these features are strongly correlated, indicating that synchrony and convergence on pitch also play an important role regarding rapport.

36
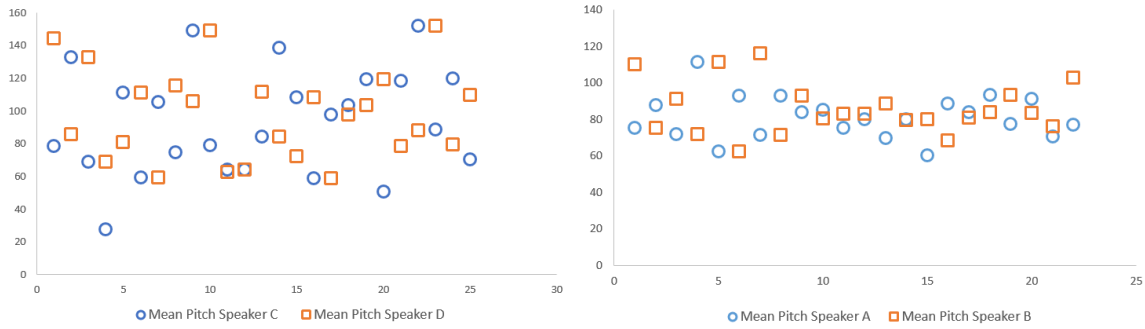
Figure 5.2. Comparison of Proximity on Pitch Mean for Two FACT Dyads.

5.3 DISCUSSION AND CONCLUSIONS

The goal of this chapter was to use the data set from Chapter Four to gain insight into the relationship between entrainment and a measure of rapport, with a focus on how different forms of acoustic-prosodic entrainment are related to rapport. I investigated three measures of entrainment, proximity, synchrony, and convergence using four acoustic-prosodic features (intensity, pitch, voice quality, and speaking rate). I found all three measures of entrainment were present in collaborative learning dialogues I used, with individuals entraining the most on intensity. All three forms of entrainment were also correlated with rapport. Entrainment on pitch and voice quality were the most highly correlated. Matching the pitch of one's speech turn-by-turn, in a measure of proximity, appeared to have the most significant relationship to rapport.

There are a few explanations for why proximity on pitch was the most highly correlated to rapport. Pitch has been found to be important to the emotional coloring of utterances and is known to convey significant metadata about a user's current state and understanding of the conversation. It is possible that turn-by-turn entrainment on pitch is

more susceptible to the context of the conversation and combines more effectively with other rapport-building behaviors as opposed to general conversation wide entrainment which is less likely to be influenced by context within the moment and more likely to represent a general similarity between two individuals. It is also possible that the measure of rapport played a role; rapport was measured as a third-party observation of slices of the interaction. This approach might have been more conducive to capturing turn-by-turn rapport which may be more related to measures of entrainment as proximity.

In this chapter, I focused on the general relationship between prosodic entrainment and rapport, regardless of other social behaviors, to obtain general insights into modeling prosodic entrainment in a dialogue system. I explore the relationship between social dialogue and entrainment in the next chapter but leave explorations of other modalities, entrainment, and rapport to future work.

CHAPTER 6

ENTRAINMENT, SOCIAL DIALOGUE, AND

GROUNDING IN COLLABORATIVE DIALOGUES

In Chapter Five, I established that the theoretical relationship between entrainment and rapport is observable in collaborative learning. I found that entrainment does appear to be more significant for individuals who are observed to have higher rapport. This finding supports exploring prosodic entrainment to build rapport within a dialogue system and provided insight into potential models for implementing entrainment. However, individuals can build rapport in multiple ways. For example, individuals can build rapport through rapport-building dialogue or social dialogue, in which they make a social connection with their partner by engaging them in off-task, social conversation. In designing a dialogue system which utilizes entrainment, it is also important to know how entrainment might differ or influence responses depending on the context in which it is introduced. It is unclear from prior work how prosodic entrainment might combine with verbal behaviors known to build rapport or even behaviors known to facilitate learning. Knowing how entrainment might interact with social and knowledge building dialogue is critical to the design and evaluation of entrainment. In this chapter, I pose the following research question: **How does entrainment differ with social and knowledge building dialogue?**

To explore how entrainment might differ for social, rapport-building dialogue and dialogue behaviors pertinent to learning, I focus on social dialogue and knowledge building dialogue such as grounding. I break down the overarching research question into three

more explicit questions focused on these specific dialogue features: (1) Do students entrain with their conversation partners while engaging in grounding behaviors? (2) How does entrainment relate to social dialogue versus problem-solving dialogue? (3) How does the interaction between grounding behaviors and social dialogue influence the ways in which students entrain on each other's speech? I explored these questions with the dialogue corpus and features described in Chapter Four. In the next section, I give an overview of the method for measuring entrainment and evaluating how it relates to grounding and social dialogue. In 6.2 I describe the results and end the chapter with a discussion of the findings.

6.1   METHOD

For this analysis, I utilized a measure of entrainment adapted from the approach by Thomason, Nguyen, and Litman (2013). Using the prosodic features described in Chapter Four, I calculated an entrainment score by first dividing the dialogues into N exchanges. An exchange was composed of two consecutive turns by different speakers. For each acoustic-prosodic feature, I identified a sequence of exchanges $(n_1, \ldots, n_N)$ where each exchange consisted of the raw feature values for each speakers' turn within that exchange:

$$n_i = (f_{a_i}, f_{b_i})$$

f_{a_i} is speaker A's feature value at exchange n_i

f_{b_i} is speaker B's feature value at exchange n_i

To measure entrainment, Thomason and colleagues used the r^2 correlation between two speaker's acoustic prosodic features within each exchange to compute an entrainment score across the entire dialogue. One limitation of this method was that a single entrainment score across the dialogue will not capture the dynamics of the turn-by-turn coordination. Since

Topic / Grounding Behavior	Mean (turns)	Standard Deviation
Problem-Solving	111.5	78.9
Activity-Related	20.5	20.3
Social	12.9	11.0
Backchannels	12.1	16.5
Grounding Turns	18.2	26

Table 6.1. Breakdown of On-Task/Off-Task and Grounding Behaviors

the dynamics of social and knowledge coordination happen at a turn-by-turn level, I modified this approach to obtain a score at each exchange. Starting with the fifth exchange to occur in the dialogue, I calculated the similarity score of A and B using the current exchange and the four exchanges preceding it. The similarity score was calculated as the linear fit coefficient, resulting in a set of r^2 coefficients for each feature which served as the entrainment scores for each turn pair. I interpreted a high score to mean that the turn had higher entrainment.

6.2 RESULTS

I used regression for the analysis, and for all tests, I considered the result to be significant if $p < 0.05$ and approaching significance if $p < 0.1$. For grounding and backchannels, we used the features described in Chapter 4.4. In the dialogue, 37% of the turns across the entire corpus contained grounding contributions and 25% were backchannels. Table 6.1 illustrates the distribution of backchannels and grounding turns across the different types of dialogue as well as the distribution for social, activity, and problem-related dialogue. In the next section I describe how I analyzed whether entrainment was indicative of grounding

		Dyad	Intensity	Pitch	Jitter	Shimmer	Chi-square (overall model)	p-value (overall model)
Grounding Contributions	*B*	-0.08	-0.34	0.51	0.16	-0.27	**11.763**	**0.038**
	p	0.013	0.22	0.05	0.58	0.37		
Backchannels	*B*	0.02	0.35	-0.07	-0.56	-0.25	5.132	0.400
	p	0.67	0.24	0.81	0.10	0.44		

Table 6.2. Logistic Regression on Grounding Contributions with Entrainment

and the dialogue types. I then investigated the co-occurrence of grounding and entrainment

in problem-solving, activity-related, and social dialogues.

6.2.1   Entrainment Differences for Grounding and Social Dialogue

To first understand the relationship of grounding and entrainment, I analyzed whether

entrainment at the mean of intensity, pitch, jitter, and shimmer had any predictive power

in terms of the two grounding behaviors, backchannels and grounding contributions. I

performed a hierarchal logistic regression analysis to control for the effect of measuring

the entrainment features across different dyads or pairs of students. I ran the analysis with

dyadic differences first and added in the entrainment features as intensity, pitch, jitter, and

shimmer. The results are given in Table 6.2; only the final models are given.

In examining backchannels, I found that dyadic differences were not a significant

predictor. Examining grounding contributions, I found dyadic differences alone

significantly predicted grounding contributions versus non-grounding contributions, with

$p = 0.013$. In the full model, predicting grounding from intensity, pitch, jitter, and shimmer

while controlling for dyadic differences, pitch mean also approached significance at $p = 0.053$. This implies that entraining on pitch may have some potential when considering the

		Intensity	Pitch	Jitter	Shimmer	Chi-square (overall model)	p-value (overall model)
Activity Related	*B*	0.10	0.17	-0.09	-0.27		
	p-value	0.89	0.81	0.91	0.76	**145.1**	**P < 0.01**
Social	*B*	-1.09	0.07	0.24	-0.14		
	p-value	0.04	0.66	0.80	0.88		

Table 6.3. Differentiating Topics in On-Task/Off-task Dialogue

behavior of grounding contributions. While the model overall for predicting grounding contributions was significant, I cannot conclude that entrainment significantly predicts grounding as it was primarily the contribution of the dyadic differences which contributed to the significance of the model. I conclude that while I observed that pitch appears to have a relationship with grounding, students did not appear to be significantly entraining with their partner when engaged in grounding behaviors.

I also investigated the relationship of entrainment to topics controlling for dyadic differences. I again used regression, this time multinomial regression to identify if entrainment differentiated between topics. I treated problem-solving as the base for the analysis, comparing how well entrainment differentiated activity-related dialogue and social dialogue from problem-solving dialogue. Controlling for dyad resulted in 8 coefficients for the 8 dyads. The results are presented in Table 4. I found that the overall model was significant at p < 0.01, and within the model I found that the coefficient of intensity mean was a key differentiator for social dialogue in comparison to the base class of problem solving dialogue. This indicated that when individuals increased entrainment on intensity mean by one-unit, there was 1.093 decrease in the relative log odds of social

dialogue in comparison to problem solving dialogue. Individuals were more likely to increase entrainment on intensity when they were engaging in problem-solving dialogue.

6.2.2 Relationship of Entrainment, Grounding, and Social Dialogue

I investigated whether entrainment differentiated between the grounding behaviors students utilized when they were engaged in on-task versus off-task dialogues. For this analysis, I utilized only grounding contributions. I again used regression to analyze grounding in terms of entrainment for each type of dialogue, identifying whether acoustic-prosodic entrainment differentiated grounding contributions within different topics.

I performed logistical regression analyses on grounding contributions with entrainment as the independent variable for all three topics, controlling for dyadic differences. The results for all three are in Table 6.4. I found that for problem-solving dialogue, the final model was not significant. However, entrainment on shimmer did contribute significantly to the discernment of grounding versus non-grounding behaviors in problem solving dialogue. When students were entraining on shimmer, the odds that their individual contributions were grounded was $\exp(\beta) = 0.512$. In examining grounding

		Dyad	Intensity	Pitch	Jitter	Shimmer	Chi-square (overall model)	p-value (overall model)
Problem-Solving	B	-0.04	-0.08	0.45	-0.09	-0.67	7.968	0.158
	p-value	0.20	0.80	0.14	0.78	0.03		
Activity Related	B	-0.01	0.10	0.17	-0.09	-0.27	0.239	0.77
	p-value	0.91	0.89	0.81	0.91	0.76		
Social	B	-0.29	-1.55	1.73	-1.75	2.35	**11.69**	**0.04**
	p-value	0.04	0.17	0.11	0.05	0.17		

Table 6.4. Differentiating Grounding in On-Task/Off-Task Dialogues

44

in activity-related dialogues, I did not observe any significant contributions, either from dyadic differences or from entrainment. For social, off-task conversation, the final regression equation was significantly predictive of grounding. Notable features which contributed to the overall significance of the equation included dyadic differences and entrainment on shimmer at $p = 0.055$.

6.3   DISCUSSION AND CONCLUSIONS

This chapter focused on how entrainment differed for different types of dialogue behaviors, particularly social dialogue and grounding, posing the following questions: (1) Do students entrain with their conversation partners while engaging in grounding behaviors? (2) How does entrainment relate to the topic of conversation (i.e., whether it is problem-solving or social)? (3) How does the interaction between grounding behaviors and topic of conversation influence the ways in which students entrain on each other's speech? I found that students do not appear to entrain differently when in engaged in task-based conversation, but that entrainment differed if they were engaged in social dialogue.

One interesting aspect of the results was that I did not find entrainment related to backchannels. This was contrary to previous findings (Levitan, Gravano, & Hirschberg, 2011). This contradiction may be the result of different approaches for measuring entrainment. Levitan et al. employed a global measure of entrainment across the dialogue, measuring how speakers entrained in the first half of the conversation as compared to the latter. I employed a moving window of correlational analysis which represents the local dynamics of entrainment. The results for this analysis suggest different behaviors may be

pertinent depending on the type of measure utilized. Future work should explore the repercussions of using different measures more thoroughly.

Finally, these results suggest that speech signal features like entrainment have potential for identifying knowledge building dialogue. In addition, detecting entrainment may also help identify when social interaction is occurring in real-time. This technique could be of large benefit to an adaptive system that attempts to assess and support face-to-face collaboration. Future work should explore how entrainment can be detected real time to facilitate and support collaboration.

The study was limited by the small sample size (only 8 dyads), and the fact that the mathematics task students completed may have been relatively easy and thus not an effective trigger for collaborative discussion and productive grounding behaviors. For these reasons, the analysis is exploratory. One key element of the analysis was the emphasis on relating features of speech to collaborative process on a turn-by-turn basis. By examining the relationship between entrainment, grounding, and on-task/off-task dialogue for each turn, I was able to build a potential process for observing how dialogue evolves over the course of an interaction.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

Overall, this dissertation is focused on four research questions addressing how entrainment might be modeled in a system, the effects modeling entrainment can have on social responses like rapport and outcomes like learning, and any insights we might gain by modeling entrainment. Part I of this thesis examined existing work on human-human entrainment and presented new work on human-human entrainment to provide insight into potential models of entrainment and its effects. The insights gained addressed two open questions pertinent to modeling entrainment:

How are different forms of acoustic-prosodic entrainment related to rapport?

How does entrainment differ for social and knowledge building dialogue?

In exploring existing work on human-human entrainment, results clearly indicated entrainment was related to positive social factors such as liking and engagement. However, the relationship between entrainment and rapport itself remained unclear. I presented new work on human-human entrainment using a corpus consisting of eight dyads working collaboratively together to solve math problems. I found that entrainment does appear to be related to rapport, and particularly turn-by-turn entrainment on pitch appeared to be stronger between dyads who were observed to have more rapport. Entrainment also appeared to differ between problem-solving and social dialogue depending on the feature of entrainment explored. These results both support and extend existing research on entrainment by providing evidence that entrainment is related to measures of rapport and suggesting that people may entrain differently when engaging in social dialogues versus

problem solving dialogues. The results contribute to the body of research that suggests that acoustic-prosodic entrainment is an indicator of productive collaborative interactions during both problem-solving and social conversation.

Towards the design of entrainment, these results suggest several potential directions for modeling automated entrainment. Given the prominence of entrainment on pitch for dyads with high rapport, pitch may be a promising feature to implement entrainment on in a system. Modeling entrainment based on proximity or turn-by-turn adaptation may also have potential for fostering rapport. In terms of understanding the repercussions of implementing entrainment in a system that uses both social, off-task dialogue while engaging in on-task problem-solving conversation, the findings suggest that implementing entrainment during both on-task and off-task dialogues should be acceptable. If the type of entrainment were to be altered depending on dialogue type, intensity appears to be a promising feature for entrainment during on-task, problem solving dialogue while pitch may have potential in off-task dialogue. Because of this, adapting to both intensity and pitch may produce interesting effects depending on the dialogue content.

There were also considerable dyadic variances on entrainment. It was rare that an individual dyad exhibited entrainment on all features, and the features and types of entrainment which were significant for an individual dyad varied – for example, three dyads exhibited significant proximal entrainment on pitch for three other dyads, pitch convergence was significant. Given the size of the data set, it is hard to say what role these individual dyadic differences played on feelings of rapport, but it is possible that individual differences may contribute to different responses depending on the features and methods used to implement entrainment.

48

The focus of this thesis is on modeling entrainment to explore its effects; however, the findings from Part I also hold implications for systems which can automatically detect entrainment. I note the possibilities regarding detecting entrainment for potential future work. Systems which can detect entrainment may provide real-time information regarding interactions between people and between people and computers, such as the degree of rapport felt between two interacting partners. Automatically detecting rapport in human-to-human and human-to-computer interactions has real-world implications. In the classroom, automatically detecting rapport can serve as a guide for teachers when students are engaged in collaborative activity. In tutorial dialogue systems, detecting rapport has implications for improving dialogue success and quality. With the knowledge that entrainment on pitch and voice quality is more likely to occur when rapport is present, this suggests systems which can detect entrainment or a lack of entrainment on these features may be able to provide support and interventions more effectively.

Future work focusing specifically on human-human entrainment should include developing a comparison of entrainment measuring approaches to verify how modeling entrainment as a dynamic versus a linear phenomenon contributes to theoretical understandings of entrainment and its relationship to learning, rapport, and the collaborative process. Future work should also incorporate the abundant information which is available at the dyadic level, including accounting for the differences in acoustic-prosodic entrainment which appear within and across dyads.

PART II:

ENTRAINMENT IN A ROBOTIC LEARNING COMPANION

CHAPTER 8

MOTIVATION AND RESEARCH GOALS FOR PART II

We focus in Part II on human-computer entrainment; specifically, I explore how we can model entrainment in the dialogue systems of robotic learning companions. Part I revealed that entrainment on pitch and proximity appears to be more relevant to feelings of rapport; with these insights, I revisit the four research questions posed in the introduction:

RQ 1: How can acoustic-prosodic entrainment be modeled in a system to positively influence social responses?

RQ 2: How does automated entrainment influence rapport when combined with content-based approaches for building rapport?

RQ 3: How does entrainment influence learning in a robotic learning companion?

RQ 4: What insights regarding human-human and human-agent interactions can we gain by manipulating social behavior in a robotic learning companion?

I introduce a series of design iterations conducted to answer these questions. The iterations followed the methodology shown in Figure 8.1, and the very first iteration in this process was inspired by the results from Part I. Motivated and informed by analysis of human-human interaction, I identified several possible models or designs for entrainment. These models were iterated on at a micro-level to identify a design appropriate for a human-robot interaction. The final design emerging from that process was then evaluated in a larger human-robot interaction study with mixed methods analyses. The outcome of that evaluation was then fed back into the process as input for a new iteration. Overall, three micro and three macro iterations were conducted, with three unique learning companions.

Each of these iterations is described in detail in the following chapters. Chapter 9 begins with an overview of related work on human-computer entrainment, learning companions, the development of dialogue systems, and the potential role individual differences might play when implementing social behaviors in a robot. Chapter 10 describes the system features that the three robotic companions had in common, including the dialogue system structure. The first micro and macro iterations using the robotic learning companion Quinn are described in Chapters 11 and 12. Chapter 13 describes enhancements made to the dialogue of the general system to better support learning and foster self-efficacy and rapport. The second macro iteration is described in Chapter 14 with details on Nico, a teachable robot for middle school mathematics. The final iterations, performed with the learning companion Emma are summarized in Chapters 15 and 16.
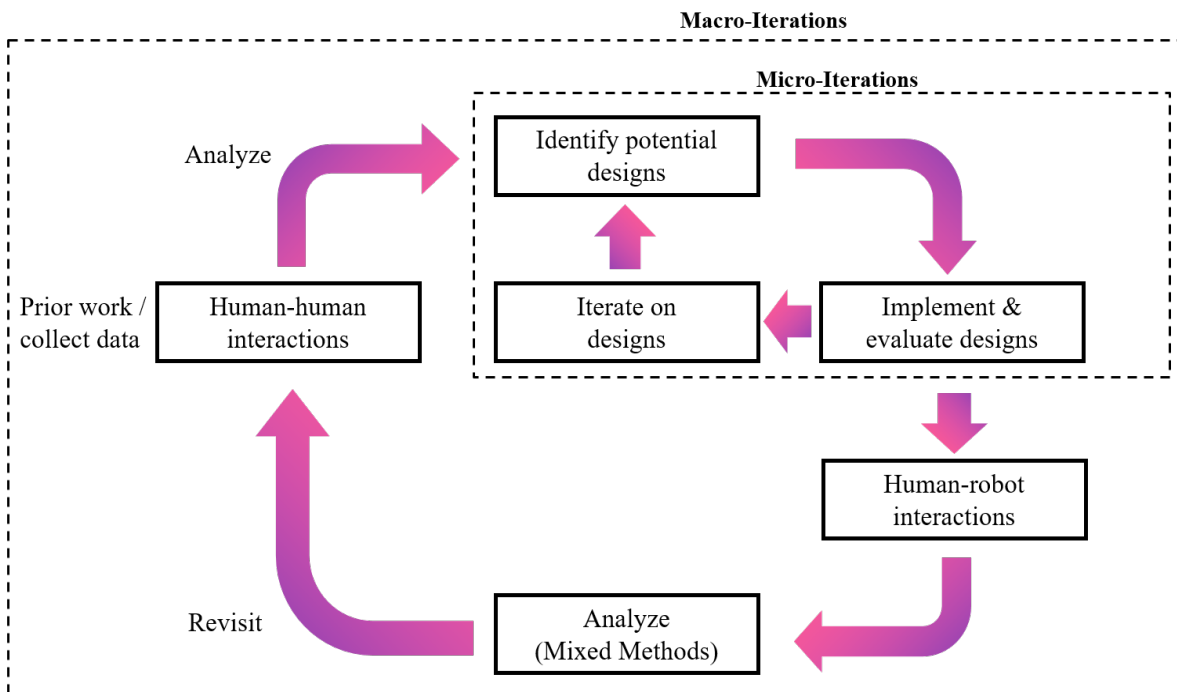
Figure 8.1. Overall Research Methodology

CHAPTER 9

RELATED WORK

Related work dealing with human-computer entrainment has primarily focused on how people will entrain to a computer (Coulston, Oviatt, & Darves, 2002), though it has been found that individuals prefer computer voices which are more like their own (Nass and Brave, 2005). Explorations of how a computer can entrain to a person are still in the early stages, and I describe this early work in the next section. In Section 9.2, I detail existing work on learning companions and how this informed the design of an entraining robotic learning companion. In Section 9.3, I touch on the potential role individual differences might play when implementing such conversational, social behaviors in a robotic learning companion. Finally, section 9.4 describes current work on the development of dialogue systems and provides baseline insight into the design of the entraining dialogue system built for this work.

9.1   ENTRAINMENT IN HUMAN-COMPUTER INTERACTIONS

Explorations of automated entrainment are still in the early stages. Only two systems to date have explored implementing acoustic-prosodic entrainment in an agent or robot. Sadoughi and colleagues (2017) built a system for a social, human-like robot which adapts on a turn-by-turn basis to a child's pitch and intensity. In their approach, they utilized a Bayesian network to select the verbal response with the most appropriate prosodic manipulation at run-time. This means the entrainment was somewhat restricted because they were selecting a pre-recorded audio clip which was closest to ideal entrainment. In

playing a game with the robot, children who interacted with the entraining robot had higher levels of engagement. Sadoughi and colleagues did not explore the effects of manipulating the prosodic features real-time and it remains unclear whether pitch or intensity or both resulted in the positive effects on engagement.

Levitan and colleagues (2016) explored the effects of adapting intensity and speaking rate in a turn-by-turn manipulation on perceptions of a virtual agent's likability and reliability. In pilot evaluations, they found positive effects for English speakers and their approach validates real-time adaptations. They did not explore the effects of real-time pitch adaptation or adaptation on multiple features at once. Neither Sadoughi and colleagues nor Levitan and colleagues explored the mediating effects of gender, effects on rapport, or how entrainment combines with social dialogue to influence social responses.

Outside of acoustic-prosodic entrainment, Lopes, Eskenazi, and Transcoso (2012) proposed a form of lexical entrainment in a spoken dialogue system to improve task success. The system adapted to the user's lexical choices; if the user's choices appeared to be degrading performance than the system would propose words for the user to adopt. System performance was improved, and error rates were reduced by 10%. I focus solely on prosodic entrainment in this work, leaving lexical entrainment to future studies.

9.2   ROBOTIC LEARNING COMPANIONS

In this work, automated entrainment is implemented in a robotic learning companion. Learning companions are based on theory that learning is social (Vygotsky, 1979) and provide both socio-motivational support and cognitive support. Learning companions come in several forms; in this work, I explore the learning companion as teachable robotic

agent, where the learner teaches the robot about a domain. By teaching, learners may attend more to the problem, reflect on their own misconceptions when correcting errors, and elaborate on their knowledge as they construct explanations (Roscoe & Chi, 2007), leading to learning. Teachable agents have demonstrated success in influencing learning (Leelawong & Biswas, 2008; Pareto et al., 2011), and teachable robots have demonstrated similar positive effects (Tanaka & Matsuzoe, 2012; Hood, Lemaignan, & Dillenbourg, 2015). Indeed, due to their physical presence and rich channels of communication, robots have under some circumstances socially engaged users more than agents (Liu et al., 2013), and this may be the case with teachable agents as well.

It has been hypothesized that there is a social component to the success of teachable agents in influencing learning. Some research has shown that when learners feel rapport or a sense of closeness for their teachable agent (Ogan et al., 2012a) they are more likely to benefit. Others have demonstrated what is called the protégé effect; that learners at once feel more responsible for their agent, are more motivated to learn for their agent, and believe the onus of failure belongs to the agent, easing the negative repercussions of failure (Chase et al., 2009). When learners feel more responsible for the agent (Biswas et al., 2010), they benefit more from teaching the agent. All these social factors may be enhanced by learners' feelings of rapport, and thus it is likely that within a teachable agent context, greater feelings of rapport may facilitate learning.

A popular way of enabling social robots and agents to build rapport with users is through rapport-building behaviors or social behaviors which support social connection; for example, a gesture which conveys 'friendliness' such as waving when one says hello, facial expressions such as smiling, or dialogue such as politeness. These behaviors have

been shown to increase rapport and learning when used by robotic tutors or robots that can teach students. Kanda and colleagues (2004) conducted a two-month trial in an elementary school with a social robot for learning English. The robot, called Robovie, could express various social behaviors, such as calling children by name. The social behaviors engaged the students and students who interacted with Robovie longer learned more.

Saerbeck and colleagues (2010) investigated how a socially supportive robotic tutor (iCat) influenced the task of language learning. iCat exhibited a variety of rapport-building, social behaviors which were both verbal and non-verbal, including dialogue, gaze, and facial expressions; they found students had higher learning performance with the socially supportive tutor. Westlund and colleagues (2015) introduced Tega, an affect-sensitive robotic tutor which pre-school children interacted with to learn vocabulary. Tega demonstrated that adaptation can increase positive valence (Gorden et al., 2016). In one of the few explorations of prosodic manipulations, Tega was also used to explore engaging preschoolers in active reading (Westlund & Breazeal, 2015). In this exploration, the robot's voice was either expressive, including a wide range of intonation and emotion, or flat, like a classic TTS engine. Their findings suggested an expressive robot is more beneficial; the expressive robot resulted in more concentration and engagement.

Social dialogue has been extensively explored as a rapport-building behavior with virtual agents. Bickmore & Cassell (2001) demonstrated that social dialogue such as small talk can have rapport-building effects, significantly enhancing feelings of trust in interactions with a virtual real estate agent. In later work, they also found that removing nonverbal cues available through facial expression and gesture negatively influenced the effects of social dialogue, and that individual differences indicated by personality played a

role in these effects, with individuals preferring and trusting an ECA which matched their own personality more (Bickmore & Cassell 2005). Gulz, Haake, and Silvervarg (2011) demonstrated that students who interacted with a teachable agent which engaged in social dialogue in the form of 'off-task' dialogue reported having a more positive experience and learned more. In work with a virtual agent tutor for multi-party dialogues, Kumar and colleagues (2010) showed that a tutor which uses social dialogue to show solidarity, trigger tension release, and exhibit an agreeing attitude can significantly influence learning. Bickmore, Vardoulakis, and Schulman (2013) showed that virtual agents which exhibited solidarity through common ground and self-disclosure improved engagement.

Rapport building behaviors like social dialogue and prosodic cues have been less explored with teachable robots. The results found with other forms of agents and robots provide support that both social dialogue and prosodic entrainment will enhance learning and rapport. For this work, I design social dialogue in line with this prior work. I use some of these same social dialogue behaviors as a baseline in the exploration of how entrainment combines with social dialogue to foster rapport.

Overall, the existing work on teachable robots is small. Co-Writer, is a Nao robot that learners teach about handwriting. The focus of Co-Writer has been on mastery experiences through adaptation of the robot's learning behavior. Studies have shown the robot can engage learners in the task and potentially promote motivation and self-confidence (Jacq et al., 2016). Tanaka and Matsuzoe explored a Nao robot that learners can teach about vocabulary through physical demonstration (Tanaka and Matsuzoe, 2012). The interaction showed teaching a robot can support learning and children are just as likely to teach verbally and by gesture, suggesting verbal and gestural communication is natural and

intuitive for teaching robots. rTAG, a Lego Mindstorms-based robotic learning environment where learners teach coordinate geometry (Walker et al., 2016), explored how physical embodiment affects social engagement; while social engagement increased in low prior knowledge learners, learning gains decreased. Collectively, this work indicates that teachable robots have potential to support learning, but the role of social factors remains unclear. In rTAG, physical embodiment alone was not enough to foster social engagement, while mastery experiences in Co-Writer was. I build on this prior work by introducing several versions of teachable robots throughout this thesis and exploring how social behaviors influence social engagement as well as learning.

9.3 INDIVIDUAL DIFFERENCES

Explorations of gendered responses in the human-robot literature are limited; as of 2014, only 21 of 190 HRI papers published from 2006 to 2013 provided any form of gender-based analysis (Wang and Young, 2014). However, there is evidence which suggests males and females might respond differently to rapport-building behaviors from a robot. Strait and colleagues (2016) found females were more sensitive to verbal communication while males were more sensitive to multiple behaviors and consistency.

In prior work on gender differences in human-agent interactions, females tended to respond more positively to social behavior from virtual agents, while males tended to respond negatively (Burleson and Piccard 2007; Vail et al. 2015; Arroyo et al. 2013; Lutfi et al. 2013; Jokinen and Hurtig 2006; Kramer 2016). Burleson and Picard introduced a multimodal, real-time affective agent which exhibited emotional intelligence in response to a user's affect. The agent's behaviors included speaking, nodding, smiling, fidgeting,

and shifting its posture forward or backward; these behaviors were adapted to mirror the user and to give evidence of 'active' listening. In analyzing responses from 76 girls and boys aged 11 to 13, girls responded much more positively to the affective tutor, expressing a stronger social bond, persevering longer, and exhibiting higher gains in meta-affective skills. Boys responded more positively on these measures to the task support only tutor. In other work, Vail and colleagues (Vail et al., 2015) explored gender responses to an agent which exhibited cognitive and affective support through verbal feedback; females reported significantly more engagement and less frustration with an agent which exhibited motivational and engaging support. Arroyo and colleagues (Arroyo et al., 2013) supported these findings with an extensive analysis of an affective pedagogical agent deployed in several public schools; female students had significantly lower frustration, excitement, self-efficacy in mathematics, and liking of mathematics when interacting with the affective agent. These agents were generally designed to exhibit rapport affectively through dialogue and physical gesture.

Given this prior work, it is possible females may respond with more rapport to the social, entraining robot than males. It is also possible males and females will differ in their use of behavioral rapport. As described in Chapter 2, behavioral rapport will be measured as linguistic politeness and males and females have been found to differ in how they exhibit politeness. Women have been found to be politer in general, often being more likely to give praise and engage in formal politeness (Chalupnik 2017, Brown 1980). Empirical studies by Holmes (1995), Coates (2015), Tannen (1994), and Hong (2012) point to women using conversation to establish, nurture, and develop relationships while men are more likely to see conversation as a tool for obtaining and conveying information. It has been suggested

that differences in linguistic strategies may be a result of an individual's experience and their role in the conversation as either a peer, expert, or sub-ordinate. If differences in linguistic strategies between males and females are observed, this could be due to males and females using politeness for different purposes (i.e. as a rapport builder vs. information conveyer) or it could be indicative of differences in their interpretation of their role as a peer versus an expert.

9.4   BACKGROUND ON DIALOGUE SYSTEMS

Dialogue systems, or systems which enable users to have a conversation with a computer, have entered mainstream society as digital assistants on cellphones and home controllers and as conversational agents on the web and in call centers. In the work, the dialogue system plays a crucial role as spoken language is the main form of interaction, the source of the social manipulation, and the dialogue system is the underlying technology which enables us to explore the research questions. This section contains an overview of background work on dialogue systems and the different ways systems are typically designed, providing an informative baseline for the design of the dialogue system in this work. This section also contains a description of the other two dialogue systems to have incorporated some form of entrainment.

9.4.1   General Dialogue System Design

Dialogue systems are typically composed of several modules, including an automatic speech recognition (ASR) module, which detects the user's speech and translates it to text, a dialogue manager to identify an appropriate response, and a text-to-speech module to

take the response and convert it to speech. The increasing popularity of these systems as a practical medium for human-computer interaction has largely been due to dramatic improvements in ASR over the past decade as deep learning approaches have reduced ASR errors. Even before these improvements to speech recognition, two common classes of dialogue systems emerged: task-oriented dialogue systems and chatbot systems. Task-oriented dialogue systems are designed for a task within a restricted domain and are suitable for short interactions with a goal-based focus. For example, searching for a restaurant, getting directions to a location, or making a reservation. Chatbots are designed for extended conversations (Jurafsky and Martin, 2018). Chatbot dialogue managers have been in existence since the 1960s, when Weizenbaum introduced ELIZA, a chatbot designed to simulate a Rogerian psychologist (Weizenbaum, 1966). Given the multi-turn nature of the teachable robot system I am interested in building and the social nature of chatbot designs, I base the dialogue system on a chatbot framework.

9.4.2   Chatbot Systems

Chatbot systems, first introduced in the 1960's, have increasingly been applied to practical applications within education, information retrieval, business, and e-commerce (Shawar & Atwell 2007). As noted by Gulz and colleagues, chatbot dialogue systems for educational applications like intelligent tutoring systems and teachable agents can be beneficial because they enable the combination of elements from task-oriented dialogue in a restricted domain with elements from the broader but shallower dialogues chatbots are known for and are ideal at producing (Gulz et al., 2011). This combination enables a system to produce social

dialogue opportunities while still maintaining domain knowledge representation with acceptable dialogue responses.

Chatbot systems fall into two classes: rule-based systems and corpus-based systems. Rule-based systems include the well-known ELIZA and PARRY as well the more recent ALICE chatbot, which makes use of Artificial Intelligence Markup Language (AIML). Rule-based systems take the user's utterance and identify a response given a set of rules. Corpus-based systems require existing conversational human-human data. The human-human data is used to identify suitable system responses either by information-retrieval algorithms or the corpus can be used to generate a mapping from user utterances to system responses via machine translation techniques.

Rule-based systems make use of pattern/transform rules where by a user's utterance can be mapped or transformed to generate the system's response. Each pattern or rule is linked to a keyword in the user's utterance; keywords possess a rank, with more specific words having a higher ranking. Responses are identified based on the keywords found in the user's utterance. Generic or non-committal responses exist in cases where no keywords can be found in the user's utterance. Both ELIZA and PARRY were developed using a rule-based chatbot design and were extremely successfully in deeply engaging individuals and passing early versions of the Turing test (Colby et al., 1972). ALICE, a generic chatbot with source code openly available, has been successfully applied to several domains (ALICE, 2002). I utilized a rule-based chatbot system using AIML in this work.

Both rule-based and corpus-based chatbots are susceptible to errors; these errors may originate either in the speech recognition module or in the design of responses. For example, with rule-based chatbots, ASR errors may result in keywords that cannot be

correctly identified while response design errors may occur when potential utterances or keywords are missed or incorrectly identified. Despite the susceptibility to these types of errors, chatbots are still successful, largely because individuals have been shown to accept a number of errors from a dialogue system and still respond socially if the system's responses are generated with an appropriate degree of interpretability. Social chatbot agents can still elicit disclosure, build rapport, increase trust, and improve learning, all while experiencing some level of ASR errors (Turkle et al. 2006; Huang et al. 2011; Levitan et al. 2016; Forbes-Riley & Litman 2005). Dialogue systems for the tutoring domain have shown that perfect automatic speech recognition and natural language understanding are not a requirement for functional, effective systems. For example, users interacting with ITSPOKE, a spoken dialogue system based on the Why2-Atlas tutorial dialogue system, could experience a certain degree of ASR failure without correlation or effect on learning gains (Litman & Forbes-Riley 2005, Litman & Silliman, 2004). D'Mello, Graesser, and King (2010) explored to what extent ASR errors affected learning gains with AutoTutor, an intelligent tutoring system for computer literacy. Comparing a speech-based version to a text-based version, they found that there were no significant differences on learning gains across modalities up to a word error rate of 0.46. D'Mello and colleagues identified the fact that performance did not degrade considering speech recognition errors as an indicator of the robustness of AutoTutor's natural language processing capabilities. AutoTutor uses multiple modes of input to identify and produce responses within an appropriate degree of interpretability.

This idea that individuals can still respond positively to systems which generate responses based on flawed input or flawed processes was originally put forth by

Weizenbaum, the creator of ELIZA. He proposed that speakers will make assumptions about their interaction partner. If their partner's responses are in line with those assumptions, the speaker's image of their partner remains unchanged, undamaged. In the case that a response is difficult to interpret, this does not mean a speaker's image must change. Rather a speaker may rationalize the response to arrive at complicated interpretations which maintain the reasonableness of the response. If such rationalizations become too massive or self-contradictory, then the image will crumble and be replaced by another. If I am careful in how I design a system's responses when errors may be present, I may be able to stay within the boundaries of rationalized reasonable interpretations.

I considered the repercussions of ASR and response-design errors in introducing a phenomenon like entrainment to a spoken dialogue system. I know from prior work that social interventions can still produce social responses as well as concrete outcomes in individuals despite system errors; this appears to be dependent on generating responses which fit within a reasonable rationalization of the system's capabilities. To facilitate the production of responses which could fit within a reasonable interpretation of the system's capabilities, I incorporated multiple modalities of input to help provide additional context when generating responses. This is described more in Chapter 10.2.

9.4.3 Dialogue Systems for Entrainment

To produce acoustic-prosodic entrainment in a dialogue system, two additional components are required beyond the typical dialogue system structure: (1) a module to extract the user's prosodic features and (2) a method for manipulating the system's text-to-speech output to entrain to the user's features. Extraction of the user's prosodic features is

easily accomplished with a variety of tools including Praat (Boersma, 2002) and OpenSmile (Eyben, Wollmer, & Schuller, 2010). For (2), manipulating a system's output, there are three possible approaches – using TTS provided manipulation tools like Speech Synthesis Markup Language (SSML) to augment the TTS output, transforming the properties of the TTS after it has been synthesized, or selecting pre-recorded audio responses which most closely match the acoustic-prosodic properties desired to simulate entrainment. Selecting pre-recorded audio is the most limited approach, restricting both the number of responses to those which can be pre-recorded as well as potentially limiting the extent to which a response can actually 'entrain' to the user. SSML and transforming the TTS after it has been synthesized are more flexible approaches.

As mentioned, there are only two systems which have incorporated methods of automatic entrainment based on the user's prosody and each took a different approach to adapting the system's output. Levitan and colleagues utilized Speech Synthesis Markup Language or SSML (Levitan et al., 2016). After extracting the user's prosody, the planned TTS output was augmented with the SSML markup tags identified by an 'entrainment' module. In contrast, Sadoughi and colleagues took the approach of pre-recording audio and selecting the audio clip with prosodic features predicted to be the most appropriate given the user's prosody (Sadoughi et al., 2017). They trained a dynamic Bayesian network (DBN) using a human-human corpus; the model identified by the network was then used to identify appropriate 'entraining' responses from a large pool of pre-recorded audio. For producing acoustic-prosodic entrainment with a teachable robot, I utilized two approaches. I explored transforming the TTS after it has been synthesized and I introduced algorithms which can be applied to systems with options to augment the TTS as it is generated.

CHAPTER 10

THE ROBOTIC LEARNING COMPANION

The dialogue system and interaction process of the robotic learning remained consistent across the six design iterations and learning companion versions. This chapter, utilizing the background knowledge on dialogue systems provided in Chapter 9, presents the general dialogue system and interaction process; any differences due to the three learning companion versions are summarized.

10.1 GENERAL SYSTEM

Learners interacted with the learning companion using spoken language and a touch-screen interface on a tablet computer (Microsoft Surface Pro). The touch-screen interface displayed each math problem, presenting each problem separately. The problems were additionally broken down into solvable steps. The three learning companions did contain slightly different domain content. Given these differences in domain content, the user interfaces were visually different. For Quinn, the mathematical content was based on literal equations, for Nico and Emma the content focused on ratios, proportions, and unit rates. Visuals of the interfaces are shown in Figure 10.1.

For all companions, the tablet interface supported speech recognition and displayed visual progress to the learner as they taught the companion. For Quinn this consisted of a progress bar shown at the bottom of the screen. For Nico and Emma, the current step was highlighted and enlarged on the screen. When the companion 'answered' a step, the corresponding table cell was updated from question marks (see Figure 10.1) to the correct

answer. Learners were encouraged to move through the problems at their own pace using buttons on the UI to advance forward.

Problem One: $2x + 7y = 5.$ **Solve for** $x.$

Step One: $2x + 7y - 7y = 5 - 7y$

Step Two: $\dfrac{2x}{2} = \dfrac{5 - 7y}{2}$

Step Three: $x = \dfrac{5 - 7y}{2}$

Hold microphone
down to talk to
Quinn

Prior Problem

Next Problem

User Interface Design for Quinn

Problem 1

Nico wants to go swimming with friends at the pool! Sadly though Nico's body isn't waterproof so Nico needs to prepare first. The plan is to use waterproof paint to protect Nico's body but Nico isn't sure how much waterproof paint covers three square inches. Help Nico figure out how much waterproof paint is needed using the table below

Touch and hold the image of Nico
to talk to Nico.

Step	Body Part	Surface Area (sq. inches)	Volume of Paint (fluid oz)
Step 0	Feet	6	2
Step 1	Legs	12	???
Step 2	Torso	???	6

Next Step

User Interface Design for Nico and Emma

Figure 10.1. User Interfaces for the Three Learning Companions

67

To speak to the learning companion, the learner pressed and held a button on the interface while they spoke. The speech interaction was real-time, and the dialogue was recorded as the student spoke. After speaking and explaining a step, learners were instructed to pause, giving the companion a chance to respond. A gif depicting that the robot was "thinking" would appear on the screen to indicate that the companion was occupied. During this period, the system would process the input and generate a response. The average response time for all companions was less than four seconds.

10.2   DIALOGUE SYSTEM

The dialogue system developed was capable of both entrainment and social dialogue for the purposes of exploring rapport. The overall structure follows that of typical dialogue systems as summarized in Chapter 9. The user's speech was recorded via a microphone on the tablet interface and once they were done speaking, the user's audio was passed into the dialogue system which consisted of four main modules: (1) an automatic speech recognizer, (2) a dialogue manager, (3) an acoustic-prosodic feature extractor, and (4) a module for prosodic manipulation and text-to-speech generation.

10.2.1 System Overview

For the **automatic speech recognizer**, I utilized the HTML5 Speech API available in Chrome[1]. Overall, across all three robotic learning companion implementations, I found the word error rate on average was 23.4%.

---

[1] https://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API

The **dialogue manager** consisted of several components including a module for basic functionality and several modules with additional functionality to enhance responses. All three implementations of the learning companion incorporated the basic functionality but some of the supplementary modules are found only in Nico and Emma. I describe the design of the dialogue manager in more detail in 10.2.2.

For the **acoustic-prosodic feature extraction**, the system could extract three acoustic-prosodic features: pitch, intensity, and speaking rate. I utilized Praat for extracting all three features. I extracted pitch using Praat's pitch estimation algorithm which performs acoustic periodicity detection based on autocorrelation (Boersma 2006). Minimum and maximum fundamental frequencies for pitch estimation were set based on the gender of the speaker (i.e. for males, 75 Hz – 300 Hz and for females 100 Hz – 500 Hz). Intensity was extracted using Praat's Intensity Contour algorithm, the mean intensity was calculated across the extracted contour, resulting in a value for intensity as Sound Pressure Level (SPL) in dB. For speaking rate, I needed a real-time estimation of an individual's speaking rate. I employed de Jong and Wempe's (2009) approach to extract speaking rate by automatically detecting the syllables in a user's audio and estimating the speaking rate based on syllables per second.

Finally, the **prosodic manipulation and text-to-speech generation** was unique to each iteration of learning companion. I leave the descriptions of that module to the following chapters. The overall dialogue system infrastructure can be seen in Figure 10.2.

10.2.2 Dialogue Manager: Basic Functionality

I utilized a rule-based chatbot system with the AIML framework, making use of the
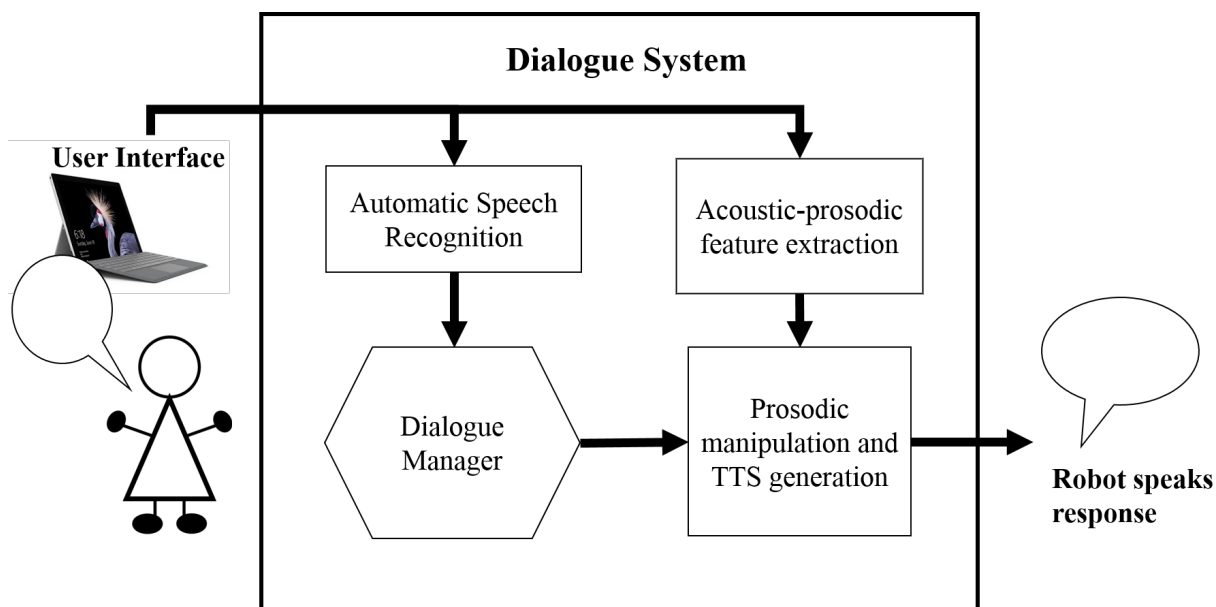
Figure 10.2. Overview of Entraining Dialogue System Structure.

PandoraBots tool for much of the implementation of the AIML (Wallace 2003). I was inspired by Gulz and colleagues (2011) who demonstrated the potential of chatbot frameworks for learning applications. The AIML framework implements a rule-based process of linking keywords to pattern/transform rules as found in seminal chatbot agents including ELIZA, PARRY, and ALICE. I utilized this process to develop responses suited to the domain content of each of the learning companions. What these means is that for each companion, I identified potential keywords in the learners' utterances which could suggest suitable responses on the part of the robot, and I designed rules and transforms using these keywords to then generate the robot's responses. For example, if a learner is trying to teach the system how to solve a literal equation such as:

$$2x + 7y = 5. \quad Solve \; for \; x,$$

70

The learner might begin by explaining, "You need to subtract 7y from each side." In this utterance, I might identify 'subtract' as a keyword and this utterance would then match the following rule/transform:

$$(* \; subtract \; *) \rightarrow$$

$$(\; Okay \; we \; subtract! \; Can \; you \; explain \; a \; little \; bit \; more \; about \; why \; we \; subtract? \;)$$

The system would then issue the response, "Okay I subtract! Can you explain a little bit more about why I subtract?" the simplest design mapped keywords back into responses that followed from the limited domain of the problem set and tended to be content-less rather than content-full. I then introduced more complex designs where possible to capture more explicit content. For example, an additional keyword mapping could capture that the learner said to "subtract $7y$". The rule/transform would be:

$$(* \; subtract \; 7y \; *) \rightarrow$$

$$(\; Okay \; we \; subtract \; 7y! \; Can \; you \; explain \; a \; little \; bit \; more \; about \; why \; we \; subtract \; 7y? \;)$$

The system would issue the response "Okay I subtract 7y! Can you explain a little bit more about why I subtract 7y?" This content-full response would be given a higher priority based on the keyword rank for "subtract $7y$". All keywords were given a rank; a higher rank increased the likelihood of a keyword being matched.

To reduce the effects of ASR errors and enable more content-full responses, I incorporated additional information from the tablet interface that learners used to interact with the system. This information included the current problem and step. I then considered each individual problem-step combination as a separate 'topic' with unique keywords, phrases, and associated pattern/transform rules. The learning companion would initiate the dialogue whenever a new problem or a new step was started. This would set the 'topic' to

that problem and step. Keywords belonging to the current problem and step were given the highest rank; general keywords and phrases not tied to the current problem and step were ranked lower and were therefore less likely to be matched first. If a student's speech could not be matched to a specific keyword, a response was selected from a set of 'generic' utterances. This set contained two types of responses: requests for clarification (i.e. "can you please repeat that?"), and general acknowledgements (i.e. "ok sounds good").

Within a problem-step, certain keywords when matched could initiate short two to three turn dialogue trees where the system would then listen for keywords associated with the system's prior utterances. An example of a short dialogue with dialogue tree is given in Figure 10.3. The dialogue is sample dialogue based on the first robotic learning companion I explored, Quinn. In the figure, Quinn initiated the dialogue at the start of the step. The learner then began an explanation telling Quinn to subtract. This initiated a dialogue tree based on the keyword subtraction. Keywords and phrases corresponding to subtraction and Quinn's prior utterances were then given the highest priority.

Finally, the ability to include social dialogue was also a part of the basic functionality. To include social dialogue, I modified the rules/transforms to include social content in addition to the domain-based content. The type of social dialogue included was dependent on the version of learning companion. I describe the design of the social dialogue more in each section on Quinn, Nico, and Emma. As an example, the social dialogue implementation for Quinn included positivity or being cheerful. With the keyword "subtract 7y" the social dialogue response that would trigger for Quinn would be:

$$(* \; subtract \; 7y \; *) \rightarrow$$

$(That \; sounds \; awesome! \; Can \; you \; explain \; a \; little \; bit \; more \; about \; why \; we \; subtract \; 7y? \;)$

10.2.3 Dialogue Manager: Supplementary Modules

All three robotic learning companions employed the basic dialogue manager functionality described in 10.2.2 to generate responses and drive dialogue. For two of the companions, Nico and Emma, I introduced several additional modules with advanced functionality to help improve system responses. A high-level architecture that summarizes how responses were matched to the user dialogue and including these additional modules is given in Figure 10.4. I describe the functionality of each these modules in more detail below.

System: Topic set to Problem One - Step One. Quinn initiates dialogue…

Quinn

So what do we do for step one of this first problem?

Student

*You need to **subtract** 7y from both sides*

(* subtract *)

Quinn

Okay we subtract! Can you explain a little bit more about why we subtract?

Student

| *Because it is positive and you need to move it to the other side* | *Because we need to get x by itself* | *Because that's what we do* | *… [other explanations]…* | *… no matched explanation…* |

(* positive * side) (* x *) (* what we do *) (* keyword *) (_)

Quinn

| *Ah okay! I get it. So we subtract because the factor is positive and we need to move it to other side.* | *Ah okay! I get it. So we subtract because we need to get x by itself.* | *Ah okay! I get it. So we subtract because that's what we do to get x by itself?* | *… [response based on explanation]…* | *Ah okay! So I'm thinking. We can subtract because we need to get x by itself?.* |

Figure 10.3. Sample Dialogue Tree Initiated by a Student's Keyword to Subtract

73

Advanced Explanation Paths

As described earlier, short dialogue trees could be initiated based on the keywords in a student's dialogue. In the basic functionality, these trees were typically designed for two to three turns and were kicked off by common-use keywords with typically one tree to one keyword. In this advanced module, I expanded on these trees given the domain content for a specific learning companion. I designed the dialogue trees based on specific explanation paths. For example, to solve a ratio problem, some students utilize a counting approach rather than setting up a ratio. They might explain to the learning companion a counting approach. I would use their dialogue to identify that they are using a 'counting' explanation approach and kick off a dialogue tree which is based on any of the keywords associated with that explanation path. Table 10.1 gives an example of the potential keywords and an alternative explanation approach for a ratio-based problem. Using this approach, I designed the dialogue such that the learning companion would 'reach' an understanding of each step and problem at the end of each 'explanation path' tree.

Prompt for Speech

In the design, I wanted the companion to exhibit awareness and I wanted the students to feel as if the robot was engaged in the interaction. To design the dialogue to help encourage this perception, I implemented additional functionality in which the robot would prompt the student to engage with them if the student had ceased interacting with it for two to three minutes. A random timer was set for between two and three minutes and if no interaction was detected, the timer would trigger the robot to initiate dialogue. For example, the robot might say "You've been quite for a while. Are you thinking?"

```
function DialogueManager(user ▯▯▯▯▯▯▯▯, current ▯▯▯▯▯▯ ▯▯▯▯) returns ▯▯▯▯▯▯▯

▯▯▯▯▯▯▯▯▯▯▯←—Pop the current explanation
▯▯▯▯▯←—Pop the last system utterance

Find the word/phrase ▯ in ▯▯▯▯▯▯▯▯ that has the highest keyword rank in
▯▯▯▯▯▯ ▯▯▯▯ given ▯▯▯▯

        if ▯ exists and ▯▯▯▯▯▯▯▯▯▯ is NOT none
                Choose the highest ranked rule ▯ for ▯ that matches ▯▯▯▯▯▯▯▯ in
                ▯▯▯▯▯▯▯▯▯▯
                ▯▯▯▯▯▯▯▯ ← apply the pattern/transform in ▯ to ▯▯▯▯▯▯▯▯
        else if ▯ exists and ▯▯▯▯▯▯▯▯▯▯ is NONE
                Choose the highest ranked rule ▯ for ▯ that matches ▯▯▯▯▯▯▯▯
                ▯▯▯▯▯▯▯ ← apply the pattern/transform in ▯ to ▯▯▯▯▯▯▯▯
                if ▯ corresponds to an explanation approach:
                        ▯▯▯▯▯▯▯▯▯▯ ← explanation approach
                        Push ▯▯▯▯▯▯▯▯▯▯ onto memory stack
        else if ▯▯▯▯▯▯▯▯▯▯ is NOT none but no keyword applies
                Choose random rule ▯ in ▯▯▯▯▯▯▯▯▯ given ▯▯▯▯
                ▯▯▯▯▯▯▯ ← apply the pattern/transform in ▯ to ▯▯▯▯▯▯▯▯
        else (no keyword applies)
                increment count of NONE responses
                if count of NONE responses > rand(3, 4)
                        ▯▯▯▯▯▯▯ ← apply pattern/transform rule from random
                                explanation approach
                        ▯▯▯▯▯▯▯▯▯▯ ← explanation approach
                        count of NONE responses← 0
                else
                        ▯▯▯▯▯▯▯←—Apply the transform for the NONE keyword to
                        sentence

return (▯▯▯▯▯▯▯)
```

Figure 10.4. Dialogue Manager Architecture with Advanced Functionality

Trigger Explanation Path

It was possible that the dialogue system would not be able to match or detect keywords,

particularly if the speech recognition for a given individual was particularly flawed. To

facilitate the flow of dialogue and avoid such issues, I designed additional functionality to

Explanation	Initial Keywords / Phrases	Dialogue Tree
Utilizing ratios	"ratio", "relationship between", "looking at how * related" …	
Utilizing counting	"count", "add up", "add together", " * plus *",…	
[Other explanations…]	[keywords / phrases…]	[different trees…]

Table 10.1. Possible Explanation Approaches for Solving a Ratio Problem

detect if a learner's speech was continuously unmatched to specific keywords for three to four turns. When this occurred, the system would initiate an explanation path for the current problem-step and trigger a dialogue tree to begin; with the dialogue tree triggered, the tree could be used to provide additional context to facilitate response identification and ensure that the companion would 'reach' an answer.

Keyword Normalization

Another modification I added to enhance the system's responses was to utilize a method of normalization to transform words that the ASR would consistently get incorrect to keywords I could identify. For example, "multiply" was occasionally recognized as "make a tie;" "surface area" was occasionally recognized as "service area." These normalizations helped to improve the generation of responses.

Gesture Generation

This module enhanced the dialogue by incorporating gesture. Gestures were managed through the dialogue manager because they were tied to the verbal responses. After

identifying an appropriate verbal response, the gesture generator would identify whether there was a corresponding gesture that matched the verbal response. There were eight emblematic or easily recognizable gestures and they included waving 'hello,' nodding head as in 'yes,' shaking head as in 'no,' putting hands on hip to make a point, raising either hand, raising hands in celebration, and shrugging. The system identified an appropriate gesture based on the content of Nico's utterance, timing the behavior to that utterance. For example, if Nico said "Hello! How are you?" the gesture 'wave' would be identified, and it would occur during the verbalization of "Hello." For Nico and Emma, I also enabled "autonomous life", a default capability which introduces a slight, swaying movement and listening behavior to indicate engagement and awareness.

CHAPTER 11

DESIGNING PITCH PROXIMITY WITH QUINN

This chapter introduces the first iteration on designing automated entrainment. Motivated by the findings in Part I, several initial models for entrainment were identified based on local proximity, a form of turn-by-turn matching. I focused on a single feature for local proximity—pitch. Local proximity on pitch prominently differentiated communicative success (Borrie, Lubold, and Pon-Barry, 2015), entrainment measures derived from pitch features were significantly higher in positive interactions (Lee et al., 2010) and entrainment on pitch was significantly related to learning (Thomason, Nguyen, & Litman, 2013). There are many ways in which local proximity on pitch might be modeled. This chapter describes the exploration of several possible designs. Designs were evaluated based on two criteria: perceived naturalness and perceived rapport.

I collected data from four individuals interacting with different entrainment designs and using crowd-sourced analysis via Amazon Mechanical Turk, compared the different adaptations regarding rapport and naturalness as perceived by third-party observers. In the next section of this chapter, I describe the companion, Quinn, in which the entrainment designs were implemented. Section 11.2 contains the descriptions of the adaptations. The method and procedure for analyzing the adaptations are in 11.3 and the results are in 11.4.

11.1   QUINN: A SOCIAL LEARNING COMPANION

I designed and built Quinn, a social learning companion. For this first iteration, Quinn is a virtual teachable agent for literal equations and is present throughout the interaction on a

Windows Surface Pro tablet. Quinn is not implemented in the robotic form as the focus of this iteration was on the design of prosodic entrainment as described in Sections 11.2.

Students interacted with Quinn using speech and the tablet interface on which Quinn was present; Quinn and the interface are shown in Figure 11.1. Quinn responded to the students with spoken language. The speech interaction was real-time, and the dialogue was recorded via microphone. Once a student was done speaking, their audio passed into the dialogue system described in Chapter 10. For Quinn, the dialogue system implementation included only the basic functionality as outlined in 10.2.2; the additional mechanisms for enhanced interaction including Advanced Explanation Paths, Prompt for Speech, Trigger Explanation Path, and Keyword Normalization were not included for Quinn and neither was the gesture generation. For further details on the system, see Chapter 10. I describe the domain content and specific interface/interaction design in the next section and the design of the social dialogue specific to Quinn in 11.1.2.

11.1.1 Domain Content and Interface Design

Students taught Quinn how to solve literal equations (i.e. "Solve $bx + gy = 14by + 6x$ for x"). The web application contained materials to guide the students in their teaching of Quinn with the worked-out solutions for each literal equation provided on the interface.
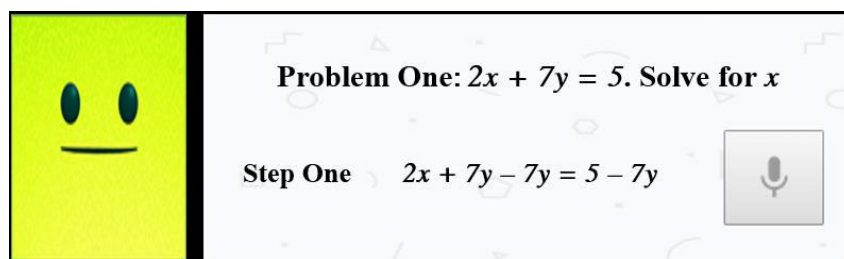
Figure 11.1. Quinn and a Sample Problem

79

Category	Description	Social Response	Non-social Response
Solidarity	*Compliments*	Ok so we add x. You're a really great teacher!	Ok so we add x. I get that we are adding x here.
Tension Release	*Being cheerful*	Ok so we add x. I'm so happy to be working with you	Ok so we add x. It makes sense that we would add x here.
	Off-topic	Ok so we add x. Do you like math?	Ok so we add x. I get adding.
Agreeing	*Comprehension*	I hear what you're saying. You're saying we add x.	We add x. It makes sense that we would add x here.

Table 11.1. Social and Non-Social Dialogue Examples for Quinn

Up to eight literal equation problems and quizzes were available in the application. The application presented one problem at a time and included the worked-out steps to reach a solution. The problems were ordered in increasing difficulty. New concepts were introduced every two problems; concepts included how to handle multi-step equations, re-arranging formulas, and factoring. Students walked Quinn through the worked-out problems using spoken language, explaining each step. Quinn responded using spoken language and had an animated facial expression when speaking, neutral otherwise. At the end of each problem, a follow-up quiz was provided. Students asked Quinn to solve the quiz, step by step. Quinn solved the quiz independent of the student. Figure 11.1 gives a sample problem as shown on the tablet interface.

11.1.2 Social Dialogue Design

Quinn was a social learning companion and had the ability to add social dialogue to responses. The social dialogue content was motivated by the social interaction strategy proposed by Bales (1950) and utilized by Kumar and colleagues with their virtual tutoring agent (2010). I chose this framework since Kumar and colleagues demonstrated that their designed social utterances had positive effects on collaborating students' communication

and because similar social behaviors were found to have positive effects by Gulz, Haake, and Silvervarg (2010) and Bickmore and colleagues (2013). Bales defined three main categories of positive socio-emotional behaviors: showing solidarity, showing tension release, and agreeing. Examples of social responses Quinn could give in each category are given in Table 11.1. Bales based his process on observations of group interactions; however, these responses and categories are also supported by human-human peer tutoring dialogue analysis which has shown that peer tutors can engage in behaviors which indicate solidarity (i.e. praise and encouragement, "come on, I can do this"), tension release (i.e. off-topic conversation such as "so what do you do for fun?") and agreeing (i.e. comprehension, "yes, okay, you are right") (Ogan et al., 2012; Bell et al., 2009). For this iteration, Quinn would issue a social response 15-20% of the time. This frequency mirrored results from analysis of human-human social responses in collaborative dialogues (Lubold 2013, Kumar 2010).

11.2   THREE METHODS OF PITCH PROXIMITY

In this section, I describe the three forms of pitch adaptation based on the form of entrainment known as local proximity and inspired by observations of how human conversation partners entrain. These pitch adaptations were implemented in the Prosodic Manipulation and TTS Generation module of the dialogue system mentioned in Chapter 10 and shown in Figure 10.1.

The three methods of pitch adaptation which operated at the turn-level to mimic local proximity are illustrated in Figure 11.2 along with a sample waveform. For all three

Waveform	control (no adaptation)	mirror partner	shift+contour	shift+flatten
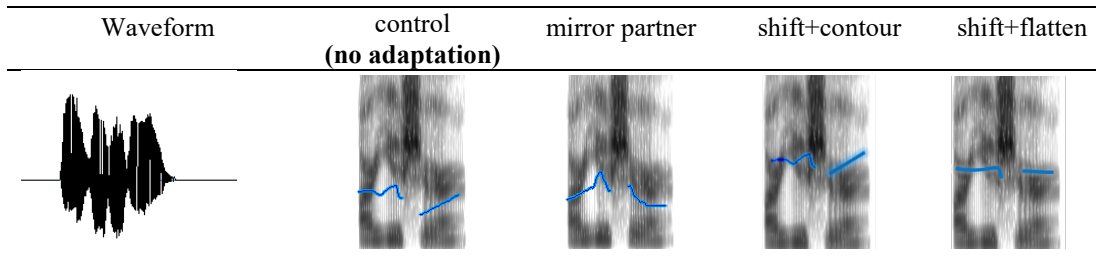
Figure 11.2. Spectrograms and Pitch Contours of Pitch-Adapted Waveforms

adaptations, the TTS output was first generated using the Microsoft Speech API. The gender of the TTS output was matched to the gender of the speaker; the female voice of "Zira" was used to adapt to female speakers and the male voice of "David" was used to adapt to male speakers. For all three methods, the system adapted its pitch based on the estimated pitch values from the human speaker's previous turn, as opposed to the longer dialogue history. The human speaker's pitch values were extracted during the acoustic-prosodic feature extraction phase described in Chapter 10.

Praat's pitch estimation algorithm based on autocorrelation was utilized to extract pitch features from the non-adapted TTS. One of the three adaptation approaches was then applied to the non-adapted TTS. Once modifications were identified and applied, the TTS was re-synthesized with Praat's version of Time-Domain Pitch-Synchronous Overlap-and-Add (TD-PSOLA). Re-synthesis with TD-PSOLA has the potential to introduce some distortion based on speaker characteristics (Longster 2003), which can lead to potential

	F1		F2	
	t	p	t	P
mirror partner	1.3	.19	1.4	.14
shift+contour	-.11	.91	.06	.95
shift+flatten	-.76	.44	-.71	.48

Table 11.2. Results Comparing Formant Values Before and After Adaptation

attenuation of some frequency values and reverberation. To identify if there would be issues regarding intelligibility for different female and male frequencies, I reviewed differences in the values of the vowel formants produced pre-adaptation and post-adaptation. Formants correspond to resonances in the vocal tract; vowels are identifiable based on formant ranges and there is a clear link between perceived vowel quality and the first two formant frequencies. For each adaptation, the resulting formants stayed consistently within expected ranges for intelligibility. Comparing pre-adaptation and post-adaptation values for each method using a paired t-test, there was not a significant difference between the formant values produced. The results of the t-tests are in Table 11.2.

The first method of pitch adaptation introduced was **mirror partner**. With mirror partner, the text to speech output was adapted to the entire pitch contour of the speaker's previous turn by replacing the original contour of the TTS with the contour of the speaker. To account and control for differences in utterance length, I resized the speaker's utterance to be the same length as the proposed text to speech output prior to applying the speaker's contour to the output. This approach to adaptation maximized the level of entrainment.

While mirroring the shape of a partner's pitch contour might strengthen automated measures of entrainment, there is the possibility of "over-adaptation" and of a mismatch between pitch contour and syntactic and semantic structure. **Shift+contour** was an alternative method of pitch adaptation that maintains the contour of the original TTS but shifts it up or down to match the mean pitch of the speaker. Since entrainment on pitch mean has been found to be correlated with learning and rapport, above and beyond other attributes of pitch, shift+contour only adapts the pitch mean. I perform this adaptation

shifting all the frequencies of the non-adapted TTS output by the difference between the mean pitch of the user and the mean pitch of the non-adapted TTS.

I introduce a third adaptation called **shift+flatten**. This adaptation serves as a minimum manipulation baseline in respect to the other two approaches. Still adapting on a single feature, pitch mean, I flatten the pitch contour of the TTS to the pitch mean of the user. The TTS output maps to the average pitch of the student. As this adaptation is intuitively the least realistic, it is expected to produce less rapport than the other two conditions. Thus, it serves as a baseline comparison for the more sophisticated pitch adaptations proposed, in addition to the control, the original synthesized waveform.

11.3   METHODOLOGY AND PROCEDURE

The three methods of pitch adaptation were evaluated with Quinn, the teachable agent described in 11.2. 32 dialogues were collected from four individuals as they interacted with Quinn. In each study, an undergraduate college student interacted with Quinn using the web application to teach Quinn how to solve eight variable equation math problems. For two problems, Quinn spoke with a non-transformed baseline speech. For the remaining six problems, Quinn alternated the type of adaptation for each problem. Two full problems were given for each type of adaptation; I collected each problem as a separate dialogue for a total of 8 dialogues per student. Statistics for the collected corpus are shown in Table 11.3. The gender of Quinn's voice was chosen to match the gender of the student. The four case studies were gender balanced with two males and two females. The gender of the speaker drove the gender of Quinn's voice. If the student was a female, then Quinn was female. If the student was male, Quinn was male.

I manually selected 40 exchanges from each of the student-Quinn dialogues. An exchange was two adjacent turns by different speakers (i.e. The student and Quinn). I selected ten exchanges for the baseline text-to-speech and ten exchanges for each of the three types of adaptation, focusing on those exchanges with maximum coherency and minimal pausing or silence, eliminating any exchanges where speech recognition may have failed. The ten exchanges were evenly split between two scenarios. In the first scenario, Quinn was the first speaker in the exchange. In the second scenario, Quinn was the second speaker and was responding. With a total of 40 exchanges per student, I utilized Amazon Mechanical Turk (AMT), a popular resource for crowdsourcing research tasks including annotations, transcripts, and subjective analysis (Buhrmester, Kwang, & Gosling, 2011). I used AMT to obtain 10 random, perceptual evaluations per exchange for a total of 400 evaluations per student or 1600 evaluations. Using third party ratings such as those collected through AMT is a standard technique in the evaluation of naturalness and social features of dialogue systems (Jurcicek et al., 2011). In addition, avoiding first-person ratings allowed us to present all dialogue approaches to each of the four individuals without concern for how their perceptions of different approaches might affect their ratings.

Through AMT, individuals, referred to as workers, were asked to listen to each exchange and answer a series of questions regarding the speakers. Each worker had access

	Mean	Std. Dev.
Dialogue length (min)	5.4	2.1
Number of turns	30	10
Turn length (sec)	10.8	4.6

Table 11.3. Dialogue and Turn Statistics for Quinn Corpus

85

to evaluate 160 exchanges (40 per student). To evaluate naturalness, I used Mean Opinion Score or MOS (ITU-T, 1994). With MOS, workers were asked to evaluate the quality of the voice on a Likert scale of 1-5, where 1 was very poor and 5 was completely natural. Workers evaluated both the human speaker and Quinn on this scale.

For evaluating rapport, I adopted a subset of questions from the rapport scale utilized by (Gratch, et al., 2007) and in the own work (Chapter 5). Workers were asked the following two questions about the relationship between the speakers on a Likert scale of 1-5, where 1 is "not at all" and 5 is "a lot." In the questions below, Alex refers to the student and Quinn refers to the virtual agent. I selected these questions because they target a shared feeling between speakers. The responses are averaged for one rapport rating.

1. Alex and Quinn understood each other

2. There is a sense of closeness between Alex and Quinn

In total, 174 workers provided evaluations of the audio. 12% or 21 workers rated 30% or more of the 160 exchanges they had access to; 40% of the workers listened to and rated only one exchange. In analyzing the results, I treated each rating as the unit of analysis. I calculated inter-rater agreement using Krippendorffs alpha (Krippendorff 2011), an alternative to Cohen's kappa designed to handle multiple raters and missing data (i.e. not all raters rated every exchange). Like Cohen's kappa, agreement is most acceptable above 0.8 and tentatively accepted above 0.66. Agreement here was measured at $\alpha = 0.69$.

11.4   RESULTS

I am interested in evaluating how the different methods of pitch adaptation performed with respect to two criteria: perceived naturalness and perceived rapport. To analyze the effects

of the pitch adaptations in terms of rapport and naturalness, I ran basic statistical analyses

on the relationship between adaptation type, naturalness, and rapport.

To assess differences in **naturalness**, I performed a one-way analysis of variance

(ANOVA) with the type of adaptation (mirror partner, shift+contour, shift+flatten, and

control) as a factor and naturalness as the dependent variable. Table 11.4 gives the means

and standard deviations for each condition. The ANOVA analysis indicated statistically

significant differences among type of adaptations, $F_{(3, 1599)} = 19.9$, $p < 0.001$. Tukey

post hoc tests indicate that shift+contour was perceived as significantly more natural than

either mirror partner ($p < 0.001$) or shift-flatten ($p < 0.001$). I expected to find mirror

partner and shift+contour to be as natural as the control. I found mirror partner was

perceived to be much less natural, on par with shift+flatten. I also found shift+contour was

not significantly different from the control, where no adaptation was performed ($p = 0.52$).

These results lead us to conclude that in pursing implementing an automatically entraining

system, shift+contour, adapting pitch by shifting the TTS contour, is the most natural of

the adaptations reviewed and is as natural as a non-manipulated TTS output.

To identify differences in how **rapport** was perceived for each of the pitch

adaptations, I performed a one-way ANOVA with adaptation type as a factor and rapport

as the dependent variable. Table 11.5 gives the means and standard deviations. I found

statistically significant differences among the types of adaptations, $F_{(3, 1599)} = 5.63$, $p <$

	Mean	Std. Dev.
control	2.22	1.36
mirror partner	1.79	1.11
shift+contour	2.39	1.33
shift+flatten	1.85	1.15

Table 11.4. Descriptive Statistics for Naturalness on Each Pitch Adaptation

	Mean	Std. Dev.
control	3.56	1.17
mirror partner	3.73	1.07
shift+contour	3.74	1.07
shift+flatten	3.35	1.15

Table 11.5. Descriptive Statistics for Rapport on Each Pitch Adaptation

0.001. The hypothesis was that mirror partner would result in the most rapport, followed by shift+contour. Shift+flatten, I expected to be the lowest. Interestingly, I observed shift+contour to be on par with mirror partner. Both indicated higher, equivalent degrees of rapport over the control. Using Tukey post hoc tests to analyze which of the pitch adaptations were significantly different, I found shift+contour generated significantly higher perceptions of rapport than shift+flatten ($p < 0.01$). Differences between shift+contour, mirror partner, and control were not significant.

Additional follow-up analyses were conducted to try to understand the effects of the different types of adaptations. I examined the average ratings for each student who participated, as shown in Table 11.6. I found that for 3 of the 4 students, the raters perceived more rapport in the exchanges where Quinn adapted by the shift+contour adaptation than in any other condition. Listening to these recordings, I identified an imbalance in terms of content spoken across the individual interactions. In most scenarios, Quinn and the student engaged in on-task, problem related conversation. However, Quinn was programmed to introduce social dialogue 15-20% of the time. For example,

Q: This is not very fun, are we almost done?

S: Math can be fun! But yeah…we're almost done

Given the possibility the raters were considering the content of exchanges in their evaluations of rapport, I annotated the exchanges as either *social* (off-topic and not about

the problem), or *not* social (on-topic and about the problem). In addition, Quinn was designed to entrain to the previous turn made by the student. In the exchanges raters listened to, I counter-balanced between exchanges where the rater would hear Quinn speak first and scenarios where the rater would hear the student speak first. In the latter, the turn to which Quinn adapted is audible. I suspected the raters were perceiving more differences in the rapport produced when they could hear the speech to which Quinn was adapting.

To explore the effect of the social exchanges versus non-social exchanges as well as the order in which Quinn spoke, I ran a 3-way ANOVA with rapport as the independent variable, including the type of adaptation, whether Quinn spoke first or second, and the social/not-social annotations as factors. The ANOVA analysis indicated statistically significant interactions between all combinations of factors except for the highest order interaction (all 3 factors). F-scores and p-values are shown in Table 11.7.

Finding significant 2-way interactions for all combinations of factors, I ran pairwise comparisons for further analysis. In social exchanges, the type of adaptation resulted in significantly different levels of rapport. When Quinn spoke second, shift+contour had significantly higher rapport than the control (p = 0.03) and shift+flatten (p < 0.001). The difference with mirror partner was nearly significant (p = 0.08). Pitch adaptation in non-

		Average Rapport				Average Naturalness			
	gender	control	mirror	s-contour	s-flatten	control	mirror	s-contour	s-flatten
S1	F	3.61	3.68	**3.74**	3.38	2.14	1.69	**2.41**	1.74
S2	F	3.58	3.70	**3.78**	3.59	2.36	1.94	**2.39**	1.93
S3	M	3.64	3.65	**3.80**	3.55	**2.28**	1.74	2.27	1.83
S4	M	**3.79**	3.36	3.29	3.25	**2.50**	1.87	2.05	1.68

Table 11.6. Descriptive Statistics for Each Student in Pitch Adaptations

social exchanges or when Quinn spoke first had less of an effect. In reviewing these results, differences between the pitch adaptations become the most notable when incorporating social/non-social annotations. This suggests that in social exchanges, shift-contour produced significantly more rapport than the other adaptations and control, and that in considering adaptations for a broader study, shift-contour may be the most optimal and effective when an agent is also social. This is supported by the findings in Chapter 6 that people tend to entrain more when engaging in off-task, social dialogue.

I found support for the hypothesis that shift+flatten would result in the least rapport. I did not find support for mirror partner as a high-rapport adaptation. Listening to the exchanges, this was mostly likely due to the original concern of mismatches between pitch contour and syntactic and semantic structure. This was supported by the finding that mirror partner is significantly less natural. Considering mirror partner did receive very low naturalness scores, the rapport perceived for this adaptation is relatively high. Overcoming issues with syntactic and semantic structure with a more nuanced adaptation which accounts for contextual dependencies may be worth exploring in the future.

Factor	F-Score	p
Type of Adaptation	6.8	< 0.001
Social/Non-Social Exchange	1.3	0.25
Quinn Speaks First/Second	3.6	0.06
Adaptation x Social Exchange	6.1	< 0.001
Adaptation x Quinn Speaking	7.5	< 0.001
Social Exchange x Quinn Speaking	12.7	< 0.001
3-Way Interaction	2.0	0.11

Table 11.7. 3-Way ANOVA with Rapport as Dependent Variable

11.5   DISCUSSION AND CONCLUSIONS

These results demonstrated that adapting to the speaker does influence naturalness and rapport and that different types of adaptations can have positive effects on perceptions of naturalness and rapport. Shifting the contour by pitch mean is one form of adaptation which sounds as natural as current text-to-speech technologies while significantly increasing perceptions of rapport, particularly during social utterances. In the next Chapter, I describe how this adaptation was used in a larger study to explore the effects of proximal entrainment on rapport and learning with a robotic form of Quinn.

CHAPTER 12

EFFECTS ON RAPPORT AND LEARNING WITH QUINN

Using the pitch shift contour adaptation from Chapter 11, a larger study was conducted to explore the effects of this form of entrainment on rapport and learning. This larger study provides insight towards the overall research questions regarding how entrainment can be modeled, the effects of entrainment on rapport and learning, and any insights entrainment might provide regarding these kinds of interactions.

For this larger study, I implemented the agent Quinn in a robotic form and conducted a between subjects' experiment. Participants taught Quinn in one of three conditions: (1) a **social plus entraining** condition in which Quinn introduced social statements and adapted its pitch via shift+contour, (2) a **social** condition in which Quinn only introduced social statements, and (3) **a non-social condition** in which Quinn did not speak socially or entrain, staying purely on task. In the next section I revisit briefly how Quinn was designed and any differences which are present from the agent version described in Chapter 11. Section 12.2 describes the procedure and measures; the results of the study are given in 12.3 and the conclusions from this study are summarized in 12.4.

12.1 QUINN (REVIEW)

For this study, Quinn consists of a LEGO Mindstorms base with an iPod mounted on top of it representing its face. Like the virtual agent in Chapter 11, Quinn's facial expressions are animated when speaking, and neutral otherwise. Students still engage Quinn via speech and a web interface; Quinn an example problem on the interface are shown in Figure 12.1.
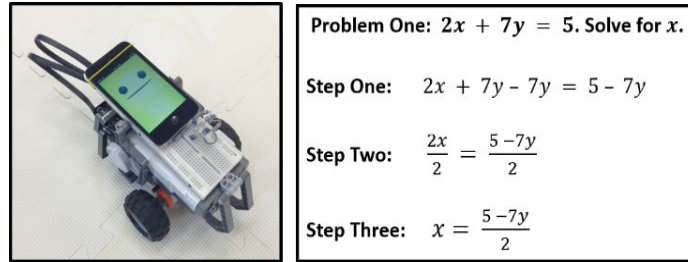
Figure 12.1. Quinn (Robotic Version) and an Example Problem

In this iteration, students teach Quinn six variable equation problems, with the application presenting one problem at a time. At the end of each problem, students ask Quinn to solve a quiz, step by step.

As with the virtual agent, Quinn can introduce social dialogue. Social responses were generated by creating two parallel dialogue options; a social dialogue and a non-social dialogue. The 'social' dialogue response was selected in the social and social-entraining conditions and resulted in social utterances 15 – 20% of the time.

12.2   METHDOLOGY AND PROCEDURE

A total of 48 individuals interacted with Quinn in the one of the three conditions described. There were 16 participants in each condition consisting of 8 females and 8 males. Participants were undergraduate students between the ages of 18 and 30; all were native English speakers. Individuals were randomly assigned to conditions, sessions lasted for 90 minutes, and were compensated $15 upon completion.

12.2.1 Procedure

Participants began by completing a 10-minute pre-test on literal equations. They were then given a practice exercise consisting of two worked-out examples of literal equations. The

participants were asked to explain the problems and how to solve each step out loud. This exercise was to help participants adjust to the tutoring task and encourage them to think about how they might explain the content. After this exercise, participants watched a short video introducing Quinn and the task.

Participants were told the task consisted of helping Quinn learn how to solve literal equations; they should walk Quinn through the steps to solve six problems and they would have the opportunity to test Quinn's understanding through quizzes. Participants were also informed they could 'reteach' Quinn if Quinn struggled on a quiz by moving back to the previous problem. After teaching Quinn all six problems, participants were given a 10-minute post-test and a short questionnaire assessing their attitudes towards Quinn. If a participant had additional availability meaning they could stay longer than the 90 minutes of allotted time, I asked them some final interview questions. Outside of availability, no other criteria were used to determine which participants were interviewed. The interviews were approximately distributed across gender (11 female, 9 male) and condition (6 control, 6 social, 8 social plus entraining).

12.2.2 Measuring Learning

Learning gains were assessed with the pretest and posttest scores. I computed normalized learning gains according to (Hake, 2002) using (1) to account for prior knowledge. If the posttest scores were lower than the pretest scores, I used (2).

$$gain = (posttest - pretest)/(1 - pretest) \quad \textbf{(1)}$$

$$gain = (posttest - pretest)/pretest \quad \textbf{(2)}$$

After removing the five participants who scored 100% on the pretest, I found of the 43 participants remaining, 23 hit a ceiling on their learning gains (scoring 100% on the posttest). With 10 individuals at zero gain, 10 individuals who gained in a normal distribution, and 23 hitting full gain, I determined analysis would be better served by grouping the learners into three groups – no gain, some gain and all gain. The results on learning gains are analyzed in this context.

I also assessed a measure regarding persistence in the interaction by collecting the number of times a student retaught Quinn. Quinn was pre-programmed to get the wrong answer on two of the quizzes. This re-teaching metric was calculated as the total number of times the student retaught Quinn, with four possible values observed: 0, 1, 2, or 3.

12.2.3 Measuring Rapport

I collected two measures for analyzing social responses: subjective, self-reported rapport and observable, verbal rapport-building behaviors. For self-reported rapport, I was interested in two kinds of self-reported rapport, general rapport related to feelings of understanding and connection and social presence, related to feeling that one's partner is real, present and attentive. For general rapport, I based the questions off work by Huang and colleagues (Huang et al., 2011) and Gratch and colleagues (Gratch et al., 2007) who developed a rapport scale over several iterations. I adapted their questions to create a nine Likert-scale questionnaire directed at capturing general rapport, i.e. feelings of connectedness, coordination, and understanding (Appendix A). Cronbach's alpha for the nine questions was .72. I averaged these nine questions to create one representative construct for general rapport, referred to as 'rapport' in the results. To measure social

presence, I utilized eight Likert-scale questions from the attentional allocation portion of the Networked Minds Social Presence Inventory (Biocca 2002, Appendix A). I utilized the attentional allocation portion of the survey because attention has been identified as a critical element of both social presence and rapport (Tickle-Degnen and Rosenthal 1990). Averaging the eight social presence questions, Cronbach's alpha was .69. I averaged the eight questions and refer to this measure as social presence in the results.

For the observable behaviors of rapport, I based the measures on prior work on linguistic rapport as discussed in Chapter 9. I assessed elements of linguistic politeness, including praise, formal politeness, inclusivity, and name usage; examples of the behaviors can be found in Table 12.1. The detailed coding scheme is given in Appendix A. In deciding on a coding scheme for linguistic politeness, I considered that different situations create unique interpretations for which linguistic structures are positively polite and may build rapport versus hinder rapport. Distinctions were made for any situations in which these behaviors may have been used to express negativity. This was rare; in those cases, the behavior was not included. Two individuals each independently coded two thirds of the dialogues for these behaviors. The average Cohen's kappa for these behaviors was 0.88. Individual kappas are reported in Table 12.1 along with the overall means and standard deviations for each condition. To assess how these behaviors differed across conditions, I

		M	*SD*	*k*
Praise	"Great job", "Good answer"	2.63	4.01	.79
Politeness	"thank you", "you're welcome"	2.98	5.49	.83
Inclusive	'we' or 'lets'	26.0	19.5	.98
Name	"That's right, Quinn", "okay Quinn"	23.5	17.8	.95

Table 12.1. Descriptive Statistics for Coding of Linguistic Rapport with Quinn

aggregated them into a single representative construct of linguistic rapport; Cronbach's alpha was .70.

12.3 RESULTS OF PITCH PROXIMITY ON RAPPORT AND LEARNING

I evaluated how individuals responded to a teachable robot as they interacted with Quinn in one of three conditions – the robot engaged in social dialogue and entrained by adapting its pitch (condition = **social plus entraining**), the robot was engaged in social dialogue but did not entrain (condition = **social**) or the teachable robot did not entrain and was not social (condition = **non-social**). I summarize the results regarding learning, self-reported rapport, and linguistic rapport in the next section. I include an analysis of how gender mediated an individual's self-reported and linguistic rapport and then compared the effects depending on whether responses were measured as self-reported or linguistic.

12.3.1 Learning Results

Analyzing learning as gain, I found 10 individuals at zero gain, 10 individuals who gained in a normal distribution, and 23 hitting full gain. I determined analysis would be better served by grouping the learners into three groups – no gain, some gain and all gain. Having grouped the students into three learning groups, I analyzed the learning gains in terms of a multinomial logistic regression. However, even with this adjustment, the overall model in the analysis including both condition and gender was not significant, $X^2(6) = 6.86$, $p = 0.33$, and I found that none of the individual predictors are significant.

I was also interested in assessing whether individuals re-taught Quinn and whether differences existed in the degree to which individuals retaught Quinn. I utilized the re-

teaching metric described as persistence Section 12.2. The means and standard deviations for persistence by gender and condition are shown in Table 12.2. I utilized multinomial logistic regression to estimate the influence of condition and gender on persistence in the interaction, given that I measure persistence in terms of total re-teaching. In the analysis, the overall model including both condition and gender was not significant, $X^2(9) =12.35$, $p = 0.19$. Looking at the predictors individually, gender is significant when controlling for condition. The likelihood of a female persisting in the interaction and re-teaching Quinn was 2.13 times more likely than a male, $p = 0.03$.

Given the significance of re-teaching in relation to gender, I explored whether re-teaching was related to learning. I ran Pearson's chi-squared correlation on the categorical learning gains described above. I found a significant correlation between re-teaching and the categorical learning gains, with $X^2(6) =17.9$, $p = 0.006$.

Finally, I assessed social presence and rapport in terms of learning. Running a multinomial regression with rapport and social presence, I found the model was not significant, $X^2(4) =4.68$, $p = 0.32$. However, in viewing the individual coefficients, social presence does approach a significant effect on learning ($p = .06$). For those individuals who

Condition	Learning			Persistence		
	Males	Females	All	Males	Females	All
Non-social	.72 (.44)	.83 (.41)	.81 (.33)	1.4 (1.1)	2.0 (.89)	1.7 (1.1)
Social	.34 (.48)	.73 (.48)	.50 (.54)	.57 (1.1)	1.4 (1.3)	1.1 (1.2)
Social plus entraining	.53 (.51)	.60 (.43)	.56 (.45)	1.0 (1.1)	1.8 (1.2)	1.4 (1.2)
All Conditions	.53 (.48)	.71 (.43)	.62 (.46)	1.0 (1.1)	1.7 (1.1)[*]	1.4 (1.2)

Table 12.2. Descriptive Statistics for Learning and Persistence with Quinn. [*] Significant at $p < 0.05$, [**] Significant at $p < 0.01$

gained but did not hit ceiling on their gain, social presence is 1.38 times higher than for those individuals who did not gain.

12.3.2 Self-Reported Rapport Results

I utilized multivariate analysis of variance (MANOVA) to explore how gender mediated self-reported rapport, measured as general rapport and social presence, to a social, pitch-entraining teachable robot. A two-way MANOVA with general rapport and social presence as dependent variables and gender and condition as independent variables revealed significant main effects for condition (Wilks' $\lambda = .80$, F $= 4.41$, p $= .02$) and gender (Wilks' $\lambda = .77$, F $= 2.54$, p $= .04$, partial eta squared $= .124$). The interaction between gender and condition was not significant. The means and standard deviations are in Table 12.3.

Performing an analysis of univariate effects to understand the effect of condition, I found individuals reported significantly less social presence when Quinn was social but did not adapt its pitch, F $(2, 42) = 4.0$, p $= .02$. While there was not a significant interaction between gender and condition, simple pairwise comparisons of gender indicated that for males, the social condition differed significantly from both the social plus entraining (p =

Condition	Social Presence			Rapport		
	Males	Females	All	Males	Females	All
Non-social	4.63 (.35)	4.75 (.68)	4.69 (.52)	4.58 (.61)	5.35 (1.11)	5.04 (1.06)
Social	4.05 (.74) [+]	4.49 (.51)	4.27 (.65)	4.90 (.75)	5.58 (.58)	5.27 (.91)
Social plus entraining	4.88 (.29)	4.71 (.79)	4.75 (.60)	4.92 (1.36)	5.59 (.78)	5.36 (1.2)
All Conditions	5.18 (.79)	5.55 (.70)	4.57 (.62) [*]	4.70 (.97)	5.60 (.71) [**]	5.22 (1.05)

Table 12.3. Descriptive Statistics for Social Presence and Rapport with Quinn. [*] Significant at p < 0.05, [**] Significant at p < 0.01

0.001) and the non-social (p = 0.01) conditions, with males reporting significantly less social presence in the social condition, suggesting the males were driving the difference.

Analysis of the univariate effects of gender revealed that regardless of the robot's behavior within conditions females felt significantly more rapport overall than males, $F (2, 42) = 8.86$, $p = 0.006$. The effect size η^2 for this difference is large at 0.18. In considering social presence, the difference between males and females approached significance, $F (2, 42) = 3.76$, $p = 0.06$, with males reporting less social presence overall.

One explanation for these results is the potential speech recognition errors made by the dialogue system. To analyze the effect of speech recognition errors on rapport and social presence, I focused on the output of the dialogue manager. As described in Chapter 10, if the dialogue manager could not match the student's words to a specific response, the dialogue manager would return either a request for clarification (i.e. "can you please repeat that?") or a general acknowledgement (i.e. "ok sounds good"). I classified the number of generic responses Quinn returned when Quinn could not match an exact pattern to a precise response and ran an ANCOVA with gender and condition as independent variables and social presence and rapport as dependent variables, with the percentage of turns where Quinn requested clarification or gave a general acknowledgement as the covariate. I found this did not have a statistically significant effect on rapport and social presence (p = 0.41) and did not alter the significance of gender and condition on social presence and rapport.

To summarize, I found individuals self-reported significantly less social presence in the social condition. These findings suggest individuals found social dialogue without the presence of prosodic entrainment on pitch to be less engaging, indicating consistency and balance of design is critical when incorporating social behaviors to build rapport. In

the next section, I explore whether these findings are supported by linguistic measures of rapport and whether the alignment of social dialogue with entrainment appeared to facilitate more rapport-building behaviors in the social plus entraining condition.

12.3.3 Linguistic Rapport Results

I measured linguistic rapport through individuals' use of four rapport-building linguistic politeness behaviors: name usage, inclusive language, praise, and formal politeness (for example, "please" or "you're welcome") while interacting with Quinn. I combined all four linguistic rapport behaviors into a single construct of linguistic rapport. I expected individuals' use of linguistic rapport to reflect similar findings as with the self-reported rapport. I conducted a two-way ANCOVA to examine the effect of condition and gender on use of linguistic rapport while controlling for dialogue length. There was a statistically significant interaction between the effects of gender and condition on the presence of linguistic rapport, $F(2, 42) = 5.45$, $p = .008$. In terms of main effects, there was not a statistically significant difference in linguistic rapport for the different conditions, $F(2, 42) = 1.26$, $p = .29$. However, I do observe significant differences by gender, $F(1, 42) = 10.6$, $p = .002$. The means and standard deviations are given in Table 12.4.

I explored the significant interaction effect; simple main effects analysis showed that females used on average significantly fewer linguistic rapport behaviors in the social plus entraining condition as compared to both the nonsocial ($p = .03$) and social conditions ($p = .002$). Males however did not change significantly in the number of linguistic rapport behaviors they used between conditions. In addition, females used significantly more linguistic behaviors than males in the nonsocial ($p = .03$) and social conditions ($p = .001$).

		Non-social	Social	Social + Entraining
Linguistic Rapport	**Females**	57.9 (29.9)	72.3 (31.0)	28.3 (24.7)
	Males	29.1 (17.3)	19.6 (16.2)	36.3 (32.7)
	Overall	43.5 (27.9)	45.9 (36.2)	32.2 (28.3)

Table 12.4. Descriptive Statistics for Linguistic Rapport with Quinn

To summarize, these findings indicated that females utilized rapport-building behaviors significantly more in the social and nonsocial conditions when compared to males and when compared to themselves in the social plus entraining condition. This suggests that the robot's social behavior did influence individuals' use of these behaviors but that it was mediated by gender and potentially that these behaviors may be more informative of female responses. If these behaviors are positively related to self-reported rapport as I have hypothesized, this may mean that the entraining mechanism failed for females, something I did not necessarily observe in the self-reported rapport results. However, if linguistic rapport is negatively related to feelings of rapport, then these findings align with the self-reported findings while also suggesting that there may be a mismatch between how females self-report rapport versus their behavior. It is also possible linguistic rapport is not related to self-reported rapport. If this is the case, then these behaviors may be indicative of another underlying factor which was influenced by the robot's social behavior in significantly different ways for males and females. I investigated the relationship between self-reported and linguistic rapport in the next section.

12.3.4 Relating Self-Reported and Linguistic Rapport

I utilized the Pearson product-moment correlation coefficient to explore whether there was

a relationship between self-reported general rapport, social presence, and the measure of linguistic rapport. I had hypothesized a relatively simple, positive relationship between self-reported measures and linguistic measures. I found that social presence was significantly, negatively correlated with linguistic rapport, $r(46) = -.44$, $p = .002$; self-reported general rapport was not significantly correlated with the linguistic rapport, although approaches a positive relationship, $r(46) = .23$, $p = .10$.

Breaking out the correlations by gender, I found that females may be the driving force behind the significantly negative correlation between self-reported social presence and linguistic rapport, $r(22) = -.49$, $p = .001$. For males, social presence and linguistic politeness were not correlated at all, $r(22) = .05$, $p = .80$. In addition, males approached a significant positive relationship between self-reported general rapport and linguistic rapport, $r(22) = .36$, $p = .08$; for females there was no relationship between general rapport and linguistic politeness, $r(22) = -.12$, $p = .57$. Figure 12.2 summarizes these findings.

Gender also appeared to be a significant indicator of individual differences in how self-reported rapport and social presence related to use of linguistic politeness. Regardless of condition, when males used more praise and politeness, they self-reported feeling more rapport. Females on the other hand reported lower feelings of social presence and rapport when they used more inclusive language and used Quinn's name more often. These findings suggest that there are individual differences present in how behaviors reflect an individual's internal rapport state and how these behaviors are influenced by the robot's use of social dialogue and entrainment.
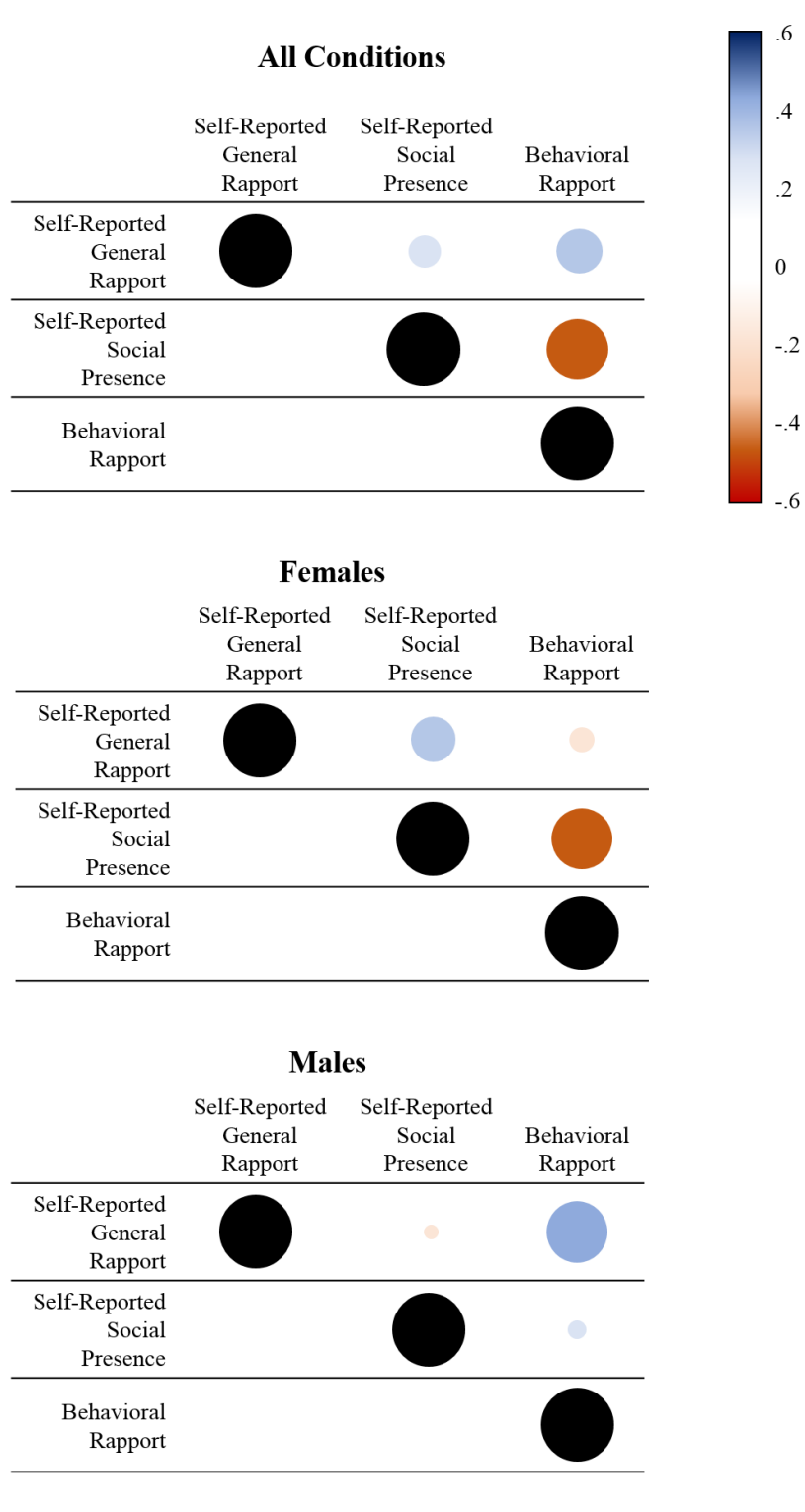
Figure 12.2. Correlations of Self-Reported and Linguistic Rapport.

12.3.5 Understanding Rapport State

Several interesting observations emerge from the exploration of pitch proximity with a social, robotic learning companion. First, it appears that an entraining robot improved perception of social dialogue. When social dialogue was present without the pitch adaptation, individuals perceived the robot as significantly less socially present. I also observed that linguistic rapport as a measure of behavioral rapport was negatively correlated to perceptions of social presence, particularly for females, and that females especially engaged in more linguistic rapport in the social condition, further supporting that the social condition was socially engaging. The results suggest that Quinn's social dialogue influenced how individuals engaged in linguistic rapport and that linguistic rapport was indicative of their own underlying rapport, mediated by gender.

There are several possible explanations for the role of gender in these findings regarding linguistic rapport. It is possible females in the study increased linguistic rapport behaviors when they felt less rapport because they were attempting to increase rapport and build a relationship where they currently did not sense one. According to prior work, women are more likely to see conversation as a means for building rapport. While females used these verbal behaviors to build a relationship, males may have utilized these verbal behaviors as relationship indicators, engaging in linguistic rapport only once a positive relationship had been initiated, confirmed, and pushed by their conversational partner. This would suggest that for males, linguistic rapport emerged because they felt rapport, not because they were trying to build rapport. If this is true, it implies that verbal behavioral measures are not indicative of the same underlying rapport state for all individuals and that this rapport state may manifest differently in response to social triggers.

105

It is possible Quinn's social dialogue may have triggered transitions in an individual's underlying rapport state and that for males and females, this state manifested differently. Exploring how individuals responded on a turn-by-turn basis to Quinn's social dialogue with their own linguistic rapport may provide confirmation of this. I hope to gain insight into whether the introduction of social dialogue in the social and social plus entraining conditions caused changes in a user's underlying rapport state as indicated by their linguistic rapport behavior. I explore this in the next section, using the observed linguistic behaviors with Quinn's own social dialogue to predict a user's underlying rapport state and I evaluate how this state changed for different individuals when Quinn was social versus when Quinn was social and entrained.

12.3.6 Input-Output HMMs to Model a User's Rapport State

I used an input-output Hidden Markov Model (IOHMM), a special type of Hidden Markov Model, to explore how an individual's rapport state can be predicted from their use of linguistic rapport and Quinn's own social dialogue. Hidden Markov Models have historically been applied for understanding hidden states such as emotions, tutoring modes, and learner engagement (Nwe, Foo, and Silva 2003; Boyer et al. 2010; Beal, Mitra, and Cohen 2007). Once a model has been created, the frequency counts of the estimated hidden states can be used to understand the relationship between the hidden state (i.e. tutorial mode or learner engagement) and desired outcomes (i.e. learning). For example, Boyer and colleagues utilized an HMM to model effective tutoring modes based on observed dialogue acts. Correlating the estimated frequency counts of the different tutoring mode states with learning, they found significant learning gains associated with state sequences. Beal, Mitra,

and Cohen modeled learner engagement; relating the hidden state of learner engagement to learning, they identified learner engagement trajectories which directly related to learning gains. Bergner and colleagues explored how tutors assist tutees when tutees make a mistake (Bergner, Walker, & Ogan, 2017). Utilizing an IOHMM, they compared the assistance value of different tutor inputs in helping the tutee correct a mistaken step and found successful as well as deleterious patterns in collaborative learning. In the work, I utilized an IOHMM to explore whether there is a hidden state associated with linguistic politeness, whether I can consider that state to be 'social' or reflective of an individual's rapport, and finally, how that state was affected by Quinn's social behaviors.

Hidden Markov Models are the simplest form of a dynamic Bayesian network. In an HMM the states are unobserved (i.e. hidden), making the HMM a useful model for estimating internal conditions such as social state which is only hinted at only by observable social cues. HMMs utilize the Markov property and assume the probability of the current state depends only on the prior state. In this work, I utilize input-output HMMs because they include one additional dependency, where the current state depends not only on the probability of the prior hidden state but also on the preceding input (for example, whether Quinn is social or not). I give a summary of how IOHMMs operate below but a complete description can be found by Bengio and Frasconi (1995). Like an HMM, the joint probability distribution of a given sequence of observations ($O_{1:T}$) and hidden states ($S_{1:T}$) is based on the Markov property. The distribution is given in Equation 12.1 for a sequence of length T.

$$P(O_{1:T}, S_{1:T}) = P(O_1)P(S_1|O_1) \prod_{t=2}^{T} P(S_t|S_{t-1})P(O_t|S_t) \qquad \text{Eqn. 12.1}$$

With IOHMMs, the hidden state at time t, S_t, is dependent on both the prior hidden state S_{t-1} and the prior input I_{t-1}. This primarily affects the transition probability, or the probability of a particular state given what has already occurred. The transition probability can be described by the input (I) and the prior state as shown in Equation 12.2. Given the total number of input types (K) and the total number of state types (N), the transition probabilities can be broken into K separate $N \times N$ transition matrices, one for each input type. I report the transition probabilities as K separate $N \times N$ transition matrices.

$$P(S_t|S_{t-1}, I_{t-1}) \qquad\qquad \text{Eqn. 12.2}$$

A model of the network based on the general form of an IOHMM is given in Figure 12.3. I aggregated Quinn's social dialogue so that I had two input types ($K = 2$) consisting of whether Quinn speaks socially. I analyzed the IOHMMs across the three conditions. For the hidden state, I was interested in a state indicative of whether the student is responding socially. I proposed two hidden states ($N = 2$), corresponding to whether a student is 'socially engaged' or not. I utilized two states because I did not want to overcomplicate the representation, and the measures I am using are more interpretable with fewer states. In addition, a model with two states resulted in an acceptably high likelihood while keeping the number of parameters suitably smaller than the dataset. For observations, I labeled a turn as rapport building if the student used any one of the four behaviors, giving us two possible observations ($O = 2$), either linguistic rapport was present, or it was not.

I trained the HMM for each condition (non-social, social, social plus entraining) on sequences composed of each students' turn-by-turn dialogue with Quinn. Each student had taught Quinn six problems. A single sequence consisted of a single student's turn-by-turn exchange with Quinn on one problem. This resulted in 319 sequences with 2545-time
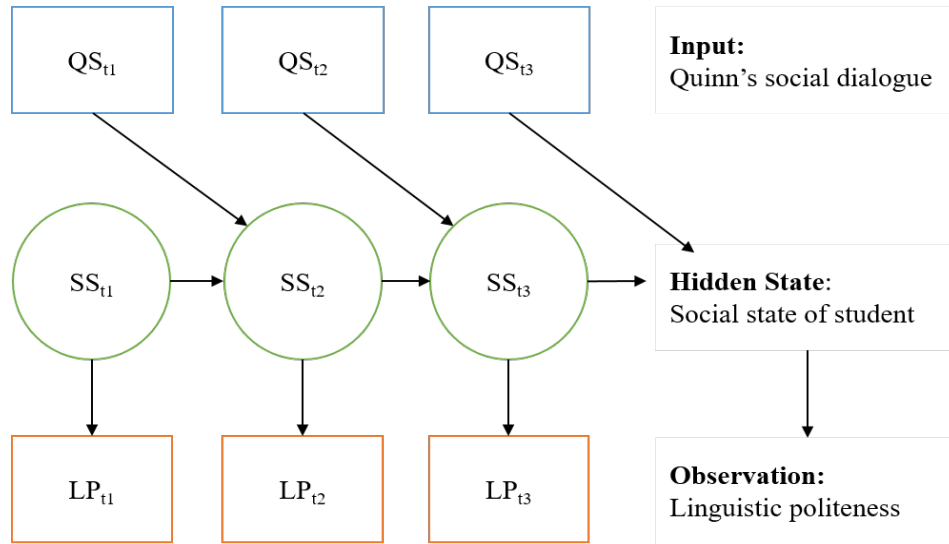
Figure 12.3. IOHMM for Exploring the Rapport State of Students

slices; each time slice consisted of an observed input and output. All parameter learning

was carried out using Murphy's Bayes Net Toolbox for Matlab (Murphy 2001), which uses

a variation of the expectation-maximization (EM) algorithm. The likelihood manifold has

local maxima, so I used multiple restarts of EM from different initial values. Using 300

restarts, I found the ten best runs, in terms of log-likelihood, resulted in values consistently

within a narrow range. Additionally, I ran models for males and females across conditions

considering how the underlying rapport state of genders might differ.

12.3.7 Results of IOHMM

The final models suggested that there is a distinguishable hidden state associated with

observing linguistic rapport. When linguistic rapport was observed, this was associated

with a distinct state (S2) which was separate from when linguistic rapport was not observed

(S1). These states were observable from the observation probabilities given in Table 12.5

broken out by gender. The state associated with observing linguistic politeness (S2) was

also clearly related to Quinn's social dialogue and adaptation; this was suggested by the results of the transition probabilities, which were broken out by males and females across conditions and shown as 2 ($K = 2$), 2 x 2 ($N \times N$) matrices in Table 12.6.

Gender differences were present in how the hidden state associated with linguistic rapport manifested, particularly when Quinn was social. In the non-social condition when Quinn was not social nor entraining, males and females responded similarly. If they were in a state which was associated with verbal rapport, they stayed in that state. However, in the social condition, when Quinn exhibited social dialogue but did not adapt, I began to see a difference in how males and females responded. For males, if Quinn was not social, males would either move to a non-linguistic rapport state or they would stay in a non-linguistic rapport state. If Quinn was social and the male student was already exhibiting verbal rapport behaviors, the male student was likely to continue exhibiting these behaviors. If Quinn was social and they were not exhibiting linguistic rapport already, a male student had a 50-50 chance of moving to a state associated with linguistic rapport. Females on the other hand were more likely to move to a state characterized by linguistic rapport when Quinn was NOT social. If Quinn was social, female students were more likely to move to a non-linguistic-rapport state. I saw these patterns intensify when Quinn was social and adapted. Males were more likely to move to a state associated with observed linguistic rapport when Quinn was social and adapted. Female students were more likely to move to a state which was NOT associated with linguistic rapport when Quinn was social and adapted.

These results suggest that (1) a hidden state exists which is associated with linguistic rapport and is clearly influenced by Quinn's social behaviors, (2) how this hidden

110

	Student is…	Non-social		Social		Social + Entraining	
		S1	S2	S1	S2	S1	S2
Males	Not exhibiting rapport	**.92**	.13	**.95**	.11	**.75**	.12
	Exhibiting rapport	.08	**.87**	.05	**.89**	.25	**.88**
Females	Not exhibiting rapport	**.77**	.06	**.88**	.18	**.92**	.12
	Exhibiting rapport	.23	**.94**	.12	**.82**	.08	**.88**

Table 12.5. Observation probabilities for IOHMM; S1 S2 represent the two hidden states

	Quinn is…	Next State	Non-social		Social		Social + Entraining	
			S1	S2	S1	S2	S1	S2
Males	**Not social** (K = 1)	S1	**.95**	.15	**.95**	**.93**	**.96**	**.92**
		S2	.05	**.85**	.05	.07	.04	.08
	Social (K = 2)	S1	---	---	.45	.12	.05	.08
		S2	---	---	.55	**.88**	**.95**	**.92**
			S1	S2	S1	S2	S1	S2
Females	**Not social** (K=1)	S1	**.95**	.01	.12	.15	.22	.26
		S2	.05	**.90**	**.88**	**.85**	**.78**	**.74**
	Social (K=2)	S1	---	---	.66	**.86**	**.92**	**.99**
		S2	---	---	.34	.14	.08	.01

Table 12.6. Transition Matrices for IOHMM to Estimate Rapport State

state manifests and the effects of Quinn's social behaviors on it are strongly mediated by gender and (3) males and females entered this underlying state on different social triggers.

12.4 DISCUSSION AND CONCLUSIONS

Forty-eight college students interacted with the teachable robot Quinn in one of three conditions, a **social** condition where the robot utilized social dialogue, a **social plus entraining** condition where the robot spoke socially and entrained using the pitch

adaptation, and a **non-social** condition where the robot neither spoke socially nor entrained. The results of these interactions provided insight into the four research questions posed by this thesis on how entrainment might be modeled, the effects of entrainment on social responses like rapport, the effects of entrainment on learning, and whether modeling entrainment in a teachable robot can provide new insight into human-human and human-agent interactions.

This study demonstrated that entrainment can be modeled as a form of turn-by-turn pitch adaptation and that this design can have a positive impact on social responses. Interestingly, the results suggest that prosodic manipulation as a form of entrainment may have served to enhance the positive perception of social dialogue while social dialogue without prosodic manipulation decreased perceptions of Quinn's social presence. In prior works, social dialogue has been shown to build rapport but, in this study, social dialogue unexpectedly produced the lowest responses, even lower than no social behavior at all. Individuals reported significantly lower social presence in the social condition and I found individuals increased linguistic rapport behaviors negatively correlated with social presence in the social condition while individuals in the social+entraining condition reported the highest feelings of social presence and rapport. These findings suggest a single channel of social behavior can fail where two channels can succeed, and strongly supports incorporating multiple channels of social behavior as an important consideration in facilitating rapport. Other work has indicated that the misalignment of multiple behaviors can harm perceptions (Meena, Jokinen, & Wilcock 2012). The results here suggest that facilitating alignment through pitch proximity can potentially improve social responses.

This study did not provide any evidence of entrainment effects on learning. Many learners hit ceiling on the posttest, suggesting that potentially the domain content was sub-optimal for enhancing learning in the given participant group. It is also possible that the design of the entrainment was too simple to influence learning. Follow-up studies will continue to explore effects of alternative entrainment designs on learning with different domain content suited to the given participant group.

The interaction results with Quinn provide interesting insights into human-human and human-agent interactions, particularly regarding gender and the degree to which gender indicates individual differences present in social responses. It is not surprising that males and females might respond differently to social behavior from a robot and exhibit different linguistic rapport when we consider the background work of Chapter 9, which suggests that females may respond more positively to social behavior from a robot and in general tend to use linguistic rapport more often. Compared to the males, females felt significantly more rapport for the robot overall. They also changed how they used linguistic rapport across the different conditions, using more linguistic behaviors associated with rapport in the nonsocial and social conditions. I found females used these verbal behaviors when they felt less rapport as opposed to more rapport for Quinn, and they were more likely to stay or move to a social state (i.e. states not associated with linguistic rapport) when Quinn was not engaging in social dialogue. Males on the other hand reported significantly low social presence in the social condition and were more likely to stay or move to a social state when Quinn was already social. These findings have several interpretations with implications regarding interactions. One interpretation is that females utilized linguistic rapport to build rapport while males used it to express rapport. This interpretation is aligned

with how males and females have been found to use and interpret conversation where females tend to view conversation as a means for building relationships while men are more likely to see conversation as a tool. This supports the theory that the verbal behaviors explored here might represent different rapport building strategies, and suggests that when these behaviors are observed, they may be utilized in different ways to assess responses.

An alternative interpretation is that males and females had different initial social inclinations towards Quinn and this resulted in different rapport responses. I measured the linguistic rapport behaviors based on theories of rapport and politeness; in human-human interactions, politeness is more commonly associated with initial encounters with strangers. As individuals get to know one another, rapport increases and politeness decreases. The longer people know each other the less polite they tend to be and the more rapport they tend to feel. I observed females used fewer of the rapport behaviors in the social plus entraining condition. Females may have been more comfortable with viewing Quinn as a 'teachable' entity that could learn, being more likely to anthropomorphize Quinn and expect Quinn to be social. As a result, when Quinn engaged in social dialogue and entrained, females were more likely to accept Quinn's social behavior as genuine and treat Quinn as a friend, dropping the social niceties of linguistic politeness I use with strangers. In contrast, males followed a more traditional path. Feeling less rapport in general, males were less comfortable with Quinn. Quinn was a 'stranger' that they could potentially develop rapport with, but they did not feel as if Quinn was their 'friend.' I found confirmation of these attitudes in the post interviews, where females were more likely to refer to Quinn as "my friend" and "we're best friends now," males were more likely to describe Quinn as "an interesting robot" and "decently complex." This suggests that

individuals who are more prone to social interaction with a robot will respond with more familiarity to a robot's social behaviors; they will be more inclined to rapport overall, and this will impact their behavioral responses accordingly. If this interpretation is accurate, this has implications for the design of social interaction for different individuals – for those who are more inclined to social behavior, social interaction models may move to more quickly to familiar behavior than for users who are less inclined to social interaction.

Towards assessing the four research questions posed by this thesis, the overall results of this study highlight the complexities inherent in measuring responses as self-reported rapport versus observable rapport. I found males and females responded very similarly to the conditions, but this was not immediately obvious from their self-reported scores. The work suggests that self-reported scores may be more informative for some individuals than for others and the addition of verbal behaviors can provide more insight into those individuals for whom the self-report is less informative. In addition, self-reported rapport measures were not aligned in the same direction as the verbal rapport behaviors I collected, particularly for females. The results emphasize the importance of assessing social responses like rapport from multiple dimensions and that when using verbal behaviors to gain insight into an individual's underlying feelings, individual differences such as gender should be considered because the underlying state indicated by their verbal behavior is not the same. That state appears to manifest and be triggered differently.

This iteration on designing entrainment focused on entrainment as local proximity on pitch. Overall, the work in this chapter demonstrated that modeling entrainment based on a form of pitch proximity can produce complex self-reported and linguistic rapport responses rooted in individual differences. However, pitch proximity did not perform

significantly better than no social behavior at all and did not observe result in effects on learning. The next few chapters line up the next iteration on the design of entrainment for a social robotic learning companion by first focusing on the dialogue of the companion and then introducing a new design for local pitch entrainment.

CHAPTER 13

ENHANCING DIALOGUE THROUGH ITERATIVE DESIGN

In this chapter, I investigate how I can enhance the design of the dialogue of the robotic learning companion. In Chapter 12, promising results on the use entrainment emerged in that it appeared to improve perceptions of social dialogue. However, the social dialogue performed poorly on its own, resulting in significantly lower social presence, and no observable effects on learning. Given the success of social dialogue in prior work, this was unexpected and presents a challenge when trying to assess the rapport-building potential of entrainment as a complementary dialogue behavior. It is possible there were missed opportunities in the robot's dialogue to foster better learning experiences as well as better social experiences; alternative designs of social dialogue may produce better outcomes. This chapter explores those alternative designs with the goal of identifying a more optimal social dialogue design to be implemented with explorations of entrainment.

As a strategy for the design of the dialogue, I concentrate on how the robot's verbal behaviors can create social self-efficacy experiences, or interactions that may aid learners in improving their self-efficacy, their belief that they can succeed in a domain. I focus on self-efficacy as a motivator for the design of the dialogue for several reasons. One, self-efficacy in STEM plays a major role in learner persistence and success and has been found to correlate with learning. Two, self-efficacy is difficult for learners to build on their own, but the teachable robot may be an ideal means for assisting learners in building self-efficacy. Three, self-efficacy has been theorized to be influenced by a series of social

experiences (Bandura, 1977). These social experiences provide an excellent guide for designing and evaluating dialogue strategies. These social experiences include:

1. **Mastery experiences**, facilitated by experiences where learners successfully teach the robot

2. **Vicarious learning experiences**, where a social model exemplifies good learning practices

3. **Social persuasion**, through social interaction learners are convinced of their success as domain experts

Bandura further hypothesizes that a learner's rapport, or sense of connection, with a collaborating partner, might enhance the effects of these pathways.

4. **Building rapport,** or a feeling of connection, with the learner

In this chapter, I examine the design of dialogue to foster these social experiences within the context of Nico. Like Quinn, Nico is a teachable robot for mathematics, but unlike Quinn, Nico is a Nao robot with a humanoid body and can introduce realistic gestures in addition to spoken dialogue. The focus of this chapter is on the design of dialogue, but Nico as a Nao robot enables future exploration of other embodied modalities.

To explore the design of the dialogue, I present the results of a multi-phase iterative design process with 14 learners; through this process, I explore how Nico's dialogue can foster social, self-efficacy-building experiences. The goal is to identify a design of dialogue that can be used more successfully with entrainment to build rapport. Towards that goal, I pose the following research questions:

RQ 1: How do different dialogue design strategies based on human-human peer tutoring and theories of rapport enhance mastery, vicarious experience, social persuasion, and rapport with a teachable robot?

RQ 2: How might individual differences, such as initial self-efficacy, influence responses to different dialogue design strategies in a teachable robot?

Motivated by the literature on peer tutoring and rapport, I iterated on the design of Nico's dialogue for each social experience. I explored how Nico's dialogue can facilitate mastery by balancing the challenge learner's face in articulating knowledge while enabling them to feel successful. I investigated how Nico's dialogue can model different learning practices such as question-asking and optimism and spark a corresponding response in the learner as vicarious experience. I explored both subtle and overt approaches to social persuasion. Finally, I explored rapport by iterating over rapport signaling behaviors of friends versus strangers. I analyzed learners' responses in each phase, yielding six design recommendations with an emphasis on how different individuals might benefit from different strategies. These design recommendations are then utilized in the design iterations for prosodic entrainment, described in Chapters 14, 15, and 16.

In the next section, I describe more about Nico and how learners interacted with Nico. In 13.2, I describe the methodology and then I describe the design of each phase and the results. I conclude with a brief discussion and a summary of recommendations.

13.1   NICO: A SOCIAL, ROBOTIC LEARNING COMPANION

For exploring dialogue design, I utilized the teachable robot Nico. Nico was a Nao robot that learners could teach about middle school mathematics. To interact with Nico, learners

utilized spoken dialogue and the tablet interface described in Chapter 10. After speaking, the learners' audio would pass into the dialogue system; for Nico, the dialogue system implementation included both the basic functionality and all the supplementary modules detailed in Chapter 10. The learning domain for the teachable robot was reasoning about ratios. The system included seven narrative-style ratio word problems; in this study, learners taught Nico four of the problems. The problems were based on the Common Core Standards for 6th grade (Common Core Standards Initiative, 2010). In addition to the narrative, the problems included a table, a common format for teaching ratios. Each row in the table was considered a problem 'step.' For each problem, Nico and the learner were given partial information; Nico would request the learner's help in how to use ratios to solve for the missing information. An image of an example problem is given in Figure 13.1.

Learners interacted with Nico using spoken language and a touch-screen interface on a tablet computer (Microsoft Surface Pro) that displayed each problem. The UI

Problem 1

Nico wants to go swimming with friends at the pool! Sadly though Nico's body isn't waterproof so Nico needs to prepare first. The plan is to use waterproof paint to protect Nico's body but Nico isn't sure how much waterproof paint is needed. Help Nico figure out how much paint will cover Nico's legs and torso!

Step	Body Part	Surface Area (sq. inches)	Volume of Paint (fluid oz)
Step 0	Feet	6	2
Step 1	Legs	12	???
Step 2	Torso	???	6

Figure 13.1. Example Ratio Word Problem with Table

displayed visual progress as the learner moved through the problems. To speak to Nico, the learner pressed a button on the interface. After they are finished speaking, a notice would appear on the interface indicating that Nico was 'thinking' while the system processed the input and generated a response. Average response time was less than four seconds. The UI tracked progress as the learner guided Nico through each problem step at their own pace, using buttons to advance forward. The current step was highlighted and enlarged on the screen.

13.2 ITERATIVE DESIGN OF DIALOGUE

With the platform described above, I can iterate over the design of a teachable robot's dialogue and explore (1) how different dialogue design strategies might enhance mastery, vicarious experience, social persuasion, and rapport, and (2) the role of individual differences in response to different strategies.

13.2.1 Method

Fourteen participants between ages 10 and 13 (M: 11, SD: 1.0, 4 female/10 male) participated across three exploratory design iterations with 5 participants (1 female/4 male) in the first phase, 5 (2 female/3 male) participants in the second phase, and 4 participants (1 female/3 male) in the final phase. All participants were native English speakers. Sessions lasted 90 minutes. Participants were recruited via flyers and emails shared during local programs offered to middle schoolers on the university campus.

The procedure for each session was the same across all phases. Participants began by completing a 10-minute pretest on ratios. Next, each participant was given a pre-survey

on self-efficacy and comfort-level towards robots. Before interacting with Nico, participants were given a few minutes to review the worked-out solutions to the problems they were to teach. Participants then watched a 3-minute video introducing Nico. After the video, Nico initiated a brief 'introduction' interaction by saying "Hello, it's nice to meet you. What is your name?" The 'introduction' gave participants an opportunity to practice talking to Nico before teaching. Participants utilized the tablet and spoken dialogue to teach Nico and moved through the problems at their own pace. Time to complete teaching the problems varied from 12 to 35 minutes. After the interaction, participants completed a post-survey; twelve participants also took a 10-minute posttest (isomorphic to the pretest). I then conducted structured interviews; the same questions were asked of every participant.

To evaluate the design and impact of each phase, I collected self-reported measures of rapport, self-efficacy, and learning gains and performed a comparative analysis on the interviews, experimenter observations, and dialogue transcripts. For the self-reported rapport, I posed a set of 14 Likert scale questions about rapport to each participant. The questions were based on a combination of prior work exploring rapport in human-human (Tickle-Degnen and Rosenthal, 1990), human-agent (Cassell, Gill, and Tepper, 2007), and human-robot interactions (Lubold, Walker, and Pon-Barry, 2016). Questions related to feelings of general rapport (i.e., "Nico and I worked well together"), positivity (i.e., "I liked Nico, Niko liked me"), attention ("Nico paid attention to me"), and coordination ("I was awkward in talking to Nico"). I asked participants in post-interviews to explain their understanding and interpretation of each survey question; these validations enabled small iterative changes to improve the wording. I aggregated the questions for each participant into an average rapport score.

Self-efficacy towards math was measured with six questions based on work by the Friday Institute for Technology (2008). Participants answered the six questions both before and after interacting with Nico. Averaging the responses, I calculated whether participants experienced a change in self-efficacy as post-score minus pre-score. Additionally, I asked how comfortable individuals were interacting with robots and human-looking robots. Finally, I calculated learning from the pre- and post-tests as a normalized learning gain score, as recommended in (Hake, 2002).

I focused the qualitative analyses on the interviews, experimenter observations, and transcripts of the interaction dialogue. Since I am interested in identifying the degree to which the dialogue can influence social experience, I coded the data for themes regarding mastery experience, vicarious experience, social persuasion, and rapport. A set of decision rules for identifying themes, as suggested by Miles, Huberman, and Saldana (1994), were identified. For example, for mastery to be present, the learner must give evidence they feel Nico learned. An example decision rule for mastery was: participant is marked as having felt a degree of mastery based on the presence of either (a) did the learner give any reference to Nico having 'learned' or (b) did the learner mention any observations regarding Nico's learning (i.e. "Nico said 'I understand' so he was learning"). Two researchers compared rules, themes and analyses to validate consistency; findings were triangulated across all data to ensure all pieces were mutually supportive.

13.2.2 Phase I

In Phase I, I began with dialogue which might foster mastery, vicarious experience, social persuasion and rapport based on human-human peer tutoring. To foster mastery, Nico

challenged participants to explain each step thoroughly before reaching an understanding of the solution. To foster vicarious experience and social persuasion, I introduced a question-answer design based on human-human peer tutoring (Graesser and Person, 1994) which found that tutees typically ask questions with low specificity. I designed Nico's dialogue as approximately 80% questions with a tendency towards low specificity. In addition, Nico was responsive, giving a sense of paying attention by acknowledging answers to questions. To build rapport, I based the dialogue design on behaviors suitable

Phase One	Phase Two	Phase Three
User: Nico, first you're going to look at the ratio between the 6 and the 2. You can set it equal to 9.	*User*: Nico, first you're going to look at the ratio between the 6 and the 2. You can set it equal to 9.	*User*: Nico, first you're going to look at the ratio between the 6 and the 2. You can set it equal to 9.
Nico: Okay but I'm not sure I get it. How do we do that?	*Nico*: Oh okay! Great! We put 6 over 2 like in a fraction?	*Nico*: So, we put 6 over 2 like in a fraction, then we can use the ratio to figure it out.
User: You put 6 over 2 equal to 9 over something.	*User*: Yeah exactly and it is equal to 9 over something	*User*: Yeah because you have two ratios, so you can multiply
Nico: Oh, okay like in a fraction?	*Nico*: These problems are hard, but I think I'm getting it, thanks to you! Now do we multiply 9 by 2?	*Nico*: Oh okay! So, we can multiply.
User: Yeah exactly		*User*: Yes, so you're gonna cross-multiply to get 18
Nico: I get it. We put 6 over 2 equal to 9. And then what do we do?	*User*: Yes, so you're gonna cross-multiply to get 18	*Nico*: Sweet! So, then we have 9 times 2 is 18.
User: Then you're gonna cross multiply 2 times 9.	*Nico*: I get it; you're so helpful! What do we do after we have 18?	
Nico: So, then we multiply and then what do we do?		

Table 13.1. Example Dialogue from Each Phase of Dialogue Design

to strangers in peer tutoring with an interest in whether it is better for Nico's dialogue to model that of a stranger or a friend. In human-human peer tutoring, tutees who strangers are ask more questions and these questions tend to be shallow (Madaio, Ogan, and Cassell, 2016); tutors and tutees who are strangers are politer (Ogan et al. 2012b). Nico's dialogue already included questions; I further designed these to be shallow and politer. Table 13.1 gives an example of the dialogue.

Five participants took part in this phase, 4 males and 1 female. The results for the rapport, change in self-efficacy and qualitative observations are summarized in Table 13.2. I found that for mastery experience four of five participants in this phase were not convinced of Nico's learning, reporting Nico only "kind of learned" (P1, P2). I observed no evidence of vicarious experience. Participants' did not appear to experience or observe Nico's model of question-asking and attention through responsiveness. The majority of the participants in this phase did not exhibit any form of social persuasion; they did not feel like Nico learned and felt that they were not successful as tutors. The one participant who felt like Nico learned did not attribute Nico's success to himself. P5 felt that Nico succeeded despite his own flaws as a tutor. Finally, participants were largely neutral in the degree of rapport they felt for Nico. Few of the participants exhibited any sense of general rapport for Nico, and while participants responded in the interviews that while they 'liked'

% of Decision Rules Met		Mastery	Rapport	Vicarious Experience	Social Persuasion	Self-reported rapport	Δ in Math Self-Efficacy
0–20%	Phase 1					3.52 (1.2)	.16 (.19)
20–50%	Phase 2					3.91 (1.2)	.50 (.20)
50–75%	Phase 3					4.67 (.29)	.68 (.17)
75–100%							

Table 13.2. Results and Observations for Each Dialogue Design Phase

125

Nico (P1, P2, P3), Nico was still 'a robot, not a person' (P2, P3). Differences in verbal behaviors emerged; unlike the others, P5 praised Nico and was more inclusive.

Individuals overall did not appear to be having productive social experiences; I found no evidence of mastery, vicarious experience, social persuasion or rapport. I did find one participant with dissimilar responses, which suggests individual differences play a pertinent role in these types of experiences.

13.2.3 Phase II

In the first design phase, I attempted to facilitate experience of mastery by challenging learners to explain each step to Nico. However, they clearly did not experience mastery. It was possible the content of the problems and act of tutoring is challenge enough; for Phase II, I increased the speed at which Nico understands and I increased the level of specificity to see how this influences mastery. For vicarious experience, I kept the question-based design and responsiveness indicating attention, but I explored whether other learning practices may be more suitable to vicarious experience with Nico. Positive social behavior during learning and staying optimistic in the face of challenging problems is correlated with learning outcomes (Pampaka, Williams, and Hutcheson, 2012). I introduced positivity (i.e. "Oh okay! Great!") and optimism (i.e. "these problems are hard, but I think I'm getting it") into Nico's dialogue. For social persuasion, I had explored a subtle approach in Phase I; in Phase II, I introduced an overt design by framing messages to give outright encouragement success as a tutor (i.e. "You're so helpful!"). While these messages could be perceived as disingenuous, participants' belief in themselves may be positively influenced. Finally, for rapport, 'stranger-like' behaviors may have been distancing but it

was also possible there were too few dimensions of behavior. I incorporated additional rapport-building behaviors while maintaining a model consistent with that of a stranger. Individuals who are strangers may introduce more positivity when building rapport (Tickle-Degnen and Rosenthal 1990). Increasing Nico's positivity to build rapport aligned with supporting vicarious experience and social persuasion.

Five participants took part in this design phase, 2 females and 3 males. I found an increased number of participants (4 of 5) exhibited mastery, feeling Nico learned and that this learning was due to their tutoring. They noted that when Nico was specific, as in "Oh I guess I divide six by three?" (P7), they felt he was learning from what they told him. I still did not see any acknowledgement from the participants of vicarious experience; they did not comment on Nico's question-asking, attention, or positivity and optimism. In terms of social persuasion, while most of the participants in this phase felt Nico had learned, few felt they were "good" tutors. It is possible I over-simplified the process for explaining steps to Nico, leading learners to feel Nico was just very smart, very intelligent, as one participant noted - "I didn't explain it very well, but he was really smart, so he got it". Also, P8 felt Nico's praise, designed to socially persuade the learners, was "undeserved." Finally, average rapport was higher in this phase. All five participants in post-interviews expressed higher engagement and two expressed feelings of accountability. These two learners also had the largest corresponding changes in self-efficacy, and praised Nico, "Good job" and "Nice job."

In this second design phase, evidence of productive social experiences increased. More participants expressed a sense of mastery, higher social engagement, and behaviors expressive of rapport. However, experiences of social persuasion in the form of feeling

127

responsible for Nico's success and especially evidence of vicarious experiences of good learning practices were not substantial.

13.2.4 Phase III

In the second design phase, I observed positive responses to mastery. However, I also observed that individuals felt Nico was "very smart" and they did not experience social persuasion. The prior phase may have over-simplified the tutoring task in a way that could not be overcome by either subtle or overt persuasion and contributing to feelings that Nico's praise was disingenuous. I re-adjusted the level of challenge required in explaining to Nico how to solve the problems where Nico is slightly slower to understand than in phase II. I also did not see substantial evidence of vicarious experience; for this phase, I focused on how increasing rapport might influence vicarious experience. For rapport, I emphasized behavior which would typically be found between friends rather than strangers. For example, Madaio, Ogan, and Cassell (2016) found that tutees who are friends tend to verbalize problem-solving statements more often than asking questions. I modified Nico's dialogue to incorporate statements about problem solving.

For this final phase, I had four participants take part, 3 males and 1 female. Again, I found evidence that participants experienced mastery, implying that slowing Nico's understanding did not influence whether participants felt like Nico learned. I did find evidence of vicarious experience. Participants commented on how teaching Nico was like teaching a friend and three participants noted the positivity of Nico's learning behavior. For example, P11 stated that he "doesn't get mad" and "Nico doesn't get frustrated at you" (P13), he stays "positive." Participants also noted Nico "doesn't get distracted as people

128

tend to do" (P13) and was a good listener (P12). All four participants in this phase gave evidence of feeling socially persuaded that they taught Nico. It was "because they explained it well that he understood it" (P11, P12, P13). In post interviews, participants' comments reflected feelings of accountability in helping Nico learn as well. This phase had the highest rapport compared to the two previous design iterations; participants commented Nico "reminds me of my friend," is "pretty cool," "funny," and "cute." Verbal behaviors when interacting with Nico showed two out of the four participants praised Nico. I found their use of praise similar to prior phases. However, one participant gave Nico a little sass. (P13: "Thank you, Nico. Now get back to the questions!").

This iteration resulted in the most evidence of social experiences for enhancing self-efficacy. Participants continued to express a sense of mastery and high expressions of rapport. I finally saw evidence that individuals vicariously experienced models of good learning practices, and individuals not only felt success in the task, but they expressed feelings of responsibility for that success.

13.2.5 Cross-Phase Trends

In addition to qualitative observations, I measured rapport, self-efficacy, and learning. The average rapport and change in self-efficacy are summarized for each design phase in Table 13.2. While participants in different design phases experienced different interactions, I explored cross-phase trends for insight into overall design directions. I found a significant correlation ($r = .71$, $p = .02$, $n = 10$) between rapport ($M = 4.0$, $SD = .9$) and learning gains ($M = .47$, $SD = .2$). I also found that rapport is significantly correlated with change in self-efficacy ($M = .46$, $SD = .3$) across all participants ($r = .62$, $p = .03$, $n = 14$). Change in self-

efficacy was not correlated with learning gain (r = .46, p = .17, n = 10). These results support the theoretical argument that rapport is related to learning and self-efficacy and imply designing to enhance rapport may result in positive outcomes.

I did not observe a correlation between self-efficacy and learning. While this may be due to the iterative design and the small sample size, it is possible the relationship is obscured by individual differences. Within each phase I observed a single individual who was very socially engaged, from their self-reported rapport to their interview responses and verbal behaviors. Regardless of phase, these individuals praised Nico more, included Nico in the learning process with inclusive language, and were more likely to anthropomorphize Nico. Viewing Nico as socially and cognitively capable, these learners had high social responses, a low bar for social experience, and higher gains. Comparatively, individuals with the lowest rapport and the lowest change in self-efficacy (P3, P6, and P8) responded to Nico with less inclusive language, little to no praise, and spoke of Nico as "the robot." I found two other individuals, who interacted with Nico in phase III, reported initial self-efficacy scores as low as P3, P6, and P8; however, their change in self-efficacy was much higher. This suggests the third design phase may have been more effective for individuals with low self-efficacy.

13.3 SIX DESIGN RECOMMENDATIONS

In this paper, I described Nico, a fully autonomous teachable Nao robot for mathematics learning that can interact with learners using natural language. I explored through iterative design (1) how different dialogue design strategies can foster four social, self-efficacy experiences: mastery, vicarious learning, social persuasion, and rapport and (2) how

individual differences influence responses to different dialogue design strategies. The final design, which yielded the highest self-reported feelings of self-efficacy and rapport, was the most successful at fostering these four experiences. It consisted of human-robot dialogue based on two human learners who are friends and introduced a moderate level of difficulty for achieving mastery experience. Overall, I found several design suggestions:

1. For mastery, dialogue design should provide the learner with the impression they are effective; if the robot reaches an answer too quickly, this reduces feelings of effectiveness. Design that incorporates equal question-asking with problem-solving statements can facilitate mastery.

2. For both vicarious experience and social persuasion, the analysis suggests if learners do not feel adequate rapport, they are less likely to have genuine social experiences, and this will influence their overall self-efficacy. This implies initially focusing design on fostering rapport.

3. To foster rapport, designing dialogue based on that of friends may produce stronger responses. I am not suggesting designing a robot to act like a long-time friend from the first interaction but targeting initial design strategies to incorporate 'friend-like' moves in initial interactions.

4. Different dialogue designs interact with learners' social predispositions and attitudes towards robots. Problem-solving statements, positivity, and high specificity may increase positive effects for individuals who are less inclined to social interaction with robots and may influence a positive change in self-efficacy.

5. For individuals with initially low self-efficacy, design for fostering social experiences is more critical. Individuals with initially high self-efficacy responded

positively across all phases while individuals with initially low self-efficacy responded positively only to the third design phase.

6. Gesture design should potentially differ depending on the learner's initial level of self-efficacy. I did not perform a full analysis of the design of gesture, keeping emblematic gestures and autonomous life consistent. However, individuals with lower self-efficacy strongly disliked Nico's autonomous life movement, while individuals with high self-efficacy preferred it.

13.4 CONCLUSIONS

With Nico and the work performed in this chapter, some of the limitations observed in Quinn's dialogue in Chapter 12 may have been addressed. With the iterative design performed here, there is support for the use of different dialogue strategies to create a more optimal learning experience. With the next iteration on entrainment, these dialogue design strategies may perform more optimally than the design used previously with Quinn.

Moving beyond the application of this dialogue design and the specific insights presented in this chapter, the results of this work demonstrate that Nico can be a suitable platform for exploring a larger space of design questions in learning companion interactions. Unlike Quinn, Nico has the potential as a humanoid robot to explore both the effects of dialogue and gesture on social learning experiences. For example, we can explore how small variations in the design of dialogue and the combination of dialogue and gesture create social experiences that might have a large impact on learning and motivation. In the following chapters, this platform is used to iterate on the design of entrainment and explore how it influences rapport and learning.

CHAPTER 14

EFFECTS ON RAPPORT AND LEARNING WITH NICO

In Chapter 12, the exploration of pitch entrainment with Quinn, a Lego Mindstorms robot, resulted in positive effects of entrainment on perceptions of social presence but there were no learning effects. There are several possible explanations for the lack of effect on learning. On the one hand, the implementation of entrainment may have been overly simple. On the other, many students were at ceiling on the posttest, and thus the domain content may have been too easy and prevented us from detecting effects.

In this chapter, I iterate on the design of entrainment taking into consideration the results from Quinn as well as insight from human-human interactions. With Quinn, entrainment was modeled as proximity on pitch, meaning the robot matched its pitch to the user turn-by-turn. Human-human entrainment and theory on rapport suggest that as an alternative design, convergence may be more optimal for building rapport. Convergence occurs when speakers grow more similar over a series of dialogue turns, over time. While both proximity and convergence on pitch have been found to be related to learning (Thomason, Nguyen, & Litman, 2013; Ward & Litman, 2007; Sinha & Cassell, 2015) and rapport, rapport is often defined according to three constructs: attention, positivity, and coordination (Tickle-Degnen & Rosenthal, 1990). Theoretically, as coordination increases, so does rapport. Entrainment as convergence is an ideal model of increasing behavioral coordination, suggesting that an agent which converges may build more rapport. A learner who feels more rapport may learn more as described in the background work on learning companions in Chapter 9.

This study introduces the next design iteration on entrainment as pitch convergence. Pitch convergence is implemented as part of the interaction mechanisms of Nico and I explore its effects on learning with middle school students. Nico was introduced in Chapter 12 and is a Nao robot that learners can teach how to solve ratio problems. It interacts with the learner using spoken dialogue and realistic gesture. In addition, Nico uses social dialogue based on the most successful dialogue design introduced in Chapter 12. This design includes social dialogue strategies found in other AIED systems like praise, enthusiasm, politeness, and inclusive language (Saerbeck et al., 2010; Lane et al, 2015; Maldonado et al., 2005). Dialogue of this sort has been shown to influence learning and was the most successful design in the results from Chapter 12.

I evaluate the influence of acoustic-prosodic entrainment with Nico using three conditions: **a social-entraining** condition, where Nico entrains and speaks socially, and two baseline conditions: a **social** baseline, where Nico speaks socially but does not entrain, and a **non-social** baseline, where Nico neither speaks socially nor entrains. These three conditions are analyzed in terms of effects on social responses and learning, and, prompted by prior work and the findings from Chapter 12, the responses of different genders are also analyzed. For this chapter, three research questions are proposed which are related to the overall goals of this thesis:

RQ 1: How does entrainment as pitch convergence influence learning in interactions with a social, robotic learning companion?

RQ 2: How does a social robotic learning companion which entrains via pitch convergence build rapport?

RQ 3: How are the effects of pitch convergence mediated by user gender?

Overall, this thesis is focused on understanding how we can model entrainment, the effects of entrainment on rapport and learning, and increased understanding of human-human and human agent interactions. Answering these proposed questions will provide insight into these over-arching questions. It seems highly probably that when Nico entrains via convergence and speaks socially, learners will report feeling more rapport and achieve greater learning gains when compared to the social and non-social baselines. With the enhanced social dialogue from Chapter 12, the social baseline should result in higher rapport and learning gains than the control.

The next section briefly reviews the design of Nico as a learning companion. In 14.2, I describe the implementation of acoustic-prosodic entrainment as pitch convergence. The third section describes the evaluation study at two middle schools with 72 participants. The results of this study are given in Section 14.4, and I end the chapter with a discussion of the implications of these results.

14.1   NICO (REVIEW)

This study made use of the same version of Nico as described in Chapter 13. As described in Chapter 13, Nico was an autonomous, Nao teachable robot for middle school mathematics. The domain content and method of interaction are the same as described previously. In this study, Nico's dialogue is based on the results from Chapter 13. The baseline, non-social dialogue pulls from the dialogue design strategies for encouraging mastery and was meant to foster learning by encouraging students to attend to the problem and elaborate on their knowledge (Roscoe & Chi, 2007). Nico's responses encourage learning by prompting for explanations and providing feedback to encourage learners to

	Non-Social Dialogue	Social Dialogue	Strategy
Nico:	Okay so I setup a ratio with 2.5 over 10. But then what do I do?	*Awesome*! So, *we* setup a ratio with 2.5 over 10. But then what do *we* do?	*Enthusiasm Inclusive*
Learner:	So, then you can reduce it, or I guess simplify it to 1 over 4	So, then you can reduce it, or I guess simplify it to 1 over 4	
Nico:	Oh, I simplify? How do you know I can do that?	*Sounds great*! So, *we* simplify? How do *we* know *we* can do that?	*Enthusiasm Inclusive*
Learner:	Because we know that 2.5 can go into 10 four times	Because we know that 2.5 can go into 10 four times	
Nico:	Oh, I get it. So, then I have 1 over 4?	Oh, I get it! *Thank you for explaining, [learner name]. You're a great tutor.* So, then *we* have 1 over 4?	*Politeness Praise Name usage*

Table 14.1. Example of Non-Social and Social Dialogues with Nico

feel that they are succeeding in helping Nico.

Nico can also speak socially. This social dialogue is designed to be social based on the third design phase of Chapter 13. The dialogue is also augmented with social content introduced in prior work (Saerbeck et al., 2010; Lane et al, 2015; Maldonado et al., 2005) and thought to be successful in building rapport. These behaviors included using the learner's name, inclusive language, praise, enthusiasm, and off-task dialogue. Table 14.1 depicts social and non-social dialogue.

14.2   DESIGNING PITCH CONVERGENCE

In this iteration, entrainment is implemented as local convergence on pitch. Convergence is a form of entrainment where speakers gradually grow closer in their speech features over time; they adapt over the course of a conversation. Local convergence refers to this phenomenon happening on a local, turn-by-turn level. Individuals converge towards one another over a series of turns and then 'reset,' moving apart, typically when there is a
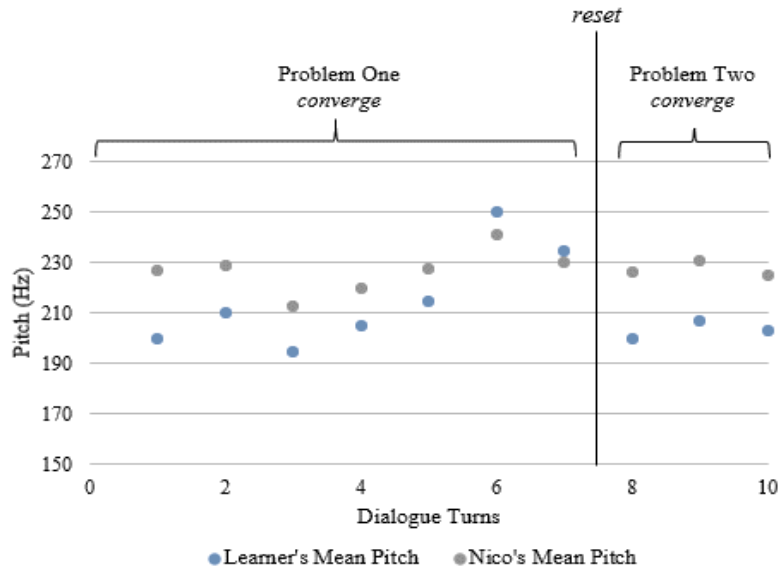
Figure 14.1. Mean Pitch Values for a Learner and Nico with Entrainment

change in topic or context. I continue, as in prior work, to focus on pitch for the implementation of convergence as pitch is known to be an important feature for conveying metacommunicative information. Pitch convergence has also been found to be related to learning in prior work and was weakly correlated with rapport.

I explored local convergence on pitch by gradually matching Nico's mean pitch to the learner's mean pitch over a series of turns. The learner's mean pitch was extracted from their immediate prior utterance. Nico would speak with a mean pitch that was closer and closer to the learner's pitch at each turn. To adapt Nico's pitch, the entrainment algorithm builds on results of Chapter 11, using the same method of shifting the pitch contour found to perform successfully in that work. This method involved shifting the text-to-speech (TTS) output up or down such that the mean of the fundamental frequency of the TTS utterance matches a target value. That target value is calculated using the mean pitch or the mean fundamental frequency of the learner's turn immediately prior.

137

This work differs from the entrainment calculation for Quinn as described in Chapter 11 with respect to the target value calculation. As described in Chapter 11, the robot mirrored the learner's mean pitch, meaning the target value was the mean pitch from the learner's utterance, target value = learner pitch. In this chapter, the calculation mimics local convergence by considering the number of turns which have passed, whether this is a new problem, and Nico's current mean pitch. Within a single problem context, the distance between Nico's mean pitch and the learner's mean pitch is gradually reduced. The target value to shift Nico's pitch is determined by the learner's pitch and the number of exchanges that have passed (one exchange = learner speaks, Nico speaks). Depending on the number of exchanges that have passed, Nico's pitch is shifted to be within a certain range of the learner's pitch (e.g., 0-1 exchanges: 50 Hz, 2 exchanges: 40 Hz, …, > 8 exchanges: 0 Hz). Thus, after 8 exchanges, Nico's mean pitch will equal the learner's mean pitch. I identified 8 as the number of exchanges to which Nico should converge based on the average number of exchanges per step in 18 pilot evaluations. When Nico and the learner moved to a new problem, Nico would 'reset' and temporarily stop converging for one turn. Nico has a baseline pitch of approximately 230 Hz. To 'reset,' Nico speaks with a pitch at that baseline. Figure 14.1 depicts the changing mean pitch values as Nico converged and reset to the learner over a series of turns across two problems.

One additional restriction was placed on the adaptation. Nico will only adapt up to ±75 Hz, to reflect a realistic entrainment distance. Nico speaks with the same voice for both males and females, a version of the default Nao text-to-speech voice, with a baseline pitch of 230 Hz. This means Nico will adapt within the range of 155 Hz – 305 Hz. I tested

the pitch convergence with four middle school students (2 female/2 male) and validated that the mechanisms of entrainment were consistent.

14.3   METHODOLOGY AND PROCEDURE

Using the pitch convergence approach, I conducted a between-subjects experiment in which learners taught Nico how to solve ratio-based problems in one of three conditions: (1) **non-social**: Nico exhibits dialogue meant to foster a learning experience and does not introduce social dialogue or entrainment, (2) **social**: Nico encourages social interaction and rapport through social dialogue, and (3) **social + entrainment**: Nico introduces equivalent social dialogue and additionally entrains via convergence on pitch. Across all three conditions, the experimenter instructions and the content of the activity were held constant. These three conditions mirror the experiment performed with Quinn in Chapter 12 to provide additional insight into how this new iteration on entrainment performs with an enhanced version of social dialogue.

14.3.1 Procedure for Exploring Pitch Convergence

Participants were 72 middle-school students from two public middle schools in the Southwestern United States. 51% of the students were recruited from one school and 49% from the other, with a mean age of 11.25 (SD = 0.47). The gender breakdown is given in Table 14.2 along with statistics regarding the dialogue. Sessions lasted 60 minutes and took place at the participant's school. As shown in Figure 14.2, students sat a desk with a Surface Pro tablet in front of them. Nico stood on the desk next to the Surface Pro, to the right of the participant. Three participants experienced technical issues during the experiment and

	Females	Males	Total Turns M (SD)	Words per Turn M (SD)
non-social	13	11	141.7 (37.0)	8.13 (4.5)
social dialogue	13	11	124.9 (28.6)	8.56 (3.4)
social dialogue + entrainment	13	11	123.8 (26.5)	10.7 (4.9)

Table 14.2. Gender Breakdown and Dialogue Statistics per Session

were excluded from the results. Thus, 22 participants remained in the non-social, 23 participants in the social condition, and 22 participants in the social-entraining condition.

Participants began with a 10-minute pretest and a short pre-survey to evaluate their initial self-efficacy towards math and tutoring. The participants were then given a few minutes to review the ratio problems and the worked-out solutions. After watching a short video depicting how to interact with Nico, students engaged in a teaching activity with Nico for 30 minutes. After the activity, they completed a 10-minute posttest and a survey on self-efficacy, rapport, and their goals.

14.3.2 Measuring Learning

To measure learning, I utilized a pretest-posttest design with an A and B form of the test. The two forms were isomorphic and counter-balanced within condition (half of the

Figure 14.2. Students Interacting with Nico at the Two Middle Schools

participants in each condition received test A as the pretest with test B as the posttest, and vice versa). The tests consisted of 10 questions around ratios, mostly procedural with some conceptual. Examples are shown in Figure 14.3. The full pretest and posttests can be found in Appendix B. I piloted and iterated on the design of the questions through 18 pilot studies, evaluating timing and applicability of questions. To analyze learning, I used both the pretest and posttest scores in statistical analyses and I also calculated the normalized learning gains according to Hake (2002). If the posttest was lower, I used (2):

$$\text{gain} = (\text{posttest} - \text{pretest}) / (1 - \text{pretest}) \qquad (1)$$

$$\text{gain} = (\text{posttest} - \text{pretest}) / (\text{pretest}) \qquad (2)$$

14.3.3 Measuring Rapport

I measured both self-reported and linguistic rapport in this study. Self-reported rapport was assessed with 12 questions designed and developed based on Tickle-Degnen and Rosenthal's understanding of rapport as being composed of three parts: attention, positivity, and coordination. Questions for positivity were developed based on measures of rapport proposed by Gratch and colleagues as well as the definition of positivity. For

Problem 1	
Activity	**Time (min)**
Treadmill	17
Weight Lifting	22

What is the ratio of time lifting weights to all activity time?

Problem 2			
Hours	46	72	
Guests	2.3	3.6	5

Can you think of a rule to describe the data in the chart?
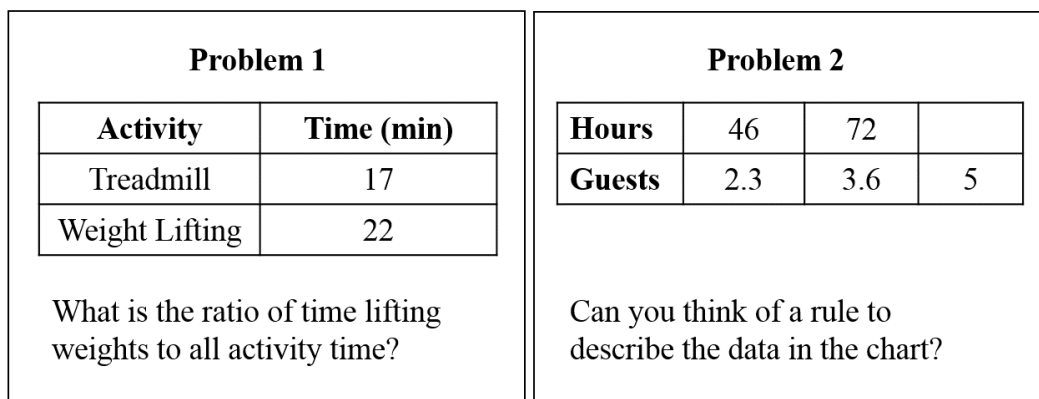
Figure 14.3. Example Problems from Pretest and Posttest

		M	SD	k
Praise	"Great job", "Good answer"	1.2	2.5	.75
Politeness	"thank you", "you're welcome"	.42	.9	.80
Inclusive	'we' or 'lets'	9.8	11.9	.97
Name	"That's right, Nico", "So Nico…"	1.8	4.4	.96
Empathy	"Me too, Nico", "I can help you"	.54	1.0	.71

Table 14.3. Descriptive Statistics and Kappa Ratings for Linguistic Rapport

attention, I designed the questions based on the attentional component of the Networked Minds Social Presence Inventory which was used to measure social presence with Quinn in Chapter 12. I drew upon measures proposed by Sinha and Cassell (2015) for coordination. Given the age group, I designed and iterated over the questions in a series of 14 pilot studies, adjusting the questions to target the desired measures while still being understandable to middle schoolers. I finalized four questions assessing positivity, four questions measuring attention, and four questions for coordination (Appendix B). I averaged the rapport questions to create a single representative construct with an acceptable internal reliability (Cronbach's α = 0.83).

To assess linguistic rapport, I coded for similar verbal rapport-building behaviors as in Chapter 12. I assessed the learners' dialogue for elements of linguistic politeness, including praise, formal politeness, inclusivity, and name usage. I utilized the same coding scheme as given in Appendix A with one modification. In addition to praise, politeness, inclusivity, and name usage, some learners exhibited empathy for Nico through their dialogue. For example, Nico would prompt the student for help by saying "I'm not sure what to do" and some students responded to this with statements such as "That's okay! I can help you." In addition to the verbal rapport behaviors for praise, formal politeness, inclusivity and name usage, I also coded for instances where the student exhibited empathy

for the robot. Examples of the behaviors can be found in Table 14.3. Two individuals each independently coded the dialogues for these behaviors. The average Cohen's kappa for these behaviors was 0.84. Individual kappas are reported in Table 14.3 along with the means and standard deviations for the behaviors as they occurred across all conditions. For analysis, I summed the total observed behaviors into a single representative construct of linguistic rapport for each participant; the statistics for the total linguistic rapport are given in Table 14.4.

14.3.4 Self-Efficacy and Other Measures

In addition to rapport and learning, measures of self-efficacy around math and tutoring were also collected. The measures were based on work by the Friday Institute for Technology (2008) and guidelines set forth by Albert Bandura on measuring self-efficacy. The questions were on a Likert scale of 1 to 5 and included four questions on their tutoring self-efficacy, such as "I can help others learn", and four questions on their self-efficacy towards math and ratios, such as "I am good at math." The full set of questions can be found in Appendix B. Participants answered all eight questions both before and after interacting with Nico. I averaged the tutoring and math questions to obtain four scores: tutoring self-efficacy (Cronbach's $\alpha = 0.47$) and math self-efficacy (Cronbach's $\alpha = 0.58$) prior to interacting with Nico and tutoring self-efficacy (Cronbach's $\alpha = 0.69$) and math self-efficacy (Cronbach's $\alpha = 0.62$) post interacting with Nico.

The results of the work with Quinn in Chapter 12 suggested that individual differences such as those hinted at by gender may influence how one approaches and perceives social behavior from a robotic learning companion. These individual differences

may encompass an individual's prior experiences, bias, and interaction goals. In an initial exploration of these kinds of variables which lead to individuals perceiving Nico's social behavior differently, I asked participants what their goal was in teaching Nico. In learning interactions, individuals with social goals may learn more (Ogan et al., 2010) and respond to social behavior from the robot differently. Three people hand-coded the goal responses for social goals versus task goals.

14.4 RESULTS OF PITCH CONVERGENCE ON RAPPORT AND LEARNING

I report the results for learning and rapport where individuals interacted with Nico, the teachable robot, in one of three conditions: a **social-entraining** condition where Nico was both social and entrained, a **social** condition where Nico was only social, and a **non-social** where Nico was neither social nor entraining. Studies were conducted across two schools. After analyzing differences between schools, there were no significant differences or interactions with school by condition or gender on learning or rapport. I therefore report the results without the additional factor of the school.

14.4.1 Learning Results

With learning, I explored whether the social-entraining condition resulted in greater learning than the social and non-social baselines by analyzing learning gains in a two-way analysis of variance (ANOVA) with condition and gender as the independent variables and gain as the dependent variable. Table 14.4 gives means and standard deviations for gain by condition and gender. I found the gain was significantly different across conditions, $F_{(2, 63)} = 6.06$, $p = 0.004$. Partial eta squared was .16, a medium effect size. Gender was not

	Non-Social	Social	Social-Entraining
Learning Gain	.04 (.08)	.15 (.04)	.23 (.04)
Pretest	.20 (.04)	.20 (.03)	.24 (.04)
Posttest	.27 (.04)	.33 (.04)	.42 (.04)
Self-Reported Rapport	4.1 (.12)	3.8 (.18)	4.1 (.13)
Linguistic Rapport	10 (2.0)	22 (3.3)	26 (3.9)
Pre-Tutoring Self-Efficacy	3.5 (.72)	3.2 (.70)	3.2 (.68)
Post Tutoring Self-Efficacy	3.4 (.77)	3.1 (.94)	3.3 (.90)
Pre-Math Self-Efficacy	3.3 (.55)	3.2 (.89)	3.3 (.81)
Post Math-Self-Efficacy	3.3 (.78)	3.2 (.92)	3.4 (.74)
Dialogue Error (WER)	22.2 (8.3)	24.9 (10.4)	25.4 (8.1)

Table 14.4 Descriptive Statistics for Learning and Rapport Across Conditions

significant, $F (1, 63) = .05$, $p = .82$ and the gender by condition interaction was not significant $F (2, 63) = 2.13$, $p = 0.12$. Tukey post-hoc analyses indicated significant pairwise differences for the social-entraining condition, with the social-entraining condition resulting in significantly more learning than the non-social ($p=.005$). The social condition approached a significantly higher gain than the non-social ($p = .06$). I did not find any differences between the two social conditions, ($p = .6$).

The findings regarding learning were further supported with a repeated measures ANOVA with pretest and posttest as the dependent variables and condition and gender as the independent variables. I observed a significant effect of condition on learning, $F (2, 63) = 3.56$, $p = .03$. Gender was not significant, $F (1, 63) = .9$, $p = .34$. The condition by gender interaction on learning was also not significant though it does appear to be approaching significance, $F (2, 63) = 2.77$, $p = .07$. Evaluating the differences between conditions with

post-hoc analyses reveals similar outcomes with the social-entraining resulting in significantly higher learning than the non-social condition, p = .04.

Additionally, I analyzed differences in posttest scores while controlling for the pretest in an ANCOVA. I found a significant main effect for condition, replicating the previous analyses, $F (2, 62) = 5.19$, $p = .008$. Gender is not significant, $F (1, 62) = .99$, $p = .32$; however, the gender by condition interaction is significant, $F (2, 62) = 4.84$, $p = .01$. Comparison of the means shows that significant differences in the posttest for condition follow the same trend as the other analyses, with the social-entraining condition resulting in a significantly higher posttest than the non-social. In comparing the gender-by-condition interaction, it appears that males learned the most in the social-entraining condition (mean = .491) and females learned the most in the social condition (mean = .347).

Based on the theoretical relationship between learning and social motivation, it seems possible that the social-entraining condition resulted in higher learning gains due to enhancing social responses. This is explored further with the analysis of rapport in the next sections. There is a potential gender-by-condition interaction regarding the performance on the posttest, suggesting that males may have responded more positively to the social-entraining condition but there is no general learning by gender interaction.

14.4.2 Self-Reported Rapport Results

Given the results found with Quinn in Chapter 12 and the theoretical relationship between rapport and entrainment, we can hypothesize rapport to be higher for the social-entraining condition. The self-reported means and standard deviations for rapport by condition are given in Table 14.4. I first explored if rapport and learning gain were correlated. They were

not: Pearson's r = -.115, p = .347. I then explored if rapport differed by condition by analyzing rapport in an ANOVA with condition and gender as independent variables. I found the hypothesis was rejected. There were no significant differences in rapport across conditions, $F_{(2, 63)}$ = .751, p = .48, $\eta2$ = 0.02. There was also no effect of gender, $F_{(1, 63)}$ = .04, p = .84, $\eta2$ = .001 or gender by condition, $F_{(2, 63)}$ = 1.49, p = .23, $\eta2$ = 0.04. I found the lack of difference in self-reported rapport across conditions surprising, especially given that there were significant differences in learning. Prior work has suggested that the length of dialogue turns may play a role in learning and potentially rapport (Rose et al, 2003; Litman et al., 2006). While every learner interacted with Nico for thirty minutes regardless of condition, there may have been differences in number of turns and the number of words per turn issued by each learner. The means and standard deviations for the total number of turns and total number of words are given in Table 14.2. I explored whether the total number of dialogue turns and the average number of words per turn for each learner played any role in their responses. However, I did not find any differences across conditions in the number of turns exchanged, $F_{(2, 63)}$ =1.22, p = .30, or words used, $F_{(2,63)}$ =1.7, p = .19. I also did not find any significant influence of turns or words on rapport or learning.

There are several possible explanations for why self-reported rapport did not differ across conditions as one might expect it would. One explanation may be that Nico was very successful in building rapport across all conditions and that the measure 'hit ceiling.' An alternative possibility is that a single post-session survey may not capture the changes in rapport which would reveal it increasing within and across conditions. Self-reported

rapport is also notoriously hard to measure. Analysis of linguistic rapport, given in the next section, may provide insight into whether one of these explanations is possible.

14.4.3 Linguistic Rapport Results

Like self-reported rapport, we might hypothesize linguistic rapport to be higher in the social-entraining condition. I analyzed differences in linguistic rapport with an ANCOVA, controlling for the length of dialogue as a covariate. I found significant differences across conditions, $F (2, 62) = 7.39$, $p = .001$. I did not observe significant differences between males and females, $F (1, 62) = 2.3$, $p = .137$ nor was there a significant gender by condition interaction, $F (2, 62) = .98$, $p = .38$. Post-hoc tests on the estimated marginal means with Bonferroni correction indicated that linguistic rapport in the social-entraining condition was significantly higher than in the non-social control, $p = .001$. The social condition also had significantly more linguistic rapport than the non-social control, $p = .02$. These findings suggest that, given the linguistic rapport changes across conditions, social factors are being influenced by the robot's social behavior. In the next section, I explore the relationship between linguistic rapport, self-reported rapport and learning.

14.4.4 Relating Self-Reported Rapport, Linguistic Rapport, and Learning

In the exploration of entrainment with Quinn in Chapter 12, I found that for some individuals, self-reported rapport was not always positively reflected in their linguistic behavior. For females, higher use of linguistic rapport was related to lower feelings of social presence. I investigated whether the same phenomenon re-occurred here by first exploring whether the self-reported measures for rapport were correlated with use of

linguistic rapport. I found a significant, positive correlation between linguistic rapport and self-reported rapport, $r(65) = .317$, $p = .007$. The strength of the correlation is moderate but present. Individuals who self-reported higher rapport also used more linguistic rapport. Analyzing the relationship between self-reported rapport and linguistic rapport by gender, the correlation trends positive for both males and females. For males, the correlation is significant, $r(30) = .39$, $p = .02$; it is not significant for females, $r(35) = .26$, $p = .10$.

Turning to learning, linguistic rapport was significantly correlated with posttest score, $r(65) = .25$, $p = .03$. However, a partial correlation between linguistic rapport and posttest controlling for pretest was not significant, $r(64) = .16$, $p = .19$. I analyzed this further by exploring by gender. I found that for males, the relationship between linguistic rapport behaviors and posttest scores while controlling for pretest was significant, $r(29) = .39$, $p = .02$. For females, it was not significant, $r(34) = .04$, $p = .80$, suggesting that for males, the use of linguistic rapport behaviors may be more indicative of their learning.

14.4.5 Self-Efficacy

I measured self-efficacy on tutoring and math as a part of the pre and post surveys given to participants. The means and standard deviations are reported in Table 14.4. Traditionally, self-efficacy is not measured pre to post on such micro-scale interactions (where the intervention is not recurring and is of short duration); however, I was curious whether responses might change from pre to post interaction with Nico. I explored this with a repeated measures ANOVA with the pre and post self-efficacy scores as the dependent variables and condition and gender as the independent variables. I performed this analysis with both the math self-efficacy and tutoring self-efficacy scores. I did not observe

significant effects for math on condition, $F (2, 63) = .08$, $p = .9$, or gender, $F (1, 63) = 1.04$, $p = .31$. I also did not observe significant effects for tutoring on condition, $F (2, 63) = .40$, $p = .67$, or gender, $F (1, 63) = .64$, $p = .43$.

I also explored whether the post self-efficacy scores appeared to be related to any of the other measures I collected, including self-reported rapport, linguistic rapport, and learning gain. Post math self-efficacy was positively correlated with rapport, $r (67) = .24$, $p = .04$ but not linguistic rapport, $r (67) = .01$, $p = .9$, or gain, $r (67) = .10$, $p = .41$. Post tutoring self-efficacy was highly correlated with rapport, $r (67) = .53$, $p < .001$. Like math self-efficacy, tutoring self-efficacy was not correlated with the gain, $r (67) = .01$, $p = .42$, or with linguistic rapport, $r (67) = .15$, $p = .21$. I explored whether the math and tutoring self-efficacy scores collected prior to interacting with Nico were also correlated with rapport to explore whether it was simply individuals who had higher self-efficacy in general who experienced more rapport. However, neither score was correlated with rapport (pre-tutoring self-efficacy: $r (67) = .10$, $p = .39$, pre-math self-efficacy: $r (67) = .004$, $p = .97$). This suggests that there was a deeper connection between experiencing higher rapport and reporting higher self-efficacy than can be explained by an individual's initial self-efficacy and suggests that the interaction may have potential for building self-efficacy.

14.4.6 Validating Dialogue Errors

Lucas and colleagues found that when an agent makes a conversational error, the negative effects can be boosted or exaggerated if the agent also incorporates social dialogue. I examined the role of dialogue errors in the above results to determine if potentially speech

recognition errors and incorrect dialogue responses were the contributing to the significant differences observed for learning and linguistic rapport. I measured the word error rate as

$$WER = \frac{S + D + I}{N}$$

where S is the number of substitutions, D is the number of deletions, I is the number of insertions, and N is the total number of words. The average word-error rate for Nico was 24.15; the WER for each condition is given in Table 14.4. Analyzing if the WER was significantly different across conditions or by gender with an ANOVA, we find that it is not, $F(1, 63) = .734$, $p = .601$. Additionally, treating the WER as a covariate in the analyses of gain and linguistic rapport, the WER does not significantly reduce the significant difference of learning gains across conditions, $F(2, 63) = 5.53$, $p = 0.006$, nor does it alter the significance of linguistic rapport, $F(2, 63) = 5.25 = .008$. Dialogue errors potentially introduced by the low WER were not responsible for the differences in learning gain and linguistic rapport observed across conditions.

14.5   DISCUSSION AND CONCLUSIONS

In this chapter, I explored prosodic entrainment on pitch with the teachable robot, Nico. This exploration built on the previous iteration of entrainment evaluated in Chapter 12; instead of simply matching, Nico mimicked local convergence on pitch. Nico also utilized a new form of social dialogue as explored in Chapter 13. Exploring the effects of this new design iteration, entrainment with social dialogue significantly improved learning when compared to the condition where Nico was not social. This was the first time that an

implementation of acoustic-prosodic entrainment in an agent has shown positive effects on learning and suggests that entrainment may be a useful mechanism for enhancing learning.

Entrainment and social dialogue were also expected lead to higher feelings of rapport. However, learners did not self-report feeling more rapport for the social, entraining companion; instead I found that they were overwhelmingly positive across the board, self-reporting at an average of 4.0 (.7) on a Likert scale of 1 to 5, which is higher than might be expected with such a scale. The mean would be expected to be closer to 3 if individuals were feeling more negative regarding the interaction (Johns, 2005; Garland, 1991). Analyzing linguistic rapport, individuals engaged in verbal rapport-building behaviors differently across conditions. Their use of these behaviors was significantly correlated with their self-reported rapport and with their post-test scores. Individuals who were the most engaged in linguistic rapport tended to be those who interacted with a social, entraining Nico. It appears that Nico's social behaviors may have facilitated greater social engagement which then led to greater overall engagement. In the next section, I look in more detail at the dialogues to find further support for this and discuss an alternative theory that entrainment may have facilitated cognitive factors. I also touch on the results regarding gender and possibilities for why we do not see the same results we saw with Quinn.

Differences in High Gain Learner Dialogue vs. Low Gain Learner Dialogue

I compared the dialogues of individuals who had higher learning gains to the dialogues of individuals with lower gains. There are obvious differences in how they interacted with Nico. The individuals with high gain exhibited many more of the rapport building behaviors, encouraging Nico with statements such as "Yes, you got it!" and praising Nico,

"Good job, you did it!" The individuals with high gain were also much more likely to respond to Nico's questions about how and why a step was performed as opposed to low gain learners who were more likely to side-step Nico's questions. Table 14.5 gives examples of two dialogues from a high gain learner and a low gain learner. The high gain learner was also in a social-entraining condition, so Nico's responses include social content. The high-gain learner attempts to explain to Nico why Nico is supposed to divide while the low-gain learner does not explain why Nico should multiply. The high-gain learner also exhibits more inclusive language, uses Nico's name, and responds with a confirmation when Nico gets the answer. It seems possible Nico's social behaviors may have positively influenced social factors. These social factors may have contributed to individuals engaging more with Nico which may have influenced their cognitive behavior.

Entrainment and Cognitive Factors

There is an alternative theory regarding the origin of entrainment which suggests that entrainment might influence cognitive factors through means such as grounding. The Interactive Alignment Model (Pickering and Garrod, 2012) suggests entrainment is an observable outcome of individuals aligning on multiple levels of dialogue from their acoustic-prosodic features to their understanding and knowledge of a situation. With this model, Nico's entrainment could have facilitated learning by prompting learners to converge towards Nico and Nico's understanding of how to solve the problem. For example, the learner might explain to Nico that Nico "needs to multiply by two." Nico may elaborate on this statement, recognizing the need to multiply by two because there are twice as many bags, "Oh because I have two more bags? I have twice as many and multiply by

153

two?" Even though learners were given the worked-out solutions, they may not always have had full domain knowledge and so Nico's queries led to deeper understanding of the problem. With local convergence, learners were invited to converge to Nico and Nico's understanding. As Nico and the learner converged, this may have facilitated deeper understanding of the domain content.

There is evidence that this occurs in human-human peer tutoring. Sinha and Cassell (2015) explored relationships of learning, convergence and rapport in dyads of peer tutors with a mean age of 13. The found relationships between learning and convergence and suggest that a virtual peer that both converges to its human partner and invites convergence may be a more effective learning partner.

High-Gain Learner	Low-Gain Learner
User: You need to divide twelve by three to get your answer	*User*: Ten times two point five is twenty-five.
Nico: Okay! So, with the twelve. You think we divide twelve? What do we divide it by?	*Nico*: Okay I multiply by ten. Why do I multiply by ten?
User: We divide it by three	*User*: Yes.
Nico: Got it! You're on top of this! But can you explain, why do we divide by three?	*Nico*: So, ten times two point five is twenty-five. Then what do I do?
User: We divide by three because six…if you have two then you must figure out what's, um, to- how to get the six. So, if you do six divided by three it's two and then you're going to do that down to 12. 12 divided by three.	*User*: So, your answer is twenty-five.
	Nico: I can go ten miles in one hour and twenty-five miles in two and a half. I will put the answer is twenty-five.
Nico: Sweet! Thank you! So, if we divide twelve by three, that's four.	*User*: [moved to next step]
User: Correct, Nico!	

Table 14.5. Dialogues from Two Learners Interacting with Nico

154

I explored whether this might be occurring in the dialogues with Nico by exploring how learners entrained lexically to Nico across the different conditions, with the theory that if learners were entraining lexically to Nico more in the social plus entraining condition, this could be evidence for Nico's prosodic entrainment encouraging convergence in response on multiple levels and potentially facilitating learning.

Lexical Entrainment

I used a measure for lexical entrainment introduced by Nenkova, Gravano, and Hirshberg (2008) which is thought to be more robust to potential word sparsity issues. Given the size of the individual conversations between learners and Nico, sparsity was a potential challenge in measuring lexical entrainment with the corpus. The entrainment score was calculated by first identifying a score for individual words. This score, $entr(w)$, was calculated as the negated absolute value of the difference between the fraction of times a word w is used by the learner and by Nico:

$$entr(w)_{learner} = -\left|\frac{count_{nico}(w)}{ALL_{nico}} - \frac{count_{learner}(w)}{ALL_{learner}}\right|$$

ALL_{nico} and $ALL_{learner}$ refer to the total number of all words spoken by Nico and the learner. I calculated $entr(w)$ for the top 25 occurring words in each individual dialogue, ignoring stop words. I then obtained a single lexical entrainment score for each learner by generalizing the above measure across the top occurring words for each learner as:

$$ENTR_{learner} = \sum entr(w)_{learner}$$

This resulted in an entrainment score for each learner ranging from 0 to $-\infty$, with scores closer to 0 indicating higher lexical entrainment.

155

I first analyzed whether the lexical entrainment scores were correlated with self-reported rapport, linguistic rapport, and learning. I found that entrainment was positively correlated with linguistic rapport, r (67) = .31, p = .009. When individuals exhibited greater lexical entrainment, they also exhibited more linguistic rapport behaviors. I observed no evidence of a relationship between lexical entrainment and self-reported rapport or lexical entrainment and learning with all other correlations showing p-values above 0.7.

I then analyzed whether entrainment differed by condition to answer the question of whether learners entrained more to Nico when Nico exhibited prosodic entrainment. A two-way ANOVA with condition and gender as factors and the lexical entrainment score as the dependent variable indicated that entrainment did not differ statistically across conditions, F (2, 63) = .01, p =.9, or by gender, F (2, 63) = 2.01, p < 0.16. The condition by gender interaction was also non-significant, F (2, 63) = 1.2, p = .3. The means and standard deviations for the lexical entrainment scores are given in Table 14.6.

This result suggests that it is more likely that social factors were at the heart of the learning differences between conditions then cognitive factors. I had hypothesized that if Nico's prosodic entrainment had triggered cognitive factors which contributed to learning, I should observe differences in how individuals entrained to Nico lexically. I cannot reject outright that Nico's prosodic entrainment may have influenced cognitive factors based on the results, but the relationship between lexical entrainment and linguistic rapport suggests social factors may be more pertinent.

One explanation regarding the relationship between lexical entrainment and linguistic rapport may be that individuals picked up on Nico's social behaviors and the measure of lexical entrainment and linguistic rapport simply identify how much individuals

	Non-Social	Social	Social-Entraining
Males	-.19 (.08)	-.21 (.09)	-.18 (.06)
Females	-.23 (.09)	-.21 (.11)	-.24 (.12)

Table 14.6. Descriptive Statistics for Lexical Entrainment

reflected Nico's social behavior back. However, I did not observe significant differences across conditions on lexical entrainment which I would expect had that been the case. In addition, the top words on which individuals entrained were not the same verbal rapport building behaviors which Nico introduced. I analyzed lexical entrainment based on high-frequency words; the highest occurring words were largely task-based, referring to how to solve problems and the content of the problems. It is more likely that lexical entrainment and linguistic rapport provide evidence of individuals' complex social responses.

Understanding Gender Effects

Given the prior work on gender differences, it seemed likely females might respond to the social behaviors of Nico more favorably. In my first iteration on entrainment with Quinn, females felt greater rapport for Quinn, and both males and females disliked the robot when it exhibited social behavior but did not entrain but they expressed this in different ways. Males had significantly lower self-reported scores while for females, this was more evident in their use of linguistic rapport. In this work, I did not observe gender differences to the same extent. I did find when analyzing differences on the posttest while controlling for prior knowledge that females scored higher on the posttest in the social condition while males appear to have scored higher in the social plus entraining condition. I do not see this

reflected however in their self-reported rapport or linguistic rapport behaviors.

One of the possible explanations for why I did not see gender differences with this version of a teachable robot may be differences in the age group. In the work with Quinn, the age group consisted of college students while Nico was explored in interactions with middle school students. Gender is used an indicator of individual differences which have developed over time due to one's community, culture, experiences, and exposure to gender stereotypes. It is entirely possible that with middle schoolers the individual differences indicated by gender were not as evident because they have not yet developed to the extent observed in young college-aged adults.

Another potential explanation may be that with Quinn, the robot was gendered via its voice – females interacted with a female-voiced robot and males interacted with a male-voiced robot. There was not clear evidence in the results with Quinn on how, if at all, this influenced responses. Prior work suggests that the gender of the robot should not have negatively influenced responses. In contrast to Quinn, I left Nico's gender unspecified in this study. I did not use personal pronouns in referring to Nico. Regarding the name 'Nico,' I had surveyed the 18 pilot participants and found they did not find Nico to be overtly male nor female. In the study, I avoided all reference to gender, letting the participants make up their own mind regarding the gender of the robot and the voice was gender neutral falling into either male child or female child ranges. I then surveyed participants after the study regarding whether they thought Nico was male or female. 87% of the participants reported that Nico was male. The interpretation of Nico as a male robot did not appear to influence responses but there was a large skew. It is unclear from these results whether the interpreted gender of the robot was responsible for these results, but it is possible that this may have

contributed to different responses. In the next study, we specify the robot's gender and consider how this may have influenced responses.

Conclusions

In this second, macro-iteration on entrainment with a social, robotic learning companion, the first evidence emerges that automated entrainment in a learning companion can have a significant effect on learning. Revisiting the four proposed research questions, I am interested in exploring how we can model entrainment, the effects of entrainment on social responses, on learning, and any insights into human-human and human-agent interactions. The results of this study suggest that regarding RQ1, a model of pitch convergence is more optimal than pitch proximity, for influencing social responses like linguistic rapport as well as learning. With respect to RQ2, how does entrainment perform when combined with social dialogue, entrainment on pitch appears to enhance responses to social dialogue. With RQ3, this study suggests entrainment can facilitate learning in interactions with a teachable robot. In the next chapter, I build on these findings, iterating on a new design of entrainment and building on this work to evaluate further models of entrainment, to understand the effects of entrainment, and explore insights automating entrainment might give us.

CHAPTER 15

DESIGNING MULTI-FEATURE ENTRAINMENT WITH EMMA

In this chapter, I iterate once again on the design of entrainment, revisiting and incorporating what we know about human-human entrainment to explore how entrainment can be modeled. The study with Nico in Chapter 14 established that a design of entrainment as convergence on pitch can have positive effects on learning and linguistic rapport. However, in human-human entrainment, speakers often entrain or adapt their prosody to one another on multiple features over the course of a conversation. Multi-feature entrainment has been found to be highly correlated with rapport and task-success, and the analysis of human-human entrainment in Chapter 6 high-lighted that entrainment on different features can be important at different points in dialogue. While I found successful results with convergence on pitch, entrainment on multiple prosodic cues has the potential to influence social factors even more strongly. I introduce several models for combining entrainment on multiple prosodic cues and as in Chapter 11, I evaluate these designs based on two criteria: perceived naturalness and perceived rapport.

I collected data from four individuals interacting with the different multi-feature entrainment designs and using crowd-sourced analysis via Amazon Mechanical Turk, compared the different adaptations on rapport and naturalness as perceived by third-party observers. In the next section of this chapter, I describe Emma, the companion used for this evaluation. Emma is a Nao robot that learners teach how to solve math problems. Like Nico and Quinn, students teach her how to solve ratio problems using spoken dialogue. In the next section of this chapter, I describe Emma in more detail. Section 15.2 contains the

descriptions of the different adaptations I explored. Section 15.3 describes the method and procedure for analyzing these adaptations, and the results are given in 15.4.

15.1   EMMA: A ROBOTIC LEARNING COMPANION

Emma is an autonomous social, teachable Nao robot for middle school mathematics; in design she is very similar to Nico. The dialogue system for Emma is just like that of Nico, incorporating both the basic and advanced functionality described in Chapter 10. Emma also can be social, employing the same social dialogue as evaluated with Nico. Sample dialogue from an interaction with Emma is given in Table 15.1.

In contrast to Nico, Emma is deliberately gendered to be a female robotic learning companion. The motivations and possible considerations behind this choice are given in the following section. In addition, learners teach Emma a slightly expanded domain, incorporating content related to proportions and proportional relationships for the 7th and 8th grade. This domain content is described more in 15.1.2.

15.1.1 Gendering a Robotic Learning Companion

In interactions with a computer agent or a robot which exhibits social behavior, individuals are going to make assumptions and interpretations of that behavior and these are going to be colored by the individual's community, culture, gender stereotyping, and gender assumptions regarding the robot. While many people may consciously acknowledge that robots do not have a 'gender,' individuals will still assign a gender to the robot, particularly if they are engaging in an interaction which is inherently social (Wang & Young, 2014).

This makes the robot's gender an important consideration when designing interactions. In prior work, I had left the gender of the robot unspecified and let the user interpret the gender. 87% of the users interpreted the gender of the robot to be male. This did not appear to influence results; however, evidence indicates that individuals will let the gender of the robot guide how they respond to social behavior (Nomura 2016) so in this study, I decided to control the gender of the robot by clearly assigning it a gender.

I gender the learning companion to be female, calling the robot Emma and using female personal pronouns. I do this for several reasons. Math and robots are stereotypically associated with males. Gendering the robot opposite to the stereotypical gender of the task can have positive effects on engagement (Eyssel & Hegel, 2012). Additionally, in human-

	Non-Social Dialogue	Social Dialogue	Strategy
Emma:	Before I go to the store, can you help me figure out how much of my battery I will use over time?	*Sweet*! Before I go to the store, can you help me figure out how much of my battery I will use over time?	*Enthusiasm*
Learner:	And we need- we need to figure out how much for one hour. So, let's multiply two thirds by three over two.	Yes, I can!	
Emma:	So, I know I use 1/20th of the battery in 2/3rds of an hour. I need to figure out how much I use in one hour. So, I multiply?	*Yay! Thank you! You are awesome.* Where do *we* start?	*Enthusiasm* *Politeness* *Praise* *Inclusive*
Learner:	We multiply two thirds by three over two.	Do you see that the time is one hour? So, we must try and s- either times or divide it by the, um, battery usage of one hour.	
Nico:	I think I get it. But can you explain your thoughts about why I multiply?	*Oh right! You're so smart!* So, *we* know I use 1/20th of the battery in 2/3rds of an hour. *We* need to figure out how much I use in one hour. So, *we* multiply?	*Praise* *Inclusive*

Table 15.1. Sample Dialogues with Emma

162

Problem 1

Emma's friends have been arguing over who is fastest at making s'mores! Emma has an equation for how fast Tasha is at making s'mores. Help Emma figure out an equation for how fast Zach can make s'mores.

Step	S'more Maker	Minutes (y)	S'mores (x)	Setup (b)	Slope (m)
Step 0	Tasha	8	2	4	2
Step 1	Zach	9	2	1	???

Figure 15.1. Example Problem on Functions

human peer tutoring scenarios, the tutor has been found to learn for all gender pairings of tutor to tutee except for when the tutor is female, and the tutee is male (Topping & Whitely, 1993; Falchikov, 2001; Sobral, 2002). In this scenario, the tutee tends to learn a great deal, but the tutor does not. As I am interested in fostering learning in the 'tutor' I can potentially avoid a negative pairing by making the robot tutee female.

15.1.2 Domain Content and Interface

The learning domain for Emma was middle school mathematics with a focus on 7th and 8th grade Common Core concepts regarding advanced ratios and proportions such as recognizing and representing whether two quantities are in a proportion, identifying the constant of proportionality in tables, and using and applying advanced math to understand relationships. With these concepts, I designed nine narrative-style word problems such as shown in Figure 15.1. Like the UI for Nico, the problems are designed in a table-based format and Emma requests the student's help in how to solve for the missing information.

163

15.2 DESIGNING MULTI-FEATURE ENTRAINMENT

In human-human conversations, people will prosodically entrain, or adapt their prosody to their speaking partner, on multiple features at once and this has been shown to be correlated to social factors and task success. It is challenging to model multi-feature entrainment as there tends to be inter-dependencies between features; manipulating pitch can influence intensity and manipulating speaking rate can influence pitch. Building on the work from Chapters 11, 12 and 14 which explored entrainment on just pitch, I introduce an approach for multi-feature entrainment which combines entrainment on pitch with entrainment on loudness and entrainment on pitch with entrainment on speaking rate. To model multi-feature entrainment, there are several design decisions which need to be made including what type of entrainment the system should model, how the system should implement the adaptation – as a part of the TTS system or on the output of the TTS, and in what order multiple features should be adapted. I describe decisions made and why in the next section, followed by the algorithm to model multi-feature entrainment.

15.2.1 Design Decisions

In deciding which type of entrainment to model, I based the approach on the success found with modeling local convergence. Adapting pitch as a form of local convergence, where the text-to-speech (TTS) was adapted over a series of turns, produced positive effects on learning and rapport with Nico. An approach of local convergence on multiple features may prove successful, given that adaptation over time has been theorized to be strongly related to rapport and convergence in human-human dialogues on multiple features has been related to pertinent social factors. For this work, I explore multi-feature local

convergence using a similar model as discussed in Chapter 14. I adapt the TTS to grow closer to the user on multiple features over a series of turns and then reset when a new topic is introduced.

In terms of how the system should implement the adaptation, the approach with Quinn and Nico involved transforming the TTS after it was synthesized. Using Praat, the synthesized text-to-speech was processed, manipulated, and re-synthesized to create a new TTS. In this iteration, I take an alternative approach using TTS augmentation tools; these tools usually come with a TTS engine and can be used to manipulate pitch, speaking rate, and loudness as the TTS is synthesized. The Nao robot comes with a form of these tools and I use these to adjust the TTS as it is synthesized, to modify the loudness, pitch, and speaking rate.

The order in which features are modified is important because the manipulation of one feature can influence another. Levitan and colleagues found that the degree of inter-dependencies between features varied depending on the tool utilized to transform the TTS. With Praat, manipulation of the speech rate influenced pitch by up to 10Hz; manipulating pitch with Praat influenced loudness by around 1dB. I additionally explored these interdependencies, both with Praat and the TTS system which accompanies the Nao robot. With Praat I found similar results as Levitan and colleagues, that manipulation of speech rate influenced pitch by up to 10Hz and that manipulation of pitch influenced intensity by up to 1dB. With the TTS system which accompanies the Nao robot, I found the effects of manipulating pitch on loudness appeared to be lower than with Praat, resulting in average of only 0.5dB change. The effect of speech rate on pitch was sometimes as high as 12 Hz; however, the TTS system was much faster in performing the adaptation on speaking rate.

This is an important consideration when implementing entrainment as latency has been found to negatively impact both the influence and degree of entrainment in human-human dialogues (Levitan et al., 2015). If the system is slow to produce a response, the positive effects of the entrainment may be reduced, and users may become disengaged. In addition, with the Nao TTS system, manipulating pitch after adapting speaking rate somewhat mitigated the effects of speaking rate as the pitch is re-adjusted to the specified target. In the following algorithm, pitch is manipulated after speaking rate and prior to loudness.

15.2.2 Algorithm for Multi-Feature Entrainment

The algorithm described here has the potential to be applied to either using TTS augmentation tools or transforming synthesized TTS with Praat. The algorithm is utilized here with the Nao TTS system. The high-level architecture is in Figure 15.2.

The algorithm accepts as input the user's mean as calculated by Praat, the robot's prior mean, the type of feature which is being adapted, and the number of exchanges (one exchange = user speaks, system speaks) which have passed. Within the program, the following parameters must be set: the maximum number of exchanges to which converge to, after which point the system will adapt to the user 100% unless triggered to reset, the minimum and maximum values for realistic synthesized output for each feature type, to ensure that the TTS is not adapted higher or lower than perceived natural limits, and the conversion equation to convert a Praat feature value to a value interpretable by the TTS system (if Praat is being used to transform the TTS after synthesis, this is not required). The conversion equation can be identified by collecting a set of user data, extracting a data

```
function entrainment(▯▯▯▯ ▯▯▯▯▯▯▯▯ ▯▯▯▯▯▯▯▯ ▯▯▯▯, ▯▯▯▯▯▯▯▯)
returns ▯▯▯▯▯▯▯▯

▯▯▯▯▯▯▯▯ ▯▯▯  ← convert(▯▯▯▯ ▯▯▯▯▯▯▯▯ ▯▯▯)

▯▯▯▯▯▯← percentage to adapt depending on ▯ ▯▯▯ ▯▯  ▯▯▯▯▯▯▯▯ for 100%
        convergence and ▯▯▯▯▯▯▯ the number of exchanges which have passed

▯▯▯▯▯▯▯▯ ▯▯▯▯ ← if (▯▯▯▯▯ ▯▯▯ ± (diff (▯▯▯▯▯▯▯ ▯▯▯▯▯▯▯▯ ▯▯▯)  ×  ▯▯▯▯▯▯)

if      ▯▯▯▯▯▯▯▯ ▯▯▯▯ < min (▯▯▯▯▯ ▯▯▯▯) : ▯▯▯▯▯▯▯▯ ▯▯▯▯ ← min(▯▯▯▯▯ ▯▯▯▯)

else if ▯▯▯▯▯▯▯▯ ▯▯▯▯ > max (▯▯▯▯▯ ▯▯▯▯) : ▯▯▯▯▯▯▯▯ ▯▯▯▯ ← max(▯▯▯▯▯ ▯▯▯▯)

▯▯▯▯▯▯▯▯ ← TTS (▯▯▯▯▯▯▯▯ ▯▯▯▯)

return (▯▯▯▯▯▯▯▯)
```

Figure 15.2. Algorithm for Calculating Entrainment

set for the target feature using Praat, and mapping this data set to the range of possible

values in the TTS system. The conversion equations for the Nao TTS are in Table 15.2.

Upon receiving the inputs, the user's mean is converted to a value interpretable by

the TTS system given the conversion equation. This value is then used with the robot's

previous mean to identify which direction the TTS should be adapted. If the system's prior

mean was higher than the user's current mean, the TTS is adapted down towards the user.

If the system's prior mean was lower than the user's current mean, the TTS is adapted up

to the user. The degree of adaptation is identified as a percentage of the difference between

	Conversion Equation
Pitch	$0.5x - 10$
Intensity	$3x - 149$
Speaking Rate	$33x - 49$

Table 15.2. Equations to Convert Praat Feature Value for Nao TTS

the user's mean and the system's prior mean, given the number of exchanges which have passed, and the number of exchanges allowed to pass before maximum convergence. For example, at the beginning of the conversation when zero exchanges have passed, the TTS will be generated at the specified baseline values for the system (i.e. 230 Hz, 68 db, 110 words per minute). If the number of exchanges for maximum convergence is set to five, then after one exchange, the system will adapt the TTS in the direction of the user by 20%, after two exchanges by 40%, three exchanges: 60%, and so on until after five exchanges, the system mean feature will equal the learner's mean. The maximum number of turns prior to convergence can be set as a parameter and the percentage increments will be automatically calculated. For this work, I identified five as the number of exchanges to which Emma should converge based on the average number of exchanges per step in four pilot evaluations. To model local convergence, the number of exchanges passed was be reset to zero whenever a new step or new problem was introduced. To model multi-feature convergence, the system calls this algorithm for each feature.

15.3   METHODOLOGY AND PROCEDURE

Perceptions of naturalness and rapport were evaluated for two multi-feature entrainment combinations: pitch with speaking rate and pitch with intensity. I compared the perceptions of these multi-feature entrainment approaches to single feature entrainment (on pitch, intensity, and speaking rate) and to non-transformed text-to-speech. Twenty-four dialogues were collected from four individuals to evaluate and compare the entrainment designs. In each interaction, a middle school student interacted with Emma, teaching Emma how to solve six math problems. For one problem, Emma spoke with a non-transformed baseline

	Mean	Std. Dev.
Dialogue length (min)	9.6	3.4
Number of turns	38	9
Turn length (sec)	7.3	2.6

Table 15.3. Dialogue and Turn Statistics for Corpus

speech and for each of the other five problems, Emma utilized each one of the different forms of entrainment. I collected each problem as a separate dialogue for a total of 6 dialogues per student. Statistics for the collected corpus are shown in Table 15.3. The four case studies were gender balanced with two males and two females interacting with Emma.

The longer dialogues were separated into shorter audio clips based on the individual steps for each problem so that each audio clip was the length of solving a single step. This resulted in 60 short dialogues for each student, with an average length of 1.03 minutes (std. dev .12). The entrainment was designed as local convergence with Emma resetting at the beginning of each step; for the audio clips where Emma was entraining, the clip contained one complete example of convergence. With a total of 60 audio clips per student, I utilized Amazon Mechanical Turk (AMT), a popular resource for crowdsourcing research tasks including annotations, transcripts, and subjective analysis (Buhrmester, Kwang, & Gosling, 2011). I used AMT to obtain 10 random, perceptual evaluations per clip for a total of 600 evaluations per student or 2400 evaluations. Using third party ratings such as those collected through AMT is a standard technique in the evaluation of naturalness and social features of dialogue systems. In addition, avoiding first-person ratings allowed us to present all dialogue approaches to each of the four individuals without worrying about how their perceptions of one approach might affect their ratings of a different approach.

Through AMT, individuals, referred to as workers, were asked to listen to each exchange and answer a series of questions regarding the speakers. Each worker had access to evaluate 240 exchanges (60 per student). To evaluate naturalness, I used Mean Opinion Score or MOS (ITU-T, 1994). With MOS, workers were asked to evaluate the quality of the voice on a Likert scale of 1-5, where 1 was very poor and 5 was completely natural. Workers evaluated both the human speaker and Emma on this scale.

For evaluating rapport, workers rated the degree of connection between the human speaker and Emma on a Likert scale of 1-5, where 1 was "no connection" and 5 was "a strong connection." In addition, workers rated on a Likert scale of 1 to 7 the degree of rapport they observed between Emma and the human speaker given the following definition of rapport:

> "Rapport is a term used to describe a combination of qualities that emerge from a particular kind of interaction. These interactions are characterized by such statements as 'we really clicked,' or 'we experienced real chemistry.' Terms like "engrossing," "friendly," "harmonious", "involving", and "worthwhile" describe interactions high in rapport."

This definition and approach to measuring rapport has been used in prior work to rate the degree of rapport between peer tutors and tutees (Madaio, Ogan, & Casssell, 2016). Finally, I had workers listen to two of the audio clips with two different type of entrainment (or no entrainment) and select which audio clip they believed had higher rapport.

In total, 236 workers provided evaluations of the audio. Only four workers rated 30% or more of the possible 240 exchanges they had access to while 63% of the workers listened to and rated at most two exchanges, meaning many of the ratings came from unique workers. In analyzing the results, I treated each rating as the unit of analysis. I calculated inter-rater agreement using Krippendorffs alpha, α = 0.67 (Krippendorff, 2011).

15.4   RESULTS

To analyze the effects of the different multi-feature adaptations in terms of rapport and naturalness, I ran a basic statistical analysis of the relationship between type of adaptation, naturalness, rapport, and connection. The marginal means and standard deviations for naturalness, rapport, and connection for the different adaptations and the non-transformed text-to-speech, referred to as the control, are given in Table 15.4.

To assess differences in **naturalness**, I performed a one-way analysis of variance (ANCOVA) with the type of adaptation (pitch, loudness, speaking rate, pitch + loudness, pitch + speaking rate, control) as a factor, naturalness as the dependent variable, and the amount of time the worker spent on the task as a covariate. Controlling for the time workers spent is common (Buhrmester, Kwang, and Gosling, 2011); workers can vary significantly in the amount of time they spend on individual ratings and this can be indicative of the quality of their responses and significantly predictive of differences. After controlling for time spent, the ANCOVA analysis indicated statistically significant differences among type of adaptations, $F_{(5, 2394)} = 12.71$, $p < 0.001$. Post hoc tests indicated that pitch + speaking rate was perceived as significantly less natural than every other form of adaptation

	Naturalness	Rapport	Connection
Control	3.14 (1.1)	4.8 (1.2)	3.6 (.93)
Pitch	3.00 (1.1)	4.7 (1.2)	3.6 (.87)
Intensity	3.00 (1.1)	4.7 (1.2)	3.6 (.96)
Speaking rate	2.90 (1.1)	4.6 (1.2)	3.6 (.96)
Pitch + speaking rate	2.51 (1.3)	4.4 (1.3)	3.4 (1.1)
Pitch + intensity	3.00 (1.1)	4.8 (1.2)	3.6 (.97)

Table 15.4. Descriptive Statistics of Perceptions for Different Adaptations

and the control (p < .001 for all factors). Outside of this, no other significant differences in naturalness emerged.

To assess **rapport**, I introduced two different questions, asking individuals about the degree of connection they perceived between Emma and the speaker as well as the degree of rapport, given a definition of rapport. I analyzed these two questions in a MANCOVA with the type of adaptation (pitch, loudness, speaking rate, pitch + loudness, pitch + speaking rate, control) as a factor, rapport and connection as the dependent variables, and the amount of time the worker spent on the task as a covariate. There was a statistically significant difference between the types of adaptations on the combined rapport variables after controlling for time workers spent, $F (10, 4706) = 3.07$, $p = 0.001$, Wilk's λ = .98, $\eta^2 = .012$. Univariate analyses revealed significant effects of adaptation type on both rapport, $F (5, 2354) = 4.9$, $p < 0.001$, and connection, $F (5, 2354) = 3.7$, $p = .002$. Post-hoc analyses of differences on rapport indicated that the control was significantly higher on rapport than pitch + speaking rate (p = .002). I also found that pitch + loudness was significantly higher in rapport than pitch + speaking rate (p < .001). I observed similar findings for connection, with both the control (p = .001) and pitch plus loudness (p = .01)

resulting in higher ratings of connection than pitch plus speaking rate. Outside of this, no other significant differences in the measures of rapport emerged.

I had also asked the workers to listen to two clips and report whether they thought one interaction held more rapport than another, given the definition of rapport. In analyzing the number of times any adaptation was preferred over any other, including the control, I only observed significant preferences when individuals heard a clip with pitch + loudness next to a clip with entrainment on speaking rate. Pitch + loudness was selected significantly more, $p = 0.001$, than the clip with entrainment on speaking rate. Outside of that comparison, none of the selections significantly exceeded a binomial test which assumed a 50-50 chance of picking one adaptation over the other, meaning that workers were just as likely to pick any adaptation or the control. There were no significant preferences.

15.5   CONCLUSIONS

From these results, I can conclude that the adaptations performed did have an effect, although small, on perceptions of naturalness and rapport. The differences between different adaptations and the control were minor, with a few exceptions. The adaptation of pitch + speaking rate resulted in surprisingly low perceptions of rapport and naturalness, especially when compared to the other adaptations and the control. I had investigated in the design of the multi-feature entrainment whether I needed to control for any unintended amplification of entrainment due to how manipulating one feature might influence another. It is unlikely the low performance of pitch plus speaking rate was due to the multi-feature manipulation. I explored whether the low performance was due to some undesirable combination of pitch plus speaking rate which was resulting in negative perceptions. For

example, entraining to low pitch combined with entraining to a slower speaking rate combines to produce a very unnatural perception and lower observations of rapport. However, different combinations of entrainment direction did not appear to significantly influence perceptions. In listening to the audio clips, I did observe that even though I had chosen minimum and maximum values for entrainment which fell within natural human parameters, these minimum and maximum values resulted in very synthetic text-to-speech output. Future work is needed for exploring how to produce a more acceptable speaking rate manipulation and exploration of why the addition of pitch seemed to spark lower responses. For this chapter, pitch plus loudness resulted in the highest rapport and rivaled the control in significant differences compared to the lowest performing adaptations. I utilize this multi-feature approach in a larger study, described in the next section.

CHAPTER 16

EFFECTS ON RAPPORT AND LEARNING WITH EMMA

With the multi-feature adaptation of pitch and loudness, this chapter describes a larger study conducted to explore the effects of multi-feature entrainment on rapport and learning. 48 middle school students interacted with Emma in one of three conditions: (1) **Non-social**. Emma did not behave socially. (2) **Entraining**. Emma entrained. (3) **Social plus entraining**. Emma entrained and spoke socially, adding social dialogue to the conversation. In contrast to the explorations with Quinn and Nico, I explore entrainment without the addition of social dialogue. With Quinn and Nico, social dialogue with entrainment consistently performed better than social dialogue alone. I am interested with this study in exploring how entrainment performs as a social behavior independent of social dialogue. I collected subjective self-reported rapport measures and coded for observable behaviors related to rapport. I also collected measures of learning, self-efficacy, and interaction goals.

I hypothesize that when Emma entrains, learners will report feeling more rapport and exhibit greater learning gains than when Emma does not entrain. I further hypothesize that when Emma entrains and speaks socially, learners will feel the most rapport and exhibit the greatest learning. In both human-human and human-robot interactions, the combination of multiple channels has been found to be important to rapport and communication (Argyle, 1988; Richmond and McCroskey, 1995; Swerts and Krahmer, 2008). Bruce and colleagues (2002) found that a robot which combined gaze and aligned appropriate movement was more compelling to passers-by. It seems likely that the effects

of two channels of social behavior, social dialogue plus entrainment, will continue result in more rapport than a single channel of prosodic entrainment alone.

This study utilizes the same version of Emma described in Chapter 15; no additional changes were made to the system or domain content. The next section describes the methodology and procedure. Section 16.2 contains the results of this exploration followed by a discussion and conclusions regarding the effects of multi-feature entrainment.

16.1   METHODOLOGY AND PROCEDURE

Participants were 48 middle-school students from one public middle school in the Southwestern United States with a mean age of 13.1 (SD = 0.75). The gender breakdown is given in Table 16.1 along with statistics regarding the dialogue. Sessions lasted 60 minutes and took place at the participant's school. As shown in Figure 16.1, students sat a desk with a Surface Pro tablet in front of them. Emma stood on the desk next to the Surface Pro, to the right of the participant. Two participants experienced technical issues and were excluded from the results. Thus, 15 participants remained in the non-social, 15 participants in the entraining condition, and 16 participants in the social-entraining condition.

Participants began with a short pre-survey on their initial self-efficacy towards math and tutoring and any prior experience with robotics. Participants then completed a 10-minute pretest. The pretest included the first three problems participants were to teach Emma. After completing the pretest, they were then given a few minutes to review the worked-out solutions to the problems pertaining to Emma. After watching a short video on how to interact with Emma, students taught Emma for 30 minutes. After the activity, they completed a 10-minute posttest and a short survey on self-efficacy, rapport, and their goals.

176

Figure 16.1. Student Interacting with Emma

16.1.1 Measuring Learning

To measure learning, I utilized a pretest-posttest design with an A and B form of the test. The pretest versions of A and B consisted of 10 questions with three of the questions mapping to the first three problems the participants would be teaching Emma. The posttest versions of A and B also consisted of 10 questions; three of the questions on the posttest were written in a similar format to the problems the students taught Emma. The remaining questions on the tests were isomorphic and consisted of questions on advanced ratios and proportions. The questions were mostly procedural with one conceptual. The tests were

	Females	Males	Total Turns M (SD)	Words per Turn M (SD)
non-social	8	8	116.3 (24.0)	7.13 (2.5)
entraining	9	7	125.4 (25.9)	9.23 (3.2)
social dialogue + entraining	9	7	119.3 (21.4)	8.9 (3.3)

Table 16.1. Gender Breakdown and Dialogue Statistics

counter-balanced within condition (half of the participants in each condition received test A as the pretest with test B as the posttest, and vice versa). The full pretest and posttests can be found in Appendix C. I piloted and iterated on the design of the questions with four pilot studies, evaluating timing and applicability of questions. To analyze learning and maintain consistency with the prior work, I used both the pretest and posttest scores in statistical analyses and I also calculated the normalized learning gains according to Hake (2002). If the posttest was lower, I used (2):

$$\text{gain} = (\text{posttest} - \text{pretest}) / (1 - \text{pretest}) \qquad (3)$$

$$\text{gain} = (\text{posttest} - \text{pretest}) / (\text{pretest}) \qquad (4)$$

16.1.2 Measuring Rapport

I measured both self-reported and linguistic rapport in this study. Self-reported rapport was assessed with the same 12 questions used to measure rapport with Nico. While I did not find significant differences in the study with Nico, these measures of rapport had high internal reliability, were correlated with linguistic rapport, and appear to be reasonable assessments of rapport. I averaged the rapport questions to create a single representative construct; reliability remained high (Cronbach's $\alpha = 0.81$).

To assess linguistic rapport, I coded for similar verbal rapport-building behaviors as with Quinn and Nico, including praise, formal politeness, inclusivity, and name usage. I utilized the same coding scheme as given in Appendix A and included the modification described in Chapter 13 regarding empathy for Emma. Examples of the behaviors can be found in Table 16.2. Two individuals each independently coded the dialogues for these behaviors. The average Cohen's kappa for these behaviors was .85. Individual kappas are

		M	SD	k
Praise	"Great job", "Good answer"	7.4	9.7	.70
Politeness	"thank you", "you're welcome"	.84	.91	.89
Inclusive	'we' or 'lets'	8.9	12.3	.95
Name	"That's right, Emma", "So Emma…"	1.3	2.1	.99
Empathy	"Me too", "I can help you"	.39	.71	.71

Table 16.2. Descriptive Statistics and Kappa Ratings for Linguistic Rapport

reported in Table 16.2 along with the means and standard deviations for the behaviors as they occurred across all conditions. For analysis, I summed the total observed behaviors into a single representative construct of linguistic rapport for each participant.

16.1.3 Self-Efficacy Measures

Measures of self-efficacy around math and tutoring were also collected; the same questions as posed with Nico based on the measures on work by the Friday Institute for Technology (2008) were used with some slight modification. The number of questions on tutoring self-efficacy was reduced from four to three, removing a question with particularly low agreement. I also reduced the number of questions on math self-efficacy from four to three. Participants answered all six questions both before and after interacting with Emma. I averaged the tutoring and math questions to obtain four scores: tutoring self-efficacy (Cronbach's α = 0.39) and math self-efficacy (Cronbach's α = 0.61) prior to interacting with Nico and tutoring self-efficacy (Cronbach's α = 0.45) and math self-efficacy (Cronbach's α = 0.71) post interacting with Nico. Unfortunately, Cronbach's α for the tutoring self-efficacy questions is even lower than in the previous study despite removing questions which had low agreement in the study with Nico.

16.1.4 Comfort-Level Around Robots

Individual differences appeared to be important in the study with Quinn, suggesting that an individual's prior experiences, bias, and interaction goals may influence their response to social behavior. With Nico, gender was not a significant factor, but it still seems possible that prior experiences may influence responses. With Emma, I explore the role of comfort-level around robots and human-looking robots along with prior experience with robots. Two questions were posed regarding how comfortable participants around robots; these questions were averaged (Cronbach's $\alpha = 0.79$) and then split into a high comfort / low comfort categorical variable where scores less than three were marked as low comfort and scores greater than three were marked as high comfort. This resulted in a mostly even distribution. We may observe differences in how individuals respond based on their prior experience and comfort-level around robots.

16.2   RESULTS

I report the results for learning and rapport where 48 middle schoolers interacted with Emma, the teachable robot, in one of three conditions: a **social-entraining** condition where Emma was both social and entrained, an **entraining** condition where Emma only entrained, and a **non-social** where Emma was neither social nor entraining. For the statistical analysis, I do not report gender as no gender differences on learning or rapport were found.

16.2.1   Learning Results

With learning, I had hypothesized that the social-entraining condition would result in greater learning than the entraining and non-social baselines, and that the entraining

	Non-Social	Entraining	Social-Entraining
Gain	.31 (.34)	.04 (.45)	.36 (.18)
Pretest	.48 (.18)	.28 (.19)	.29 (.19)
Posttest	.63 (.16)	.36 (.25)	.53 (.22)
Self-Reported Rapport	4.1 (.47)	3.9 (.54)	4.4 (.47)
Linguistic Rapport	16.7 (14)	15 (13)	24 (16)
Comfort-level	4.1 (.24)	4.0 (.20)	4.2 (.17)
Dialogue Errors	17.5 (6)	16.4 (9)	18.7 (9)

Table 16.3 Descriptive Statistics for Learning and Rapport Across Conditions

condition would result in greater learning than the non-social condition. I verified this hypothesis by analyzing learning as a repeated measures ANOVA with pretest and posttest as the dependent variables and condition as the independent variable. Mean pretest and posttest scores are given in Table 16.3. I observed a significant effect of condition on learning, $F (2, 43) = 3.91$, $p = .03$. Pairwise comparisons within conditions revealed significant learning in every condition: non-social improved by 14% ($p = .001$), the entraining condition improved by 9% ($p = .04$), and the social plus entraining condition improved by 24% ($p < 0.001$). This suggests that greater learning occurred in the social plus entraining condition, and surprisingly, the least learning appears to have occurred in the entrainment-only condition.

To maintain consistency with prior work, I measured the learning gain as well and analyzed differences by gain using in a one-way analysis of variance (ANOVA) with condition as the independent variable and gain as the dependent variable. Table 16.3 gives means and standard deviations for gain by condition. I found the gain was significantly

different across conditions, $F (2, 43) = 4.03$, $p = 0.03$. Partial eta squared was .15, meaning the effect size was medium. Tukey post-hoc analyses indicated significant pairwise differences for the social-entraining condition when compared to the entraining only condition indicating more learning in the social-entraining ($p=.03$). The nonsocial condition approached a significantly higher gain than the entraining-only ($p = .08$). I did not observe any indication of differences between the social-entraining condition and the non-social condition, ($p = .8$).

In reviewing pretest scores, I found significant differences between conditions. Despite random assignment to conditions, the average pretest scores for the non-social condition were significantly higher than the social plus entraining condition ($p = .02$) and the entrainment-only condition ($p = .02$). There was not a significant difference between the pretest scores for the social plus entraining and entrainment only conditions. I controlled for pretest in all other analyses and given the differences in pretest scores, I analyzed differences in posttest scores while controlling for the pretest with an ANCOVA. I found a significant main effect for condition, $F (2, 42) = 4.19$, $p = .02$, $\eta^2 = .17$. Pairwise comparisons of the means revealed a significant difference in posttest scores, controlling for pretest scores, between the social plus entraining (marginal mean = .58) and the entraining condition (marginal mean = .42), $p = .01$. There was not any difference between the nonsocial (marginal mean = .52) and social-entraining posttest scores.

I had hypothesized that the social-entraining condition would result in the most learning, followed by the entrainment-only condition and the non-social control. However, I found that the entrainment-only condition performed poorly; it resulted in significantly

lower learning than social-entraining analyzed multiple ways. The entrainment-social resulted in higher learning, but not significantly more than the non-social control.

16.2.2 Self-Reported Rapport Results

I hypothesized that self-reported rapport might increase for the entraining-only and the social-entraining conditions; if it did influence rapport, it may also enhance learning. The self-reported means and standard deviations for rapport by condition are given in Table 16.3. I first explored if rapport was correlated with the pretest and posttest scores. I found rapport was not correlated with pretest ($r (44) = -.23$, $p = .11$) nor was it correlated with the posttest ($r (44) = .02$, $p = .89$). I then explored if rapport differed by condition in an ANCOVA with condition as the independent variable and controlling for differences in pretest scores. I found that self-reported rapport did significantly differ across conditions after controlling for pretest, $F (2, 42) = 4.30$, $p = .02$, $\eta 2 = 0.17$. Exploring pairwise comparisons, I found the social plus entraining condition (marginal mean = 4.4) had significantly higher rapport than the entrainment only condition (marginal mean = 3.8). The non-social control (marginal mean = 4.2), while lower in rapport, was not significantly different from the social plus entrainment condition.

16.2.3 Linguistic Rapport Results

I explored linguistic rapport with behaviors including politeness, praise, name usage, and inclusive language as well as empathy. I first analyzed whether these behaviors appeared to be related to either self-reported rapport or learning. I found that they were moderately negatively correlated with rapport, $r (44) = -.29$, $p = .04$. However, these behaviors were

also positively correlated with pretest, r (44) = .34, p = .02, and with posttest, r (44) = .36, p = .02. I then analyzed whether the behaviors differed across conditions in a one-way ANCOVA with condition as the independent variable, linguistic rapport as the dependent variable and controlling for pretest. After controlling for pretest, I do not find a significant effect of condition, F (2, 42) = .32, p = .7.

16.2.4 Comfort-Level Around Robots Results

I explored whether there were any differences in reported comfort-level across conditions with a chi-square test for independence. We found no significant relationship between condition and how comfortable individuals reported they were with robots, X2 (2, 46) = .61, p =.74. Exploring whether comfort-level was related to other important characteristics such as an individual's prior experience with robots or gender, only eleven out of the forty-eight participants had some form of prior experience with robotics. These participants were approximately distributed across conditions (3 in the non-social, 4 in the entraining-only, and 4 in the social-entraining). Based on the means and standard deviations, individuals with prior experience with robots did not appear to differ significantly in how comfortable they were around robots. In addition, there was not a gender effect for comfort, X2 (1, 46) = .49, p = .48. Males and females did not differ in their comfort-level.

I then explored whether an individual's comfort-level around robots was related to their self-reported rapport utilizing an ANOVA with rapport as the dependent variable and controlling for pretest. Individuals who reported feeling more comfortable interacting with robots also reported significantly higher rapport, F (1, 43) = 9.1, p = .004. Analyzing whether there was a relationship between Emma's social behavior and comfort-level, I

utilized an ANCOVA with rapport as the dependent variable and condition and comfort-level as independent variables. Additionally, we controlled for differences in prior knowledge via pre-test scores as a covariate. After considering differences in pre-test, we found the interaction between comfort-level and condition approached significance, $F (2, 39) = 3.2$, $p = .05$. Condition was also significant, $F (2, 39) = 6.6$, $p = 0.003$, as was comfort-level, $F (1, 39) = 11.5$, $p < 0.002$. Exploring differences in rapport for individuals with high-comfort versus low-comfort, we found that for individuals with low-comfort, there were no differences in rapport between the social-entraining ($M = 4.1$, $SD = .51$), entraining ($M = 3.8$, $.73$), and non-social ($M = 3.8$, $SD = .41$) conditions, $F (2, 19) = .86$, $p = .4$. For individuals who expressed being very comfortable in interacting with robots, the robot's social behavior in the different conditions had a significant effect on feelings of rapport, $F (2, 21) = 6.65$, $p = .005$, with individuals in the social-entraining condition feeling significantly more rapport ($M = 4.64$, $SD = .3$) than individuals in the entraining-only condition ($M = 4.0$, $SD = .31$). The difference between individuals in the social-entraining condition and the non-social condition ($M = 4.3$, $SD = .38$) was not significant.

Finally, I explored the role of comfort-level with respect to learning. Adding comfort-level to the repeated measures ANOVA as a categorical high/low measure, there were no significant differences on learning for individuals with high versus low comfort, $F (2, 40) = 2.5$, $p = .12$. However condition and comfort-level do approach a significant interaction on learning, $F (2, 40) = 2.54$, $p = .09$. Exploring post hoc analyses for individuals with a high degree of comfort around robots, the entraining only condition results in significantly less learning than the social-entraining ($p = .006$) and non-social ($p = .04$) conditions.

16.3   DISCUSSION AND CONCLUSIONS

This study focused on how multi-feature prosodic entrainment on pitch and loudness influenced feelings of rapport and learning with 48 middle school participants as they interacted with the robotic learning companion Emma. There was a significant effect of entrainment on rapport and on learning when entrainment was introduced in the presence of social dialogue as compared to entrainment by itself. This was driven by the individuals who felt comfortable interacting with robots. The trends in the results suggest that entrainment with social dialogue is the optimal condition for individuals who are highly comfortable interacting with robots but that for individuals who not as comfortable, less social behavior may be more optimal.

The single channel of social behavior of entrainment performed poorly, resulting in significantly lower learning and self-reported rapport than entrainment with social dialogue. Lucas and colleagues found that in conversational interactions, if an agent makes a conversational error, negative effects can be boosted or exaggerated depending on the timing of the social dialogue (Lucas et al., 2018). The social-entraining condition performed strongly in comparison to entrainment alone; I analyzed however whether dialogue errors in the entraining-only condition may have produced the dip in social responses in that condition. I calculated errors as the number of utterances in which Emma asking the student to reiterate what they said (i.e. "I'm sorry, I didn't hear you. Can you say that again?"). These utterances were triggered when the student's dialogue could not be matched to an appropriate response. The means and standard deviations for the average number of errors per condition are given in Table 16.3. Running an ANCOVA with condition as the independent variable and the number of errors as the dependent variable,

the differences across conditions were not significant, $F (2, 41) = 3.94$, $p = .02$, suggesting that the dialogue errors were not likely to have contributed to the differences in responses.

The results also suggest that the multi-feature entrainment might not have succeeded as well as single-feature entrainment was found to perform with Nico in Chapter 14. There are a couple explanations for why this might be. One potential explanation is that it is possible entrainment on specific features is more pertinent at different moments during conversational dialogue, and that for different features, critical opportunities for entrainment differ. For example, I observed in Chapter 6 in the analysis of human-human collaborative dialogues that entrainment on intensity was significantly more common when individuals were engaged in problem solving versus social dialogue. Levitan and colleagues found entrainment at turn exchanges is related to the type of turn exchange; individuals tend to entrain more on multiple features during backchannels than other turn types and they entrain more on pitch when interrupting pauses. It is possible that by implementing convergence on both pitch and intensity across all dialogue, I may have negatively amplified entrainment at critical moments when entrainment on multiple features should have differed. Emma entrained on both pitch and intensity where in human-human dialogues, I more likely would have seen entrainment on pitch but not intensity or entrainment intensity but not pitch, which potentially resulted in a less ideal outcome than if entrainment had occurred on pitch alone.

An alternative explanation for why the multi-feature entrainment approach appears to have been less successful than pitch alone may be that for some individuals, it was possible Emma entrained more on pitch or more on intensity depending on the difference between Emma and the learner. For example, an individual who was substantially quieter

187

than Emma may have interacted with a more intensity-entraining companion than someone who spoke with a more average loudness but a substantially lower or higher pitch. I explored whether there was a prominent 'feature' of entrainment for different individuals and whether this played a role in responses by calculating whether Emma appeared to entrain more on intensity versus pitch for individual users. I looked at Emma's pitch and intensity values in the entraining and social plus entraining conditions. I analyzed how Emma changed between each turn on pitch and intensity for each user and identified based on the average normalized change in pitch and intensity whether Emma appeared to adapt more on pitch or more on intensity. Overall, Emma did adapt significantly on both features for all participants, but for 18 individuals Emma's change in intensity was higher than the change in pitch and for the other 14 individuals Emma's change in pitch was higher. I evaluated whether there were differences on learning, self-reported rapport, or linguistic rapport for those users for which Emma appeared to adapt more on pitch versus intensity. I did not observe any significant differences on self-reported rapport (F (1,29) = 1.9, p = .18), linguistic rapport (F (1,29) = .13, p = .71), or learning gain (F (1,29) = .15, p = .70) and significant interactions with condition. Based on these results, it is possible entrainment to one feature versus another did not play a significant role in the differences between the entraining and the social plus entraining conditions; the low performance of the entraining condition may have been due to more nuanced effects.

In terms of limitations, this study did have limited statistical power because of the modest sample size (N = 48), which may have played a role in limiting the significance of some of the statistical comparisons conducted. A post hoc power analysis revealed that based on the mean, between-groups comparison effect size observed for learning (d = .16),

an N of approximately 198 learners would have been required to obtain statistical power at the recommended .80 level (Cohen, 1988).

Finally, I found that individuals' linguistic rapport behaviors were negatively correlated with self-reported rapport, meaning that individuals who reported feeling less rapport for Emma engaged in more rapport-building dialogue. While I did not observe differences across conditions, the negative correlation was surprising. I saw a similar negative correlation in the prior work with Quinn, when females exhibited significantly more linguistic rapport and it was significantly negatively correlated with their rapport. In this case, I did not observe a gender difference regarding the negative correlation. The sample size is small, so it is possible that I needed to collect more data. An alternative explanation is that potentially the gender of the robot played a role. In the study with Quinn, females interacted with a female robot and males interacted with a male robot. In this study, Emma was gendered to be female for all participants. It is possible that the robot's gender plays a larger role than expected in influencing the presence and meaning of linguistic rapport behaviors.

Overall, the work in this chapter demonstrates that verbal and prosodic social behaviors can combine to produce positive responses but that in future work, designing entrainment on pitch and intensity may need to involve more nuanced and dialogue specific constraints as well as controls for how much a companion entrains on specific features given individual differences in users as well as potentially the assigned gender of the robot.

PART III:

CONCLUSIONS

CHAPTER 17

CONCLUSIONS

This thesis explored how the complicated phenomenon of acoustic-prosodic entrainment, where individuals adapt their acoustic-prosodic features of speech to one another over the course of a conversation, could be implemented in the dialogue system of a robotic learning companion. As a part of this work, the following research questions were posed:

RQ 1: How can acoustic-prosodic entrainment be modeled in a system to positively influence social responses?

RQ 2: How does automated entrainment influence social responses when combined with content-based approaches for building rapport?

RQ 3: How does entrainment influence learning in a robotic learning companion?

RQ 4: What insights regarding human-human and human-agent interactions can we gain by manipulating social behavior in a robotic learning companion?

I began the exploration of how to design entrainment by exploring how entrainment has been found to occur in human-human conversation. I added to the existing knowledge of entrainment by investigating its occurrence in a corpus of human-human data consisting of dyads working collaboratively together to solve math problems. This work provided initial insight into RQ 1, how can acoustic-prosodic entrainment be modeled in a system, suggesting that a model of entrainment based on pitch and proximity may build rapport and enhance learning. I explored this model and iterated on it over a series of six studies, focusing on how entrainment can be modeled, the effects it might have on rapport, particularly when combined with other rapport-building behavior, the effects entrainment

can have on learning, and whether modeling entrainment can provide any insight into human-human and human-agent interactions. I summarize the high-level results of three of these studies in Tables 17.1 and 17.2; these three studies were the largest and most defining in terms of answering the given research questions. I discuss the significant observations which emerged from these studies and the answers they provided to the proposed research questions in the rest of this chapter.

17.1 RQ 1: MODELLING ENTRAINMENT

I introduced several designs of entrainment in this work and explored three implementations of these designs in larger studies with 48 or more participants. The most successful design when considering effects on rapport and learning was a form of local convergence on pitch mean, where the robot adapted its pitch to the user over a series of turns and then reset when a new topic was introduced, such as a new problem. In Chapter 14, the exploration of Nico resulted in significantly more learning and significantly more linguistic rapport than a robot which did not entrain and did not speak socially. While this model appears to result in the most optimal results, aggregating all the results across all the implementations reveals several other interesting implications regarding modeling entrainment.

(1) Acoustic-prosodic entrainment is a social process: Theoretically, being on the receiving end of entrainment should lead an individual to feel more rapport for their entraining partner. With Quinn and Nico, I found that entrainment enhanced the effects of social dialogue; being on the receiving end of entrainment appeared to lead to individuals

	Robot Type	# of Participants	Participant Age Range	Robot Gender	Math Content
Study 1 Quinn	Lego Mindstorms	48	18 – 30	Male for males Female for females	Literal equations
Study 2 Nico	Nao	72	11 – 12	Unspecified. 87% of users viewed as male	Ratios
Study 3 Emma	Nao	48	12 – 14	Female	Ratios & Proportions

Table 17.1. Conclusions: Summary Details of Each Study

to feel more rapport for an entraining, social agent when entrainment was implemented on pitch. With Emma, where I implemented entrainment on pitch and intensity, the presence of entrainment independent of social dialogue performed poorly. However, once again, combining entrainment on pitch and intensity with social dialogue resulted in significantly higher rapport. Entrainment on pitch has been suggested to be more related to social processes and therefore may perform better in the presence of social dialogue (Levitan 2014). Observing again and again the success of entrainment when combined with another social behavior suggests that entrainment is indeed a social process and that when modeling entrainment, it is vital to keep in mind the other social behaviors being introduced.

(2) Increased complexity requires increased sophistication: In the initial design, I modeled entrainment as pitch proximity, where the robot mirrored the student at every turn. I iterated on that design using a similar method to adapt pitch but increasing the sophistication of the adaptation by incorporating the concept of convergence. With the design change from proximity to convergence, I created a more sophisticated model of

single feature entrainment. I then increased the complexity of the entrainment model by adding an additional feature, intensity. However, when iterating on the design to incorporate an additional feature, the sophistication of the model was left unchanged and

		Study 1 **Quinn** *Nonsocial* *Social* *Social+Entraining*	**Study 2** **Nico** *Nonsocial* *Social* *Social+Entraining*	**Study 3** **Emma** *Nonsocial* *Entraining* *Social+Entraining*
Learning	Condition	Not significant	**Significant** Social+entraining > Nonsocial**	**Significant** Social+entraining > Entraining*
	Gender	Not significant	Not significant	----
Self-reported Rapport	Condition	**Significant** Social+entraining > Social*	Not significant	**Significant** Social+entraining > Entraining*
	Gender	**Significant** Females report more overall rapport** Males less social presence in social*	Not significant	----
Linguistic Rapport	Condition	Not significant	**Significant** Social+entraining > Nonsocial**	Not significant
	Gender	**Significant** Females exhibit in non-social and social Males exhibit in social+entraining	Not significant	----
Self-reported and Linguistic		Negatively correlated*	Positively correlated*	Negatively correlated*

Table 17.2. Conclusions: Summary Results from the Three Studies. * p < .05, ** p < .01

individuals did not respond as positively to the model. The design of multi-feature entrainment was unsophisticated; it seems highly probable that a model which incorporates context and individual differences may improve effects. Iterating on the design in a similar way to how the social dialogue was iterated on in Chapter 13 may produce similar marked improvements in responses.

(3) Perceptual evaluations facilitate design: The work of modeling entrainment is far from finished. This thesis presents a model that can enhance responses but there are many more nuances in human-human conversation to be incorporated in future iterations. For that future work, this thesis presents an excellent methodology for designing and evaluating models using perceptual evaluations. The use of perceptual evaluations is a very effective methodology for eliminating designs and suggesting areas for iteration. The effects of a particular model hinted at by the perceptual evaluations proved meaningful when models were evaluated during longer interactions. To improve this methodology for future work, aggregating perceptions across the interactions as suggested by "thin-slicing" approaches (Madaio, Cassell, and Ogan, 2017) may help improve evaluations while keeping costs low during this form of design iteration.

(4) Multiple channels enhance social responses: Multiple channels of social behavior consistently resulted in better outcomes. Entrainment combined with social dialogue had the highest self-reported rapport or linguistic rapport in every study and resulted in significantly more learning with Nico and Emma. Not only did the combined behaviors perform the best but it also appears that combining two social behaviors can enhance an

initially low performing social behavior. A higher performing social behavior may even mask the less than ideal effects of another behavior. With both Quinn and Emma, I found the single channel of social behavior, in one case social and in the other entrainment, to perform less than optimally. The addition of a second social behavior enhanced responses significantly. These findings suggest that no single behavior is key to fostering rapport and enhancing learning; exploring only content or only gesture or only prosody is not the way if we want to truly build social relationships. However, it is important to be aware of these contrasting effects, particularly when modeling a behavior like entrainment which appears to be tightly coupled to other social behavior.

17.2 RQ 2 AND RQ 3: EFFECTS ON RAPPORT AND LEARNING

RQ 2 and RQ 3 were focused on how automated entrainment influences social responses like rapport and concrete outcomes like learning. The findings from this work suggest that entrainment and social dialogue can enhance rapport and that the presence of entrainment can influence learning. In the context of robotic learning companions, greater feelings of rapport may lead to greater engagement and in turn, greater engagement may lead to greater learning, as suggested by the protégé effect. The protégé effect occurs when a learner, feeling more rapport and social engagement for their teachable agent, engages more with the material and exhibits greater willingness to address misconceptions, facilitating learning. This effect has been found in other teachable agents with successful outcomes.

With the three versions of entrainment in Quinn, Emma, and Nico, the results suggest that learners who interacted with a social, entraining robot perceived the robot to be more socially present (Quinn), felt more rapport for the robot (Emma), learned more

(Nico and Emma), and exhibited more linguistic rapport (Nico). In exploring how individuals spoke to a social, entraining robot, I observed particularly with Nico, where the learning result was the most significant, that learners who were more socially engaged were also more cognitively engaged. In their dialogue, they made more attempts to explain to Nico the how and why behind solutions as opposed to learners who were less socially engaged. These findings suggest that entrainment can enhance rapport and social engagement, which may facilitate overall engagement and increase a participant's willingness to think through problems and explain solutions. This increased engagement in the activity of teaching may have facilitated learning.

In considering the effects of entrainment on rapport and learning, this thesis also provides some insight into the role of individual differences like those indicated by gender. The results of the studies here provide insight not only into the role of the user's gender but also how the gender of the robot may influence how individuals exhibit rapport. With Quinn, females felt more rapport overall; they also engaged in more linguistic rapport. Female use of linguistic rapport was significantly negatively correlated with their self-reported rapport while it was positively correlated for males. Females also interacted with a female robot and males interacted with a male robot, indicated by the voice. With Nico, no significant gender differences emerged; overall, linguistic rapport was significantly positively correlated with self-reported feelings of rapport. Most participants believed Nico to be male. With Emma, again no significant gender differences emerged; however, linguistic rapport was back to being significantly negatively correlated with self-reported rapport, for all participants. Emma was intentionally gendered to be female.

These results suggest that while the gender of the user may be indicative of underlying social responses in some cases, the gender of the robot may influence how individuals behave, how they exhibit rapport. Prior work would suggest that responses to gendered robots are complex but that they are not driven by stereotypes (Rea, Wang, and Young, 2015; Reich-Stibert and Eyssel, 2017). The results here would suggest that when engaging with a female gendered robot, individuals in general appeared to use more rapport-building behaviors when they felt less rapport. This could perhaps be to manage rapport. When interacting with a male robot, individuals used more rapport-building behaviors when they felt more rapport. Being positively correlated with their self-reported rapport, these behaviors were not about managing rapport but expressing rapport. These findings suggest that in evaluating effects individual differences are important to consider but that future exploration should also include how individuals behave in response to gendered robots and what this implies about how they are feeling.

17.3 RQ 4: INSIGHTS INTO INTERACTION

RQ 4 was concerned with whether there are any insights to be gained from manipulating social behaviors like entrainment in a robotic companion, particularly insights into human-human interaction or human-agent interaction. There are important implications to this research question because it can often be difficult to gain insight into some aspects of human-human interaction. For example, there are questions which require controlled responses by one of the participants, such as how does deliberate entrainment on a single feature influence social responses? Or, how do different dialogue strategies influence experiences of self-efficacy in learning-by-teaching interactions? If the platform presented

here can reduce both the challenge and the cost of gaining insight into such questions, the space of questions about interactions which can be posed and answered is expanded. In the next few paragraphs, I touch on a few of the insights gained from this work and how the teachable robot platform may enhance future research.

Prior to this work, it was unclear and difficult to detect from analysis of human-human entrainment how deliberate entrainment on a single feature such as pitch can influence social responses. The results of the studies with Quinn, Nico, and Emma demonstrated that deliberate entrainment on a single feature can make a positive impact on how individuals perceive and respond to social dialogue, and that the form of entrainment does not need to be sophisticated or extremely nuanced to positively influence responses. This finding gives value to investigating entrainment on individual features in both human-human and human-agent interactions because even a single feature can have an impact on perceptions. This is contrary to what some have proposed regarding entrainment, suggesting that it should be studied in aggregate since individuals tend to manipulate more than one feature at once and speakers perceive multiple features at once.

Exploring the impact of different dialogue strategies in human-human peer tutoring dialogues can be challenging because it is dependent on observing strategies employed by human participants. The results of Chapter 13 demonstrated that a robot can employ a variety of strategies and that these can be evaluated for useful they are in creating social experiences that will build a tutor's self-efficacy. While explored in a human-agent context, these dialogue strategies which promoted social, self-efficacy building experiences may be applicable in human peer tutoring scenarios. This methodology of exploring different

interaction strategies with a robotic tutee could be used in future work to explain why some peer tutoring dyads are more successful.

Ultimately, the teachable robot platform presented here reduces both the challenge and the cost of gaining insight into questions which may be challenging to answer in human-human interactions. The final version of the teachable robot platform includes a robust and modular dialogue system and functional robotic platform which can be used to explore questions regarding social behavior and interaction. The results of this work highlight the potential benefits of using human-robot and potentially human-agent interactions to explore and gain insight into open questions in human-human interactions. It is a significant outcome of this work to demonstrate the potential platforms Nico and Emma have for investigating and gaining insight into previously difficult questions.

17.4   CONTRIBUTIONS

An Entraining Dialogue System: I introduced a dialogue system containing an acoustic-prosodic entrainment module which can produce entrainment on multiple acoustic-prosodic features in combination with other rapport-building verbal behavior, and demonstrated successful, positive effects on rapport and learning.

Theory of Entrainment and Rapport: I contributed to the theory of entrainment and rapport demonstrating that even simple models of entrainment implemented in a teachable agent context can positively influence feelings of rapport and enhance interaction. I also found that a social behavior such as entrainment can have positive effects on learning

Understanding Individual Differences: I found interesting indications that gender is pertinent when measuring responses and that the potentially the gender of the robot may

play a vital role in determining what responses mean and how individuals may engage with a social companion.

17.5   FUTURE WORK

17.5.1 Automating Entrainment

In this thesis I explored simple, exaggerated models of entrainment. For example, the robot matched mean pitch turn-by-turn on every single turn, the robot converged over a series of turns on mean pitch, or the robot converged over a series of turns on pitch and intensity. Each model was slightly more complex than the previous; however, individuals in human-human conversations are incredibly more complex, entraining at different points on different features depending on the dialogue structure and context. I found success with these models, but future work should target systems which incorporate entrainment as a part of the text-to-speech synthesis. To achieve this, one immediate next step would be to expand on the questions raised by the multi-feature entrainment approach. How can a more sophisticated model of pitch and intensity entrainment achieve higher rapport and learning? How can we better describe the relationship between entrainment and social dialogue? There is also the open question of whether automated entrainment may be more successful as simple and exaggerated models versus more complex, as has been suggested by recent work (Benus, et al., 2018). Finally, I observed significant dyadic differences in the analysis of human-human entrainment. Understanding how and why two individuals might entrain one way while two individuals entrain in a different way, yet both have high rapport is critical to designing a successful system. Answering these questions is required to provide a solid foundation for understanding how a system can achieve successful entrainment.

In exploring models of automated entrainment in the future, another measure of interest which is particularly important to the long-term development of entraining dialogue systems is the user's prosodic responses. Future work should explore in more depth how learners changed their own prosody in response to the system. This will provide further insight into the effects of automated entrainment and can potentially provide guidance on how to use a learners' prosody to detect when a system's entrainment mechanism is working as intended or needs to adapt differently.

This thesis gives evidence that there is potential for entrainment to enhance human-computer interactions where social factors play an important role. It also supports that these social factors, enhanced through entrainment, can support outcomes such as learning. Future work could expand automated entrainment to other domains of human-computer interaction where developing rapport is important to the interaction and evaluate whether similar effects can be achieved.

17.5.2 Robotic Learning Companions

A major contribution of this thesis is that a social behavior such as entrainment has potential to facilitate learning in interactions with a robotic learning companion. This finding provides support for replicating other similar rapport-building phenomena in interactions with teachable robots. For example, future work could explore how lexical entrainment and physical entrainment combine with prosodic entrainment to enhance rapport and increase learning. Future work should also include exploring the open questions around gender and how the gender of the robot may influence responses. Finally, I explored this work with a teachable robot because of the potential robots have for

enhancing social responses in a physical space. This implies future explorations of manipulating multi-channel social behavior such as gesture. Another potential exploration of future work might be whether the findings with a teachable robotic companion extend to a virtual teachable agent and whether the influence of these behaviors differs when physical presence is no longer an attribute.

17.6   EPILOGUE

In Part I of this thesis I explored open questions regarding entrainment in human-human dialogues. In Part II I explored designs of acoustic-prosodic entrainment in spoken dialogue system for human-robot interactions where the robot was a form of robotic learning companion known as a teachable robot. I presented a social, entraining dialogue system and demonstrated that entrainment can be used to enhance rapport and learning in interactions with a teachable robot. The contributions of this work provide motivation for exploring future phenomena like entrainment to enhance factors such as rapport and learning as well as a platform with which to explore these phenomena, and add to the body of knowledge on entrainment, rapport, and learning.

REFERENCES

Acosta, J., and N. G. Ward. Achieving rapport with turn-by-turn, user-responsive emotional coloring. Speech Communication, 53(9):1137–1148, 2011.

Afzal, S. and P. Robinson, "A study of affect in intelligent tutoring," in Proceedings of the 13th International Conference of Artificial Intelligence in Education (AIED '07), Marina Del Rey, Calif, USA, July 2007

Alexander, E., Bank, C., Yang, J. J., Hayes, B., & Scassellati, B. (2014). Asking for Help from a Gendered Robot. 36th Annual Conf. Cognitive Science Society, 2333–2338.

ALICE. 2002. A.L.I.C.E AI Foundation, http://www.alicebot.org/

Ambady, N. and Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. Psychological bulletin, 111(2):256, 1992.

Andrist, S., Spannan, E., & Mutlu, B. (2013). Rhetorical robots: Making robots more effective speakers using linguistic cues of expertise. ACM/IEEE International Conference on Human-Robot Interaction, (Figure 1), 341–348. http://doi.org/10.1109/HRI.2013.6483608

Argyle, M. (1988). Bodily Communication. New York: Methuen & Co. Ltd.

Arroyo, I., et al.: Affective Gendered Learning Companions. In: Dimitrova, V., Mizoguchi, R., du Boulay, B., Graesser, A.C. (eds.) Proc. of the 14th International Conference on Artificial Intelligence in Education, pp. 41–48. IOS Press (2009)

Bales, Robert F. "A set of categories for the analysis of small group interaction." American Sociological Review 15.2 (1950): 257-263

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. Psych. Review, 84(2), 191.

Bandura, A., & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. Journal of Personality and Social Psychology, 41(3), 586-598.

Bares, W., et al. "Methods, systems, and computer program products for providing automated customer service via an intelligent virtual agent that is trained using customer-agent conversations." U.S. Patent No. 7,305,345. 4 Dec. 2007.

Beal, Carole, Sinjini Mitra, and Paul Cohen. "Modeling learning patterns of students with a tutoring system using Hidden Markov Model, In proceedings of the 13th International Conference on Artificial Intelligence in Education (AIED), R. Luckin et al.(Eds), Marina del Rey, July 2007." 238-245.

Bechade, Lucile, et al. "Behavioral and Emotional Spoken Cues Related to Mental States in Human-Robot Social Interaction." Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. ACM, 2015.

Bell, Diana Calhoun, Holly Arnold, and Rebecca Haddock. "Linguistic Politeness and Peer Tutoring." Learning Assistance Review 14.1 (2009): 37-54.

Bell, Linda, Joakim Gustafson, and Mattias Heldner. "Prosodic adaptation in human-computer interaction." Proceedings of ICPHS. Vol. 3. 2003.

Bengio, Yoshua, and Paolo Frasconi. "An input output HMM architecture." Advances in neural information processing systems. 1995.

Benus, S., Levitan, R. and Hirschberg, J. "Entrainment in spontaneous speech: the case of filled pauses in supreme court hearings." In 3rd IEEE Conference on Cognitive Infocommunications, Kosice, Slovakia, 2012

Benus, S., et al. "Prosodic entrainment and trust in human-computer interaction." Speech Prosody, 2018.

Benus, S., Levitan, R., Hirschberg, J., Gravano, A., & Darjaa, S. (2014, November). "Entrainment in Slovak collaborative dialogues." In Cognitive Infocommunications (CogInfoCom), 2014 5th IEEE Conference on (pp. 309-313). IEEE.

Bergner, Yoav, Erin Walker, and Amy Ogan. "Dynamic Bayesian Network models for peer tutoring interactions." Innovative Assessment of Collaboration. Springer, Cham, 2017. 249-268.

Bethel, C. L., & Murphy, R. R. (2010). Emotive Non-Anthropomorphic Robots Perceived as More Calming, Friendly, and Attentive for Victim Management. 2010 AAAI Fall Symposium Series, (September 2015), 4–6.

Bickmore, T. W. (2003). Relational agents: Effecting change through human-computer relationships (Doctoral dissertation, Massachusetts Institute of Technology).

Bickmore, T. W., Vardoulakis, L. M. P., & Schulman, D. (2013). Tinker: a relational agent museum guide. Autonomous agents and multi-agent systems, 27(2), 254-276.

Bickmore, T., & Cassell, J. (2000, November). "How about this weather?" social dialogue with embodied conversational agents. In Proc. AAAI Fall Symposium on Socially Intelligent Agents.

Bickmore, T., & Cassell, J. (2001, March). Relational agents: a model and implementation of building user trust. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 396-403). ACM.

Bilous, Frances R., and Robert M. Krauss. "Dominance and accommodation in the conversational behaviors of same-and mixed-gender dyads." Language & Communication 8.3 (1988): 183-194.

Biocca, F., & Harms, C. (2002). Defining and measuring social presence: Contribution to the networked minds theory and measure. Proceedings of PRESENCE, 2002, 7-36.

Biswas, G., Jeong, H., Kinnebrew, J. S., Sulcer, B., & ROSCOE, R. (2010). Measuring self-regulated learning skills through social interactions in a teachable agent environment. Research and Practice in Technology Enhanced Learning, 5(02), 123-152.

Biswas, G., Leelawong, K., Schwartz, D., Vye, N., & The Teachable Agents Group at Vanderbilt. (2005). Learning by teaching: A new agent paradigm for educational software. Applied Artificial Intelligence, 19(3- 4), 363-392.

Blanchard, E. G., Volfson, B., Hong, Y. J., & Lajoie, S. P. (2009, July). Affective artificial intelligence in education: From detection to adaptation. In AIED (Vol. 2009, pp. 81-88).

Boersma, Paul. "Praat: doing phonetics by computer." http://www. Praat. org/ (2006).

Borrie, S. A., Lubold, N. and Pon-Barry, H. "Disordered speech disrupts conversational entrainment: a study of acoustic-prosodic entrainment and communicative success in populations with communication challenges." Frontiers in psychology 6. 2015.

Boyer, Kristy Elizabeth, et al. "Characterizing the effectiveness of tutorial dialogue with hidden markov models." International Conference on Intelligent Tutoring Systems. Springer, Berlin, Heidelberg, 2010.

Brown, P. 1980. How and Why Are Women More Polite: Some Evidence from a Mayan Community. In Women and Language in Literature and Society, ed. S. McConnell-Ginet, R. Borker, and N. Furman. New York: Praeger.

Brown, Penelope and Stephen C. Levinson (1987). Politeness. Some Universals in Language Usage. Cambridge: CUP. Originally published as Universals in language usage: Politeness phenomenon. In Questions and Politeness: Strategies in Social Interaction, Esther Goody (ed.) (1978). Cambridge: CUP.

Bruce, A., Nourbakhsh, I., & Simmons, R. (n.d.). The role of expressiveness and attention in human-robot interaction. Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292), 4, 4138–4142. http://doi.org/10.1109/ROBOT.2002.1014396

Buhrmester, M., Kwang, T., & Gosling, S. D. Amazon's Mechanical Turk a new source of inexpensive, yet high quality, data? Perspectives on psychological science, 6(1), 3- 5, 2011.

206

Burgoon, J. K., Stern, L. A., & Dillman, L. (2007). Interpersonal adaptation: Dyadic interaction patterns. Cambridge University Press.

Burleson, W., & Picard, R. W. (2007). Gender-specific approaches to developing emotionally intelligent learning companions. IEEE Intelligent Systems, 22(4), 62–69. http://doi.org/10.1109/MIS.2007.69

Burns, M., Rapport and relationships: The basis of child care. Journal of Child Care, 1984. 2: p. 47-57.

Cahn, J. E., & Brennan, S. E. (1999). A psychological model of grounding and repair in dialog. In Proc. Fall 1999 AAAI Symposium on Psychological Models of Communication in Collaborative Systems.

Cassell, J., Gill, A. J., & Tepper, P. A. (2007, June). Coordination in conversation and rapport. In Proceedings of the workshop on Embodied Language Processing (pp. 41-50). Association for Computational Linguistics.

Cassell, Justine, and Timothy Bickmore. "Negotiated collusion: Modeling social language and its relationship effects in intelligent agents." User Modeling and User-Adapted Interaction 13.1-2 (2003): 89-132.

Castellano, G., et al. (2013, July). Towards empathic virtual and robotic tutors. In Int'l Conference on Artificial Intelligence in Education (pp. 733-736). Springer, Berlin, Heidelberg.

Cesta, A., G. Cortellessa, F. Pecora and R. Rasconi. Supporting Interaction in the RoboCare Intelligent Assistive Environment, AAAI 2007 Spring Symposium, 2007.

Chalupnik, Malgorzata, Christine Christie, and Louise Mullany. "(Im) politeness and Gender." The Palgrave Handbook of Linguistic (Im) politeness. Palgrave Macmillan UK, 2017. 517-537.

Chase, C. C., Chin, D. B., Oppezzo, M. A., & Schwartz, D. L. (2009). Teachable Agents and the Protégé Effect: Increasing the Effort Towards Learning. Journal of Science Education & Technology, 18(4), 334-352.

Chidambaram, V., Chiang, Y.-H., & Mutlu, B. (2012). Designing Persuasive Robots: How Robots Might Persuade People Using Vocal and Nonverbal Cues. 7th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '12), 293–300. http://doi.org/10.1145/2157689.2157798

Chin, Matthew G., et al. "Measuring individual differences in anthropomorphism toward machines and animals." Proceedings of the human factors and ergonomics society annual meeting. Vol. 48. No. 11. Sage CA: Los Angeles, CA: SAGE Publications, 2004

Chou, C., Chan, T., & Lin, C. (2003). Redefining the learning companion: the past, present, and future of educational agents. Computers & Education, 40(3), 255-269.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. Perspectives on socially shared cognition, 13(1991), 127-149.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. Cognitive science, 13(2), 259-294.

Coates, Jennifer. Women, men and language: A sociolinguistic account of gender differences in language. Routledge, 2015.

Common Core State Standards Initiative. (2010). Common Core State Standards for Mathematics. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers.

Core, M., Ishizaki, M., Moore, J., Nakatani, C., Reithinger, N., Traum, D., & Tutiya, S. (1999). The report of the third workshop of the discourse resource initiative. Chiba University and Kazusa Academia Hall.

Coulston, R., Oviatt, S., & Darves, C. (2002). Amplitude convergence in children's conversational speech with animated personas. In Proceedings of the 7th International Conference on SLP (pp. 2689-2692).

Csapó, B., Lörincz, A., & Molnár, G. (2012). Innovative assessment technologies in educational games designed for young students. In Assessment in Game-Based Learning (pp. 235-254). Springer, New York, NY.

D'Mello, S.K., Picard, R., Graesser, A.C.: Toward an affect-sensitive AutoTutor. IEEE Intelligent Systems 22, 53–61 (2007)

Darves, Charles, Shannon Oviatt, and R. Coulston. "Adaptation of users' spoken dialogue patterns in a conversational interface." Proceedings of the 7th International Conference on Spoken Language Processing. Vol. 1. 2002.

De Carolis, Berardina, et al. "Modeling and Simulating Empathic Behavior in Social Assistive Robots." Proceedings of the 11th Biannual Conference on Italian SIGCHI Chapter. ACM, 2015.

De Jong, Nivja H., and Ton Wempe. "Praat script to detect syllable nuclei and measure speech rate automatically." Behavior research methods 41.2 (2009): 385-390.

De Looze, Céline, et al. "Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction." Speech Communication 58 (2014): 11-34.

D'Mello, Sidney K., Art Graesser, and Brandon King. "Toward spoken human–computer tutorial dialogues." Human–Computer Interaction 25.4 (2010): 289-323.

Drolet, A.L. and M.W. Morris, Rapport in conflict resolution: accounting for how face-to-face contact fosters mutual cooperation in mixed-motive conflicts. Experimental Social Psychology, 2000. 36: p. 26-50.

Eckert, Penelope, and Sally McConnell-Ginet. Language and gender. Cambridge University Press, 2003.

Eyben, F., Wöllmer, M., & Schuller, B. (2010, October). Opensmile: the munich versatile and fast open-source audio feature extractor. In Proceedings of the 18th ACM international conference on Multimedia (pp. 1459-1462). ACM.

Eyssel, F., D. Kuchenbrandt, S. Bobinger, L. de Ruiter, and F. Hegel. 'If you sound like me, you must be more human': On the interplay of robot and user features on human-robot acceptance and anthropomorphism. In Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction, pages 125–126. ACM, 2012.

Falchikov, Nancy. Learning together: Peer tutoring in higher education. Psychology Press, 2001.

Fandrianto, Andrew, and Maxine Eskenazi. "Prosodic Entrainment in an Information-Driven Dialog System." INTERSPEECH. 2012

Forbes-Riley, Kate, and Diane Litman. "Designing and evaluating a wizarded uncertainty-adaptive spoken dialogue tutoring system." Computer Speech & Language 25.1 (2011): 105-126.

Forney, G. David. "The Viterbi algorithm." Proceedings of the IEEE 61.3 (1973): 268-278.

Foster, Mary Ellen, Andre Gaschler, and Manuel Giuliani. "How can i help you': comparing engagement classification strategies for a robot bartender." Proceedings of the 15th ACM on International conference on multimodal interaction. ACM, 2013.

Frager, S., & Stern, C. (1970). Learning by Teaching. The Reading Teacher, 23(5), 403-417.

Friday Inst. for Edu. Innovation. 2008. Student Attitudes toward STEM Upper Elementary School Students.

Frisby, Brandi N., and Scott A. Myers. "The Relationships among Perceived Instructor Rapport, Student Participation, and Student Learning Outcomes." Texas Speech Communication Journal 33.1 (2008).

Gallois, Cindy, and Howard Giles. "Communication accommodation theory." The international encyclopedia of language and social interaction (2015).

Garland, Ron. "The mid-point on a rating scale: Is it desirable." Marketing bulletin 2.1 (1991): 66-70.

Gerosa, M., et al. "A review of ASR technologies for children's speech." Proceedings of the 2nd Workshop on Child, Computer and Interaction. ACM, 2009.

Giles, H. (1979). A new theory of the dynamics of speech. Diogenes, 106, 119-136.

Giles, H. and Smith, P.M. Accommodation theory: Optimal levels of convergence. In H. Giles & R. St. Clair (Eds.), Language and social psychology (pp. 45-65). Oxford: Blackwell. 1979.

Girard, S., et al. (2013, July). Defining the behavior of an affective learning companion in the affective meta-tutor project. In International Conference on Artificial Intelligence in Education (pp. 21-30). Springer, Berlin, Heidelberg.

Gordon, Goren, et al. "Affective Personalization of a Social Robot Tutor for Children's Second Language Skills." AAAI. 2016.

Graesser, A., & Person, N. (1994). Question Asking during Tutoring. AERJ, 31(1), 104–137.

Graf B, Hans M, Schraft RD (2004) Care-O-bot II development of a next generation robotics home assistant. Auton. Robots 16, 193–205.

Gratch, J., Wang, N., Gerten, J., Fast, E., and Duffy, R. "Creating rapport with virtual agents." In Intelligent Virtual Agents, Springer, 125-138, 2007.

Gravano, A., Benus, S., Levitan, R., and Hirschberg, J. Three ToBI-based measures of prosodic entrainment and their correlations with speaker engagement. In Spoken Language Technology Workshop (SLT), 2014 IEEE (pp. 578-583). IEEE. 2014.

Griffin, M. M., & Griffin, B. W. (1998). An Investigation of the Effects of Reciprocal Peer Tutoring on Achievement, Self-Efficacy, & Test Anxiety. Contemporary Educational Psychology, 23(3), 298-311.

Gulz, Agneta, Magnus Haake, and Annika Silvervarg. "Extending a teachable agent with a social conversation module–effects on student experiences and learning." International Conference on Artificial Intelligence in Education. Springer, Berlin, Heidelberg, 2011.

Gweon, Gahgene, et al. "Predicting idea co-construction in speech data using insights from sociolinguistics." Proceedings of the International Conference of the Learning Sciences. International Society of the Learning Sciences: Sydney, Australia. 2012.

Hake, R. R. (2002). Relationship of individual student normalized learning gains in mechanics with gender, high-school physics, and pretest scores on mathematics and spatial visualization. In submitted to the Physics Education Research Conference, Boise, ID.

Harley, Jason M., et al. "Examining the predictive relationship between personality and emotion traits and students' agent-directed emotions: towards emotionally-adaptive agent-based learning environments." User Modeling and User-Adapted Interaction 26.2-3 (2016): 177-219.

Hess, U., & Blairy, S. (2001). Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. International journal of psychophysiology, 40(2), 129-141.

Holmes, J. (1995) Women, Men and Politeness, London, Longman.

Hong, J.-C., & Hwang, M.-Y. (2012). Gender Differences in Help-Seeking and Supportive Dialogue During On-Line Game. Procedia - Social and Behavioral Sciences, 64(2007), 342–351. http://doi.org/10.1016/j.sbspro.2012.11.041

Hood, D., Lemaignan, S., & Dillenbourg, P. (2015, March). 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 (pp. 83-90). ACM.

Huang, Lixing, Louis-Philippe Morency, and Jonathan Gratch. "Virtual Rapport 2.0." Intelligent virtual agents. Springer Berlin/Heidelberg, 2011.

ITU-T Recommendation P.85. Telephone transmission quality subjective opinion tests. A method for subjective performance assessment of the quality of speech voice output devices, 1994

Jacq, S. Lemaignan, F. Garcia, P. Dillenbourg and A. Paiva. Building Successful Long Child-Robot Interactions in a Learning Context. 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Christchurch, NEW ZEALAND, 2016.

Johns, R. (2010). Likert items and scales. Survey Question Bank: Methods Fact Sheet, 1, 1-11.

Jokinen, K., & Hurtig, T. (2006). User expectations and real experience on a multimodal interactive system. Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2(January 2006), 1049–1052.

Jurcicek, F., Keizer, S., Gašic, M., Mairesse, F., Thomson, B., Yu, K., & Young, S. "Real user evaluation of spoken dialogue systems using Amazon Mechanical Turk." Proceedings of INTERSPEECH. Vol. 11. 2011

Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial. Human-Computer Interaction, 19(1), 61–84. http://doi.org/10.1207/s15327051hci1901&2_4

Kanda, Takayuki, et al. "A two-month field trial in an elementary school for long-term human–robot interaction." IEEE Transactions on robotics 23.5 (2007): 962-971.

Kang, S.-H., Gratch, J., and Watts, J. The Effect of Affective Iconic Realism on Anonymous Interactants' Self-Disclosure. In: Proceedings of Interaction Conference for HumanComputer Interaction (2009).

Karpiński, Maciej, Katarzyna Klessa, and Agnieszka Czoska. "Local and global convergence in the temporal domain in Polish task-oriented dialogue." situations 2 (2014): 12.

Kauchak, D. P., & Eggen, P. D. (1993). Learning and teaching. New York: Allyn Bacon.

Kennedy, J., Baxter, P., Senft, E., & Belpaeme, T. (2016). Social robot tutoring for child second language learning. ACM/IEEE International Conference on Human-Robot Interaction, 2016–April 231–238. http://doi.org/10.1109/HRI.2016.7451757

Kim, Y., Baylor, A., Shen, E.: Pedagogical agents as learning companions: the impact of agent emotion and gender. Journal of Computer Assisted Learning 23(3), 220–234 (2007)

Kramer, Nicole C., et al. "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." Computers & Education 99 (2016): 1-13.

Krippendorff, K. (2011). "Computing Krippendorff's alpha-reliability." Philadelphia: Annenberg School for Communication Departmental Papers.

Kuchenbrandt D, Haring M, Eichberg J, Eyssel F. Keep an eye on the task! how gender typicality of tasks influence human–robot interactions. The sixth international conference on social robotics. October 27–29, 2014. Sydney, Australia. In: Int J Soc Robot 2014:417–427.

Kumar, R., Ai, H., Beuth, J., Rosé, C.: Socially Capable Conversational Tutors Can Be Effective in Collaborative Learning Situations. In: Aleven, V., Kay, J., Mostow, J. (eds.) ITS 2010. LNCS, vol. 6094, pp. 156–164. Springer, Heidelberg (2010)

Lakin, J. L., and Chartrand, T. "Using nonconscious behavioral mimicry to create affiliation and rapport." Psychological science 14.4 (2003): 334-339.

Lane, H. C., et al. (2013, July). The effects of a pedagogical agent for informal science education on learner behaviors and self-efficacy. In International Conference on Artificial Intelligence in Education (pp. 309-318). Springer, Berlin, Heidelberg.

Lee, C.C., et al. Quantification of prosodic entrainment in affective spontaneous spoken interactions of married couples. In: Proc. Interspeech, Makuhari, Japan. pp. 793–796. 2010.

Lee, M., Forlizzi, J., & Kiesler, S. (2012). Personalization in HRI: A longitudinal field experiment. … (HRI), 2012 7th ACM …, 319–326. http://doi.org/10.1145/2157689.2157804

Lee, Namseok, Hochul Shin, and S. Shyam Sundar. "Utilitarian vs. hedonic robots: role of parasocial tendency and anthropomorphism in shaping user attitudes." Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on. IEEE, 2011.

Leelawong, K., & Biswas, G. (2008). Designing Learning by Teaching Agents: The Betty's Brain System. International Journal of AI in Education, 18(3), 181-208.

Lelong, A., & Bailly, G. (2011). Study of the phenomenon of phonetic convergence thanks to speech dominoes. In Analysis of Verbal and Nonverbal Communication and Enactment. The Processing Issues (pp. 273-286). Springer Berlin Heidelberg.

Lemaignan, S., Jacq, A., Hood, D., Garcia, F., Paiva, A., & Dillenbourg, P. (2016). Learning by Teaching a Robot: The Case of Handwriting. IEEE Robotics & Automation Magazine, 23(2), 56-66.

Levine, M. D. (2002). Educational care: a system for understanding and helping children with learning differences at home and in school. Cambridge, MA: Educators Pub. Service.

Levitan, R., & Hirschberg, J. B. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions.

Levitan, R., Benus, S., Galvez, R. H., Gravano, A., Savoretti, F., Trnka, M., Hirschberg, J. (2016). Implementing acoustic-prosodic entrainment in a conversational avatar. Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 08–12–Sept, 1166–1170. http://doi.org/10.21437/Interspeech.2016-985

Levitan, R., Benus, Š., Gravano, A., & Hirschberg, J. (2015). Entrainment and Turn-Taking in Human-Human Dialogue. AAAI Spring Symposium on Turn-Taking and Coordination in Human- Machine Interaction, (Gravano 2009), 44–51.

Levitan, R., Gravano, A., & Hirschberg, J. (2011, June). Entrainment in speech preceding backchannels. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (pp. 113-117). Association for Computational Linguistics.

Levitan, Rivka, et al. "Acoustic-prosodic entrainment and social behavior." Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human language technologies. Association for Computational Linguistics, 2012.

Leyzberg, D., Spaulding, S., & Scassellati, B. (2014). Personalizing Robot Tutors to Individuals' Learning Differences.

Leyzberg, Dan, et al. "Robots that express emotion elicit better human teaching." Proceedings of the 6th international conference on Human-robot interaction. ACM, 2011.

Litman, D. J., Rosé, C. P., Forbes-Riley, K., VanLehn, K., Bhembe, D., & Silliman, S. (2006). Spoken versus typed human and computer dialogue tutoring. International Journal of Artificial Intelligence in Education, 16(2), 145-170.

Liu, P., Glas, D. F., Kanda, T., Ishiguro, H., & Hagita, N. (2013). It's not polite to point: Generating socially appropriate deictic behaviors towards people. 2013 8th ACM/IEEE HRI.

Longster, Jennifer Ann. Concatenative speech synthesis: A Framework for Reducing Perceived Distortion when using the TD-PSOLA Algorithm. Diss. Bournemouth University, 2003.

Lubold, N., & Pon-Barry, H. (2014, November). Acoustic-prosodic entrainment and rapport in collaborative learning dialogues. In Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge (pp. 5-12). ACM.

Lubold, N., Walker, E., & Pon-Barry, H. (2016, March). Effects of voice-adaptation and social dialogue on perceptions of a robotic learning companion. In Human-Robot Interaction (HRI), 2016 11th ACM/IEEE International Conference on (pp. 255-262). IEEE.

Lubold, Nichola, and Heather Pon-Barry. "Acoustic-prosodic entrainment and rapport in collaborative learning dialogues." Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge. ACM, 2014.

Lubold, Nichola, Erin Walker, and Heather Pon-Barry. "Relating Entrainment, Grounding, and Topic of Discussion in Collaborative Learning Dialogues." Proceedings of Computer Supported Collaborative Learning. 2015.

Lubold, Nichola, Heather Pon-Barry, and Erin Walker. "Naturalness and rapport in a pitch adaptive learning companion." Automatic Speech Recognition and Understanding (ASRU), 2015 IEEE Workshop on. IEEE, 2015.

Lucas, Gale M., et al. "Getting to Know Each Other: The Role of Social Dialogue in Recovery from Errors in Social Robots." Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction. ACM, 2018.

Lutfi, Syaheerah Lebai, et al. "I feel you: the design and evaluation of a domestic affect-sensitive spoken conversational agent." Sensors 13.8 (2013): 10519-10538

Madaio, M. A., Ogan, A., and Cassell, J. The Effect of Friendship and Tutoring Roles on Reciprocal Peer Tutoring Strategies. Intelligent Tutoring Systems Lecture Notes in Computer Science, 423-429. 2016.

Mairesse, François, and Marilyn A. Walker. "Towards personality-based user adaptation: psychologically informed stylistic language generation." User Modeling and User-Adapted Interaction 20.3 (2010): 227-278.

Maldonado, H., Lee, J., Brave, S., Nass, C., Nakajima, H., Yamada, R., Iwamura, K., Morishima, Y.: I learn better together. In: Proc. CSCL 2005, pp. 408–417 (2005)

Manson, Joseph H; Gregory A Bryant, Matthew M Gervais, and Michelle A Kline. Convergence of speech rate in conversation predicts cooperation. Evolution and Human Behavior, 34(6):419–426, 2013.

McDaniel, B.T., et al..: Facial Features for Affective State Detection in Learning Environments. In: Proceedings of the 29th Annual Meeting of the Cognitive Science Society, pp. 467–472 (2007)

Michael Natale. Convergence of mean vocal intensity in dyadic communication as a function of social desirability. Journal of Personality and Social Psychology, 32(5):790–804, 1975.

Miller, Andrew D., et al. "Partners in care: design considerations for caregivers and patients during a hospital stay." Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. ACM, 2016.

Miles, M. B., Huberman, A. M., & Saldana, J. (2013). Qualitative data analysis. Sage.

Mills, Sara. Gender and politeness. Vol. 17. Cambridge University Press, 2003

Min, Wookhee, et al. ""Thanks Alisha, Keep in Touch": Gender Effects and Engagement with Virtual Learning Companions." Artificial Intelligence in Education: 18th International Conference, AIED 2017, Wuhan, China, June 28–July 1, 2017, Proceedings. Vol. 10331. Springer, 2017.

Mullany, Louise Jane. Identity and role construction: a sociolinguistic study of gender and discourse in management. Diss. British Thesis Service, The British Library, 2003.

Murphy, Kevin. "The Bayes net toolbox for MATLAB." Computing science and statistics 33.2 (2001): 1024-1034.

Mushin, I., Stirling, L., Fletcher, J., & Wales, R. (2003). Discourse structure, grounding, and prosody in task-oriented dialogue. Discourse Processes, 35(1), 1-31.

Nakatani, C., & Traum, D. (1999). Coding discourse structure in dialogue. University of Maryland Institute for Advanced Computer Studies Technical Report, 1, 1-42.

Nass, Clifford Ivar, and Scott Brave. Wired for speech: How voice activates and advances the human-computer relationship. Cambridge, MA: MIT press, 2005.

Nenkova, A., Gravano, A., & Hirschberg, J. (2008, June). High frequency word entrainment in spoken dialogue. In Proceedings of the 46th annual meeting of the association for computational linguistics on human language technologies: Short papers (pp. 169-172). Association for Computational Linguistics.

Niederhoffer, K.G., and J. W. Pennebaker. Linguistic style matching in social interaction. Journal of Language and Social Psychology, 21(4):337–360, 2002

Nomura, T. (2016). Robots and Gender. Gender and the Genome, 1(1), gg.2016.29002.nom. http://doi.org/10.1089/gg.2016.29002.nom

Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, co-presence, and social presence in virtual environments. Presence: Teleoperators & Virtual Environments, 12(5), 481-494.

Nwe, Tin Lay, Say Wei Foo, and Liyanage C. De Silva. "Speech emotion recognition using hidden Markov models." Speech communication 41.4 (2003): 603-623.

Ogan, Amy, et al. "Intercultural negotiation with virtual humans: The effect of social goals on gameplay and learning." International Conference on Intelligent Tutoring Systems. Springer, Berlin, Heidelberg, 2010.

Ogan, A., Finkelstein, S., Mayfield, E., Dadamo, C., Matsuda, N., & Cassell, J. (2012a). "Oh, dear Stacy!" Social interaction, elaboration, and learning with teachable agents. Proc. of the 2012 ACM CHI '12.

Ogan, A., Finkelstein, S., Walker, E., Carlson, R., & Cassell, J. (2012b). Rudeness and Rapport: Insults and Learning Gains in Peer Tutoring. Intelligent Tutoring Systems Lecture Notes in CS, 11-21.

Otterbacher, J., & Talias, M. (2017). S / he ' s too Warm / Agentic ! The Influence of Gender on Uncanny Reactions to Robots. Human-Robot Interaction, 214–223.

Pampaka, M., Williams, J., & Hutcheson, G. (2012). Measuring students' transition into university and its association with learning outcomes. British Education Research Journal, 38(6), 1041-1071.

Pantic, Maja, et al. "Human computing and machine understanding of human behavior: A survey." Artificial Intelligence for Human Computing. Springer, Berlin, Heidelberg, 2007. 47-71.

Pardo, J.S.: On phonetic convergence during conversational interaction. Journal of the Acoustical Society of America 119(4), 2382–2393 (2006)

Pareto, L., Arvemo, T., Dahl, Y., Haake, M., Gulz, A.: A Teachable-Agent Arithmetic Game's Effects on Mathematics Understanding, Attitude and Self-efficacy. In: Biswas, G., Bull, S., Kay, J., Mitrovic, A. (eds.) AIED 2011. LNCS, vol. 6738, pp. 247–255. Springer, Heidelberg

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. Behavioral and Brain Sciences, 36(4), 329-347.

Pineau, J., M. Montemerlo, M. Pollack, N. Roy and S. Thrun, Towards Robotic Assistants in Nursing Homes: Challenges and Results, Robotics and Autonomous Systems 42(3–4), pp. 271–281, 2003.

Rea, Daniel J., Yan Wang, and James E. Young. "Check Your Stereotypes at the Door: An Analysis of Gender Typecasts in Social Human-Robot Interaction." International Conference on Social Robotics. Springer International Publishing, 2015.

Reich-stiebert, N., & Eyssel, F. (2017). (Ir) relevance of Gender? On the Influence of Gender Stereotypes on Learning with a Robot. Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction - HRI '17, (March 6-9), 166–176.

Richmond, V., & McCroskey, J. (1995). Immediacy, Nonverbal Behavior in Interpersonal Relations (pp. 195-217). Boston: Allyn & Bacon.

Robinson, D. R., Schofield, J. W., & Steers-Wentzel, K. L. (2005). Peer and Cross-Age Tutoring in Math: Outcomes and Their Design Implications. Educational Psychology Review, 17(4).

Roscoe, R. D., & Chi, M. T. (2007). Understanding Tutor Learning: Knowledge-Building and Knowledge-Telling in Peer Tutors Explanations and Questions. Review of Educational Research, 77(4), 534-574.

Rosé, C. P., Bhembe, D., Siler, S., Srivastava, R., & VanLehn, K. (2003). The role of why questions in effective human tutoring. In U. Hoppe, F. Verdejo & J. Kay (Eds.) Artificial Intelligence in Education. Amsterdam: IOS Press.

Sadoughi, N., Pereira, A., Jain, R., Leite, I., & Lehman, J. F. (2017, March). Creating Prosodic Synchrony for a Robot Co-player in a Speech-controlled Game for Children. In Proceedings of the 2017 ACM/IEEE Int'l Conference on Human-Robot Interaction (pp. 91-99). ACM.

Saerbeck, M., Schut, T., Bartneck, C., & Janse, M. D. (2010). Expressive robots in education: varying the degree of social supportive behavior of a robotic tutor. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1613-1622). ACM.

Schermerhorn P, Scheutz M, Crowell CR. Robot social presence and gender: do females view robots differently than males? In: Proceedings of the 2008 ACM/IEEE international conference on human–robot interaction (HRI 2008). March 12–15, 2008. Amsterdam, The Netherlands. Piscataway, NJ: IEEE; 2008:263–270.

Schuller, B.l, B. Vlasenko, F. Eyben, G. Rigoll, and A. Wendemuth. Acoustic emotion recognition: A benchmark comparison of performances. In Automatic Speech Recognition & Understanding, 2009. ASRU 2009. IEEE Workshop on, pages 552–557. IEEE, 2009.

Schweitzer, A., & Lewandowski, N. (2013). Convergence of articulation rate in spontaneous speech. Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 525–529.

Semin, G. R., & Cacioppo, J. T. (2008). Grounding social cognition: Synchronization, entrainment, and coordination. Embodied grounding: Social, cognitive, affective, and neuroscientific approaches, 119-147.

Siegel, Mikey, Cynthia Breazeal, and Michael I. Norton. "Persuasive robotics: The influence of robot gender on human behavior." Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on. IEEE, 2009.

Silvervarg, A., Gulz A. & Sjödén, B. (2010). Design for off-task interaction – Rethinking pedagogy in technology enhanced learning. In Proceedings of the 10th IEEE International Conference on Advanced Learning Technologies, July 5-7, Tunisia.

Sinha, T., & Cassell, J. (2015, November). I click, I align, I 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 (pp. 13-20). ACM.

Skelton, J. R., and F. D. R. Hobbs. "Descriptive study of cooperative language in primary care consultations by male and female doctors." Bmj 318.7183 (1999): 576-579.

Sobral, Dejano T. "Cross‐year peer tutoring experience in a medical school: conditions and outcomes for student tutors." Medical education 36.11 (2002): 1064-1070.

Spencer-Oatey, H. (2002). Managing rapport in talk: Using rapport sensitive incidents to explore the motivational concerns underlying the management of relations. Journal of Pragmatics, 34(5), 529–545. http://doi.org/10.1016/S0378-2166(01)00039-X

Spencer-Oatey, H. (2005). (Im)Politeness, face and perceptions of rapport: unpacking their bases and interrelationships. Politeness Research, 1(1) (Im), 95–119. http://doi.org/10.1515/jplr.2005.1.1.95

Strait M, Briggs P, Scheutz M. Gender, more so than age, modulates positive perceptions of language-based human– robot interactions. In: Salem M, Weiss A, Baxter P, Dautenhahn, eds. 4th international symposium on new frontiers in human robot interaction. April 21–22, 2015. Canterbury, United Kingdom. Online document at: https://hrilab.tufts.edu/publications/straitetal15aisb.pdf Last accessed October 30, 2016.

Stewart, Moira, et al. "Evidence on patient-doctor communication." Cancer 25.1999 (1999): 30.

Swerts, M. G. J., & Krahmer, E. J. (2008). Facial expressions and prosodic prominence: Effects of modality and facial area. Journal of Phonetics, 36(2), 219-238.

Syrdal, Ann, et al. "TD-PSOLA versus harmonic plus noise model in diphone based speech synthesis." Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on. Vol. 1. IEEE, 1998.

Szafir, Daniel, and Bilge Mutlu. "Pay attention!: designing adaptive agents that monitor and improve user engagement." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2012.

Tanaka, F., & Matsuzoe, S. (2012). Children Teach a Care-Receiving Robot to Promote Their Learning: Field Experiments in a Classroom for Vocabulary Learning. Journal of Human-Robot Interaction, 1(1), 78–95. http://doi.org/10.5898/JHRI.1.1.Tanaka

Tannen D. 1994. Gender and discourse. London: Oxford University Press.

Tapus, Adriana, Cristian Ţăpuş, and Maja J. Matarić. "User—robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy." Intelligent Service Robotics 1.2 (2008): 169-183.

Ten Bosch, Louis, Nelleke Oostdijk, and Lou Boves. "On temporal aspects of turn taking in conversational dialogues." Speech Communication 47.1 (2005): 80-86.

Thomason, J., Nguyen, H. V., & Litman, D. (2013, July). Prosodic entrainment and tutoring dialogue success. In International Conference on Artificial Intelligence in Education (pp. 750-753). Springer, Berlin, Heidelberg.

Tickle-Degnen, L., & Rosenthal, R. (1990). The Nature of Rapport and Its Nonverbal Correlates. Psychological Inquiry, 1(4), 285-293.

Topping, Keith, and Marjorie Whiteley. "Sex differences in the effectiveness of peer tutoring." School Psychology International 14.1 (1993): 57-67.

Traum, D. R. (1999, November). Computational models of grounding in collaborative systems. In Psychological Models of Communication in Collaborative Systems-Papers from the AAAI Fall Symposium (pp. 124-131).

Traum, D., Dillenbourg, P., & Route De Drize, B. A. T. D. (1998, August). Towards a normative model of grounding in collaboration. In ESSLLI-98 Workshop on Mutual Knowledge, Common Ground and Public Information.

Tsui, Amy BM. "Reticence and anxiety in second language learning." Voices from the language classroom (1996): 145-167.

Turkle, S., Breazeal, C., Dasté, O., & Scassellati, B. (2006). Encounters with kismet and cog: Children respond to relational artifacts. Digital media: Transformations in human communication, 120.

Vail, Alexandria Katarina, et al. "The Mars and Venus effect: the influence of user gender on the effectiveness of adaptive task support." International Conference on User Modeling, Adaptation, and Personalization. Springer International Publishing, 2015.

Van Mulken, S., André, E., & Müller, J. (1998). The persona effect: how substantial is it?. In People and computers XIII (pp. 53-66). Springer London.

Vygotsky, L. S. (1978). Mind and Society: the Development of Higher Mental Processes,(edited by Cole, M., et. al).

Walker, E., Girotto, V., Kim, Y., & Muldner, K. (2016). The Effects of Physical form and Embodied Action in a Teachable Robot for Geometry Learning. 2016 IEEE 16th - ICALT. IEEE, 381–385.

Wallace, R. S. (2003). The Elements of AIML Style. ALICE A.I. Foundation, Inc.

Wang, N., Johnson, W.L., Mayer, R.E., Rizzo, P., Shaw, E., Collins, H.: The politeness effect. Int. J. Human Computer Studies 66, 96–112 (2008)

Wang, Ning, and Jonathan Gratch. "Rapport and facial expression." 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops. IEEE, 2009.

Wang, Y., & Young, J. E. (2014, May). Beyond Pink and Blue: Gendered Attitudes towards Robots in Society. In Proceedings of Gender and IT Appropriation. Science and Practice on Dialogue-Forum for Interdisciplinary Exchange (p. 49). European Society for Socially Embedded Technologies.

Ward, A., & Litman, D. (2007). Automatically measuring lexical and acoustic/prosodic convergence in tutorial dialog corpora. In Workshop on Speech and Language Technology in Education.

Weizenbaum, Joseph. "ELIZA—a computer program for the study of natural language communication between man and machine." Communications of the ACM 9.1 (1966): 36-45

Westlund, J. K., & Breazeal, C. (2015, March). 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 (pp. 65-66). ACM.

Westlund, K, et al. (2017). Flat vs. Expressive Storytelling: Young Children's Learning and Retention of a Social Robot's Narrative. Frontiers in human neuroscience, 11, 295.

Westlund,K., J. M., Lee, J., Plummer, L., Faridia, F., Gray, J., Berlin, M., Quintus-Bosz, H., Harmann, R., Hess, M., Dyer, S., dos Santos, K., Adalgeirsson, S., Gordon, G., Spaulding, S., Martinez, M., Das, M., Archie, M., Jeong, S., & Breazeal, C. (2016). Tega: A Social Robot. In Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction: Video Presentations, pp. 561

Wheldall, Kevin, and Paul Mettem. "Behavioral Peer Tutoring: training 16-year-old tutors to employ the 'pause, prompt and praise' method with 12-year-old remedial readers." Educational Psychology 5.1 (1985): 27-44.

Zhao, R., Papangelis, A., & Cassell, J. (2014). Towards a dyadic computational model of rapport management for human-virtual agent interaction. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8637 LNAI, 514–527. http://doi.org/10.1007/978-3-319-09767-1_62

APPENDIX A

QUINN – MEASURES AND CODING SCHEME

Rapport Measures: Participants responded on a Likert scale from 1 to 5

I felt I had a connection with Quinn
I felt I was able to engage Quinn
I think that Quinn and I understood each other
I felt that Quinn was interested in what I had to say
I felt that Quinn was warm and caring
I felt that Quinn was intensely involved in the interaction
I felt that Quinn seemed to find the interaction stimulating
I felt that Quinn was respectful to me
I felt that Quinn showed enthusiasm while talking to me

Social Presence Measures: Participants responded on a Likert scale from 1 to 7

Quinn was easily distracted
I was easily distracted
Quinn tended to ignore me
I tended to ignore Quinn
I sometimes pretend to pay attention to Quinn
Quinn sometimes pretended to pay attention to me
Quinn paid close attention to me

Coding Scheme:

Politeness: "P" is polite to Quinn, follows conversational niceties (like saying hello)
        Ex 1: Thank you, Quinn
        Ex 2: ah step four please

Complimenting or praising: "P" praises Quinn
        Ex 1: good job Quinn
        Ex 2: great! Now I factor out the two
        Ex 3: nice!

Name usage: "P" uses Quinn's name
        Ex 1: Nice job Quinn (this would contain checks in both the praise column and
        the name column)

Inclusive: "P" includes Quinn, for example by using 'inclusive' language such as "us,"
"we,", "together", "let's"
        Ex 1: Let's do problem one!

Empathy: "P" expresses sympathy or empathy towards Quinn. Responds to Quinn's
complaints, responds to concerns with agreement, and empathy
        Ex 1: me too Quinn
        NOTE: Empathy coded for Nico and Emma only

APPENDIX B

NICO – MEASURES

Pre and Post Tests:

Test A

Galaxy Stats	Number of Objects
Moons	17
Stars	22

1. The table shows the numbers of moons and stars in a galaxy. What is the ratio of moons to all possible objects?

2. 2:3 and 4:6 are equivalent ratios. Write in numbers an equivalent ratio to the ratio you gave in problem one.

Use the chart below to answer questions 3 – 4.

Number of Hours Since Park Opening	2.2	3	7
Number of Guests at the Park	220	300	

3. Write a rule that describes the data in the chart.

4. Use the rule to find the number of guests at the park after 7 hours.

5. Alicia and her brother are holding a Halloween party. She has made a special drink with 4 cups of fruit juice and 8 cups of seltzer water. Alicia has also made 20 cupcakes and 15 hotdogs. She has just found out though that her brother invited 3 times as many guests as she thought! How much more does she need to prepare?

6. In a package of star stickers there are 15 red stars, 25 gold stars and 10 green stars. What is the ratio of red to gold stars (in simplest form)?

7. On a triangle, each side measures 5 cm, 10 cm, and 30 cm, respectively. In lowest terms, find the ratios of the lengths of the sides.

8. John can buy 3 books for $18.75. How many books could John buy for $54?

9. Allison can read 80 pages in 2 hours. How many hours will it take her to read 240 pages?

10. Albert owns a bakery that specializes in chocolate chip cookies. He has very specific standards for his cookies. Albert personally checks the cookies before they are sold. This is the table he uses when checking cookies.

Number of Cookies	Number of Chocolate Chips
2	30
4	60
5	75
7	105

Which of the following equations can be used to find CC, the number of chocolate chips you can find on any number of C, cookies?

a. CC = 28 + C
b. CC = 15 + C
c. CC = 15 x C
d. CC = 30 x C

Test B

Activity	Time (min)
Treadmill	25
Lifting Weights	35

1. The table shows how Levon spends his time at the gym. What is the ratio of the time lifting weights to all activity time?

2. 2:3 and 4:6 are equivalent ratios. Write in numbers an equivalent ratio to the ratio you gave in problem one.

Use the below table to answer questions 3 – 4.

Cost ($)	46	72	
Number of Hours	2.3	3.6	5

3. Write a rule that describes the data in the chart.

4. Use the rule to find the cost after 5 hours.

5. On his mother's birthday, Juan has cooked dinner for his mother and some guests. He made a huge pot of rice with 4 cups of rice and 8 cups of water. He made two pernil (pork shoulders) and baked 3 cakes for dessert. His mother has decided to triple (invite three times as many) guests. How many more rice, water, pernil, and cakes should Juan cook?

6. At a putt-putt course there are 50 yellow golf balls, 45 red golf balls, and 65 blue golf balls. What ratio compares the number of blue golf balls to the total number of golf balls (in simplest form)?

7. On a triangle, each side measures 5 cm, 10 cm, and 30 cm, respectively. In lowest terms, find the ratios of the lengths of the sides.

8. Genevieve spent $56.25 to fill her 15-gallon tank. How much did she pay per gallon?

9. Leo buys 5 DVDs for $60. At this rate, how much would he pay for 3 DVDs?

10. Zylfina is an elf who likes flowers. Every year she counts the number of flowers that appear in the meadow outside of her house. She's been keeping track of the number of flowers for many years and she has found that there is a relationship between the number of flowers and the amount of rain that falls. Using the table below, which of the following equations explains the total number of flowers (flowers) per inch of rainfall (rain)?

Number of flowers	Inches of rainfall
48	4
24	2
72	6

    a. flowers = 22 + rain
    b. flowers = 22 x rain
    c. flowers = 12 + rain
    d. flowers = 12 * rain

Rapport: Participants responded on a Likert scale from 1 to 5

Emma and I understood each other
Emma and I had a connection

Emma was easily distracted
Emma paid close attention to me
I was easily distracted
I paid close attention to Emma

Emma liked me
I liked Emma
I was unfriendly to Emma
Emma was unfriendly to me

Emma was awkward in talking to me
My conversation with Emma was easy
I was awkward in talking with Emma
Emma's conversation with me was easy

Self-Efficacy: Participants responded on a Likert scale from 1 to 5
I am good at math
I struggle with explaining how to solve math problems to others
I can help others learn
I can explain math problems to another student my age
Math is hard for me.
Ratio problems are hard for me

Mastery and Social Persuasion:
Emma learned because I explained the problems well
Emma would want me to help again because she thinks I'm good at ratios
I can help Emma learn ratios in the future
I can help Emma learn math in the future
Emma learned because I am good at math

Comfort-Level Around Robots
I feel comfortable interacting with human-looking robots
I feel comfortable interacting with robots

APPENDIX C

EMMA – MEASURES

Pretest

Problem 1:

Finish filling in the gray boxes below by using the relationship between A and B.

Step	A	B
0	6	2
1	3	
2		3

Problem 2

Lucy can eat more ice cream faster than anyone else in her class. She eats $4\frac{1}{2}$ gallons every $\frac{1}{3}$ of an hour. She eats the ice cream a constant rate. How many gallons of ice cream can she eat in one hour?

Give your answer as improper fraction rather than a mixed fraction (for example, 5/4).

Problem 3.

Finish filling in the gray boxes below by using the relationship between A and B.

Step	A	B
0	2.5	25
1		30
2	7.25	

Write the following as mixed fractions:

Problem 4.   22/6   _____

Problem 5.   18/4   _____

Problem 6.   13/5   _____

Problem 7.

Finish filling in the gray boxes below by using the relationship between A and B.

Step	A	B
0	$\frac{1}{20}$	$\frac{2}{3}$
1		1
2		3

Problem 8.

Select the brand with the least expensive corn per ounce.

    a. **Brand A:**

Ounces	Price
18	$1.50
36	$3.00
54	$4.50

    b. **Brand B:**

$0.75
Brand B

15 oz
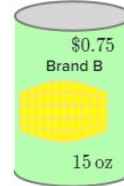
    c. Maia buys an **11** ounce can of **Brand C** corn for **$2.50**

Problem 9.

Which table has a constant of proportionality (or consistent ratio) between y and x of $\frac{1}{3}$?
Circle the letter of the correct table.

A		B		C		D	
x	y	x	y	x	y	x	y
8	2	2	1	6	1.5	6	2
10	2.5	8	4	7	2	9	3
12	3	6	12	8	2.5	15	5

231

Problem 10.

The following table shows the number of tickets purchased for a popular concert for every hour they are on sale. What is the relationship between the number of hours the tickets are on sale and the number of tickets sold?

Number of hours	1	2	5
Tickets sold	15	30	75

Posttest

Problem 1:

Finish filling in the gray boxes below by using the relationship between A and B.

Step	A	B
0	12	6
1	4	
2		15

Problem 2

Ben drinks tea at an incredible rate. He drinks $3\frac{1}{2}$, liters of tea every $\frac{2}{3}$ of an hour. Ben drinks tea at a constant rate. How many liters of tea does he drink in one hour?

Give your answer as improper fraction rather than a mixed fraction (for example, 5/4).

Problem 3.

Finish filling in the gray boxes below by using the relationship between A and B.

Step	A	B
0	3.1	6.2
1		12
2	9	

232

Write the following as mixed fractions:

Problem 4.   19/8   _____

Problem 5.   29/4   _____

Problem 6.   7/3   _____

Problem 7.

Finish filling in the gray boxes below by using the relationship between A and B.

Step	A	B
0	$\dfrac{1}{10}$	$\dfrac{1}{2}$
1		2
2		5

Problem 8.

Select the house with the best price per square foot.

a. **House A:**

Room Sq. Footage	Price
20	$180
30	$270
6.5	$58.5

b. **House B:**

$ 930
310 sq. feet
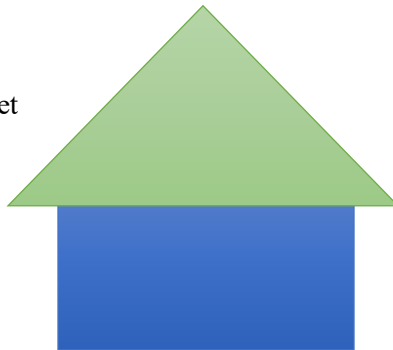
c. **House C:** Maia buys a **400** square foot house for **$1600**

Problem 9.

Which table has a consistent ratio between y and x of $\frac{1}{4}$? Circle the letter of the correct table.

	A			B			C			D	
X	**y**		**x**	**y**		**x**	**y**		**x**	**y**	
8	2		3	1		6	1.5		2	1	
10	2.5		6	2		7	2		4	2	
12	3		9	3		8	2.5		8	4	

Problem 10.

The following table shows the number of tickets purchased for a popular concert for every hour they are on sale. What is the relationship between the number of hours the tickets are on sale and the number of tickets sold?

Number of hours	1	2	5
Tickets sold	12	24	60