DESIGNING EFFECTIVE COMMUNICATION STRATEGIES FOR HUMAN-ROBOT COLLABORATION

by

Allison V. Sauppé

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Computer Sciences)

at the

UNIVERSITY OF WISCONSIN–MADISON

2015

Date of final oral examination: June 10, 2015

The dissertation is approved by the following members of the Final Oral Committee:

Bilge Mutlu, Assoc. Prof., Computer Sciences, University of Wisconsin–Madison
Martina Rau, Asst. Prof., Educational Sciences, University of Wisconsin–Madison
Julie Shah, Assoc. Prof., Aeronautics and Astronautics, Massachusetts Inst. of Technology
Benjamin Snyder, Asst. Prof., Computer Sciences, University of Wisconsin–Madison
Xiaojin Zhu, Assoc. Prof., Computer Sciences, University of Wisconsin–Madison

To William, Louise, Elizabeth, and John Terrell, for nurturing a love of learning,

&

to Jason Sauppé, for his endless love and encouragement.

ACKNOWLEDGMENTS

This work would never have been possible without the patient guidance of my advisor, Bilge Mutlu. His invaluable mentorship has shaped me not only as a researcher, but as a designer, statistician, public speaker, and instructor. I would also like to thank my committee, Martina Rau, Julie Shah, Ben Snyder, and Jerry Zhu, for their thoughtful feedback and advice on this work.

I would like to especially thank my labmates over the years from the Human-Computer Interaction laboratory, including Sean Andrist, Faisal Kahn, Toshikazu Kanaoka, Erdem Kaya, Yun Kim, Chien-Ming Huang, Steve Johnson, Margaret Pearce, Tomislav Pejsa, Irene Rae, Victoria Schröder, Dan Szafir, Zhi Tan, and Kohei Yoshikawa. You have each been an invaluable sounding board for my work, and a great source of joy and friendship. I would also like to thank all of the staff in the HCI lab who helped support my research.

I am grateful to my parents, Doug and Louise Terrell, for the love of learning and creativity they inspired in me from a young age, and for the unwavering love and support from my parents, siblings, extended family, and in-laws, even when they were unsure of what it is I do. Also, my husband, Jason Sauppé, has been an invaluable source of love and support throughout this journey—I'm so glad we traveled this road together.

I would also like to thank all those students I encountered during my studies that provided their friendship and guidance, including Eric Alexander, Aubrey Barnard, Aaron Gember-Jacobson, Emily Gember-Jacobson, Bryan Gibson, Matthew Gombolay, Lena Olson, Ian Rae, Elizabeth Soechting, and Danielle Albers Szafir. A special

thanks to Drew Houvener, Mike Mrozek, and Irwin “Snooze” Reyes for their support from afar during this time.

I would like to thank Matt Boutell, Curt Clifton, and Sriram Mohan for their advice and encouragement as I pursued graduate school. I had no idea what I was getting into, and your guidance was invaluable.

I appreciate the support I received from David DeWitt during my second year of graduate school which was crucial in allowing me to continue my studies for my Ph.D. Also, I would like to thank the sponsors of this work, including the National Science Foundation and Mitsubishi Heavy Industries, Ltd.

Finally, I am grateful to God for His persistent faithfulness and inspiration in my life. *Deo gratias.*

CONTENTS

Contents iv

List of Figures vii

Abstract xi

1 Introduction 1

1.1 Context and Scope 4

1.2 Research Approach 10

1.3 Research Platforms 15

1.4 Outline 17

2 Related Work 19

2.1 Communication and Collaboration 19

2.2 Behavioral Cues for Collaboration 30

2.3 Programming Social Behavior on Robots 39

2.4 Chapter Summary 41

3 Formative Study: Understanding the Reality of Collaborative Robots 42

3.1 Robots in Manufacturing 43

3.2 Study Outline 44

3.3 Method 47

3.4 Results 51

3.5 Discussion 69

3.6 Study Conclusions 75

4 Behavioral Cue Study: Speech Patterns 77

4.1 Formative Study 78

4.2 Interaction Blocks 90

4.3 Discussion 101

4.4 Study Conclusions 105

5 Behavioral Cue Study: Instruction and Repair 107

5.1 Formative Study 108

5.2 Implementation 116

5.3 Experimental Evaluation 123

5.4 Discussion 129

5.5 Study Conclusions 133

6 Behavioral Cue Study: Deictic Gestures 134

6.1 Understanding Deictic Gestures and Settings 135

6.2 Implementation 139

6.3 Experimental Evaluation 142

6.4 Discussion 147

6.5 Study Conclusions 158

7 General Discussion 160

7.1 Summary and Synthesis 160

7.2 Limitations and Future Considerations 166

8 Conclusion 178

References 180

LIST OF FIGURES

1.1 The three robotic platforms used in this study: a) Baxter, b) NAO, and c) Wakamaru. 16

3.1 Examples of a collaborative robot at work packaging medicine cups at Company 1 (left) and moving plastic tubes at Company 2 (right). 45

3.2 An example of the workflow from a work cell for packaging medicine cups adapted for the robot. (1) The conveyor belt has predefined placement for cups in the form of an array of cups drilled into the surface of the conveyor belt. (2) The industrial robot removes cups from the injection-molding machine and places them on the conveyor belt until enough cups are added. (3) The conveyor belt moves forward, and the robot picks up a stack of cups. (4) The robot places the stack in a set of guides for the automatic bagging machine. (5) When enough cups have been added to the guides, the robot sends a signal to the automatic bagging machine to package the cups and deploy a new bag. 52

3.3 The robot from Company 2, dressed up in a wig and jester hat. Previously, the robot was adorned with a rainbow clown wig. 60

3.4 Two operators working near the robot. The vantage point of the operators makes it difficult to monitor the robot's status. Additionally, the tasks the operators are completing requires their visual attention. 64

4.1 Examples of the experimental setup for all five scenarios. From left to right, participants are sharing in a storytelling experience, engaging in a conversation, conducting an interview, learning how to configure a set of pipes, and collaborating on how to sort foodstuffs. 80

4.2 Models for the conversation, collaboration, and interaction scenarios. In the instruction scenario, dark-colored states are for the dominant role (instructor), while light-colored states are for the respondent's role (student). In the conversation and collaboration scenarios, the roles for each participant are not fixed throughout the interaction. E.g., during a conversation, a participant may ask the other a question, which may then be followed by the second participant asking the first participant a similar question. Here, dark-colored and light-colored states indicate that two different participants occupy each state. States in dashed outlines can be occupied by any participant. 83

4.3 Models for the interview and storytelling scenarios. In the interview and storytelling scenarios, dark-colored states are for the dominant role (instructor), while light-colored states are for the respondent's role (student). States which have a dashed outline can be occupied by any participant. . 85

4.4 Models for the seven patterns I discovered. The dark and light-colored states indicate when one participant occupies the dark-colored state, a second participant must occupy the light-colored state. For example, for pattern 5, if one participant asks a question, the second participant gives an answer. 86

4.5 A screenshot of *Interaction Blocks*, my authoring environment for using a pattern language for synthesizing human-robot interactions. 91
4.6 A researcher demonstrating the setup of the design sessions with the Interaction Blocks authoring environment and a NAO humanoid robot. . 94

5.1 The *instructor* (participant on the left) directing the *student* (participant on the right) in assembling a predetermined pipe configuration. 108
5.2 Examples of how the two factors found in my modeling, *instruction grouping* and *instruction summarization*, can be jointly used. 113
5.3 On the right, the setup used in my experimental evaluation. After the robot gave an instruction, the participant retrieved the pieces necessary from behind them and assemble the pieces on the workspace in front of the robot. A camera above the workspace captured the configuration of the pieces. On the left, an example of the robot autonomously guiding a participant in assembling pipes. 124
5.4 Results from my evaluation. Significant and marginal results were found for total task time, number of breakdowns encountered, participants' perceived rapport with the robot, and their overall experience with the task. (†), (*), and (**) denote $\mathrm{p} < .10$, $\mathrm{p} < .050$, and $\mathrm{p} < .010$, respectively. 127

6.1 Instances of a human performer and the NAO robot demonstrating the deictic gestures studied in this work. 135
6.2 A model of the gesture-contingent gaze behavior implemented in my study. Start and end times are relative to the onset of speech. 140

6.3 The two workspaces used to represent the six settings I explored. The left workspace displays the ambiguity (left) and the no visibility (right) settings. The right workspace displays the clustered objects setting (left), the distant from referrer setting (top right), and the neutral setting (bottom right), which was used for both the neutral and noise settings. 142

6.4 A participant evaluating the robot touching the blue block. 145

6.5 Results for both the accuracy of each gesture and the perceived effectiveness of each gesture across all settings. (***), (**) denotes $p < .001$, $p < .01$, respectively. Exhibiting and touching gestures were more accurate than the two baselines and the sweeping and grouping gestures and were perceived to be more effective than the two baselines and the other gestures. 148

6.6 Results for the communicative accuracy of each gesture, displayed by setting. (***), (**), (*) denotes $p < .001$, $p < .01$, and $p < .05$, respectively. Exhibiting and touching were consistently more accurate than sweeping and grouping across the majority of settings. 149

6.7 Results for the perceived effectiveness of each gesture, displayed by setting. (***), (**), (*) denote $p < .001$, $p < .01$, and $p < .05$, respectively. Exhibiting and touching were consistently perceived to be more effective than presenting, pointing, sweeping, and grouping across the majority of the settings. 150

ABSTRACT

Technological advancements are enabling robots to begin working together with humans as partners on physical tasks. To design these *collaborative robots* to work effectively with their human counterparts, we must first understand what expectations and perceptions people have of these robots. With this understanding, we can then design collaborative behaviors that meet those needs, enabling collaborative robots to be more effective partners.

In this dissertation, I aim to address questions that will allow us to design more effective collaborative robots. For example, what do interactions look like when a collaborative robot is introduced into a human environment? How do human partners perceive this robot? What collaborative behaviors would be useful for these collaborations? For those behaviors we would like to design, can we build models by hand from human-human data? From these models, can we make recommendations for how collaborative robots should employ these behaviors?

This dissertation seeks to answer these questions through four studies. The first study examines three manufacturing sites that have adopted collaborative robots in their workflow, using interviews and observations to assess the current status of collaborative robots and provide recommendations about future designs. The remaining three studies, inspired by scenarios similar to the one in the first study, each focus on a specific behavioral cue: speech patterns, teaching and repair, and deictic gestures. Those studies which focus on a specific behavioral cue use human-human data to inform models of behavior that can then be implemented on a robot and tested in a human-robot evaluation, examining the impact of the model on

multiple task outcomes.

The contributions of this work are an understanding of real-world collaborative behaviors, conceptual models of human collaborative behaviors, a contextualization of these behaviors and an understanding of their role in facilitating interactions, and tools to facilitate developing and testing human-robot collaborations. These contributions help to inform the design and implementation of future iterations of collaborative robots.

1 INTRODUCTION

Robots have long been anticipated to serve as ubiquitous assistants that work together with humans in a variety of day-to-day environments. From helping assemble furniture in the home to assisting with tasks in manufacturing settings, robots are envisioned as providing complimentary skill sets to human partners, helping perform tasks that are dangerous, repetitive, or require precision. Technological advancements are beginning to make these *collaborative robots*—a robot that is capable of working together with a human partner on some physical task—a reality. However, while these robots might be technically capable of performing a task with a human partner (e.g., handing an object to a person), there is little work on how the robot should behave when interacting with its human partner during the task.

When humans engage in a physical task together, each person often employs specific behavioral cues or communication strategies to create a more fluid and effective interaction with their partner. People might use their gaze (Griffin, 2001; Meyer et al., 1998; Tomasello, 1995), gestures (Clark, 2005; Kendon, 2004), speech (Clark, 2009; Croft, 2000), body posture (Argyle, 2013; Ekman and Friesen, 1967), or head nods (Helweg-Larsen et al., 2004; Stivers, 2008) to indicate their thoughts or intent regarding the task. However, the use of these cues to convey a particular intent is not always uniform. Rather, which cues are employed, and when they are employed, might vary based on the context of the task. For example, if two people are assembling furniture together, one person might need to inform their partner about which piece to use next. If the piece is located close to the person, they might just pick it up and hand it to their partner, perhaps commenting "You need this

piece next." However, if the person is incapable of acquiring the piece (e.g., their hands are full, the piece is closer to their partner), the person might instead choose to verbally describe the piece, saying "The next piece you need is the bracket to your right." While both approaches effectively convey the next object needed, they differ in which cues are used and how those cues are employed. The use of different approaches results from an understanding of contextual differences and leveraging this understanding to select how to communicate.

To enable collaborative robots to successfully serve as partners to human workers, robots will need to be capable of understanding how to use cues or communication strategies when collaborating on physical tasks, particularly if their use of cues might vary depending on context, as shown in the above example. Prior work has examined how to enable robots to interact socially with humans, including studying many of the social cues outlined previously (Breazeal and Scassellati, 1999; Breazeal et al., 2005; Huang and Mutlu, 2013; Mutlu et al., 2006; Scheeff et al., 2002). However, these contexts are often more conversational in nature, such as a robot helping teach a class (You et al., 2006) or answering or acting as a museum guide (Kopp et al., 2005; Kuno et al., 2007). The introduction of a physical task that the robot is working on with a human partner raises several new questions: What does collaboration look like when a collaborative robot partner is introduced? How do humans view a collaborative robot as a partner on a physical task? What social cues and behavioral skills do humans want the robot to have? And based on the answer to the previous question, how should these behaviors be designed for collaborative robots? How can collaborative robots use these social cues effectively when working with their human

partner? Finally, how can we make implementation of these cues accessible to a wide range of end users?

This dissertation aims to address these questions through a combination of ethnographic field studies, qualitative and quantitative lab studies, and tool design. The contributions of this dissertation can be categorized as follows:

- **Understanding of Real-World Collaborative Behaviors**: Through studying an environment where collaborative robots have already been introduced, a grounded understanding of the needs, desires, and expectations users have for collaborative robots.
- **Computational Models of Human Collaborative Behaviors**: Models of how specific social cues are used, drawn from the literature and human-human observations, that can be implemented on a robot.
- **Contextualization of Behaviors and an Understanding of their Role in Facilitating Interaction**: An understanding of how the use of social cues might differ depending on contextual factors, and how this variation supports a range of collaborative outcomes (e.g., time spent on task, rapport).
- **Tools to Facilitate Developing and Testing Human-Robot Collaborations**: The synthesis of tools to enable a wide range of users to implement social behaviors on collaborative robots.

This chapter outlines the context and questions this dissertation is concerned with, the research approaches used to answer these questions, and describes the robotic platforms considered in this work.

1.1 Context and Scope

This dissertation aims to address two fundamental components of designing effective behaviors for human-robot collaboration. The first component is developing an understanding of the needs, demands, and expectations users will have for these collaborative robots. The second component is creating effective behaviors to meet those needs, demands, and expectations. An understanding of both components is necessary to build a foundation for human-robot collaboration that can translate across settings, from robots working with humans in the home to the workplace. To demonstrate how this dissertation addresses these two components, I provide motivating examples below in the context of human-robot collaboration in manufacturing settings.

1.1.1 Understanding Real-World Needs

> Lyndhurst Manufacturing, a producer of build-it-yourself home furniture, has recently acquired Baxter, a collaborative robot platform. As part of the decision to acquire Baxter, Lyndhurst Manufacturing sought out information about the impact adding a collaborative robot would have on their social environment. In particular, the company was interested in learning more about how their workers would view the robot, whether their workers would feel threatened by the robot's presence, and best practices for integrating Baxter into the workflow of the workers.

The introduction of robots into day-to-day environments has required developing

an understanding of the real-world perceptions and interactions these robots elicit. For example, the introduction of a courier robot in a hospital setting helped to clarify how different stakeholders classified the robot's role and distinguish why owners might attribute humanlike characteristics to the robot (Mutlu and Forlizzi, 2008). Similarly, introducing collaborative robots that work together with humans on a variety of tasks raises many questions about their integration into these environments, perceptions of humans working near or with them, and how interactions might be structured to accommodate the robot. Chapter 3 will present field work conducted at three manufacturing sites that own a collaborative robot called Baxter[1]. The remainder of the research questions regarding human-robot collaboration presented in this dissertation are inspired by scenarios in a manufacturing plant similar to the ones observed as part of this field work.

1.1.2 Creating Effective Behaviors

Inspired by the context and findings illustrated in the ethnographic study, I identified three different social behaviors for effective human-robot collaboration to focus on for the remainder of my dissertation: speech patterns, deictic gestures, and verbal instruction and repair. Below, I outline scenarios in a manufacturing facility inspired by the observations and interviews from the ethnographic study discussed above. Each scenario motivates a particular study contained in the dissertation.

[1]This work was originally published in Sauppé and Mutlu (2015b).

1.1.2.1 Speech Patterns

> Laura, who works as part of the team that maintains equipment at the manufacturing plant, has been tasked with making Baxter more personable for workers to interact with. As part of updating Baxter, Laura chooses to explore some limited social scenarios that Baxter could engage in to build rapport with its co-workers. Laura uses a tool called *Interaction Blocks* to try out a few different social scenarios. After testing these scenarios with co-workers, Laura deploys an update to Baxter, enabling it to engage in limited conversations with co-workers.
>
> Mark has been working with Baxter for several weeks at the same work cell in the manufacturing plant. After the introduction of Laura's updates, Mark is able to greet Baxter as he arrives at the work cell today, and Baxter returns the greeting. Later, Mark asks Baxter a question about which tool he needs to complete his work, and Baxter responds that Mark needs a smaller wrench.

Speech interactions, such as the one described above, are often the most explicit behavior humans use during an interaction, helping to convey information or build social rapport. Goffman argued that human interactions follow a specific "order" and characterized a number of *speech patterns* in which people interact, such as how greetings unfold and how people leave an interaction (Goffman, 1983). As robots are expected to handle an increasingly diverse range of scenarios, such as helping give instructions to patients on physical therapy or asking questions of a student to better

understand problems they are encountering, designers of human-robot interactions, who might be unfamiliar with the formal constructs surrounding dialogue, might be able to leverage these speech patterns to prototype and test interactions on robots. Chapter 4 will explore designing a set of fundamental speech patterns that can be used as building blocks for human-robot interactions[2].

Additionally, in the scenario outlined above, there is a desire to update Baxter with additional social functionality. While robots are often programmed prior to release by people with a specialized background (e.g., computer science), the introduction of robots into real-world settings will likely create scenarios where their everyday users, such as Laura in our example, would like to modify or add to their behavior to customize the robot for their needs. This new set of everyday users will require new authoring environments for implementing robot behavior that are designed for novice users. Chapter 4 also details the design and evaluation of an interface, called *Interaction Blocks* that allows novice users to prototype and test interactions using the speech patterns discussed above.

1.1.2.2 Deictic Gestures

> Joe is a new worker at the manufacturing plant, working with Baxter to affix brackets to a cabinet. For each cabinet, Joe places the brackets, while Baxter rotates the cabinet to ensure Joe can easily reach all the locations where brackets are needed. Different cabinets use different brackets, requiring Joe to refer to a guide for each type of cabinet. While

[2]This work was originally published in Sauppé and Mutlu (2014a).

placing brackets on one cabinet, Baxter notices that Joe is adding the bracket in the wrong place. Baxter stops, explaining to Joe where the most recent bracket should be placed while simultaneously pointing toward the new location. Joe has difficulty understanding Baxter due to the noisy environment and his hearing protection, and asks Baxter for clarification. This time, Baxter physically touches its end effector to the new location while verbally giving an explanation. Joe is able to understand where the bracket should be placed, moves the bracket to the correct location, and then finishes placing the remaining brackets.

Humans use a variety of social cues, including gaze (Griffin, 2001; Meyer et al., 1998; Tomasello, 1995), speech (Clark, 2009; Croft, 2000), and gestures (Clark, 2005; Kendon, 2004), to refer the objects and spaces around them to others. Which cues humans choose to use and how heavily they rely on some cues to communicate depends on a variety of factors, including the environmental context (Lozano and Tversky, 2006) and emphases the speaker may want to place on their communicative act (Goldin-Meadow, 1999; McNeil, 1992). Among these social cues are deictic gestures, a type of gesture that is used to refer to an object or space (Fillmore, 1982; McNeil, 1992). While pointing is considered the canonical deictic gesture, several other types of deictic gestures, such as touching and presenting, are used (Bolinger, 1983; Kendon, 2004; Lempers, 1979). With multiple deictic gestures to choose from, there is little guidance for how robots should select which deictic gesture to use in a given the context. Scenarios where robots might need alternate channels of communication, such as giving instructions in noisy workplaces or working with preverbal children, raise

multiple questions surrounding this space. Are there differences in the effectiveness between different types of deictic gestures? Does the effectiveness of the deictic gesture vary depending on the environmental factors present? Chapter 6 will present a study that examines different types of deictic gestures available, the challenges of implementing those gestures on a robot, and the tradeoffs of using different gestures in different environmental conditions[3].

1.1.2.3 Instruction and Repair

> An employee at the manufacturing plant, Joanie, has been assigned to a new task, working with Baxter to build the frame of a cabinet. During Joanie's first day at her new work station, Baxter verbally walks her through the steps necessary to build the cabinet, letting Joanie build and complete each step prior to giving the next instruction. After Joanie feels comfortable with the sequence of steps, Baxter resumes its role in the process. Baxter moves the completed parts of the cabinet frame, allowing Joanie to add in additional components. When working on one cabinet, Joanie cannot remember where a particularly component fits, so she asks Baxter for help. Baxter explains where the component should go; Joanie adds the component correctly, and they resume building the cabinet.

Humans often engage in instructing one another on a task. While instruction is commonly thought of as occurring as part of formal training, such as part of training for a new job, humans might informally offer instruction during a number of everyday

[3]This work was originally published in Sauppé and Mutlu (2014c).

tasks, from teaching a new recipe to a friend to helping a co-worker with an unfamiliar task, such as in the example scenario presented. Collaborative robots present a unique opportunity for instructional use due to their capacity for a breadth of task instructions. However, while their limitations (e.g., dexterity, precision, adaptability) make it difficult for the collaborative robot to actually complete the task, humans are well-equipped to help a robot carry out a physical task. These differing capabilities lead to pairing humans and collaborative robots to work together, with robots serving in an instructional role, enabling the pair to leverage the strengths of each participant, such as a robot providing instructional support to an astronaut attempting to perform repairs. As a result of this new dynamic, there are many questions regarding how a robot should give instructions that must be addressed. Are there different strategies regarding how the robot should give instructions? If there are multiple strategies, are some strategies better than others? How will the robot correct mistakes the human makes during the task? Chapter 5 will address these questions, presenting a study that examines how human instructors give instructions and correct mistakes, developing models of human instruction-giving that can be implemented on a robot, and evaluating a human-robot instructional task to show how these different strategies affect task performance outcomes, such as the number of mistakes made[4].

1.2 Research Approach

Designing effective behaviors for human-robot collaboration requires both understanding the needs, demands, and expectations users will have for these robots, and then

[4]This work was originally published in Mutlu et al. (2015); Sauppé and Mutlu (2014b, 2015a).

creating effective behaviors to meet those needs, demands, and expectations. To accomplish both of these goals, a multi-faceted approach is needed to both capture the reality of working with a collaborative robot in a day-to-day environment, as well as carefully examining the impact of social behaviors in controlled settings. Below, I discuss the methodology used in this dissertation that enables these approaches.

1.2.1 Ethnographic Field Studies

Ethnographic studies originated as a technique in the field of anthropology, enabling researchers to better characterize the groups they were studying (Taylor and Bogdan, 1998). This characterization occurs through the process of exploring social phenomena by the researcher placing themselves in the context of the group, requiring the researcher to go into *the field* (i.e., to where the group is located). Ethnographic field work has been adopted by other areas outside of anthropology, including the area of human-robot interaction, primarily as an open-ended inquiry into previously unexplored domains. For example, prior ethnographic field work in human-robot interaction has explored the integration of Roombas into home environments (Forlizzi, 2007; Forlizzi and DiSalvo, 2006; Sung et al., 2007), the adoption of a courier robot in a hospital setting (Mutlu and Forlizzi, 2008), and the perceptions of a snack delivery robot in a workplace (Lee et al., 2012). An ethnographic field study is used as the data collection method for the work described in Chapter 3.

While there are multiple data collection methods that can be used in ethnographic field studies, the following discussion will focus on the two data collection methods used in the fieldwork presented in this dissertation: fly-on-the-wall observations and

semi-structured interviews. Below, I present a brief overview of each data collection method.

1.2.1.1 Fly-on-the-Wall Observations

To better understand the environmental context in which a group functions, researchers conduct *fly-on-the-wall observations*, a technique where the researcher observes the environment, as well as the actions and interactions that take place in that environment (Taylor and Bogdan, 1998). The researcher takes detailed field notes, which might include drawing the layout of the physical space, indicating interesting or repeated phenomena, or taking pictures of the environment (Emerson et al., 2011).

1.2.1.2 Semi-Structured Interviews

In addition to observations, key stakeholders in the environment are identified for *interviews* with the researcher. These interviews are *semi-structured*, meaning that some questions are prepared prior to the interview centered around a few themes, but that the researcher might ask questions based on their experiences during fly-on-the-wall observations, or might ask follow-up questions as appropriate (Taylor and Bogdan, 1998). Interviews are recorded with a voice recorder and transcribed after the interview has concluded.

1.2.2 Grounded Theory

Grounded theory is a methodological approach for constructing a theory or theories from the analysis of qualitative data, such as the observational and interview data

from an ethnographic field study (Glaser et al., 1968). For the purposes of this dissertation, grounded theory has three main steps: (1) open coding, (2) axial coding, and (3) selective coding. In *open coding*, data from all sources (e.g., interview data, observational data) is coded for concepts. For example, a line from observational data that reads "Worker ran over to clean up cups robot spilled 10 minutes after the incident" might be coded as "mistake recovery" and "delay in noticing mistake." Open codes are developed organically as additional data is coded, such as a single code being divided into two codes later in the coding process, after discovering that the single code is insufficiently detailed to encompass all instances. *Axial coding* then uses the resulting open codes to make connections amongst them. Examples of connections might be repeated phenomena. Finally, *selective coding* is used to identify relationships among the axial codes, such as multiple axial codes that concern the social perceptions of the robot. In this dissertation, grounded theory is used for analysis for the ethnographic work described in Chapter 3 and for the user study of an authoring environment I created in Chapter 4.

1.2.3 Formative Studies

Since collaborative robots are envisioned to assume roles previously taken on by humans partners, an understanding of human behaviors helps to inform and inspire the design of effective collaborative behaviors for robots for this work. A *formative study* uses data collected from human-human interactions, often in a laboratory setting, to aid in inspiring models of human behavior. Those studies which occur in a laboratory setting enable the researcher to better control some of the variance

that naturally occurs in human-human interactions by giving a predefined task, while also enabling the researcher to capture a rich data set through the use of additional recording equipment. In this dissertation, the work on speech patterns in Chapter 4 and the work on how robot's can give instructions in Chapter 5 both use formative studies as means for better understanding and modeling behavior in human-human collaborations.

1.2.4 Building Computational Models

Insights from human-human collaboration, through a combination of data from formative studies and information from prior literature, can be used to build computational models that allow implementation of collaborative behaviors on robots. In this dissertation, these models take the form of finite state machines or rules for decision-making. This dissertation builds computational models in the work on speech patterns in Chapter 4, the work on robot deictic gestures in Chapter 6, and the work on how robot's can give instructions in Chapter 5 all use the process outlined above for building models of human behavior that can be implemented on a robot.

1.2.5 Experimental Evaluation

After developing the computational models, these models can be evaluated in controlled experimental settings, allowing the researcher to better understand the nuances of each model through objective, subjective, and behavioral measures. Experimental evaluation begins with the researcher implementing the model on the robot for a specific experimental task, such as a robot teaching a human how to connect a series

of pipes. As part of implementation, multiple strategies for using the model might be implemented, to better study the impact of varying approaches. Implementations might result in an autonomous robot system, which can interact with a person in the given context without control or intervention from the researcher, or the implementations might involve predefining the interaction that a person would observe. Participants are then recruited to complete a task with the robot, allowing them to participate in an immersive experience that helps to situate reactions to and perceptions of the robot. After participation, participants complete a questionnaire to capture subjective perceptions of the robot. Recorded video and audio data from the interaction is examined for objective and behavioral measures. In this dissertation, the work on robot deictic gestures in Chapter 6 uses a set of predefined interactions to evaluate the gesture models, while the work on speech patterns in Chapter 4 and the work on how robot's can give instructions in Chapter 5 both use an autonomous experimental setup to evaluate the respective behaviors in each chapter.

1.3 Research Platforms

My work focuses on three different humanlike robotic platforms: the Baxter robot, the NAO robot, and the Wakamaru robot. These three platforms were each chosen for the specific design characteristics and capabilities they provide. A picture of each robot can be seen in Figure 1.1.

The Baxter robot[5] (Figure 1.1, a) is designed as a collaborative robot suitable for manufacturing facilities. Designed and released by Rethink Robotics in October 2012,

[5]www.rethinkrobotics.com/products/baxter/

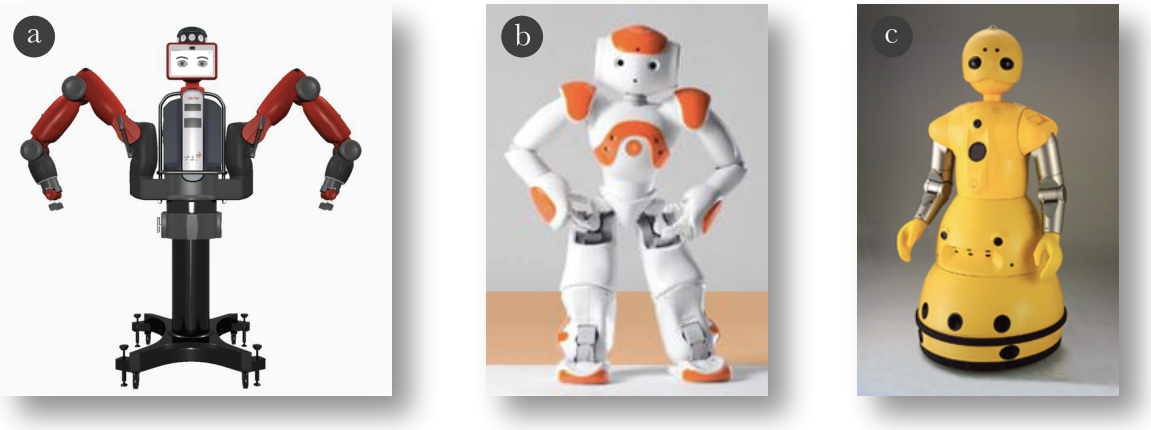

Figure 1.1: The three robotic platforms used in this study: a) Baxter, b) NAO, and c) Wakamaru.

the Baxter robot is one of the first collaborative robot platforms intended for use in day-to-day environments. The robot has become one of the most widely adopted collaborative robot platforms in US manufacturing facilities, making it an ideal focus for my ethnographic field work. Baxter is considered safe to operate around humans, and is trainable for a variety of tasks using a visual programming interface. The robot's design follows a humanoid morphology, including two manipulator arms and a screen used as a "face" through the display of eyes, and is the standing height of a human. The robot itself is 3 ft (0.914 m) tall, but deployed robots are often mounted on a platform that raises its height to be between 5 ft 10 in (1.778 m) and 6 ft 3 in (1.905 m).

Baxter's hardware and software is optimized primarily to perform *pick and place* tasks. In these tasks, Baxter acquires an object or objects using either one or both of its "hands," i.e., grippers attached to the end of its arms, from a bin or a moving conveyor belt. The object(s) are then deposited in another bin or a workbench. The

robot might be integrated with nearby systems with which it can communicate. For example, the robot might pick a finished product from the moving conveyor belt of an assembly line and place it in a shipping container. After the product is appropriately placed and stacked in the container, the robot would communicate to the assembly line that it is ready to pick and place the next product.

The remaining two robotic platforms, the NAO[6] (Figure 1.1, b) and Wakamaru[7] (Figure 1.1, c) robots, were each used in laboratory experiments as part of this dissertation. Both robots feature a humanoid morphology and are designed for personal use in the home. The NAO robot is 1.9 ft (0.58 m) tall and features three degrees of freedom in its head, as well as six degrees of freedom in each arm, including articulated fingers: shoulder pitch, should roll, elbow yaw, elbow roll, wrist yaw, and finger pitch. The NAO is also equipped with speakers, a microphone, and both speech detection and text to speech algorithms. The Wakamaru robot is 3.28 ft (1 m) tall and features three degree of freedom in its head, as well as four degrees of freedom in each arm: shoulder pitch, should roll, elbow yaw, and elbow roll.

1.4 Outline

In the remainder of this dissertation, I will examine relevant prior work, including the ramifications of a robot in everyday human environments, methods used to implement social behavior on robots, and a review of literature concerning specific social cues for both humans and robots. Next, I will present a formative study that highlights

[6]http://www.aldebaran.com/en/humanoid-robot/nao-robot
[7]http://www.mitsubishi.com/mpac/e/monitor/back/0602/story.html

the needs, desires, and expectations users have from collaborative robots. Inspired by the context of the formative study, I will then outline three laboratory studies that focus on understanding how a specific behavioral cue can best be utilized by a robot for effective collaboration. I will then present a general discussion of the dissertation work, including implications, limitations, and future directions. Finally, I will conclude by summarizing the contributions and of this work.

2 RELATED WORK

The two primary goals of this dissertation are to better understand how people perceive and interact with collaborative robots, and designing robot behaviors for more effective human-robot collaboration. While prior work has examined robot behavioral cues (Andrist et al., 2013b; Bickmore and Cassell, 2005; Brooks and Breazeal, 2006; Fong et al., 2003; Liu et al., 2013; Staudte and Crocker, 2009), these studies have traditionally been contextualized in conversational scenarios, rather than collaborative scenarios. My work builds upon our understanding of behavioral cues to extend them for human-robot collaboration.

In this chapter, I begin by reviewing how humans have collaborated with one another as well as with computers and machines, highlighting the social repercussions of this collaboration, particularly in regards to robots. Next, I review prior work in both the human-human and human-robot literature concerning behavioral cues for interactions, focusing on the three particular behavioral cues examined in this dissertation. Finally, I discuss how behavioral cues are implemented on social robots and why new interfaces are needed for this process that cater to a wider population.

2.1 Communication and Collaboration

The process of *collaboration* requires two or more people to work together towards a shared goal. Such an effort can encompass a range of activities, such as discussing the planning of a trip or working together to assemble a piece of furniture. These endeavors require the effective use of *communication* between participants in order

to share thoughts, intentions, and uncertainties in an effort to reach a mutually satisfactory completion of the collaboration (O'Daniel and Rosenstein, 2008; Olson et al., 2001). Communication can take on both verbal and non-verbal forms, ranging from a description of a person's thoughts or intentions to the use of *social cues*. These social cues are the verbal and non-verbal cues humans use to communicate with others and to reveal insights into a partner's thoughts or intentions (Blakemore and Decety, 2001; Brennan et al., 2008; Bryan et al., 2012; Clark, 2005; Doshi and Trivedi, 2009; Meltzoff, 1995). Examples of social cues include gesturing at an item needed or gazing toward where a person will move next.

Humans are adept at processing these social cues, with children in particular often relying more heavily on social cues than speech as a means for communication (Caselli, 1990; Rutter and Durkin, 1987). As a result, social cues serve an important role in communication, replacing or augmenting verbal communication by providing additional information about a person's thoughts or action (Brennan et al., 2008; Goldin-Meadow, 1999; McNeil, 1992). For example, a gesture might more clearly indicate a specific object than speech alone, or a person's tone of voice might betray their seemingly neutral speech on a matter. Prior work has found that the use and observance of these social cues helps facilitate communication, such as clarifying a person's intent (Jokinen, 2009; Scherer et al., 1972).

For the purposes of this work, I will be focusing on the use of collaboration by two people in physical contexts, such as repairing a piece of equipment or cooking together. By communicating or clarifying intentions through social cues, others working with this person can effectively adapt their future actions, taking into account the other's

intent. Consider the following scenario:

> Mike and Marie are building a bookshelf together. During assembly, Mike is considering what action he should take next. He notices Marie beginning to reach for a piece of hardware that needs to be affixed to the bookshelf's case. Recognizing what step she intends to complete, he realizes that they will need a screwdriver to complete the step, and he moves to acquire the screwdriver to attach the hardware. As she holds the hardware in place, he tightens the screws to secure the hardware to the case.

In this example, Marie communicates her intent to complete the next step to Mike by physically moving towards acquiring the next piece needed. To complement her next action, Mike selects the tool needed to complete the step. For the purposes of this dissertation, this process of organizing one person's actions to complement another's actions is called *coordination*.

Below, I outline the theoretical foundations for how social cues facilitate human-human collaboration, highlight how these theories extend to collaboration with machines, and discuss previous work in human-machine collaboration, noting the perceived sociality of these technologies and the particular role of robots in day-to-day environments.

2.1.1 Theoretical Foundations

Communication and collaboration between humans has been widely studied, resulting

in numerous theories and models to better explain these phenomena. Below, I discuss three theoretical frameworks—*mental models*, *theory of mind*, and *embodied cognition*—that motivate and contextualize the ideas presented in this work.

2.1.1.1 Mental Models

The theory of mental models is used to explain people's thought process (Craik, 1967; Johnson-Laird, 1983). Under this theory, people's minds develop representative models of reality, including dialogue they might hear or actions they might observe (Johnson-Laird, 1983). These mental models then serve as references for sense-making, evaluation, and decision-making (Johnson-Laird, 1983). With regards to communication and collaboration between humans, mental models can be used to understand social cues from others in the current context and to aid oneself in deciding which social cues should be used to communicate.

In the context of HCI, mental models are a form of user models, where the models inform a user's perception of how a system works and, in turn, how the user will interact with the system to accomplish their goals (Carroll et al., 1987; Staggers and Norcio, 1993; Wilson and Rutherford, 1989). Using mental models to represent a user's understanding of a system can be beneficial for studying users in a variety of contexts, such as communication using computer-mediated communication (Mantovani, 1996), use of mobile phone menus (Ziefle and Bay, 2004), and human-robot interaction (Steinfeld et al., 2006).

Prior work in the field of HRI has studied how mental models shape interaction with robots, leveraging mental models users might already have to ease encounters between

humans and robots (Hegel et al., 2009; Lee et al., 2005; Mirnig and Tscheligi, 2015). For example, Hegel et al. (2009) found that robots which appeared more humanlike were evaluated as more appropriate for settings that require more communication and collaboration, such as healthcare, personal assistance, and teaching. Thus, the idea of mental models suggests that developing robots with humanlike social behaviors will facilitate more natural and fluid interactions between humans and robots.

2.1.1.2 Theory of Mind

The concept of intentionality is defined as the commitment of a person to executing a particular action (Malle and Knobe, 1997). The formulation of an intent is often driven by the individual's desire to achieve a particular goal (Astington, 1993). This formulation requires a variety of other skills, including forethought and planning, to appropriately fulfill an intention (Bratman, 1987). What differentiates an intent from a desire is this level of planning to turn the intent into an achievable reality (d'Andrade, 1987).

From an early age, children begin to attribute intent to the actions of others. For example, children at 15 months of age are capable of understanding the intentions of others in physical tasks, even when the goal is not achieved (Meltzoff, 1995). Later, children learn how behaviors are driven by intent (Feinfield et al., 1999), contributing to the development of an ethical system where intentionality is used as a factor to establish the culpability of an individual.

Prior work suggests that, after developing a capacity for understanding intent, humans also develop a theory of mind—the ability to attribute mental states to

others (Baron-Cohen, 1997; Leslie, 1987). The development of theory of mind enables people to understand that other humans they interact with may have intents that can differ from their own (Leslie, 1987; Blakemore and Decety, 2001). Theory of mind then shapes the way people interact with one another in a way that is most easily observable in physical tasks, such as moving a table together or navigating through a crowd (Sebanz and Knoblich, 2009). In these scenarios, humans rely on theory of mind abilities to attribute intent to other participants and to adapt their own behaviors to accommodate the intent of others, resulting in seamless interactions.

While the use of mental models discussed above is based on the assumption that a robot employing more humanlike behaviors (e.g., social cues) will evoke responses similar to working with a human partner, theory of mind instead examines how humans will respond to intentional beings. Developing robots that people will attribute a theory of mind to will allow people to collaborate with robots as they would human partners, recognizing their robot partner's intent and efficiently adapting their own future behaviors to accommodate the robot's goal. Implementing a theory of mind on a robot requires the robot to both understand the social cues displayed by human partners as well as express their own social cues that can communicate an intent or desire (Scassellati, 2002). This dissertation is concerned with the second goal of robot expression of social cues in collaborative contexts.

2.1.1.3 Embodied Cognition

Embodied cognition posits that human thought processes are influenced by interaction with the environment via embodiment (Rosch et al., 1992; Wilson, 2002). Scholars

cite the way we use metaphors grounded in embodied language as evidence for embodied cognition (Lakoff and Johnson, 1980). For example, adjectives that express a physically higher position are considered powerful, while adjectives that express a physically lower position are considered submissive, such as saying "I have control *over* it," or "He was feeling *down*." Under embodied cognition, stimuli from the environment and a person's subsequent motions are not simply inputs and outputs, but critical components of one's thought process. As a result, collaboration under embodied cognition is not just an orchestration of two peoples' thought processes, but a set of reactions in response to peoples' embodiments and environments.

When extended toward HRI, embodied cognition suggests that communication by a robot is not simply about the robot producing the correct output, but rather about the robot producing output that considers embodiment, such as the robot's capabilities and morphology, their partner's capabilities, and the surrounding environment. This dissertation addresses how to choose and exhibit social cues given the environment. Chapter 7 also discusses some of the challenges of varying robot morphologies with using particular social cues.

2.1.2 Collaboration with Machines

Computers and machines have traditionally been employed to mediate collaboration, rather than as collaborative partners. For example, computer applications, including email, videoconferencing technology and groupware, have changed the way people relate to others and complete their work (Grudin, 1988, 1994). However, rather than collaborating with the computer, people in these contexts use the computer as a tool

to collaborate with others.

Technological advancements now allow computers to take on some level of autonomy and agentic behavior, enabling them to monitor and respond to human actions. For example, tutoring programs allow the computer to monitor a student's state or understanding of a topic to improve learning outcomes (Aleven and Koedinger, 2002; Szafir and Mutlu, 2013). Virtual agents appearing on a computer screen can be used to provide humanlike interactions in a variety of settings, such as answering questions in an online stores (Aggarwal and Vaidyanathan, 2003; Benbasat and Wang, 2005). Although these technologies allow computers to provide some level of humanlike, collaborative behavior, they still require interacting in the realm of the computer, rather than the physical realm that humans occupy.

Robotic technologies now provide even greater levels of autonomy and agentic behavior than traditional computer interfaces. Rather than requiring human users to interact with the computer in the virtual realm, these robots are capable of interaction in the physical realm. Although this introduction of physical embodiment now allows interactions and elicits perceptions that are more pronounced or effective than virtual agents (Wainer et al., 2007), the challenges of creating a robot that is safe for humans to be near has necessarily limited the applications of robots. As a result, interaction with robots has typically involved more conversational or command-based interactions, such as a robot teaching a class (Kanda et al., 2007) or serving as a waiter in a restaurant (Franklin et al., 1996).

However, recent advances in technology, driven primarily by the manufacturing industry, have enabled a new class of robots, called *collaborative robots*, that are

safe for humans to be near and flexible enough for humans to collaborate with on a variety of applications (Kock et al., 2011; Matthias et al., 2011). Collaborative robots are poised to follow a long trend of technologies—from desktop computers to virtual agents—that are perceived by human users as having social qualities, in addition to introducing this new functionality of being capable to work in the physical realm. These robots also represent a shift from the traditional use of robots in either strictly conversational roles or as autonomous and separate from humans. In the paragraphs below, I briefly review prior literature on the perception of prior technologies as having sociality, and the introduction of robots into day-to-day environments and their perceived sociality.

2.1.2.1 Perceived Sociality of Technologies

The design of computer technologies follow metaphors that shape the way their users perceive and interact with them. Fogg (1998) proposed "tool," "media," and "social actor" as three such metaphors that respectively result in perceptions of computers as providing new abilities, conveying content, and playing social roles. When computers follow the metaphor of a social actor, particularly displaying aspects of human language, offering interactivity, and playing roles that humans play, their users "mindlessly" apply social rules and expectations despite explicitly acknowledging that these machines have no social qualities (Nass and Moon, 2000). Even computers that minimally follow this metaphor elicit attributions of gender, ethnicity, personality, and expertise to them, displays of politeness and reciprocity toward them, and disclosure of information when they divulge information first (Moon, 2000; Nass and Moon,

2000).

When computer technologies are designed to more closely follow the metaphor of a social actor, as speech interfaces, virtual agents, and social robots do, people's interactions with them more closely resemble human-human interactions (Bickmore and Cassell, 2005; Kopp et al., 2005; Nass and Lee, 2000). Research on embodied conversational agents (ECA) has demonstrated that people employ dialogue strategies from human-human conversations, such as greetings, farewells, small talk, and insults, and display elements of a human conversational style, such as disfluencies, in their interactions with ECAs that engage in social dialogue with them (Bickmore and Cassell, 2005; Kopp et al., 2005).

While the design of robotic technologies vary in how closely their designs follow the metaphor of a social actor, we expect much of these attributions and responses to be present in people's interactions with them. For example, users of robots designed with minimal cues for sociality, such as the Mars Rover, perceive the robot as a social actor, identifying with the unique qualities and abilities of the robot as well as a social resource for the human team (Vertesi, 2008).

2.1.2.2 Robots in Day-to-Day Environments

The last decade has witnessed the widespread introduction of robotic technologies into day-to-day environments. Prior work has studied how these products changed workplace and domestic practices and how their users perceived them. Studies of the use of the robotic vacuum cleaner Roomba in domestic environments found that users attributed lifelike characteristics, such as personality, gender, intentions, and feelings,

to their cleaning (Forlizzi, 2007; Forlizzi and DiSalvo, 2006; Sung et al., 2007) and developed a sense of unique ownership and intimacy with their products that led them to promote their robots in their social networks (Sung et al., 2007).

Robotic technologies have also been introduced to organizations, most prominently to perform transportation and delivery tasks at hospitals. Prior work studying the effects of the introduction of these robots on work and social practices found differences in how different groups responded to and worked with the robot (Ljungblad et al., 2012; Mutlu and Forlizzi, 2008; Siino and Hinds, 2005). Hospital workers whose jobs were less demanding benefited from the help that the robot provided and perceived the robot more positively, while others who worked in a more demanding environment found the robot to be a burden and a disruption to their social environment due to the robot's inability to recognize and adapt to those demands (Mutlu and Forlizzi, 2008). Based on their familiarity and experience with the robot, different stakeholders applied different cognitive frames to the robot, including "alien," "machine," "worker," and "colleague" (Ljungblad et al., 2012). Prior work has also examined how people interact with robots integrated into organizations and the broader social processes involved in these interactions. A study of the deployment of a snack-delivery robot in a university building found that users develop social relationships with the robot and that individual interactions with the robot result in a "ripple effect" in the social environment, engaging non-users in these interactions and promoting socializing (Lee et al., 2012).

2.2 Behavioral Cues for Collaboration

Collaboration is the cornerstone that allows humans to work together to achieve some shared goal, whether it is exchanging information, teaching a friend how to complete a task, or helping someone move furniture. Humans use a number of behavioral cues, such as gaze (Griffin, 2001; Meyer et al., 1998; Tomasello, 1995), gestures (Clark, 2005; Kendon, 2004), speech (Clark, 2009; Croft, 2000), body posture (Argyle, 2013; Ekman and Friesen, 1967), or head nods (Helweg-Larsen et al., 2004; Stivers, 2008), to indicate their thoughts or intent regarding the task. These cues help to make the interaction more effective and fluid, minimizing the number of misunderstandings, the total task time, and improving subjective measures, such as rapport and perceived teamwork, between participants (Argyle, 2013; Clark, 2005; Tomasello, 1995).

Developing similar types of communicative behaviors for robots is necessary to achieve fluid human-robot collaboration. In the realm of conversational encounters, prior work has found extensive support for emulating humanlike communicative behaviors on robots to achieve effective interactions. Various properties of humanlike gaze—one of the preeminent social cues used in communication—have been extensively studied and implemented on robots, including where to direct gaze during a conversation (Kobayashi et al., 2010; Mutlu et al., 2006, 2009; Staudte and Crocker, 2009), how to bring attention to a particular object (Imai et al., 2003; Scassellati, 1999; Huang and Mutlu, 2012), and appropriate times to look away from a conversational partner (Andrist et al., 2013a). Other studies have explored the use of other humanlike social behaviors, including gestures (Brooks and Breazeal, 2006; Huang and Mutlu, 2013; Liu et al., 2013), motion profiles (Dragan and Srinivasa,

2014; Zheng et al., 2013), and speech usage (Andrist et al., 2013b; Bickmore and Cassell, 2005; Fong et al., 2003) by robots. In these studies, the humanlike behavior outperformed baseline behaviors on various objective and subjective measures, suggesting that human-robot conversational encounters are more fluid when the robot employs humanlike behaviors. Following these results and the theoretical foundations outlined in Section 2.1.1, it would follow that enabling robots to use humanlike social behaviors and cues in collaborative scenarios will result in more humanlike communication, allowing humans to seamlessly transfer their habitual collaborative behaviors to working with a robot partner.

In this section, I focus on three behavioral cues that are studied in this dissertation: speech patterns, instruction and repair, and deictic gestures. For each behavioral cue, I review prior work on that cue from both the human-human and human-robot interaction literature.

2.2.1 Speech Patterns

Speech is one of the most explicit forms of human communication, often augmented by other, non-verbal social cues (e.g., gaze, gestures) (Clark, 2005, 2009; Croft, 2000). Although human speech can take on a seemingly limitless number of forms, prior work has suggested that these verbal interactions follow an implicit set of patterns (Goffman, 1983). Below, I discuss prior work on speech patterns in human interactions, and the use of patterned behaviors in human-robot interactions.

2.2.1.1 Human-Human Interaction

Human interactions follow an invisible structure, a shared *interaction order* (Goffman, 1983), that signals to its participants how they should act and interact with others. For example, the openings of encounters follow a particular "routine" that involves a greeting or a summons by a participant and an in-kind response either in the form of a greeting or an answer (Schegloff, 1972). Similarly, conversations might involve the pattern "question-answer pairs," where one participant is posing a question followed by an answer from a different participant (Clark, 1992). These routines may be combined to produce a more elaborate interaction, as shown in the example below.

> *Participant 1*: <upon seeing a friend at work> Oh, good morning! Are you attending the meeting this afternoon?
> *Participant 2*: Yes, I was planning on going.

In this brief encounter, Participant 1 opens the interaction with a greeting and subsequently poses a question, which Participant 2 answers. The participants of this encounter employ both the greeting and question-answer pair routines to seek and provide information and coordinate activities. Research on human communication has shown that the use of such routines or "interaction patterns" facilitates effective communication, specifically processes such as fluency (Tannen, 1989) and grounding (Clark, 1996). On the other hand, breakdowns occur in communication when these patterns are violated (Schegloff, 1972).

2.2.1.2 Human-Robot Interaction

Research on human-robot interactions has also considered how robots might display patterned behaviors, such as gaze patterns that facilitate conversational turn-taking and joint attention (Huang and Mutlu, 2012), to enable more effective human-robot interaction. Researchers have also proposed the use of such patterned behaviors as building blocks for constructing human-robot interactions (Kahn et al., 2008; Peltason, 2010), such as an "initial interaction" pattern proposed by Kahn et al. (Kahn et al., 2008)—an interaction pattern analogous to greetings in human interactions. However, this body of work does not yet offer a pattern language that might serve as building blocks for interactions across a wide range of scenarios. Furthermore, this research has not developed tools or environments to support design exploration or prototyping for constructing human-robot interactions.

2.2.2 Instruction and Repair

Effective communication relies on both the ability to use language and other behavioral cues, and the ability to present information, such as what information to disclose at a particular moment, how the information is presented, and how miscommunication is handled. In particular, as robots enter a number of instructional roles, such as training workers and aiding in the home, they will need to be capable of both aspects of effective communication during instruction. While prior work has focused on the ability of robot's to use language and their behavioral cues during tasks (Andrist et al., 2013b; Boucher et al., 2012; Huang and Mutlu, 2012; Staudte and Crocker, 2009), I focus on how robots can present information effectively to their human partners.

Below, I review prior work on how one human would instruct another, and work in human-robot interaction on how a robot might most effectively give instructions.

2.2.2.1 Human-Human Interaction

Effectively communicating a series of instructions is a complex task that has been studied at a number of levels, including how human instructors develop and communicate instructions for their students. Prior work has suggested that instructors follow a discourse planning process based on iterative refinement, where the instructor first picks a subgoal to complete and then further decomposes the subgoal into atomic actions (Blaylock et al., 2003; Grosz and Kraus, 1996). Instructions are then ordered based on logical segmentations of steps to help students contextualize the task (Grosz and Sidner, 1986). These models provide important insights into how instructors break task goals into a set of instructions.

Successfully directing a student in a task also relies on feedback from the student. Despite the best efforts of instructors, there will inevitably be instances of *breakdowns*—misunderstandings or miscommunication concerning task goals—that can either impede progress in the interaction or lead to further breakdowns in the future if they are left unaddressed (Zahn, 1984). To correct breakdowns, humans engage in *repair*, a process that allows participants to correct misunderstandings and helps ensure that all participants have a similar understanding of the relayed information (Hirst et al., 1994; Zahn, 1984). The process of engaging in repair is often context-sensitive (Seedhouse, 1999). For example, when a topic is being discussed in a classroom, the instructor frequently initiates repair to clarify students' statements. However, when

the classroom is engaged in a task, students are more likely to initiate repair with their peers.

2.2.2.2 Human-Robot Interaction

Prior research in robotics has explored how robots might function in instructional settings, such as daycare facilities and classrooms (Kanda et al., 2007; Tanaka and Kimura, 2009; Tanaka and Movellan, 2006), and aid in task instruction, such as offering assistance in a hand washing task (Hoey et al., 2005) and giving directions in a cooking task (Torrey et al., 2007).Among these studies, work on task instruction has focused on how robots might adapt task instructions to user needs and instructional goals. For instance, Torrey et al. (2006) explored how adapting the comprehensiveness of the robot's instructions to its user's expertise might affect task outcomes and user experience. They found that more comprehensive instructions resulted in fewer mistakes among novices, while experts rated the robot as more effective, more authoritative, and less patronizing when it provided brief descriptions. Foster et al. (2009) studied the effects of the order in which the robot provided task goals along with instructions on student recall of task steps, showing that providing task goals prior to issuing task steps resulted in fewer requests for repetition by the student later in the task.

Ideally, an instructional task would proceed without the need for clarification or intervention. However, just as repair is necessary in human instruction, robots must also be capable of identifying breakdowns and offering repair to improve the effectiveness of human-robot instruction. Prior work in enabling robots to provide

repair has identified a variety of techniques to lower the need for repair, such as taking into account the speaker's perspective (Trafton et al., 2005) or alleviate the negative impact of breakdowns through framing (Lee et al., 2010).

While work in human-human teaching has developed many theories concerning effective instruction, models of how robots should provide task instruction and repair to maximize task outcomes and student experience are needed. In the following section, we detail our work on developing such a model.

2.2.3 Deictic Gestures

Deictic gestures serve a crucial role in communication, augmenting or even replacing complex verbal descriptions of objects or spaces (Goldin-Meadow, 1999; McNeil, 1992). Below, I present prior work on how humans use deictic gestures in interactions as well as how deictic gestures have been used in human-robot interactions.

2.2.3.1 Human-Human Interaction

Deictic gestures are often used to augment or replace verbal descriptions of the object being referred to, also called the *referent* (Kobsa et al., 1986). The importance of deictic gestures in communication is shown in pre-verbal children, who will use deictic gestures as a way of communicating with their caretakers prior to their ability to form utterances to describe their wants and needs (Caselli, 1990). Once humans have the ability to verbally communicate, the use of deictic gestures increases and becomes more nuanced, serving to support more complex utterances (Jancovic et al., 1975). At this point, deictic gestures are used to decrease cognitive burden, allowing for

complex verbal descriptors to be eliminated in favor of a deictic gesture toward the referent (Goldin-Meadow, 1999). The replacement of fully articulated speech with a combination of partially articulated speech and deictic gestures reduces cognitive load for the speaker, by requiring less processing to form an utterance, and the listener, by requiring less processing to interpret the utterance. For example, a speaker might replace the description of an object with "this" and a deictic gesture that indicates the referent. Gestures might even fully replace utterances in settings such as a noisy factory environment (Harrison, 2011).

Traditionally thought of as "pointing gestures", deictic gestures are comprised of a more diverse set of gestures that are used to draw attention to an object. Caretakers often use touch to more concretely focus the attention of infants on an object (Lempers, 1979), while students or instructors in an instructional block building task might hold up a piece to implicitly confirm whether it's correct (Clark, 2005). Prior research has demonstrated that the speaker's use of gestures affects information recall and rapport in listeners (Breazeal et al., 2005; Huang and Mutlu, 2013).

Work by Clark demonstrated that deictic gestures are much broader than pointing (Clark, 2005). However, while the use of these additional deictic gestures has been mentioned in relation to other research (Alibali et al., 1997; Bolinger, 1983; Clark, 2005; Lempers, 1979), a more thorough understanding of the breadth of deictic gestures and why they are chosen for particular settings is needed.

2.2.3.2 Human-Robot Interaction

Prior work in human-robot interaction recognizes the need for robots to gesture naturally in order to communicate in a more humanlike fashion. Much of this work has focused on enabling robots to use deictic gestures to enhance task outcomes, such as the robot's use of gestures improving user performance in manipulation tasks (Breazeal et al., 2005; Okuno et al., 2009; Salem et al., 2012). In general, robots use deictic gestures similarly to humans to help bring attention to objects of joint interest and achieve common spatial ground (Brooks and Breazeal, 2006). When the environment may make using deictic gestures difficult or impossible, robots are also able to use perspective-taking to ensure that their deictic gestures are used in ways that are interpretable by the listener (Trafton et al., 2005). To ensure gestures are used appropriately, research has focused on enabling robots to use pointing gestures in socially appropriate ways (Liu et al., 2013).

Deictics are often thought of as referring to an object, but can also be used to refer to a region of space, such as the opposite end of a room. Prior work in robot deictics has shown that referring to a region of space—which is often more difficult to verbalize than an object—results in only a marginally worse accuracy rate than referring to an object (St Clair et al., 2011). Robots are able to use visual differences in spaces in combination with deictics to help listeners identify the correct region in the space (Hato et al., 2010). St Clair et al. (2011) demonstrated that a robot using a combination of deictic gestures and gaze to refer to a space resulted in higher accuracy than using just one modality. Additionally, how the gesture was implemented and executed was significant, with human pointing using a bent arm

producing significantly worse accuracy than pointing with a straight arm.

While it is desirable for robots to use deictic gestures effectively, prior work has also focused on allowing robots to recognize and understand these same gestures from the humans they interact with (Schauerte and Fink, 2010; Sugiyama et al., 2007). This includes understanding when a deictic gesture has been made and correctly identifying the referent (Brooks and Breazeal, 2006).

Although HRI research has successfully implemented human pointing behaviors in numerous applications, there is still much to understand about what other deictic gestures robot behavior designers should consider using, what properties are most effective, and how the particular setting should shape gesture choice.

2.3 Programming Social Behavior on Robots

Traditionally, implementing social behavior on robots has been tasked to programmers to complete via writing source code, making it difficult for end-users and owners to modify the robot's behavior for their own need. The need for programmers to implement social behaviors on robots has been due to two main issues: (1) the complexity of programming robots, including processing sensor input and working with expansive application programming interfaces (APIs), and (2) anticipating and accounting for unexpected input, such as unrecognized speech or an unfamiliar environment. However, as robots become more commonplace products in a variety of everyday settings, increasingly novice users might want to modify or add to the robot's social behavior. Traditional programming by writing source code is a highly

specialized skill that most users do not posses, requiring a new approach to enabling novice users to adapt a robot's behaviors to their own preferences.

A body of work has developed several authoring tools and environments to help both expert developers and novice users better control and interact with robots (Humphrey et al., 2007) and to evaluate human-robot interactions (Kooijmans et al., 2006). Environments for developers include the Robot Operating System (ROS), which offers an architecture for abstracting and reusing specific functionalities across different robot platforms (Quigley et al., 2009). A particular ROS module, the Robot Behavior Toolkit, offers the ability to specify robot social behaviors based on a repository of "rules" or patterns (Huang and Mutlu, 2012). However, the use of these environments requires a substantial amount of development expertise as well as effort to build suitable pattern repositories.

Authoring environments for novice users include a number of commercial and research tools for programming robot behaviors (Glas et al., 2012). For example, Interaction Composer (IC) involves a graphical interface that enables users to coordinate multiple facets of a robot's behavior, such as dialogue and gestures, by choosing from a set of "behavior blocks" to compose the interaction (Glas et al., 2012). RoboStudio offers an expert authoring environment to build graphical interfaces that enable novice users to customize the behaviors of the robot to their needs or preferences (Datta et al., 2012). Finally, Lohse et al. (2014) has developed a framework for iterative design and evaluation of robot behaviors. While these tools and environments involve easy-to-use interfaces that are accessible to both expert and novice users, they require the designer to have knowledge of a pattern language and to map this knowledge to

specific robot behaviors that can be authored using the tools.

2.4 Chapter Summary

This chapter began by providing an overview of how humans have traditionally worked with and perceived technology, particularly robotics technology, in the past. Next, I reviewed prior work from both human-human and human-robot interaction literature on three specific behavioral cues that are covered in this dissertation: speech patterns, instruction and repair, and deictic gestures. Finally, I discussed the programming of social behavior on robots, focusing on authoring environments that aid novice users.

3 FORMATIVE STUDY: UNDERSTANDING THE REALITY OF COLLABORATIVE ROBOTS

The introduction of collaborative robots as task partners that work with humans represents a drastic shift in how people perceive and engage with robotic technology. However, this adoption poses many new questions, as demonstrated in the scenario in Section 1.1.1. For example, how are these robots integrated into human spaces, such as the home and at work? What challenges exist in working with this technology? How are these robots perceived by the humans they work with? The study presented in this chapter aims to answer these questions.

In this work, I present an ethnographic field study conducted at three different manufacturing plants throughout the US that had adopted and integrated Baxter, a collaborative robot platform previously described in Section 1.3. Using a grounded theory analysis of observations and interviews, I discovered that results fell into two main categories: how the integration of robot co-workers shapes the physical layout and workflow of a workspace, and the social impact a robot co-worker has in industrial settings. Results from this work suggest several areas where robot co-workers could be improved for more efficient collaboration, such as improved sensing capabilities to allow the robot to more effectively complete work and enabling the robot to handle social relationships so as to build rapport with human co-workers.

3.1 Robots in Manufacturing

Traditional industrial robots are characterized as being inflexible and dangerous, making it difficult for humans to work nearby safely. These robots often operate with a high degree of control, and multiple robots are orchestrated to work together without the interference of humans. Advances in robot kinematics have lead to new collision reaction control strategies and to the development of safer robots. For example, relying on torque control and gravity compensation in the event of a collision can lessen the amount of force on co-located human workers (Haddadin et al., 2008), while actuators with variable impedance may lead to safer robots by reducing joint stiffness while the robot is moving at a higher speed (Tonietti et al., 2005).

As a result of improved safety mechanisms, new robotic products are being released that are safe to operate around humans. Examples of these robots are Rethink Robotics' Baxter, which uses Series Elastic Actuators to minimize the force of impact (Rethink Robotics, 2012), and ABB's Dual-Arm Concept Robot, which contains software-based collision detection (ABB Group, 2013). Additionally, new technologies, such as ABB's SafeMove, are enabling external sensing capabilities on robots to help monitor the robot's speed and location (Kock et al., 2013).

The introduction of collaborative robots into industrial settings requires reshaping approaches to robot control. Work at ABB has examined how these robots might be integrated into the environment as flexible tools (Kock et al., 2011) as well as how such robots might be certified to safely operate around people (Matthias et al., 2011). Prior research also highlighted the differences in needs between large industrial enterprises and small and medium enterprises (SME) (Brogårdh, 2007). This work

suggested that SMEs would benefit from modular robots that can be configured for a variety of pre-existing work cells, can quickly be set up and trained, and require minimal expertise to program and troubleshoot. These features promise the ability to accommodate the current staff at SMEs, allowing them to use collaborative robots with minimal training and negating the need for outside support or contractors.

3.2 Study Outline

To better understand how collaborative robots are affecting the work practices and perceptions of manufacturing employees, I conduced an ethnographic study at three manufacturing plants that had acquired a particular robot platform. Below, I describe the study sites I visited, my data collection methods, and the analysis of the data.

3.2.1 Study Sites

We recruited three manufacturing companies, each of which owned at least one Baxter robot, to participate in my study. Below are brief descriptions of the companies.

Company 1 is a small family-owned business of about 40 employees. It specializes in plastic injection molding and produces plastic parts for different clients. These parts are often components of products that the company's clients manufacture or assemble.

Company 2 is a small business of about 50 employees. It produces and sells components used for securing electrical connections and is known for its outdoor waterproof electrical components.

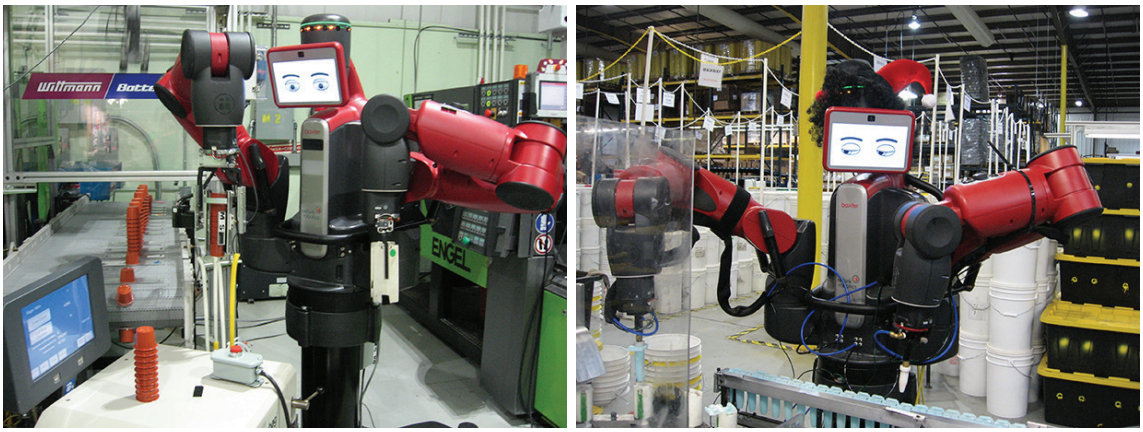

Figure 3.1: Examples of a collaborative robot at work packaging medicine cups at Company 1 (left) and moving plastic tubes at Company 2 (right).

Company 3 is a large international company of several thousand employees spread out across multiple facilities that produce office furniture. Four of these facilities have purchased a single Baxter. I included one of these facilities with approximately 150 employees in my sites.

Each company had owned their robot for four to eight months at the time of my visit. In addition to phone and e-mail correspondence over several months, I spent four days at Company 1, two days at Company 2, and one day at Company 3 conducting fieldwork. All sites were located within the continental United States.

3.2.2 Workspace Setup

The robot was integrated into its own *work cell*—an area defined by its resources and the equipment dedicated to a single task—at each site visited (see Figure 3.1). Below are brief descriptions of Baxter's work cell at each site.

At *Company 1*, Baxter was assigned to move and automatically bag stacks of medical cups. In this setup, a traditional industrial robot was responsible for moving finished medical cups from the plastic injection molding machine to a conveyor belt, which moved stacks of finished medical cups toward Baxter, which then picked up a stack and placed it in the automatic bagging machine. After placing two stacks in the machine, Baxter sent a signal to the bagging machine to place the two stacks in a bag and deploy a new bag. Periodically, an operator gathered completed bags and packaged them in a larger shipping box. Operators and maintenance staff also attended to the robot when problems arose.

At *Company 2*, Baxter set up plastic tubes to be filled with a silicone material for waterproofing wires. A hopper of tubes gradually dispensed tubes into a second container and eventually onto a line that moved the tubes towards Baxter. Baxter then picked up a tube in each hand and placed them in a second machine to be processed. After processing, the second machine deposited the completed tube into a bucket. Operators were responsible for periodically filling the hopper with more tubes and replacing the filled buckets. Operators and maintenance staff attended to the robot when problems occurred.

At *Company 3*, Baxter was set up in a separate area for maintenance workers to retrain it for a new task. During my visit, the robot was being trained to package hardware for furniture, such as brackets, into a box. The maintenance worker arranged the hardware to be boxed to the right of Baxter in a particular configuration, flat pieces of cardboard to stack between hardware to the left of Baxter, and the box to package the hardware in front of Baxter. When activated, Baxter picked up

each piece of hardware individually and placed it in the box. At predefined points, Baxter used its left hand to acquire a piece of cardboard and place it on top of the hardware currently in the box. Then, additional hardware was added. In the targeted workflow, the box and hardware would be delivered to Baxter via a conveyor belt, and the completed box would continue down the conveyor belt. An operator would be responsible for periodically giving Baxter additional flat pieces of cardboard.

3.3 Method

During the course of the study, I collected data on a number of different facets of the integration of the robot into the manufacturing environment, including motivations for purchasing a collaborative robot, the process of integrating it to the existing manufacturing workflow, organizational changes to accommodate the use of the robot, and worker perceptions of and interactions with the robot. In this work, I focus on the design elements and factors that shaped worker perceptions of and interactions with the robot, including the robot's appearance, its social behavior, and its introduction into the work environment.

3.3.1 Fly-on-the-Wall Observations

We conducted fly-on-the-wall observations of activities around and involving the robot at both Company 1 and Company 2. While Company 3 had a collaborative robot that had previously been deployed at an assembly line, they had decided to re-train the robot in a separate area for a new task. While I did observe that the

robot was located in a separate area for retraining at Company 3, I did not have the opportunity to see the robot being trained. Experiences about the retraining process were instead gathered from interviews. We also observed nearby human-operated work cells to better understand what made the robot's work cell unique. At Company 3, data collection also included observations of the setup of the human-operated work cells and interviews about how the robot was or would be integrated into these tasks.

At each site, observations focused on the robot and its surrounding environment, including how the robot was completing its task, the robot's interactions with nearby equipment, how the robot reacted to unexpected situations, and how workers interacted with the robot. Observations also included understanding the more general environment of the company, including how workers interacted with one another and the organizational structure of the company.

3.3.2 Semi-Structured Interviews

In addition to the fly-on-the-wall observations, I conducted semi-structured interviews with key stakeholders at each site, including *management*, *maintenance*, and *operators*. While there were differences in the task Baxter was assigned at each site, the roles and experiences of the stakeholders was uniform, allowing us to consider the same organizational roles across all three sites. Below, I describe the organizational roles of these stakeholders and their involvement with the robot.

Management staff included employees who were responsible for decisions regarding obtaining the robot, for setting and overseeing company goals, or for high-level human-resource issues. These employees had varying degrees of interaction with

the robot, including helping with troubleshooting work cell problems or contacting technical support, depending on the company's size. However, management relied on maintenance staff for the integration of and troubleshooting with the robot. Management staff were asked questions concerning the size, organization, and mission of the company; the demographics of their workforce; the decision-making process behind purchasing the robot; and the effect of the robot on various metrics, such as productivity and profit.

Maintenance staff included employees who handled the upkeep and troubleshooting of machines in the manufacturing environment. Additionally, these employees were responsible for the integration of the robot, programming the robot, training other employees on using the robot, and troubleshooting the robot as necessary. These employees were often the first to handle day-to-day issues with the robot and its work cell. Maintenance staff were asked questions concerning their roles and responsibilities at the site; how they prioritized their work; their involvement with the integration of the robot; what skills they acquired during integration; their interaction with the robot; and troubleshooting the robot.

Operators included employees who worked at one or more workstations at the manufacturing facility. Although different operators might be assigned to a particular workstation at different times, each operator was solely responsible for meeting the quota at their workstation, and often developed a unique workflow for that particular workstation. Operators typically resolved minor troubleshooting tasks, but would contact maintenance staff when additional help or expertise was necessary. One operator was always assigned to work alongside and monitor the robot. These employees

were not trained on how to program the robot, but they knew how to handle common mistakes in the work cell and how to reset the robot if necessary. Some operators were assigned to work alongside the robot every day, while others only worked with the robot every other day. Operators were asked questions concerning their previous manufacturing experience; whether they had prior experience completing the robot's task manually; their perceptions and interactions with the robot; and their process for troubleshooting problems that the robot encountered.

Across the three sites, I interviewed a total of 17 informants, including six managers, eight maintenance employees, and three operators. The interviewees were identified from among employees suggested by the authors' contacts at each site, workers observed during fly-on-the-wall observations, and employees mentioned during previous interviews. The interviews started with the researcher seeking and obtaining informed consent and proceeded with a semi-structured interview involving an initial set of questions at Company 1 and growing sets of questions at Company 2 and Company 3 that built upon knowledge from previous site visits. The interviews were captured as written field notes and audio recordings. Each interview was approximately 30 min in length (M = 27 min, 22 sec; SD = 5 min, 33 sec), and employees received a $5 USD gift-card to a local coffee shop as compensation.

3.3.3 Analysis

A Grounded Theory approach (Glaser et al., 1968) was used to analyze textual data obtained from field notes and interview transcriptions. I first conducted an open coding process in which codes were assigned to significant events or references. Open

coding was completed for all field notes and transcriptions. To establish inter-rater reliability, a second researcher then used provided codes to code 10% of the data. The inter-rater reliability showed substantial agreement between the primary and secondary coders (82% agreement, Cohen's $\kappa = .79$). Next, axial coding was used to identify phenomena, such as repeated events or interactions, among the codes. In total, 11 axial codes were developed that relate to worker perceptions of and interactions with the robot. Finally, I used a selective-coding process to understand the relationships among axial codes.

3.4 Results

Results from this work showed two main themes: implications for the design of the work cell and workers' perceptions of it, and the social implications of a robot co-worker. Below, I explore results for each theme.

3.4.1 Work Cell Implications of an Industrial Robot Co-Worker

The first theme was regarding the work cell implications of a robot co-worker. In particular, this theme involved the challenges of *integrating* a robot co-worker, the changes in *workflow* required to enable a human and collaborative robot to work together, and how additional iterations of collaborative robots might changes *future configurations* of work cells.

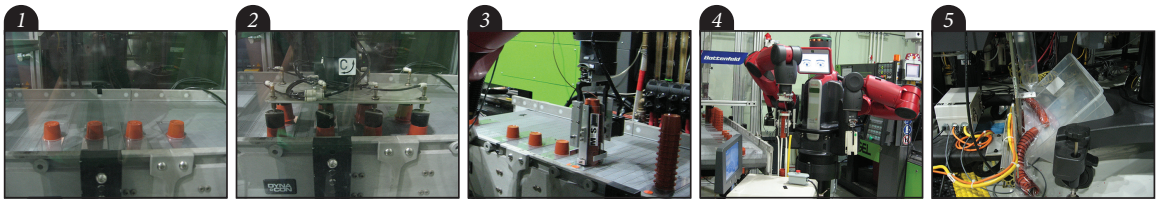

Figure 3.2: An example of the workflow from a work cell for packaging medicine cups adapted for the robot. (1) The conveyor belt has predefined placement for cups in the form of an array of cups drilled into the surface of the conveyor belt. (2) The industrial robot removes cups from the injection-molding machine and places them on the conveyor belt until enough cups are added. (3) The conveyor belt moves forward, and the robot picks up a stack of cups. (4) The robot places the stack in a set of guides for the automatic bagging machine. (5) When enough cups have been added to the guides, the robot sends a signal to the automatic bagging machine to package the cups and deploy a new bag.

3.4.1.1 Integration

Integrating the robot was an extensive process that involved selecting a task for the robot, iteratively modifying and refining the work cell to accommodate the robot, and training personnel to program the robot for its new environment. Figure 3.2 shows the workflow at a modified work cell.

Adapting a Work Cell for Robots: A prominent finding from my observations and interviews with all stakeholders is that the robot was integrated into work cells that had initially been designed for human operators. Maintenance staff were often responsible for reconfiguring the workspace to accommodate the robot. Adaptations included moving nearby equipment to accommodate the robot's larger footprint, designing mechanisms to deliver materials or product to the robot rather than the robot retrieving them from a container, and developing a PLC to allow the robot to communicate with pertinent equipment, such as signaling a conveyor belt to stop

product delivery if the robot experiences a problem.

In addition to the minimum changes necessary to integrate the robot, maintenance and management staff worked to remove potential sources of error in the workflow. For example, early integration at Company 1 involved the robot manually triggering an automatic bagging machine by pulling a lever with its hand. Eventually, the company chose to have the robot send a signal to the bagger instead of using its hand. Staff believed that a signal would be more reliable and would require less troubleshooting once the robot was active. Additionally, removing potential sources of error eliminated additional programming of less common or unforeseen situations, resulting in less design and testing work during integration. **Personnel**: Maintenance staff primarily performed the programming and integration of the robot. Depending on company size, management also assisted when needed. Smaller sites, such as Company 1 and Company 2, hired outside contractors for PLC work. Management citepd the lack of time and knowledge among maintenance staff to perform the complex PLC work required for the new work cells.

Although operators were familiar with day-to-day interactions with the robot, they were not trained on how to program the robot for new tasks, primarily due to management not envisioning moving the robot after perfecting its operation on its task. In interviews, management staff citepd several additional reasons why operators would not be trained in the immediate future on how to program the robot. For example, operators often see high turnover, both across the company as well as in what work cells operators work on, making the cost of training employees too high. Additionally, since operators have the potential to be assigned to any work cell,

management staff try to keep tasks as simple as possible to cater to the lowest common denominator. Many older operators do not have the technical knowledge or desire to work with robots. Additionally, hiring requirements for jobs in manufacturing plants are often minimal, and some operators may not have the computational thinking skills required to train the robot. Finally, with the additional environment engineering and PLC programming required, management believed that asking operators to retrain the robot for a new task would be too complex.

3.4.1.2 Workflow

The integration of the robot brought about two main changes in workflow, which I discuss below.

Operator Multitasking: my observations at both Company 1 and Company 2 indicated that operators managed one or two additional work cells in addition to the robot. While it was common for some operators to be assigned to multiple work cells, interviews with operators indicated that working on multiple automated work cells is more predictable than working with the robot. Operators often develop a personal workflow for handling all their assigned work cells, such as knowing how often to check the production of a particular work cell. In contrast, when working with the robot, there can be instances where the robot or work cell shuts down unpredictably. This disruption might be due to an inability to access or find a given object or the incorrect placement of an object. Operators report their workflow being disrupted by the robot with varying frequency on a given day. These disruptions created stressful situations for operators if they felt they were not achieving sufficient production levels

for the day. Additionally, if the operator also worked on an automated work cell that continued production despite the operator's absence, operators noted that the disruptions from the robot could cause significant backups on the automated work cell, posing problems in handling of the product.

Need for Maintenance to Troubleshoot: Although operators became aware over time of how to troubleshoot common problems, I observed that maintenance staff were frequently called upon by operators to help troubleshoot issues. Maintenance staff were often needed if the issue was technical, or if a problem occurred with the equipment with which the robot interacted. I also observed operators discussing a problem with maintenance staff if operators believed they had a new insight about the cause of recurring problems. Maintenance staff and operators were sometimes observed working together to identify the source of issues.

3.4.1.3 Future Configurations

While my data collection primarily focused on understanding current integration and workflow practices, several interviewees speculated as to how the future of manufacturing might change with the large scale introduction of collaborative manufacturing robots, which I discuss below.

Robots as Temporary Skilled Employees: Company 3 noted the need for temporary skilled employees. At their plant, they often have spikes in production that require hiring temporary labor. The work at this plant requires significant training before the employees can be productive, making managers reluctant to release the temporary labor at the end of the production spike if they are aware of another

production spike in the near future. However, keeping these employees on during slower production periods is costly. To solve this problem, management at Company 3 envisioned the use of a fleet of robots to serve as temporary skilled employees at their nearby plants. Each plant could schedule the use of the robots on an as-needed basis. Robots could be quickly trained, or even trained beforehand, for the necessary tasks, reducing the need for temporary human labor.

Operators as Robot Managers: Employees at Company 3 described a future where operators will manage robot co-workers, rather than multitasking between their own work cell and the robot's work cell. Management staff noted that the robot's strengths included precision, repeatability, and constantly being online. However, the robot is currently not mobile and cannot easily troubleshoot its own problems. Since the operator's skill set complements that of the robot, the operator could be responsible for clusters of robot workers. The operator could train robots for new tasks, provide and move materials as necessary for the robot, and troubleshoot problems as they occurred. Management commented that these new roles would be more challenging for operators and require a level of technical competency not currently expected. The increased demands and responsibilities were seen as a positive development in manufacturing workflow, as operators would likely feel more engaged in their jobs.

3.4.2 Social Implications of an Industrial Robot Co-Worker

A second theme of the results involved the social implications of a robot co-worker. In particular, this theme involved the unique *operator-robot relationship*, the *attribution*

of humanlike characteristics to the robot, the desire for *social interactions with the robot*, and *responses to the the robot's design* from co-workers. Each theme is supported with observations or quotes from interviews where applicable. Stakeholder perspectives are labeled with either "MG" for "Management," "MT" for "Maintenance," or "OP" for "Operator," followed by a "C" and the company number corresponding to the stakeholder's affiliation.

3.4.2.1 Operator-Robot Relationship

A prominent theme that emerged from my analysis was the differential perceptions of the robot by operators and by maintenance and managerial staff. Operators who worked directly with the robot regularly characterized their relationship with the robot in collegial or personal terms, referring to the robot as their "work partner" (OP2C2) or "friend" (OP1C2). Even other operators who worked at nearby work cells perceived these relationships as unique, one operator noting that her co-worker at a nearby station often referred to the robot as her "son" due to their ability to communicate and work well with each other.

> **OP1C2:** People call him my son. They don't like [the name] "Baxter" and think it's funny how much I like working with him, that I understand him.

Although operator-robot relationships were usually cordial, operators also characterized their relationships with the robot in negative, yet familial or relational, terms.

> **OP3C3:** He [the robot] just has a hard time doing work a lot. Especially when he goes down, I'm like "What's up?" ... Feels like babysitting my grandkids.

> **OP1C2:** Sometimes I write down on my [time] sheet "Baxter was not a team player today."

Operators also noted that they talked about the robot as a "friend" outside of work with their acquaintances. Some operators reported that their friends sometimes asked "how the robot was doing" (OP2C2).

While operators characterized their relationship with the robot in collegial and personal terms, maintenance and management staff viewed working with the robot to be similar to working with other industrial equipment. These employees often referred to the robot as "monotonous" and "error prone," describing their interactions with the robot as involving "fixing it" when problems arose.

We believe that these differences result from different formative experiences with the robot. Maintenance and management staff indicated that the initial demonstrations of the robot during its acquisition had set high expectations that were challenged during the integration process due to difficulties with enabling the robot to quickly and reliably sense its environment. Addressing these difficulties required these employees to iteratively create a static and predictable work environment for the robot and the intelligent sensing features of the robot to be underutilized. This gradual shift away from the initially-envisioned use-cases may have disillusioned these employees and resulted in perceptions that were similar to those of other equipment. While some operators had worked with the robot during this transition period, they

had little knowledge of why the robot was transitioned into a more static environment, potentially maintaining their initial frames of the robot.

> **MT1C1:** It [Baxter] is easy to program, it's the precision of everything else [around Baxter] that's difficult.

> **MT2C1:** my biggest thing is to tie Baxter into the bagger, tie Baxter into the conveyer, and tie Baxter there; and again it comes down to inputs, outputs, and there's not a lot of versatility. ... Right now, I have to sit here meticulously and program every little spec of dust where vision would be boom, boom, boom.

Even operators showed awareness of the differences in how they perceived their relationship to the robot compared to how maintenance and management staff did, as expressed in the following excerpt:

> **OP1C2:** He [MT3C2] likes to come tinker around. It's like his little toy. I'm like "Don't touch anything! You'll screw him up."

3.4.2.2 Attribution of Humanlike Characteristics

My analysis revealed a second theme that centered around the operators attributing humanlike characteristics, such as personality and intent, to the robot. Operators frequently described Baxter as having personality traits, such as "cheerful" (OP1C2), "happy" (OP1C2, OP2C2), "fun" (OP2C2, OP3C3), and "perky" (OP1C2), as illustrated in the following excerpt:

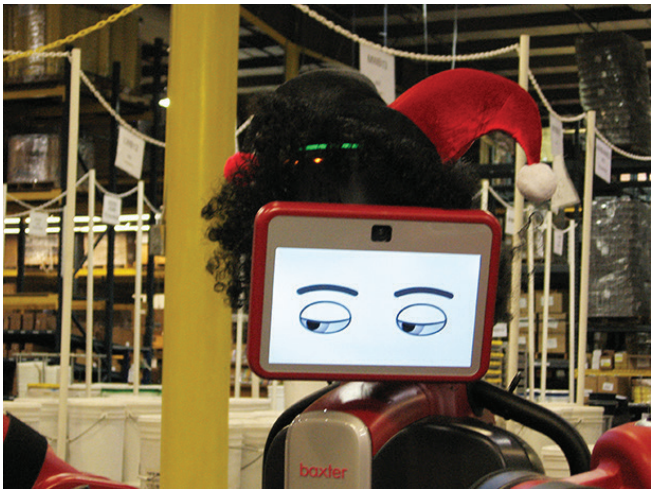

Figure 3.3: The robot from Company 2, dressed up in a wig and jester hat. Previously, the robot was adorned with a rainbow clown wig.

> **OP2C2:** Yeah, he's a lot of fun to be around. He can improve my day.

At Company 2, the robot's physical appearance had been altered to include a wig and jester hat (see Figure 3.3). Operators at this site felt that the robot's new appearance fit well with its perceived personality, as described below:

> **OP1C2:** To me, it totally fits him. ... So if he's in a good mood, it fits him. Or sometimes he has an attitude, "Whatever," you know? That look is just him.

Operators also described the robot and its actions as triggering a range of feelings, reporting feeling happy or pleased while working with the robot—sometimes more so than with a human co-worker—as illustrated in the excerpt below.

> **OP3C3:** Yeah, it can be nice to work by him when I just want a quiet day and he's working well. Lot less hassle then trying to tell someone you don't want to talk.

Other times, operators reported feeling upset or angry with the robot for its actions and expressing resent towards the robot for its mistakes. An operator expresses stress and frustration in the following excerpt:

> **OP1C2:** If he's [the robot is] having a bad day, that...is...very frustrating. Cuz there's no numbers getting out on my job or his job, he's just a mess. ... It's a little stressful.

Some operators believed that the robot expressed intent in its actions, most commonly when the work cell or the robot malfunctioned, particularly when the operator had just left the work cell moments before. For example:

> **OP1C2:** I know that [the robot makes mistakes] and I understand that, so certain things don't bother me. Now sometimes, if I have 8 hours of that, and I'm like, that's when I think he knows what he's doing on purpose, or he's beeping, and as soon as I turn my back to look at him, he stops, I turn around again he beeps, and I'm like, "Really Baxter? Are you doing that on purpose? Cuz you're driving me nuts!"

As a result, these operators expressed the need for someone to watch or supervise the robot in their absence. They felt that supervision might help prevent the robot

from making a mistake as well as allow mistakes to be corrected promptly. This feeling is illustrated in the excerpt below:

> **OP1C2:** I'm going to use the bathroom. If something weird happens, I'm only going to be gone 30 seconds. Sometimes I want to get someone to watch him, like he's a kid.

Finally, regardless of the type of emotions the robot regularly elicitepd, all operators reported that the robot inspired interest. Even if some operators were initially reluctant to work with the robot, they eventually became engaged with the robot. Operators reported asking maintenance staff questions and suggesting improvements in Baxter's programming, how the work cell should be organized, and how to optimize the way operators interacted with the robot. For example:

> **OP3C3:** I noticed it didn't search so good sometimes, ... so I told them [maintenance] to see if they could make it better.

3.4.2.3 Social Interactions with the Robot

Another prominent finding from my analysis was that operators reported having a wide range of social interactions with the robot. The most pertinent day-to-day interaction for operators was listening to the sound of the robot's work to monitor the robot's activities. Many work cells are not designed in such a way that nearby operators can visually monitor the robot, as illustrated in Figure 3.4. Even when the the design of the work cell allowed visual monitoring, directing attention toward the robot would mean taking attention away from their own work. Thus, operators

learned to interpret the sound and rhythm of the robot's work in order to identify patterns of mistakes that demanded their attention. For example, no sound may indicate that the work cell had shut down, the sound of objects being incorrectly placed may suggest that a part of the cell had shut down and the robot continued to operate, and the sound of the robot acquiring new missing objects may indicate that the objects that the robot would later acquire were not being reliably moved toward the robot. The excerpts below provide examples of the practices that operators developed to interpret the sound of the robot's work.

> **OP1C2:** Now I don't look, I mostly listen. He's like a child, if he's been quiet for too long, I know something's wrong. ... We've developed an understanding.

> **OP2C2:** I mean maybe just because I'm accustomed to working with him everyday, maybe now I can anticipate the problems a little bit more. I know what's going on.

Operators also reported sometimes finding themselves talking to the robot. These expressions were often musings while trying to understand why the robot had stopped working. At other times, operators were upset with the robot and were admonishing it or yelling at it out of anger or frustration, as illustrated in the excerpts below.

> **OP1C2:** I find myself just wondering aloud sometimes when something is wrong, hoping he'll give me an answer.

> **OP3C3:** Sometimes he really tees me off, and I let him know it.

Figure 3.4: Two operators working near the robot. The vantage point of the operators makes it difficult to monitor the robot's status. Additionally, the tasks the operators are completing requires their visual attention.

While operators reported already engaging in social interactions with the robot, they and some maintenance staff expressed a desire for the robot to display more social behavior. These employees appreciated the robot's use of its eyes to convey sociality (discussed in the next theme), but felt that the robot could be more socially interactive, for example, by making small talk. Operators explained that although their work required them to focus on their workspace, they frequently engaged in small talk with one another during shifts to help pass the time and establish and maintain relationships. They wished that the robot could similarly engage in basic small talk, mimicking the sociality of working with other nearby operators during shifts where operators are assigned work with the robot and no other operators are nearby, as shown in the excerpts below.

> **OP3C3:** They're [humans are] quicker and I need somebody to talk to. I couldn't teach him to talk. ... I tried to teach him to deal cards but that

didn't work either.

OP2C2: It would be nice if he could just shoot the breeze.

OP1C2: I want it to say "Good morning, [informant's name], my favorite co-worker" and display a little bouquet of flowers.

Maintenance staff agreed with operators that the addition of speech to the robot would be beneficial, but for other reasons. The robot's work cell can stop for a variety of reasons, including equipment malfunction, the lack of a necessary material or object, or a problem with the robot. Unsure of how to remedy the situation, operators often turn to a nearby maintenance worker for help. In such situations, neither operators nor maintenance workers usually have sufficient context for diagnosing the problem, requiring them to check many different components of the environment. Maintenance staff believed that the addition of speech capabilities to the robot may enable verbal troubleshooting, such as providing specifics on the problem (MT2C1) or giving step-by-step instructions on how to correct the problem (MT3C2, MT6C3).

MT3C2: It would be nice if Baxter could fix his own problems, but I would settle for him telling us how to do it.

MT5C3: I started working here long after he [the robot] got here, and sometimes I have no clue what to do [to fix him]. Him helping would be good.

Workers also suggested that the robot's face, which doubles as a screen, could provide more information. Employees felt that the screen could offer redundant

information in addition to the speech content. Manufacturing plants often contain background noise that might at times grow progressively louder. While employees believed that speech would be the easiest way to communicate with the robot, the addition of using the screen as an information display would provide employees with an alternative channel of communication should speech be impaired by the environment.

3.4.2.4 Responses to a Robot's Design

The last theme that emerged from my analysis was worker responses to the robot's appearance, focusing on two features: the robot's overall form and its eyes. Both elements of the robot's design were considered important for staff to feel comfortable working near the robot. Workers described the humanlike design of the robot as "familiar," giving them a sense of security and comfort when working in close proximity to the robot. This familiar design was in stark contrast with many industrial robots that are distinctly non-human and dangerous to be around. In the excerpt below, a maintenance worker highlights that this familiar design provides the robot with predictability.

> **MT4C2:** I like that it looks kinda like a human. ... It's familiar, ya know? I feel like I know what to expect.

Two other employees described the robot not as humanlike, but still as bearing a resemblance to other lifelike forms, specifically to a "praying mantis" due to the way it arms rotate. These employees still felt that this appearance induced feelings of safety when compared to other industrial robots, as illustrated in the excerpt below.

> **MT3C2:** Those arms, they remind me of a, what is it? Praying mantis? Yeah. Still, looks very calm, deliberate.

Employees also expressed a preference toward the dark-gray-and-red color and the industrial design of the robot. Compared to traditional industrial robots, employees felt that the robot's design suggested a friendly and non-threatening, even a "playful," interaction, as indicated below.

> **OP1C2:** Yeah, I'm like [to my friends], "He's like a Rock 'em, Sock 'em robot!" That red, it's so playful.

At Company 2, the addition of the wig and jester hat shown in Figure 3.3 further emphasized the humanlike appearance of the robot, adding to its "personality."

One particularly well-liked design element was the robot's eyes. The robot was equipped with a screen for a face, which displayed a pair of graphical "eyes." The face or the eyes served no sensing purpose (i.e., vision capabilities) but instead provided a way for nearby workers to understand the robot's current status in a "natural" way that did not require additional training, as expressed by a managerial employee below.

> **MG5C3:** I like them [the eyes]. ... Because, I love that, I mean, it, because it's the nonverbal communication. ... I think that it's just natural.

These eyes were pre-programmed to follow the trajectory of its arms, allowing employees to better anticipate where the robot's hands were likely to move next. As

illustrated in the excerpt below, this feature was considered particularly useful for new employees who might still be learning about the task.

> **MG1C1:** They [new employees] don't usually understand what the robot will do next, where it will go... This [the eyes], this helps them get it.

The robot also had a set of pre-programmed facial expressions: "confused" for when it is has trouble completing a task, "sad" for when the robot has given up on a task, and "surprised" for when a human had entered the workspace of the robot.

> **MG1C1:** They [the operators] know what the "surprise" look is, they know what the "sad" look is, they...they know it.

Employees noted that the robot's eyes and facial expressions were particularly useful when glancing at the robot from farther away. At a distance, the robot's eyes and facial expressions provided some context as to its otherwise indiscernible task status.

Additionally, employees felt that the robot's eyes conveyed intelligence. This perceived intelligence gave employees who worked in or around the robot's work cell confidence in the robot's actions and intentions, as expressed by the maintenance worker below.

> **MT4C2:** The eyes make him seem smart. Like he knows what he's doing.

3.5 Discussion

The themes that emerged from my analysis suggest two key implications for the design of collaborative robots: the importance of designing for sociality and the need to support a diverse set of relationships between the robot and different stakeholders. While this study helps answer the types of questions encountered in the motivating example in Section 1.1.1, these implications can be extended to other domains, such as hospitals, space shuttles, and the home, as well as other types of agentic technologies, including speech-based and embodied virtual assistants, such as a speech-based task guidance system building on its prior relationship with its user. Below, I discuss the implications of my work and the limitations of this study.

3.5.1 Designing for Sociality

Many of my results highlight the importance of sociality in a robot playing the role of a co-worker. I did not expect the social elements of the robot's design or social relationships people established with it to be important factors in its integration into a manufacturing environment, due to my naive presumption that there is little need for sociality in completing manufacturing tasks. Workers across three organizations in my study repeatedly brought up sociality in characterizing their relationship with the robot, in discussing the characteristics of the robot, and in suggesting improvements for the robot. My observations suggest that this desire for increased sociality stems from current social practices operators engage in amongst one another during their own work, such as two operators at adjacent workstations engaging in small talk.

The design of future collaborative robots must take into account the benefits of supporting user expectations of sociality to improve work practices as well as the social environment in these settings. Although they offer the same safety and flexibility benefits as Baxter does, many collaborative robots are designed only as single robotic arms with little or no elements to support sociality. My results indicate that social features that are already included in Baxter's design, such as its overall humanlike morphology and the behaviors displayed by its eyes and face, not only provide users with a positive experience by eliciting feelings of safety and comfort but also improve manufacturing work by communicating cues that are necessary for successful coordination. However, increased sociality has the potential to create false expectations that may risk user safety. Although collaborative robots are becoming increasingly safe for nearby users, designers must strive to match the perceived safety of the robot with its actual safety. Future designs of collaborative robots must build on the success of these features and further expand their use of design elements that support sociality while understanding and balancing this increased sociality against the expectations and the needs of its users.

Based on my findings, I believe that future designs could improve the robot's sociality to achieve two design goals. First, collaborative robots must be designed to support and enrich the social environment in the settings to which they are introduced. I found that social interactions and relationships are key elements of collaborative work even in task-oriented, safety-critical settings such as manufacturing. Supporting expectations for basic conversational skills, such as greeting co-workers and nearby workers at the beginning and end of their shifts, might enhance the social environment

in these settings. Second, collaboration by definition requires a coordination of actions for which communication and social cues are critical. Therefore, future collaborative robots must be designed with the necessary communicative functions to facilitate this coordination. For example, basic language capabilities could be added to allow users to ask the robot questions such as "What's wrong?" when there is a problem or to seek guidance from the robot in addressing it by asking "How do I fix it?" Additionally, future designs could draw on the cues that users currently rely on to monitor the robot, such as the sound of its operation or the direction of its gaze, to support this implicit form of communication.

3.5.2 Supporting Different Relationships

I found that different stakeholders made different attributions to the robot, maintenance and management staff perceiving the robot in more mechanical terms and operators viewing it as an agent with whom they can build a relationship. These different attributions resulted in different behaviors toward the robot and different characterizations of relationships with it.

Prior studies of the introduction of technology into organizations found similar differences in the responses of different stakeholders to the technology. For instance, studies of the introduction of delivery robots into hospitals found varying perceptions among stakeholders based on worker familiarity and time spent with the robot (Ljungblad et al., 2012), on the organizational role and gender of the workers (Siino and Hinds, 2005), and on the workload of and emotional demands on the workers (Mutlu and Forlizzi, 2008). The design recommendations made by these studies

included creating different behaviors that are better suited to the communication needs and context of different stakeholders, such as employing subtle light displays to alert high-workload employees at an oncology unit and using entertaining, pre-recorded voices that contribute to the cheerful social environment of a postpartum unit (Mutlu and Forlizzi, 2008).

The design of future collaborative robots must similarly accommodate different perspectives and interactions with the robot, such as robots recognizing the different stakeholder needs of doctors, patients and nurses in a hospital, or of parents and children in the home. In regards to the manufacturing setting explored here, future designs could draw on the social elements of the robot's design to improve the robot's sociality for maintenance and management staff to help reshape perceptions of the robot from industrial equipment to a more sociable co-worker. Social behaviors could be built into the types of interactions that these stakeholders are engaged in, such as integration, programming, and troubleshooting. Additionally, in developing relationships with operators, the robot could become aware of the operator's unique workflow, and incorporate that awareness into communicating with the operator (e.g., if the robot recognizes it is almost out of raw materials, asking the operator about acquiring more at a convenient time). While many existing collaborative robots including Baxter are equipped with capabilities for interactive programming, such as learning from demonstration, these capabilities could be augmented to include conversational elements for input and feedback.

3.5.3 Human-Machine Interfaces

My analysis suggests that while the current interface design is promising, it lacks the level of intuition desired to allow use by users with a variety of backgrounds and skill levels. The challenges presented by PLC programming present an opportunity to develop a more intuitive interface for communicating with external equipment. Such an interface could allow those who have computational thinking skills but lack background in PLC programming to more easily integrate the robot with other systems. Particularly in small companies, where PLC programming is often contracted out, this interface would allow companies to save money otherwise spent on hiring external support. Additionally, future interface designs might prompt users to consider possible points of failure in the system and ask users to provide guidelines for the robot to follow when the robot encounters these situations.

3.5.4 Future Work Cell Design

As collaborative-manufacturing-robot technologies converge on the vision of being easy to retrain for new tasks, the role of operators and maintenance workers will likely be redefined. Maintenance staff will shift toward designing robot work cells, ensuring that the physical layout of the plant is suitable for the capabilities of robots. Operators will become more technically competent, ensuring their cluster of robots is operating correctly and efficiently and perhaps retraining and troubleshooting as necessary. Such a shift will likely require improvements in two areas. First, operators will need a higher degree of computational thinking in order to effectively retrain robots for new tasks. Second, robots will need a more accessible interfaces for operators to work

with, whether that be an authoring environment available via a computer screen, or through interacting socially with the robot to provide instructions and training.

3.5.5 Limitations

While my findings offers many interesting insights into the integration of collaborative robots into manufacturing settings, my study has limitations that point to follow-up studies and analyses and future research directions. First, many of my results highlight the importance of sociality in worker interactions with and perceptions of the robot, which in large part might have been shaped by the robot's humanoid form (Breazeal, 2003). However, prior work suggests that people's responses to robots are shaped by a broader set of design elements, such as how the robot's appearance matches its task (Goetz et al., 2003). Future work should examine worker interactions with collaborative robots with different morphologies, such as robotic arms, performing different types of tasks. Second, my study sites included some of the very first manufacturing facilities to own and use collaborative robots, who as early adopters, might have experienced integration issues that may become a rarity as the technology and integration practices mature. A smoother integration process might change some of my observations, such as the mechanical view that maintenance and management staff had of the robot. Future studies that focus on the integration process could clarify the role of integration problems (or lack thereof) on worker perceptions of collaborative robots.

3.6 Study Conclusions

The introduction of collaborative robots into human spaces is poised to revolutionize how work is done in these settings and how workers adapt to and interact with a robot "co-worker." To better understand these changes and guide the future design of these technologies, I conducted an ethnographic study at three manufacturing sites located in the continental United States that were early adopters of a particular type of collaborative robot intended for industrial settings. I conducted fly-on-the-wall observations and interviews at each site with different stakeholders, including managerial employees, maintenance staff, and operators. The Grounded Theory analysis found seven themes, centered around how the introduction of a robot co-worker changes the work cell the robot is used in, and the social component a robot co-worker introduces.

These findings have broad implications for robot co-workers, both within manufacturing settings as well as for other collaborative settings. Drawing on my findings, I recommended that future designs augment the social capabilities of collaborative robots, specifically to support the coordination necessary to perform collaborative work and to enrich the social environment in the robot's particular settings, whether it be the home or a workplace. I also suggested that future designs accommodate the expectations and needs of different stakeholders, such as improving the social capabilities of the robot not only for immediate collaborators, but also for those who have less frequent and different types of interactions with the robot. These improvements will help both individuals and organizations to more smoothly integrate collaborative robots into their setting and the corresponding social environment. I

also make longer term recommendations concerning the programming of these robots to realize a future where operators service clusters of robots. The findings of this study contribute to our broader understanding of interactions with robotic products in real-world settings, and these recommendations offer designers concrete guidelines for better supporting collaboration and improving user experience in these settings.

4 BEHAVIORAL CUE STUDY: SPEECH PATTERNS

As collaborative robots become a more commonplace products in a range of settings, designers will need materials and tools that will enable them to explore and prototype a range of interactions that these collaborations might take on. For example, a robot working together with a human might offer greetings, answer questions, and give instructions, such as in the scenario described in Section 1.1.2.1. These interactions will be comprised of a fundamental set of communicative acts. This work explores how these communicative acts might serve as design patterns and how a pattern language can be used to enable design exploration and prototyping of human-robot interactions.

In this work, I aim to build a pattern language that might serve as building blocks for human-robot interactions across different forms of interaction, focusing on five scenarios in which robots are expected to engage, including collaboration Fong et al. (2003), conversation Hsiao et al. (2003), instruction Huang and Mutlu (2012), interviews Fussell et al. (2008), and storytelling Mutlu et al. (2006). I present a formative study of human interactions across these five scenarios and describe seven patterns that appear across these scenarios. These patterns informed the design and implementation of *Interaction Blocks*, an authoring environment that enables users to explore and create interactions for social robots. I anticipate Interaction Blocks having three primary groups of users: (1) interaction designers who have little programming background, (2) designers or programmers who would like to rapidly prototype interactions, and (3) designers or programmers who might not have experience in working with human behavior as a design element. Through a

qualitative evaluation with interaction designers and developers, I demonstrate the feasibility of the use of a pattern language for designing human-robot interactions and the usability of our authoring environment.

4.1 Formative Study

To build a pattern language that enables design exploration and prototyping for human-robot interactions, I conducted a formative study of human interactions that involved observations of human interactions and identifying and modeling patterns in which these interactions unfolded. While a variety of behaviors and scenarios have been studied by linguists and psychologists (Schegloff, 1972; Clark, 1992; Goffman, 1983), the results have not been constructed into design patterns that can be translated and implemented on a robot. To achieve natural, humanlike robot behaviors, I chose to ground my development of interaction design patterns in observations and detailed analyses of human interactions. I collected data from eight dyads performing in five social interaction scenarios that are representative of the types of interactions robots are envisioned to encounter in their future roles in society: *conversation*, *collaboration*, *instruction*, *interview*, and *storytelling*. I then followed an iterative modeling process to construct models that represented each of the scenarios. These models revealed common interaction design patterns that appeared across multiple scenarios. The paragraphs below detail and discuss the process of discovering and formalizing these patterns.

4.1.1 Data Collection

4.1.1.1 Participants and Data Corpus

A total of 16 native-English-speaking participants from the University of Wisconsin–Madison took part in this study. Participants studied a diverse range of fields, and their ages ranged 19–62 ($\mathsf{M} = 25.44$, $\mathsf{SD} = 10.39$). Participants were assigned into dyads to jointly interact in the social interaction scenarios. I randomly assigned participants into dyads and conversational roles such that they were fully stratified by gender. The instructor and storyteller roles were never held by the same participant, allowing the instructor to learn the task to be instructed while the storyteller learned the story. The scenarios were executed in the same order for each dyad.

A single video camera equipped with a wide-angle lens was used to capture the entire upper body of the participants for all scenarios. The final data corpus consisted of five scenarios for each dyad, which amounted to 3 hours and 31 minutes of audio and video data. The average lengths for the collaboration, conversation, instruction, interview, and storytelling scenarios were 3:47, 4:47, 5:02, 7:32, and 3:59, respectively.

4.1.1.2 Procedure

After arrival in my laboratory, a researcher provided the participants with an overview of the interaction scenarios. Following informed consent, the participants were then led into a usability laboratory, which was initially configured for the first scenario in which they would be participating. The room was reconfigured for each scenario, as shown in Figure 4.1. Following the completion of all five scenarios, participants completed a demographic questionnaire. The researchers then debriefed both participants.

Figure 4.1: Examples of the experimental setup for all five scenarios. From left to right, participants are sharing in a storytelling experience, engaging in a conversation, conducting an interview, learning how to configure a set of pipes, and collaborating on how to sort foodstuffs.

The five scenarios and post-experiment questionnaire took approximately one hour. Participants were compensated $10 each for their time.

4.1.1.3 Interaction Scenarios

I designed five interaction scenarios, shown in Figure 4.4, that showed the characteristics of the scenarios in which robots are expected to interact with people and that followed previous modeling research, which will be discussed in more detail below. Each scenario was intended for small group interaction. For data collection, I engaged two participants in each interaction as the minimum number required to realize the scenario.

1) Conversation: In the conversation scenario, I aimed to observe how participants would handle a set of unstructured exchanges about a given topic. Framing, turn-taking, and recovering from errors are all important aspects of a successful conversation (Tannen, 1989). In particular, participants who have not yet established any level of rapport may favor question-answer pairs, rather than conversational dialogue (Tannen, 1989).

In the scenario, the participants discussed their educational experiences and goals.

They were instructed to continue this conversation until the researcher re-entered the room when the conversation naturally subsided.

2) Collaboration: The collaboration scenario targeted gaining a better understanding of how participants collaborate toward a common goal. Previous research has shown that joint activity requires flexibility and communication in order to effectively coordinate differing opinions from participants (Clark, 1996).

In my collaboration scenario, the participants worked together to sort six grocery bags of foodstuffs onto two tables. The tables were divided into three areas meant to represent common food storage locations in the home: the *pantry*, the *fridge*, and the *countertop*. These areas were further subdivided into areas based on types of food. Participants were asked to place the empty grocery bags on the table to indicate that they were done.

3) Instruction: In the instruction scenario, I aimed to observe how an expert conveys knowledge to a non-expert and, when that knowledge is miscommunicated, how the expert might help the non-expert to recover (Skehan, 1996).

My instruction scenario involved first training one of the participants (the *instructor*) in assembling a particular pipe configuration to allow two sinks to drain into a single system. The instructor was given as much time as they needed to learn how to configure the pipes. Upon learning the necessary steps to build the pipes, the instructor walked the second participant (the *student*) through each step of the assembly.

4) Interview: The goal of the interview scenario was to capture the process by which one participant questioned a second participant to obtain information. Interviewers

may sometimes request sensitive information, requiring substantial rapport between the interviewer and interviewee. The structure and the tone of an interview can help determine the level of rapport between participants (Matarazzo and Wiens, 1972).

In my interview scenario, one of the participants (the *interviewer*) was told that they would be conducting a job interview for a generic position. The interviewer was given a list of 14 questions and time to review them. After reviewing the questions, the second participant (the *interviewee*) was asked all 14 questions by the interviewer. The researcher reentered the room when all the questions were answered.

5) Storytelling: In storytelling, I were interested in observing how one participant relayed a story to a second participant and what discourse patterns (e.g., asking questions) helped participants communicate more effectively (Langellier, 1989).

In my storytelling scenario, one participant (the *storyteller*) watched a seven-minute video of a cartoon story. The storyteller was given as much time as they needed to feel comfortable with retelling the story, after which they had three to five minutes to retell the story to the second participant (the *listener*). After the storyteller finished, participants frequently had a conversation concerning the story. The experimenter re-entered the room when the conversation subsided.

4.1.2 Analysis

For each scenario, I followed an iterative process of data coding and modeling. The coding process involved a researcher iteratively coding all video data and a second researcher annotating 10% of the videos to confirm the reliability of the coding process (81% agreement, Cohen's $\kappa = .76$). In the coding of the data, I annotated the

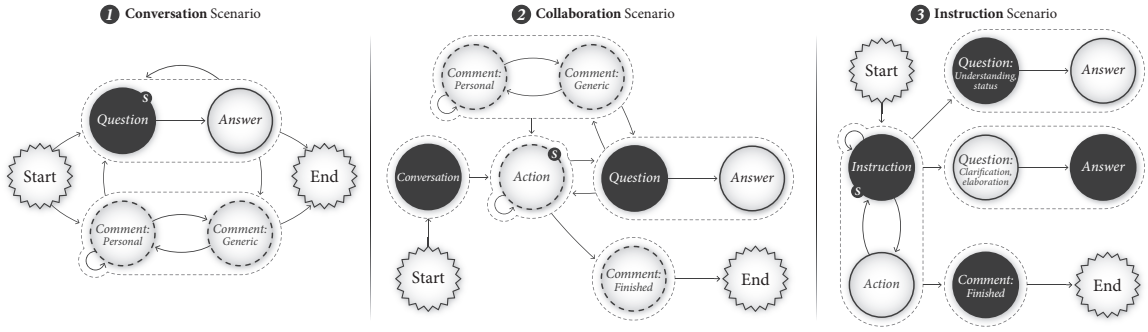

Figure 4.2: Models for the conversation, collaboration, and interaction scenarios. In the instruction scenario, dark-colored states are for the dominant role (instructor), while light-colored states are for the respondent's role (student). In the conversation and collaboration scenarios, the roles for each participant are not fixed throughout the interaction. E.g., during a conversation, a participant may ask the other a question, which may then be followed by the second participant asking the first participant a similar question. Here, dark-colored and light-colored states indicate that two different participants occupy each state. States in dashed outlines can be occupied by any participant.

videos for the set of *states*—significant events in the interaction—and the *transitions* between these states, generating a *model* of the interaction scenario. From these models, I extracted what appeared as the core part of the interaction: the states that were essential to characterize the interaction. For example, in an interview, the interviewer asking a question and an interviewee answering it serve as two core states of the interaction. Finally, I examined the models from all five scenarios for states or sequences of states that frequently appeared to identify what might serve as *interaction design patterns*.

My consideration of significant events in the interaction as states follows prior studies of social interaction (Tannen, 1989; Clark, 1996; Matarazzo and Wiens, 1972). When connected with other states, the resulting model represents the flow of the

interaction from one event to the next. The idea of states translates well to social interaction, where one participant usually holds the floor and is engaged in an event (e.g., sharing a comment) (Sacks et al., 1974). For the purposes of an interaction with multiple participants, the states of all participants can be mapped to a single model. When a state that is held by only one participant is entered, all other participants implicitly wait. The use of states and transitions also fits well with a common paradigm used in robot programming where "nodes" are used to control functionalities of the robot (Quigley et al., 2009). The use of states provides a flexible representation for the flow of interaction for multiple participants.

After establishing the core states, I reviewed each video and noted any deviations from this core part of the interaction in both kind and number. An example deviation in the interview scenario is the interviewee asking for clarification regarding the interviewer's question. State models for each scenario were then constructed from the core part of the interaction, any deviations from the core part of the interaction, and notes collected from the video data. The resulting scenario models were then compared against the interactions in the videos for any mistakes or inconsistencies.

The resulting state models for collaboration, conversation, instruction, interview, and storytelling are illustrated in Figures 4.2 and 4.3. Each model uses a set of standard conventions to convey the flow of the scenario. All models have a start and end state. In the instruction, interview, and storytelling scenarios, dark-colored states represent the dominant role (instructor, interviewer, and storyteller), while light-colored states represent the respondent's role (student, interviewee, and listener). In the conversation and collaboration scenarios, the roles for each participant are not

fixed throughout the interaction. For example, during a conversation, one participant may ask a second participant a question, which may then be followed by the second participant asking the first participant a similar question. Here, dark-colored and light-colored states indicate that two different participants occupy each state. States which have a dashed outline can be occupied by either participant.

4.1.3 Models

After constructing a model for each scenario, I identified common interaction structures, which served as *design interaction patterns*, or patterns. For example, a question being asked and answered is comprised of two separate states (a question state and an answer state). However, questions are almost always followed by answers, and thus the interaction between both states is codified into a pattern. I identified seven common patterns (Figure 4.4) across all five scenarios:

1) Introductory Monologue: The introduction that begins the interaction is often

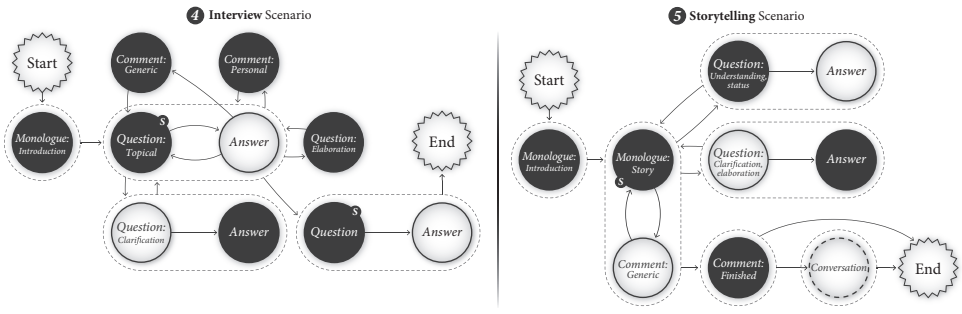

Figure 4.3: Models for the interview and storytelling scenarios. In the interview and storytelling scenarios, dark-colored states are for the dominant role (instructor), while light-colored states are for the respondent's role (student). States which have a dashed outline can be occupied by any participant.

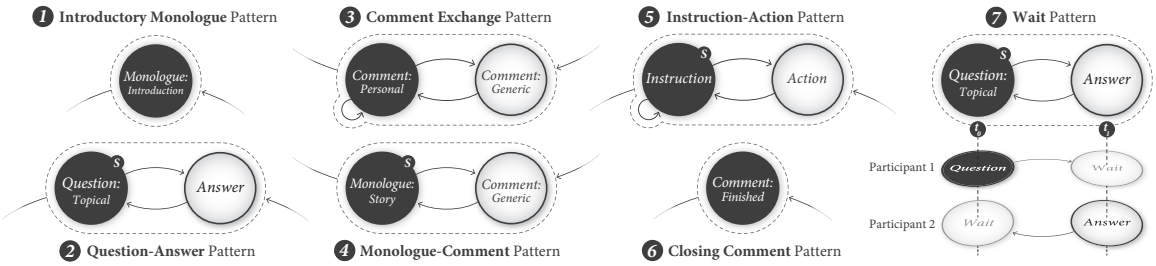

Figure 4.4: Models for the seven patterns I discovered. The dark and light-colored states indicate when one participant occupies the dark-colored state, a second participant must occupy the light-colored state. For example, for pattern 5, if one participant asks a question, the second participant gives an answer.

the most important indicator of how well the remainder of the interaction will play out (Schegloff, 1967). A short introduction can be used to introduce other participants to a scenario by giving an overview of the remainder of the interaction. In the interview scenario, some participants started with a short introduction of themselves, describing the interview as part of the "hiring process." In the storytelling scenario, all participants started with a short introduction that set the stage for the story, such as identifying the primary characters and setting. An example of an introduction used in the interview scenario is shown below:

> *Interviewer*: Hi, welcome. ... So, today I'll be asking you a few questions to gauge your compatibility for this job.

Previous work on spontaneous encounters notes that these introductions can take different forms (Schegloff, 1972). It is likely that for scenarios where I did not observe an introduction state (conversation, collaboration, and instruction), an introduction state would have occurred in a natural setting. For example, a person may pass by

their friend while shopping and summon the friend with a short introduction ("Hello Jane! How are you doing?") before commencing with a conversation.

2) Question and Answer: The idea of a *question-answer pair* is a well-studied component of discourse management (Clark, 1992). A *question* is a sentence meant to elicit information from other participants. Questions may introduce a topic of interest (e.g., "What has been yMy most difficult job?"), request elaboration (e.g., "Could you elaborate on that?"), ask for clarification (e.g., "What do you mean by that?"), confirm understanding (e.g., "Does that make sense?"), check background (e.g., "Have you ever watched Looney Tunes?"), and request status (e.g., "How are you?"). In an archetypal question-answer pair, a question is be followed by an answer. An *answer* is the response to a question that aims to satisfy the questioning participant's curiosity (Schegloff, 1972). The excerpt below from the storytelling scenario is an example of a question-answer pattern:

Storyteller: Do you know who Marvin the Martian is?

Listener: Oh yeah, from Looney Tunes.

3) Generic Comment and Personal Comment: A *comment* is a short statement offering the speaker's opinion. Comments are either generic (e.g., "Wow") or personal (e.g., "I tried that and didn't like it"). In my data, participants engaged in exchanging comments frequently move between sharing personal insights—either of their own volition or after being prompted—and offering generic comments. The following excerpt illustrates an exchange of comments from the conversation scenario:

Participant 1: Wow.

Participant 2: Yeah...I had never done anything quite like that before.

Participant 1: I had a similar experience once, but it wasn't nearly that exciting.

Participant 2: Interesting.

In this example, both participants offer both generic and personal comments, highlighting the fluidity between these two types of comments for all participants of the interaction.

4) Monologue and Generic Comment: A *monologue* is a longer form of speech during which no response is expected. Monologues may involve the telling of a story (e.g., "...Once Marvin had reached Bugs Bunny, he chose to..."). Although monologues expect no response, listeners may occasionally offer unsolicited commentary, as illustrated by the excerpt below from the storytelling scenario:

Storyteller: ...and then, all of a sudden, thousands of these aliens appear on Earth.

Listener: That's a lot.

5) Instruction and Action: An *instruction* is a command offered by one participant to direct the actions of another participant. The proper response to this instruction is often an *action*, although the action might follow the instruction with a delay depending on whether it is an appropriate time to perform that action (Brandstätter et al., 2001). Instruction-action pairs are commonly found in teaching scenarios where the teacher is directing the student. Below is an example of an instruction-action pair from the instruction scenario:

Instructor: Now connect the long pipe with the one shaped like an "S".

Student: <locates both the long and S-shaped pipe, and then connects them>

6) Finished Comment: Upon the completion of the goals of the scenario, one or more of the participants will note that the scenario is completed by offering a *finished comment*. For some scenarios (the interview, instruction, and storytelling scenarios), only one of the participants is able to end the scenario (e.g., only the storyteller knows when the story is finished). In the collaborative scenario, either participant is able to end the scenario, as illustrated in the following example:

Participant: I think that's it, so I should be done.

Previous work has shown that conversations frequently have a definite ending initiated by either participant, whether through an unexpected interruption (e.g., a phone call), a forced ending (e.g., a train stop forces participants to go separate ways), or an achievement of the goals of the conversation (e.g., obtaining some piece of information) (O'Leary and Gallois, 1985). However, my instantiation of a conversation scenario lacked a comment confirming the end of the interaction, due to the conversation scenario in my study providing no concrete end goal, except to converse for three to five minutes until the experimenter interrupted the conversation at a natural break.

7) Wait: One pattern implicit in all scenarios involving two or more participants is the *wait* pattern. The majority of states across my scenarios are intended for a single participant. When a participant transitions into a state intended for a single

participant, all other participants enter a wait state, as shown in Figure 4.4. Data on conversations overwhelmingly supports only a single speaker at a time and other participants listening to the speaker, with multiple speakers being common but brief (Sacks et al., 1974).

My analysis of five common social interactions not only provides a deeper understanding of each scenario, but also reveals the prevalence of a number of patterns across scenarios. These patterns confirmed some previous patterns (e.g., question-answer pairs (Clark, 1992)) as well as discovered new ones (e.g., Generic Comment and Personal Comment).

4.2 Interaction Blocks

My analysis of human interactions revealed seven core patterns that appear across the five common social scenarios. To draw on these patterns in design exploration and prototyping for human-robot social interactions, I developed *Interaction Blocks*, an authoring environment that enabled interaction designers to synthesize interactions and prototype them on a NAO humanoid robot, a robot platform commonly used in research and design for human-robot interaction.

4.2.1 Authoring Environment

My authoring environment is composed of three sections: the *control panel*, the *pattern library*, and the *interaction timeline*. An example of the tool can be seen in Figure 4.5.

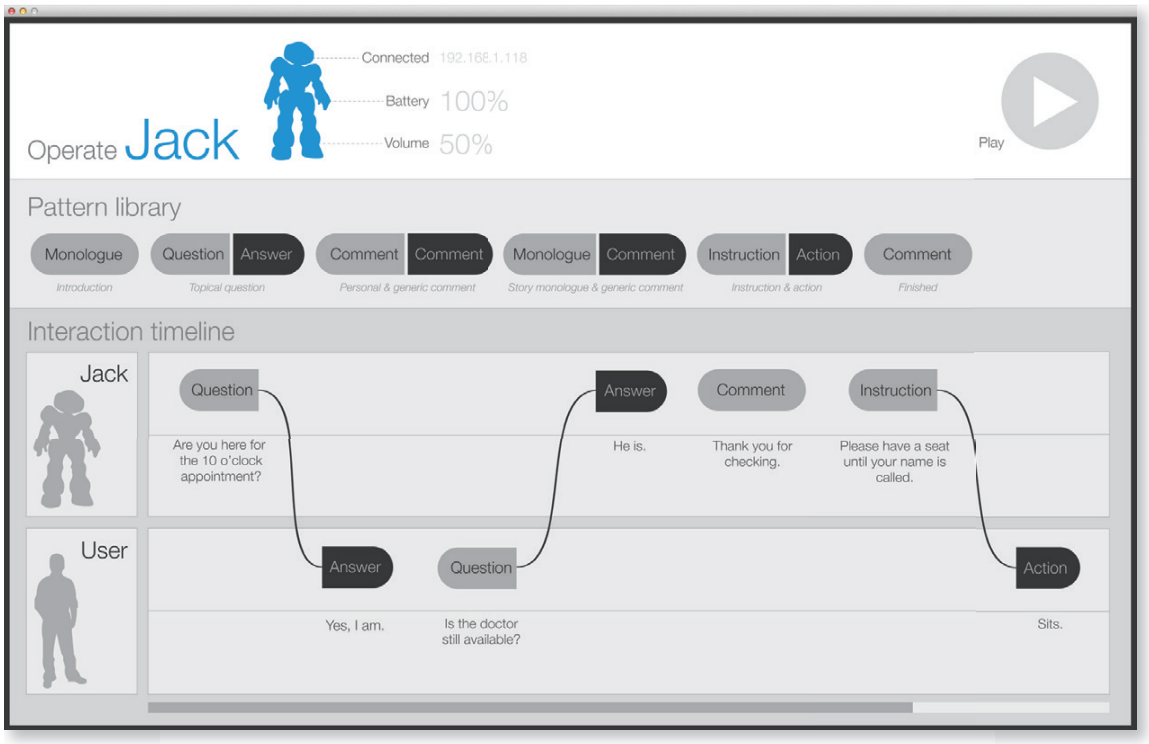

Figure 4.5: A screenshot of *Interaction Blocks*, my authoring environment for using a pattern language for synthesizing human-robot interactions.

The *control panel* displays pertinent information to the user. The silhouette of the robot provides a visual indication of connection status, with blue indicating "connected" and grey indicating "disconnected." If connected, the control panel also informs the user of which robot they are connected to, and the IP address, current battery level, and current volume of the connected robot. The "Play" button on the far right side of the interface is used to upload and execute the user's synthesized dialogue on the connected robot.

The *pattern library* contains the patterns discovered in my formative study, with

each pattern represented in a capsule shape. I chose to represent my patterns in a capsule shape to accommodate dual-colored capsules for those patterns which are comprised of two states. To compose an interaction, patterns can be dragged out of the pattern library and onto the interaction timeline.

The *interaction timeline* is where the user composes the interaction. To accommodate both the user and the robot roles, the timeline is divided in half, each half assigned to one of the roles. When the user moves a pattern from the pattern library to the interaction timeline, the pattern is automatically added in place to the appropriate role. For patterns with two states, the movement of the pattern into the interaction timeline causes the pattern to divide in half, as shown in Figure 4.5. A Bézier curve connects the two states, indicating that no patterns can be inserted between them. When a pattern is divided into two states, the user is in control of the first state in the pattern. The second state is automatically added to the opposite role. For example, if a "question" is added to the user's timeline, an "answer" will be added to the robot's timeline.

4.2.2 Implementation on the NAO Robot

To allow users to evaluate their synthesized interaction on a robot, I enabled Interaction Blocks to connect and execute the resulting dialogue on a NAO robot. All robot utterances were generated through the NAO's native text-to-speech application programming interface (API), while participant responses were recognized using the NAO's native natural language processing API.

In addition to the dialogue, socially appropriate gaze behaviors were automatically

incorporated into the robot's interaction. The NAO employed its native face-tracking capabilities to enable consistent gaze with the participant. I introduced Perlin noise (Perlin, 2002)—a technique used in animation and film to simulate randomness that appears natural—to create small head shifts in the robot and create a lifelike appearance. Additionally, I constructed socially appropriate gaze aversion behaviors for the robot, using the timings provided by Andrist et al. (2013a). For instance, when the floor was passed to the robot, the robot would gaze away before returning its gaze to the participant, resuming Perlin noise, and continuing with its part of the dialogue, all of which helped the robot display natural social behaviors.

4.2.3 Evaluation

I evaluated my tool in design sessions with local members of the design and development community from two groups: *interaction designers* and *developers*. Using a scenario for an interaction between a robot receptionist and human patient, participants were asked to construct and test an episode of interaction using my authoring environment.

4.2.3.1 Design Task

Participants were given a scenario that described an exchange at a dentist's office between a patient (the user) and the receptionist (the robot). The scenario included a set of micro-interactions, such as "the receptionist greets the patient" or "the patient asks the receptionist when their next scheduled appointment is," that guided the participant in constructing an interaction episode. Participants took as much time as

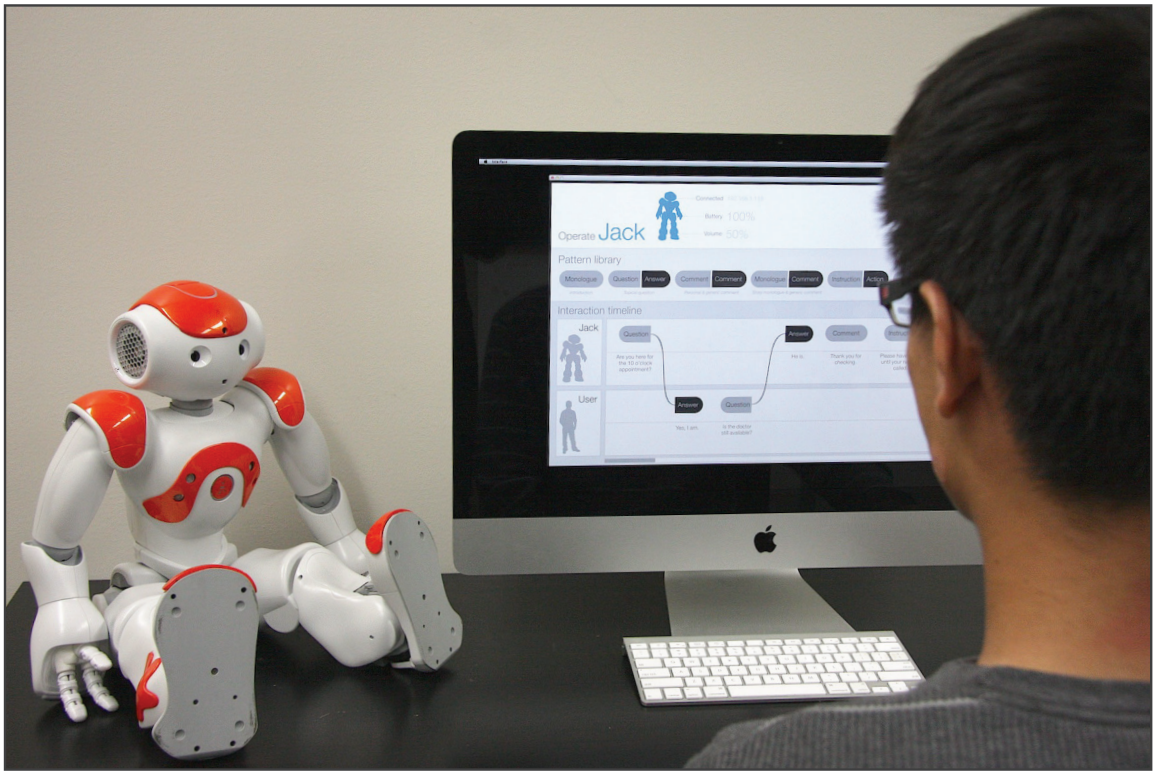

Figure 4.6: A researcher demonstrating the setup of the design sessions with the Interaction Blocks authoring environment and a NAO humanoid robot.

necessary, taking between 29 and 62 minutes (M =45 minutes, 30 seconds, SD =10 minutes, 37 seconds) to complete the design task. The setup of the study is shown in Figure 4.6.

4.2.3.2 Recruitment & Selection

I recruited 10 participants—five interaction designers and five developers—from the local campus community. The interaction designers were selected from among students who had completed the most advanced level interaction design course offered on

campus, which taught principles and methods for user-centered interaction design and provided students with hands-on experience in design exploration and prototyping. Developers were recruited from among upper-level computer science majors with web development experience with the goal of capturing the perspectives of more technical users who might use my tool. Participants were between 19 and 26 years of age ($M = 22.5$, $SD = 2.68$) and were either computer science students or were employed in positions that required a computer-science background.

4.2.3.3 Session Procedure

Following informed consent, participants were guided into a usability laboratory by a researcher. The researcher demonstrated Interaction Blocks, introducing the participant to the purpose of the tool, the layout of the interface, and the workflow necessary to explore and prototype interaction episodes and to test them with the robot. The participant was given five minutes to explore the interface independent of the researcher and was provided with the opportunity to ask questions at the end of the exploration period. At the end of this period, the researcher provided the participant with the scenario and left the room. The participant was given as much time as necessary to complete the design task. Following the completion of the task, the participant completed the System Usability Scale (SUS). The experimenter then conducted a semi-structured interview with the participant on the design of Interaction Blocks, their experience with using the authoring environment, and how it facilitated their exploration. Participants were compensated $10 USD for their time.

4.2.3.4 Analysis

Following guidelines suggested by Bangor et al. (Bangor et al., 2008), a usability score from 0 to 100 was calculated by rescaling each of the 10 items used in the SUS to range from 0 to 4, summing the scores for the 10 items, and multiplying the result by 2.5.

I used content analysis to analyze the interview data (Berelson, 1952). Each interview was transcribed, and pertinent responses from each participant were included in a spreadsheet. After reviewing the responses across all participants, significant responses were noted, and affinity diagramming was used to organize the responses into emerging themes, refining as necessary.

4.2.4 Results

In this section, I present my findings from two different sources of data: SUS scores and interview data. The SUS scores serve as a metric of the overall usability of the authoring environment in enabling the use of a pattern language for design exploration and prototyping in creating human-robot interactions. Findings from the interview data provide insights into the designers' and developers' experiences in the design sessions and guidelines for future design and development of methods and tools for supporting human-robot interaction design.

4.2.4.1 SUS Scores

The SUS scores for the authoring environment ranged from 75 to 95 ($\mathsf{M} = 84$, $\mathsf{SD} = 7.38$), interaction designers assigning an average score of 82, while the developers

assigned an average score of 86. Based on established guidelines for interpreting SUS scores (Bangor et al., 2008), a score of 80 or higher places the interface in the highest quartile when considering a survey of interfaces evaluated using the SUS.

4.2.4.2 Interview Data

Below, I highlight themes found from the interview data. Excerpts from the interview are labeled with either ID or DE for the interaction designers and developers respectively, followed by a number.

Patterns' Ease of Use: Interaction designers and developers stated that the patterns—particularly patterns with multiple states—made it easy to quickly compose and test interactions. For patterns with multiple states, participants felt that the authoring environment's ability to automatically add the second state to the opposite role alleviated the burden of mentally planning the necessary pairings of behaviors and helped prevent creating inappropriate behaviors such as composing a question without an answer. Additionally, ID2 noted that the use of patterns—rather than a generic dialogue box used to accommodate all utterances—forced him to more carefully consider his dialogue prior to adding it. The same participant also inquired about the ability to create custom patterns. The excerpts below illustrate responses related to the ease of use of the pattern language.

> *DE1: For most people, programming is not the easiest thing, and this is a very intuitive way to design dialogues.*
>
> *ID2: I think the linkages of showing questions-answer, instruction-action show the clear delineation with that. That would really help in formalizing*

the structure and kind of removing the ambiguity of real conversation from what you're working on...

However, while participants commented on the ease of use of the patterns in general, some expressed a desire for additional guidance on the nuances between some of the patterns due to their lack of domain knowledge. For example, DE2 noted that there was a "monologue" pattern, a "comment" pattern, and a "monologue-comment" pattern. The participant was initially unsure why a "monologue-comment" pattern would be different from the combination of the individual "monologue" and "comment" patterns. DE1 voiced confusion about the presence of fMy "comment" states across three patterns. While participants generally felt comfortable using the patterns to accomplish the task goals, many participants asked for some form of documentation. Suggestions ranged from a traditional documentation interface with examples or a walk-through (ID4) to tooltips for each pattern (ID1).

Design Exploration: Participants found the ability to compose interactions on the timeline of the authoring environment and to execute them on the robot to support their design exploration and iterative design of the behaviors and interactions that they were building toward achieving natural dialogue. These comments are best illustrated by excerpts below by ID3.

ID3: I think that human-human interaction can change all the time, and hearing it more and more and more would make it better in the end.

ID3: I mean it seems like this kind of interface is kind of, like, to play, so you can start using questions and answers right. Change the questions

and the answers, add a comment in if you want.

Approachable Interface: All participants commented on the clean and minimalistic design of the interface for the authoring environment, discussing it as an enabling tool that gave them the confidence that the complex task of synthesizing an interaction would be doable, as highlighted by the excerpts below.

> *ID4: It's very clean in design. It seems easy to use for that reason. It doesn't have many functions, so it seems like it's easy to learn.*
>
> *ID3: I think everything's very easy to see, like the flow. It's not easy to get lost. It's the least amount of details for someone who's maybe not majoring computer science.*
>
> *DE1: For most people, programming is not the easiest thing, and this is a very intuitive way to design dialogues.*

Error Prevention: DE1 and DE3 noted that the use of the pattern language and the visual authoring environment would aid designers who may have little to no development or programming experience in building an interactive application with a robot, as illustrated in the excerpt below. Their comments also suggested that this approach would enable developers to bypass many of the bugs frequently encountered when programming, such as syntax errors and typos, and alleviate the need for them to learn a new application programming interface (API).

> *DE3: If it has the functionality to prevent me from making mistakes, then that's good too, right?*

The Need to Support Free-Form Exploration: Some designers and developers expressed that the addition of a less structured environment might better facilitate their design exploration. Examples included a whiteboard (DE1, DE3), with paper and pencil (ID1), sticky notes (ID4), and a text editor (ID1, ID2). While these participants appreciated the structure and code generation benefits provided by Interaction Blocks, they felt a need for a less-structured initial step to the exploration that allowed them to easily create, edit, and view the behavior and dialogue components that they planned to use in the more structured construction of the robot's exchange with users. Comments from DE2 and ID2 below illustrate the need to support free-form design exploration.

> *DE2: Well honestly I'd probably write out a script in a text editor, and then drag out all the parts and fill in the text.*
>
> *ID2: This more feels like that I know what my design is and I am going to formalize it someway rather than 'this is how I lay it out and tweak it as I go kind of thing.' I almost feel this is the last stage.*

Need for Branching: While participants appreciated the simplicity of the current interface, several participants raised the question of accommodating alternative ways in which the interaction or the dialogue between the robot and its user might unfold. Questions focused on two facets of this problem: accommodating different utterances with the same semantic meaning (e.g., "Yes" versus "Yes, please"), and accommodating utterances with different semantic meanings (e.g., "Yes" versus "No"). Participants suggested adding into the authoring environment support for a list of

possible responses for accommodating utterances with the same semantic meaning. To handle multiple branches of dialogue that result from utterances with different semantic meanings, participants suggested creating multiple timelines that can be collapsed and expanded (DE5) and displaying multiple timelines at once (ID4).

In addition to these suggestions, participants commented on the possibility of additional visualizations, color coding patterns for easier identification and disambiguation, and adding additional information concerning the robot's status.

4.3 Discussion

Results from the design sessions focused on three main findings: the use of design patterns, the design of the authoring environment, and the workflow required to synthesize dialogue. These results demonstrate how users can test and deploy new behaviors on a collaborative robot, similar to the example outlined in Section ref-sec:speechPatternsExample, as well as offer many future directions for improving and expanding the pattern language and authoring environment to better support human-robot interaction design.

4.3.1 Design Patterns

The use of a pattern language was reviewed favorably by interaction designers and developers alike, with many noting the ease in which they could explore and prototype exchanges for the robot and users through their use. Additionally, participants highlighted the structure that patterns afforded as an advantage, noting that patterns

forced them to consider their design choices for dialogue and interaction elements and how they contributed to the overall goals of the interaction.

Although participants found patterns easy to use, what interaction elements some patterns represented was not always clear to them. Participants suggested various forms of providing help within the authoring environment, including tooltips and examples to help acclimate users to the pattern language. Documentation of the patterns would not only help interaction designers better understand the design elements, but would also inform researchers interested in using and extending the pattern library presented here.

4.3.2 Authoring Environment

The authoring environment was cited by all participants for its ease of use and approachable design, which was confirmed by the SUS scores. The requests and suggestions for improvement discussed in the previous section underline the importance of accommodating more advanced functionality, such as branching, in future versions of the authoring environment without sacrificing too much from the simplicity of the current design. Some participants suggested the use of a "superuser" mode or the ability for users to reveal or hide detailed options or functionality for patterns.

4.3.3 Workflow

Participants frequently cited the ability to rapidly modify and evaluate their design ideas as a useful feature that aided them in the iterative development of their final design. This finding confirms prior work that highlights the importance of rapid

design exploration and prototyping in the development of human-robot interactions to provide designers with a better understanding of how their designs would perform in real-world settings (Lohse et al., 2014). Additionally, the use of Interaction Blocks frees the designers' workflow of debugging, enabling them to concentrate on designing interactions.

While participants expressed enthusiasm for the use of Interaction Blocks as a part of their workflow for designing and prototyping human-robot interactions, they also expressed a desire for a less formal design step prior to using Interaction Blocks. Some participants viewed using the authoring environment as a last step to formalize their work, while others suggested using Interaction Blocks earlier in the design process after brainstorming to create a basic idea of how the designed interaction may unfold. Further studies of the design practices of interaction designers might inform the design of flexible authoring environments that support informal as well as formal exploration at different stages of the design process.

Finally, designers and developers voiced a need for the ability to sketch out dialogue exchanges, see how they fit within the interaction, and easily refine individual dialogue elements. To support this need, the authoring environment might provide a view of the script of the entire dialogue, enabling users to easily gain an overview of their design and make changes at each exchange from a global perspective. Additionally, this sketching environment might allow users to draft the dialogue first and use a drag-and-drop interface to add these dialogue elements to the patterned interaction elements.

4.3.4 Interaction Designers vs. Developers

While there were many topics that interaction designers and developers agreed on, there were also differences in how these two groups evaluated the use of the pattern language and authoring interface. For instance, interaction designers were more concerned with how new visualizations might enable users to better compose and manage complex interactions, while developers focused on the reduction in errors enabled by the use of the pattern language and visual authoring, as well as the desire for more information on the robot's status, the ability to batch edit and move large portions of the interaction, and the need for incorporating branching into the development of the dialogue. These trends might reflect how interaction designers with different training, expertise, or backgrounds might have different needs for authoring human-robot interactions; those with a development background might be empowered by the ability to more precisely control the robot, while those with a design focus might want more flexibility in design exploration.

4.3.5 Lessons Learned

The development of the pattern language and the authoring environment required decisions that might prove valuable for future work in this or related areas. One key challenge was selecting a diverse set of social interaction scenarios for my formative study of human interactions. While there are other roles that robots will likely fulfill, such as coaching, I needed to balance the diversity of scenarios with the workload involved in analyzing a large corpus. Another challenge was discovering an appropriate level of abstraction for my models (Figure 4.4) that would enable

their use as design patterns. I followed an iterative process of reviewing the data from human interactions and sketching, constructing, and refining my models until no further modifications could be made.

4.3.6 Limitations

The pattern language that I used in building my authoring environment relied on my observations and analysis of interactions in five scenarios. While I carefully chose these scenarios to represent many of the interactions robots will encounter, additional or more complex scenarios might reveal additional patterns. Furthermore, the evaluation of the use of the pattern language and authoring environment involved primarily student designers and developers in relatively short design sessions, due to the limited volume of interaction design practice for robotic technologies and limited number of human-robot interaction designers. Future explorations might seek to engage professional interaction designers with experience in human-robot interaction design in longer-term design sessions to better understand how the approach presented here might support design exploration and prototyping human-robot interactions. Given the difficulty and overhead involved in programming complex robot systems, design tools such as Interaction Blocks might significantly benefit such users.

4.4 Study Conclusions

With collaborative robots poised to become a ubiquitous presence in human environments, users will need the capability to rapidly prototype and deploy robot

behaviors that are suited to their unique needs and settings, such as developing a robot receptionist in a doctor's office or deploying new behaviors for a manufacturing robot, as outlined in Section 1.1.2.1. In this work, I explored how a design pattern language and visual authoring environment might enable designers to rapidly perform design exploration and prototyping for human-robot interaction. I observed and analyzed interactions from eight dyads engaged in five scenarios and developed seven interaction design patterns. I then built an authoring environment, *Interaction Blocks*, to enable interaction designers to use these patterns to rapidly construct, evaluate, and refine human-robot interaction. I conducted a qualitative evaluation of the use of the pattern language and authoring environment with a group of ten interaction designers and developers, as they prototyped an exchange between a robot and its user. My results highlight the potential for the use of design patterns and the workflow that my authoring environment promotes to design, prototype, and evaluate interactions, enabling interaction designers to take advantage of patterns to synthesize complex interactions.

5 BEHAVIORAL CUE STUDY: INSTRUCTION AND REPAIR

As part of their role as task partners, collaborative robots are expected to also fulfill an instructional role in helping train humans they might be working with, such as teaching medical students surgical procedures or a robot instructing human workers on an assembly task as described in Section 1.1.2.3. To be effective instructors, these robots will need to be cognizant of how to present task instructions such that task outcomes are maximized. For example, how many instructions should a robot give at once? Does the answer change based on the task, or how well the student is performing the task? If the student makes a mistake, how should the robot correct it?

In this work, I seek to design a strategy for how collaborative robots can best give instructions to their human partners, as well as correct any mistakes their human partner might make. Section 5.1 will begin by describing the collection of data on human instructions and building a model of instructional strategies from that data. The resulting model was then implemented on an autonomous robot, as described in Section 5.2. Section 5.3 describes an experimental laboratory evaluation using this autonomous robot in a pipe-building scenario, where the robot teaches a human how to connect a series of pipes. Finally, Section 5.4 discusses implications of this work for future iterations of collaborative robots.

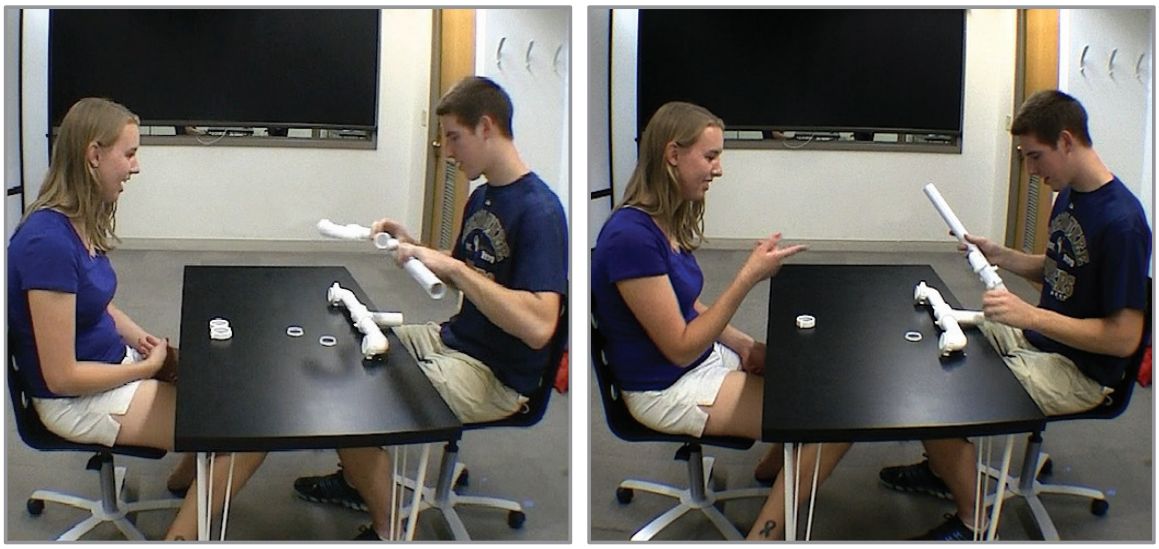

Figure 5.1: The *instructor* (participant on the left) directing the *student* (participant on the right) in assembling a predetermined pipe configuration.

5.1 Formative Study

To better understand human teaching strategies, I collected video data of human-human interactions during an instructional pipe-assembly task that resembled assembly tasks which robots might guide humans in, such as furniture assembly. Below, I discuss my data collection process, analysis, and the instruction models I constructed from the data.

5.1.1 Data Collection

I collected video data from eight instructor-trainee dyads during a pipe-assembly task. In each of these interactions, one participant (the instructor) first learned how to connect a set of pipes into a particular formation from a pre-recorded video.

Instructors were given as much time as necessary to re-watch the video and were provided use of the pipes during training. Upon learning the instructions, the instructor trained the second participant (the trainee) on how to correctly assemble the pipes without the aid of the video (Figure 5.1).

Eight males and eight females aged 18 to 44 ($\mathsf{M} = 23.75$, $\mathsf{SD} = 8.56$) were recruited from the local community. Each interaction was recorded by a video camera equipped with a wide-angle lens to capture the participants and the task space. The instructional portion of the task, excluding the time the first participant spent learning how to construct the pipes, ranged from 3:57 to 6:44 minutes ($\mathsf{M} = 5:11$, $\mathsf{SD} = 2:19$).

5.1.2 Analysis

The videos were analyzed and coded for significant events, including how many instructions the instructor gave during a single turn, whether the instructor summarized subsequent instructions, and how repair was initiated and given. To ensure reliability of all coded data, a second experimenter coded the videos. The inter-rater reliability showed a substantial level of agreement between the primary and secondary coders (79% agreement, Cohen's $\kappa = .74$) (Landis and Koch, 1977).

The analysis of my data helped us to better understand different strategies instructors use to deliver instructions and confirmed examples for my understanding of repair gained from the literature. In my data, I observed instructors organizing their instructions along two major factors: how many instructions they gave at once, and whether they gave a high-level summary of what the next few instructions

would accomplish. I coded my videos with these two factors. My analysis showed that, considering all instructions given across all dyads, 72% of instructions involved descriptions of individual steps, while 28% were grouped with one or more other instructions. Twenty-one percent of all instructions were prefaced with a summary of the following instructions, with the remaining 79% of instructions not including a summary.

My analysis also showed that participants always initiated the repair process verbally, regardless of whether they became aware of a breakdown verbally, such as a question being asked, or visually, such as noticing that the task space was not configured correctly. Here, I differentiate between *trainee-initiated* and *instructor-initiated* repair. My analysis of the data showed that 65% of repairs were trainee-initiated, while 35% of repairs were instructor-initiated.

Trainee-initiated repair—also called *requests*—always used a verbal statement to clarify or confirm instructor expectations when the trainee either did not understand or misunderstood an instruction. Verbal requests for repair ranged from generic statements (e.g., "What?") to more detailed requests for repair (e.g., "Where should the pipe go?"). I categorized each statement into one of three categories: *confusion*, *confirmation*, and *clarification*. These categories are consistent with previous work that categorizes confusion as not understanding and clarification as misunderstanding (Gonsior et al., 2010; Hirst et al., 1994; Koulouri and Lauria, 2009).

Where trainee-initiated repair was directed towards better understanding expectations, instructor-initiated repair clarified or corrected the trainee's perceptions of the task. Instructors initiated repair under one of two circumstances: *mistake detection*

and *hesitancy*. When instructors noticed the trainee performing an action that the instructor knew not to be consistent with the goals of that instruction, such as picking up the wrong piece, they verbally corrected the trainee. When instructors noticed that the trainee was hesitating to take action, which was indicated by an average delay of 9.84 seconds in following an instruction, they asked if the trainee needed help.

5.1.3 Models

My analysis informed the development of a model with two components: *instructional strategies* and *repair*.

5.1.3.1 Instructional Strategies

As noted in my analysis, instructor strategies for organizing instructions involved two factors: grouping and summarization. In *instruction grouping*, instructors vary the number of instructions given from $1 \ldots \mathtt{i}$ before the student completes the instructions. Instructors may provide one instruction at a time and allow the student to carry out the instruction before providing the next instruction or offer *grouped* instructions by conveying $\mathtt{i}$ instructions, given that $\mathtt{i} > 1$, prior to the student fulfilling the instructions. When instructors provide *instruction summarization*, they preface their instructions with a high-level summary of the goal of the subsequent $\mathtt{k}$ instructions. For example, when the next four steps will result in a set of pipes forming a U-shape, the instructor may say "Now, we'll be taking a few pipes and connecting them into a U-shape" prior to giving the first step. The integration of summarization and

grouping into the instruction process is outlined below:

```
current ← x
bool summarize?
bool grouping?
if summarize? then
    summarize(current, current + k)
end if
for y ← 0; y < i do
    instruction(current)
   if !grouping? then
    action(current)
   end if
    y ← y + 1
    current ← current + 1
end for
if grouping? then
   for z ← 0; z < i do
    action(z)
    z ← z + 1
   end for
end if
```

While I categorized instructional strategies into the grouping and summarization factors, my analysis demonstrated that all four possible combinations of these factors were exhibited, as illustrated in Figure 5.2.

Instruction Summarization	Instruction Grouping: *Not grouped*	Instruction Grouping: *Grouped*
Not summarized	**Instructor:** Now take this *[points toward pipe]* and just attach it like that *[makes connecting motion]* <student acts>. Then take this one *[points toward joint]* and put it here. <student acts>	**Instructor:** You'll now connect these two and then connect them to this piece *[points toward piece]* so they'll be pointing straight up. <student acts>
Summarized	**Instructor:** So you're going to use these two to connect them in and form a U-shape. So take one of these *[points toward pipe]* <student acts>, and then one of those *[points toward washer]* <student acts>, and you'll want the skinny side facing out. <student acts>	**Instructor:** OK and you want to start with one arm. So the arms are going to screw onto the smooth side, so they'll go onto the top of the t-piece. So you're going to want to take a washer first, and you'll want to put the fat side towards the curve of the washer and then put the washer on top of that, and then put the t-piece there. <student acts>

Figure 5.2: Examples of how the two factors found in my modeling, *instruction grouping* and *instruction summarization*, can be jointly used.

5.1.3.2 Repair

Regardless of the instructional strategy utilized, instructors frequently engage in repair. Below, I discuss the three different forms of repair that I observed in my data—*requests*, *hesitancy*, and *mistake detection*–and present model components for determining whether repair is needed and, if so, how it might be performed.

Requests: All verbal requests from the student—regardless of the type of question or statement—were considered requests for repair. I categorized the types of utterances using semantic language modeling to classify the type of question or statement, allowing the model to determine the appropriate behavior based on the type of utterance. For example, the questions "Which piece do I need?" and "What piece

should I get?" were recognized as the same question.

Hesitancy: Hesitancy in performing instructions can be determined using a number of measures, such as time elapsed since the last interaction or time elapsed since the workspace was last changed. Modeling of hesitancy might be dependent on task. For the pipe-assembly task, I chose to use the time elapsed since the workspace last changed, which provided conservative estimates of hesitancy-based breakdown, as using time elapsed since the last interaction could result in incorrectly concluding that the student is hesitant while he or she is still working. I considered 10 seconds of no change to the workspace as a hesitancy-based breakdown, based on how long instructors in my human-human study waited before determining a breakdown in interaction had occurred.

Mistake Detection: While requests and hesitancy-based breakdowns are triggered by a student's action or inaction, mistake detection requires checking the student's work. In my proposed model, I chose a simulation-theoretic approach to direct the robot's behavior in relation to the participant. This is a common approach for modeling human behavior, as it posits that humans represent the mental states of others by adopting their partner's perspective to better understand the partner's beliefs and goals (Gallese and Goldman, 1998; Gray et al., 2005), and has been used in designing robot behaviors and control architectures, allowing robots to consider their human partner's perspective (Bicho et al., 2011; Nicolescu and Mataric, 2003). In the context of an instructional task, the instructor has a mental model of an action that they wish to convey to the trainee. Following instruction, the instructor can assess gaps in the trainee's understanding or performance by comparing the trainee's

actions to their mental model of the intended action and noting the differences that occur.

Following the simulation-theoretic approach, I defined a set of instruction goals $P = \{p_1, \ldots, p_n\}$ for the robot regarding the result of the participant's action or inaction given the current instruction. Depending on the task, P may vary at each step of the instruction, as some instruction goals may no longer be applicable, while others may become applicable. As the participant engages in the task, the robot will evaluate whether the current state of the workspace is identical to the set of instruction goals P^*. If any of the individual task goals p_k do not match p_k^*, then there is a need for repair.

How repair is carried out depends on which task goal p_k has been violated. As I observed in my analysis of the human-human interactions, the instructor repaired only the part of the instruction that was currently incorrect. Additionally, there is an inherent ordering to the set P that is informed by the participant's perception of the task. The participant's ordering of P is informed by *elaboration theory*, which states that people order their instructions based on what they perceive as being the most important, and then revealing lower levels of detail as necessary (Reigeluth et al., 1980). By imposing an ordering of decreasing importance on the set P based on these principles for a given task, I can ensure that each p_k takes precedence over any p_{k+n} for $n > 0$. If multiple p_k are violated, then the task goal with the lowest k is addressed first. An example of this ordering can be seen if a participant has picked up the wrong piece and attached it in the wrong location. The instructor first repairs the type of piece needed and then repairs the location of that piece.

Although I discuss the model for detecting mistakes in terms of task steps and goals, this model can also be extended to understanding and repairing verbal mistakes. For example, if the participant mishears a question and responds in a way that is inconsistent with the answers expected, then repair is needed. The appropriate answers of the intended question can be formalized as $\mathfrak{p}_k$, and any answer that does not fulfill $\mathfrak{p}_k$ can be considered as a cause for repair.

5.2 Implementation

To create an autonomous system that implements my models, I contextualized my task in the same scenario used for modeling human-human interactions. Using my findings from the previous stage, I designed my system to enable the processing of both verbal and visual information to check the participant's workspace and to detect and repair breakdowns.

5.2.1 Hardware

I implemented my model on a Wakamaru humanoid robot (Figure 5.3). My model uses information provided by both video and audio captured at 12 frames per second using a Microsoft Kinect stereo camera and microphone-array sensor. The camera and microphone were suspended three feet above the participant's workspace, as shown in Figure 5.3. This camera setup provided a visible range of the workspace of 43 inches by 24 inches. A second stereo camera was placed behind the robot to track the participant's body and face.

5.2.2 Architecture

The architecture for my model involved four modules: *vision*, *listening*, *dialogue*, and *control*. The vision and listening modules capture and process their respective input channels. The control module uses input from these modules to decide about the need for repair and relays the status of the workspace to the dialogue module if feedback from the robot is needed.

The pipe-assembly task used in my implementation involves multiple copies of five types of pieces: three types of pipes (short, medium, and long) and two types of joints (elbow and t-joints). All pieces were marked with augmented reality (AR) tags to allow detection by the workspace camera. The orientation of each tag was used to identify object type, location, and rotation. The location and orientation of tags on pipes and joints were consistent across each type of object, and tag locations on each object were known to the system.

5.2.2.1 Vision Module

The vision module was designed to achieve two goals: to detect the status of the participant's workspace and to process information on the participant's location. Sensing necessary for achieving each of these goals is managed by a separate camera.

At each frame, the vision module processes the frame to discover which pipes are connected, creating a graph of pipe connections, C. There are three main instructions to building C: finding the AR-tag glyphs in the frame, associating those glyphs with pieces, and detecting which pieces are connected based on a set of heuristics. The description of these instructions are omitted due to space limitations.

At the completion of the participant's turn, C is checked against the correct workspace configuration, C^*. If the two graphs are isomorphic—identical in structure—then the participant has successfully completed the instruction. If the graphs are not isomorphic, then the robot will discover an inconsistency between the two graphs during the isomorphism check. The lowest p_k^* which is violated is then passed to the control module. In those cases where the system needs to check multiple instructions at once, the graph C is built incrementally by systematically eliminating possibly extraneous pieces and then comparing against C^*.

Each frame is searched for AR glyphs using a modified version of Gratf [1] to create a set of glyphs G, where each glyph in G is defined by its type t, its position (x, y), and its rotation θ. Upon discovering a glyph, the algorithm checks to see if there are any pieces of the appropriate type (e.g., long pipe, t-joint) that are missing that particular glyph. The glyph is associated with a piece if the algorithm matches the glyph to the piece based on its proximity and rotation properties. If no piece is found, a new piece is created, and the glyph is associated with the new piece. A set of pieces P results, where each piece p is described by a set of glyphs that are associated with that piece. All of the glyphs for a piece p form a bounding box that gives a rough estimate of the physical boundaries of that piece. Using these coordinate, I can confirm whether any two pieces are connected; then, I can build a graph structure that reflects the workspace.

At the completion of each participant's turn, the correct graph structure G^* is compared against the structure G of the workspace. If the two graphs are isomorphic,

[1] Gratf: http://www.aforgenet.com/projects/gratf/

then the participant has successfully completed the instruction. The robot will discover an inconsistency between p_k and p_k^* during the isomorphism check if the graphs are not isomorphic. The lowest p_k^* which is violated is then passed to the control module.

If the system needs to check multiple instructions simultaneously, the set of pipe connections C is built incrementally, starting with the first instruction that needs to be checked. Since each instruction involves the addition of a new piece to a specific location and with a particular rotation, checking the workspace for the first instruction s_1 will result in the detection of too many pieces, since pieces for instructions through s_n are also on the table. In this case, the module is responsible for systematically eliminating extraneous pieces from C. A piece is defined as extraneous if its removal does not result in a disjoint graph in C and does not reduce the count of that particular piece below what is needed to complete the instruction. Once a modified version of C resulting in a correct check of s_1 is found, pieces are added incrementally back to C such that they maintain connectivity between all pieces in C and maintain a set P that is equivalent to the number of each type of piece needed to complete the instruction s_m.

The second goal of the vision module—detecting the participant's location—is checked at every frame. When the participant is within 1 ft. of the workspace, the robot repositions its head so that it is gazing at the table, monitoring the workspace. When the participant is further away (e.g., standing back to check their work, retrieving the piece), the robot raises its head and gazes toward the participant's face. However, if the participant or the robot is talking, or if the robot is checking

the workspace in response to a prompt from the user, the robot looks toward the participant or where on the workspace changes have been made, respectively.

The second goal of the vision module—detecting the participant's location—is checked at every frame. When the participant is within 1 ft. of the workspace, the robot repositions its head so that it is gazing at the table, monitoring the workspace. When the participant is further away (e.g., standing back to check their work, retrieving the piece), the robot raises its head and gazes toward the participant's face. However, if the participant or the robot is talking, or if the robot is checking the workspace in response to a prompt from the user, the robot looks toward the participant or where on the workspace changes have been made, respectively.

5.2.2.2 Listening Module

The listening module is used to detect and categorize requests from the participant. The robot uses the capabilities of the Microsoft Kinect speech recognition hardware and API to categorize the participant's speech acts. A grammar of possible sentences was generated using examples from the data collected in my modeling study. Possible speech acts are categorized as one of the following:

- *Request for repetition*: (e.g., "What did you say?" "Can you repeat the instructions?")
- *Check for correctness*: (e.g., "Is this the right piece?" "I'm done attaching the pipe.")
- *Check for options*: (e.g., "Which pipe do I need?" "Where does it go?")

Utterances that did not belong to one of these categories, such as confirmation of an instruction, were ignored by the system.

I use a dialogue manager to coordinate responses to each type of query. Each recognized utterance has an associated semantic meaning that indicates the purpose of the utterance. For example, the phrase "What did you say?" is assigned the semantic meaning of "recognition request." These semantic meanings allow the control module to understand the type of utterance processed and to reply to the utterance appropriately given the current state of the participant's workspace. To process requests that refer to the workspace, the system first checks the state of the workspace through the vision module. For example, asking "Did I do this right?" requires the robot to determine whether the current workspace is correct.

5.2.2.3 Control Module

Decisions on the robot's next action are determined by the control module. It uses input from the vision and dialogue modules and, following a simulation-theoretic approach, makes decisions by comparing this input to actions that the robot expects in response to its instructions. According to my model, I define a set P that describes which possible expectations can be violated by the participant. Consistent with elaboration theory, ordering of task expectations are based on observations from my study of human instructor-trainee interactions, which resulted in the following categories:

- *Timely Action* (p_0): The participant acted in a timely fashion.
- *Correct Piece* (p_1): The participant used the correct piece.
- *Correct Placement* (p_2): The participant placed the piece in the correct location relative to the current workspace.

– *Correct Rotation* (p_3): The participant rotated the piece correctly relative to the current workspace.

The first expectation ensures that the participant does not hesitate for too long, which might indicate confusion, when adding the next piece. Based on my previous analysis, I defined hesitancy as the workspace remaining unchanged for more than 10 seconds since the robot's last instruction. The last three expectations ensure that the participant chose the correct piece to add, added the piece in the correct location, and rotated the piece correctly.

5.2.2.4 Dialogue Module

The control module passes three pieces of information to the dialogue module after evaluating input from the vision and listening modules: current instruction, the semantics associated with the speaker's last utterance (if any), and the result of the control module's evaluation of the workspace (if any).

Given this information, the dialogue module initiates the appropriate verbal response. Responses depend on which instruction of the task the participant is completing, the current layout of the workspace, and the type of question the participant asked. Not all responses are dependent on all three pieces of information; requests for repetition of the last instruction are independent of how the workspace is currently configured, and responses to hesitancy are independent of the current workspace and interaction with the participant. However, a request to check if the participant has correctly completed an instruction requires knowledge of both the instructions completed and the current layout of the workspace.

5.3 Experimental Evaluation

To evaluate the instructional strategies that I identified from my analysis, I conducted a human-robot evaluation that followed the same task setup as the earlier modeling study. Due to a lack of sufficient theory that would predict how these instructional strategies might affect trainee performance and experience, I chose not to pose any hypotheses and conducted my analysis in an exploratory manner.

5.3.1 Experimental Design

In order to better understand the effectiveness and tradeoffs of various teaching strategies, I designed a between-participants-design study to compare four different models of teaching strategies that fell along two factors: the *grouping* factor and the *summary* factor. The grouping factor defines how many instructions are issued during the instructor's turn. For the purposes of my study, the grouping factor has two levels: the *no grouping* level, where a single instruction is given during the round, and the *grouping* level, where a set of two or more instructions are given at once. The summary factor defines whether or not the instructor gives a summary of the objective for the next few instructions. For the purposes of my study, this factor has two levels: the *no summary* level, where the instructor does not give summaries, and the *summary* level, where the instructor does give summaries. Participants from my study of human instructor-trainee pairs exhibited all four combinations of these two factors, resulting in four conditions for my study: (1) *no grouping, no summarization*, (2) *grouping, no summarization*, (3) *no grouping, summarization*, and (4) *grouping,*

Figure 5.3: On the right, the setup used in my experimental evaluation. After the robot gave an instruction, the participant retrieved the pieces necessary from behind them and assemble the pieces on the workspace in front of the robot. A camera above the workspace captured the configuration of the pieces. On the left, an example of the robot autonomously guiding a participant in assembling pipes.

summarization.

The architecture detailed in the previous section was used in all conditions. Differences between conditions were controlled in the control module that managed decisions on how to structure instructions. Additionally, the dialogue module responded to requests in the grouping level that did not exist in the no grouping level (e.g., repeating multiple instructions).

5.3.1.1 Task

All participants were autonomously guided through assembling a set of pipes by the robot in the setup shown in Figure 5.3. Participants were given two bins—one for pipes and one for joints—that contained only the pieces necessary for completing the task, mimicking the setup in which different types of parts might be kept at a workshop. Following an introduction, the robot directed the participant in the assembly task by

issuing instructions according to the condition to which the participant was assigned, varying the number of instructions provided and whether or no high-level summaries of future instructions were provided. The robot also provided repair as necessary. Following completion of the task, the robot thanked the participant. Completing the task took between 3:57 and 9:20 minutes ($\mathsf{M} = 6:44$, $\mathsf{SD} = 1:23$).

In the *no grouping, no summarization* and *no grouping, summarization* conditions, the robot provided one instruction at a time, while the grouping condition involved two to four instructions at a time. Additionally, in the *no grouping, summarization* and the *grouping, summarization* conditions, the robot provided a high-level summary of the next few steps prior to giving instructions, while it provided no summary in the other conditions. Following instructions, the participant retrieved the pieces to complete the steps and assembled the pieces on the table. If the participant requested repetition or clarification, the robot answered. When the participant asked the robot to check the workspace, it confirmed correct actions or provided repair according to my model. If no repair was needed, it congratulated the participant on completing the task and proceeded to the next instruction or set of instructions.

Participants started the study standing three feet away from the robot, with a two foot long table between them. A second table was placed five feet behind where the participant started. A single video camera captured the entire interaction for additional data analysis.

5.3.2 Procedure

Following informed consent, participants were guided into the experiment room. The experimenter explained the task and introduced the participant to the pieces used in the task. After the experimenter exited the room, the robot started the interaction by explaining that it would provide step-by-step instructions for assembling the pipes. The robot then provided instructions until the participant completed the entire structure. At the end of the task, the robot thanked the participant. The participant then completed a questionnaire and received $5. A total of 32 native English speakers between the ages of 18 and 34 ($M = 23$, $SD = 4.9$) were recruited from the local community. These participants had backgrounds in a range of occupations and majors. All conditions were gender balanced.

5.3.3 Measures

I used two objective measures to evaluate the participant's performance in the task: total number of breakdowns and total task time. Total number of breakdowns was defined as the number of times the participant made a mistake in fulfilling the instruction that the robot provided or asked for repetition or clarification of the instruction. I also measured task completion time, since a lower number of repairs would likely indicate a faster task time. All trials of the study were videotaped, and these measures were extracted with video coding. To ensure reliability of the measures, a second experimenter coded for repairs. The inter-rater reliability showed substantial agreement (87% agreement, Cohen's $\kappa = .83$) (Landis and Koch, 1977).

I also used subjective measures that collected data on the participant's impressions

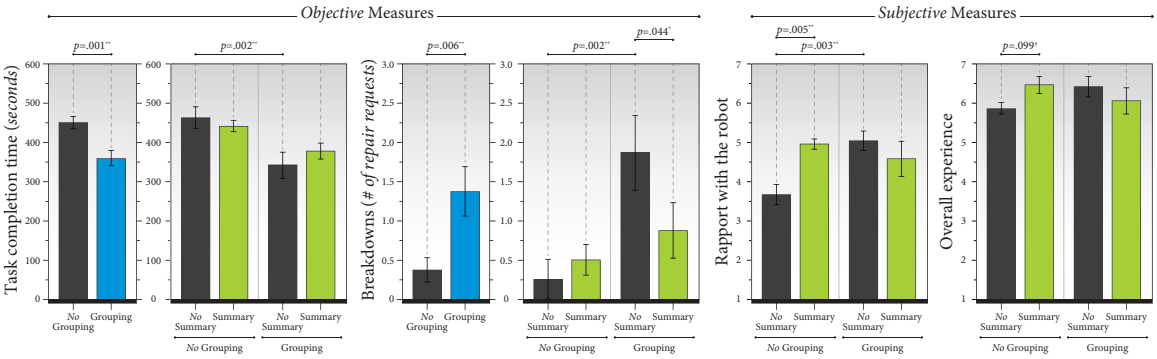

Figure 5.4: Results from my evaluation. Significant and marginal results were found for total task time, number of breakdowns encountered, participants' perceived rapport with the robot, and their overall experience with the task. (†), (*), and (**) denote $p < .10$, $p < .050$, and $p < .010$, respectively.

of the robot, including likability, naturalness, and competency, the participant's experience with the task, and their rapport with the robot. Participants rated each item in my scales using a seven-point rating scale. A confirmatory factor analysis showed high reliability for all scales, including the likability of the robot (10 items, Cronbach's $\alpha = .846$), the naturalness of the robot (6 items, Cronbach's $\alpha = .842$), the competency of the robot (8 items, Cronbach's $\alpha = .896$), the participant's experience during the task (8 items, Cronbach's $\alpha = .886$), and the rapport between the participant and the robot (6 items, Cronbach's $\alpha = .809$).

My analysis of data from these measures involved a two-way fixed-effects analysis of variance (ANOVA), including grouping, summary, and the interaction between them as fixed-effect factors. Contrast tests used Scheffé's method.

5.3.4 Results

I primarily report marginal and significant effects of the instructional strategies used by the robot on objective and subjective measures and summarize them in Figure 5.4.

I used two objective measures to evaluate the effectiveness of the robot's use of instructional strategies: the total number of breakdowns that occurred during the task and the total time taken to complete the task. The analysis of this data showed that grouping instructions significantly reduced task completion time, $\mathsf{F}(1, 28) = 13.35$, $\mathsf{p} = .001$, while significantly increasing the number of breakdowns, $\mathsf{F}(1, 28) = 8.87$, $\mathsf{p} = .006$. Summarization had no overall effect on task time, $\mathsf{F}(1, 28) = 0.07$, $\mathsf{p} = .793$, or the number of breakdowns, $\mathsf{F}(1, 28) = 1.25$, $\mathsf{p} = .274$. The analysis also showed a marginal interaction effect between grouping and summarization over the number of breakdowns, $\mathsf{F}(1, 28) = 3.47$, $\mathsf{p} = .073$, but no interaction effects were found over total task time, $\mathsf{F}(1, 28) = 1.29$, $\mathsf{p} = .266$. Contrast tests across conditions showed that, when the robot did not provide a summary, grouping instructions significantly reduced task completion time, $\mathsf{F}(1, 28) = 11.47$, $\mathsf{p} = .002$, but resulted in a significant increase in the number of breakdowns, $\mathsf{F}(1, 28) = 11.71$, $\mathsf{p} = .002$. This increase was alleviated by summarization; participants encountered significantly fewer breakdowns when the robot also provided a summary along with grouped instructions, $\mathsf{F}(1, 28) = 4.44$, $\mathsf{p} = .044$.

The subjective measures included scales for capturing the participants' perceptions of the robot including, likability, naturalness, and competency, their rapport with the robot, and their overall experience with the task. The analysis showed an interaction effect between grouping and summarization over the participants' rapport with the

robot, $F(1, 28) = 8.76$, $p = .006$. When the robot provided no summary, grouping instructions improved participant rapport with the robot, $F(1, 28) = 10.81$, $p = .003$. When the instructions were not grouped, summarization also improved rapport with the robot, $F(1, 28) = 9.54$, $p = .005$. Additionally, the analysis showed a marginal interaction effect between grouping and summarization over participants' ratings of their overall experience with the task, $F(1, 28) = 3.68$, $p = .065$. Contrast tests showed that, when the robot did not group its instructions, summarization marginally improved participants' overall task experience, $F(1, 28) = 2.91$, $p = .099$.

5.4 Discussion

The data from my objective and subjective results provided a number of findings to guide the design of collaborative robots in their role as instructions, the implications of which I highlight below.

My objective results showed that grouping instructions resulted in a tradeoff between task completion time and the number of breakdowns that the participants encountered. I found that participants completed the task significantly faster when the robot grouped its instructions than when the robot provided instructions one-by-one. I observed that when participants received multiple instructions, they retrieved all parts necessary to complete these instructions from the bins at once, proceeded with assembling multiple pieces in a sequence, and sought confirmation of the correctness of the whole sequence from the robot, completing the overall assembly significantly faster. When participants received instructions one-by-one, they instead retrieved pieces

one-by-one and proceeded to the next instruction only when the robot confirmed the successful completion of an assembly, which resulted in overall longer task completion times. Contrary to the improvement in task completion times, participants encountered significantly more breakdowns when the robot grouped its instructions than when the robot provided individual instructions. I speculate that grouped instructions required participants to retain a greater amount of information, which might have impaired their understanding or recall of the instructions, resulting in mistakes in the assembly that had to be repaired by the robot.

Further analysis into the breakdowns occurred with grouped instructions showed that 60% of breakdowns occurred in the first set of instructions, which contained four instructions, 25% occurred in the second, third, and fifth set of instructions, which all contained three instructions, and 15% occurred in the fourth set of instructions, which contained two instructions. This distribution of breakdowns indicates an increase in the number of breakdowns as the number of grouped instructions increases, which might indicate a greater cognitive load placed on the participant by the introduction of more pieces (Sweller, 1988). Additionally, participants may have demonstrated selective attention when the robot provided grouped instructions, causing them to miss information (Sweller, 1988). This explanation is supported by the effects of summarization on the number of breakdowns; providing a summary of subsequent steps significantly reduced the number of breakdowns that the participants encountered in carrying out grouped instructions. The summary provided by the robot might have consolidated the participants' understanding of the grouped instructions. However, I also note that some of the breakdowns that occurred early in the interaction

might have been caused by the participant acclimating to the task and interaction with the robot, or the task involving a greater variety of pieces to choose from at the beginning.

My analysis of the subjective results showed interaction effects between grouping and summarizing over participant rapport with the robot and their overall experience with the task. I found that participants reported higher rapport with the robot when it grouped instructions with no summary than when the robot used neither grouping nor summarization. This improvement might be due to the quicker, less monotonous experience that the robot offered when it delivered instructions all at once and spent no time on summarizing them. The results also showed that participants reported higher rapport with the robot and overall experience with the task when the robot provided a summary of subsequent steps along with individual instructions than when it neither grouped its instructions nor provided a summary. I speculate that, when the robot provided a summary of what is ahead in the task, as a summary involved information on upcoming steps in the instruction, participants might have perceived the robot as more invested in the instruction and felt more informed, although this information did not improve task performance.

5.4.1 Design Implications

These results have a number of implications for implementing instructional capabilities on collaborative robots. My results suggest that, despite resulting in more mistakes, grouping significantly improves task completion times, making it ideal for interaction scenarios in which faster task times are critical and mistakes are not costly. Further-

more, coupling summarization with grouping alleviates some of the mistakes caused by providing students and trainees with a large number of instructions. However, there are many scenarios where providing instructions one-by-one might be preferable. For example, with more complex tasks or students who might have trouble keeping up with the robot's instructions (e.g., novices), providing instructions one-by-one might help the student complete the task with fewer breakdowns. Additionally, in situations where making a mistake could be dangerous or costly, giving instructions individually would reduce the chance of these mistakes occurring. In these scenarios, including summaries of upcoming instructions might also improve student experience and result in an improved rapport with the robot.

5.4.2 Limitations

The work presented here has two key limitations. First, although my model considers two structural components of instruction-giving, there may be other structural components I did not observe in human interactions and thus did not include in my model. An analysis of human interactions in a more diverse set of instructional scenarios might enable the development of richer models of instruction. Second, while my repair model offered repair when prompted, the system did not proactively offer repair due to the difficulty of accurately discerning when mistakes occurred. The structure of the task and available methods for perception made it difficult to continuously update a model of the workspace and determine whether it was being modified, as participants obstructed the camera's view when modifications were occurring. Third, my evaluation focused on testing only the immediate effects of the

proposed instructional strategies on student performance and perceptions, and did not explore how these strategies might be used for instruction that takes place over time or how much students retain the instruction in the long term.

5.5 Study Conclusions

As robots are introduced as working with human partners, they have the potential to serve as instructional guides, such as teaching in labs, giving patients directions in a doctor's office, or guiding co-workers through a task, as described in Section 1.1.2.3. To enable robots to serve as instructors, we need to enable them to provide instructions effectively. In this work, I describe two key instructional strategies—grouping and summarization—for organizing instructions based on observations of human interactions in an instructional pipe-assembly task. I implemented these strategies on a humanlike robot that autonomously guided the participant through connecting a set of pipes. I evaluated the effectiveness of these strategies for student task performance and experience in a human-robot evaluation. My results showed that, when the robot grouped instructions, participants completed the task faster but encountered more breakdowns. These breakdowns were alleviated by summarizing the grouped instructions. I also found that summarizing instructions increased participant rapport with the robot and overall experience with the task. My findings show that grouping instructions results in a tradeoff between task time and breakdowns, and that summarization has benefits under certain conditions, suggesting that robots selectively use these strategies based on the goals of the instruction.

6 BEHAVIORAL CUE STUDY: DEICTIC GESTURES

This study examined how robots might use *deictic gestures*—gestures that direct attention to collocated objects, persons, or spaces—that include pointing, touching, and exhibiting to help their listeners understand their references, such as the context described in Section 1.1.2.2 where the robot physically touched a specific area to indicate its importance to a human worker. In particular, as robots collaborate with humans in increasingly diverse environments, they will need to effectively refer to objects of joint interest and adapt their references to various physical, environmental, and task conditions (Lozano and Tversky, 2006) and use gestures to augment or replace complex verbal descriptions (Goldin-Meadow, 1999; McNeil, 1992). In this work, I provide a curated set of human deictic gestures, drawn from a range of prior literature, that help to clarify the variety of deictic gestures humans might use. Results from this work also provide insights as to the properties of effective gestures, and recommendations as to which gestures fare best in different environmental settings.

Section 6.1 will begin by providing an overview of the six types of human deictic gestures and the six types of environmental settings considered in this work. Translation and implementation of these gestures onto a robotic platform is then described in Section 6.2. Section 6.3 describes an evaluation of these gestures through a human observing and scoring the robot's gestures in the different environmental settings. Finally, Section 6.4 will discuss the implications and limitations of this work.

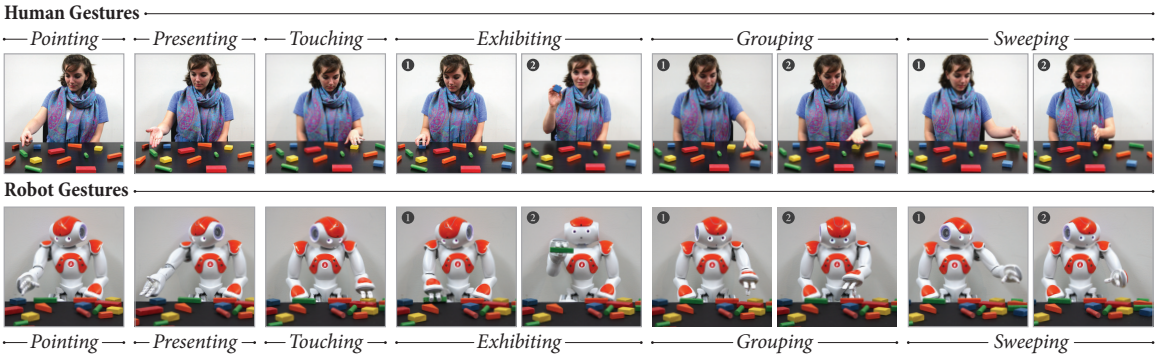

Figure 6.1: Instances of a human performer and the NAO robot demonstrating the deictic gestures studied in this work.

6.1 Understanding Deictic Gestures and Settings

Understanding both the types of deictic gestures available to robots and the settings in which they may be appropriate is necessary for a comprehensive study of the interplay between environmental setting, communicative goals, and gesture choice. In this section, I describe six gestures and six environmental settings that robots are envisioned to encounter and that might impact gesture effectiveness.

6.1.1 Deictics

There are describe six deictic gestures that I focus on in this work: pointing, presenting, touching, exhibiting, sweeping, and grouping. These gestures combine results from prior work to inform our understanding of human deictic gestures. Examples of each gesture can be seen in Figure 6.1.

Pointing – Pointing is often considered the prototypical deictic gesture, being universally understood across cultures (Kita, 2003), ages (Murphy, 1978), and even

species (Miklósi and Soproni, 2006). A pointing gesture uses an extended index finger with the hand rotated so that the palm faces toward or perpendicular to the ground to direct attention. The hand does not come into physical contact with the referent. A pointing referent may be a single object, a region of space, or no specific object or region (McNeil, 1992). Prior work has already explored implementing human pointing gestures on robots, revealing that pointing gestures which point away from the body, rather than across the body, are more accurate at communicating the referent (St Clair et al., 2011).

Presenting – Presenting uses a similar style to pointing in that the speaker gestures toward the referent without coming into contact. However, where as pointing leaves only the index finger extended, presenting extends all fingers and points the palm of the hand upwards. This gesture is often interpreted as inviting, encouraging the listener to, for example, pick up the referent (Bolinger, 1983; Kendon, 2004). Presenting gestures are also used by speakers to indicate that they are ready to receive an object that they previously requested (Bolinger, 1983).

Touching – Touching is used in similar settings as pointing and presenting; however, touching removes ambiguity in that the speaker's hand comes into direct physical contact with the referent. This absence of ambiguity makes touching ideal in situations where verbal communication is absent or impaired. For example, touching seems to be a preferred deictic gesture for mothers communicating with non-verbal infants, since verbal capabilities are required for understanding pointing gestures (Lempers, 1979). Touching may also be a preferred deictic gesture in a factory or on a noisy shop floor. Touching may also be used to refer to an object in situations where constructing an

accurate verbal description to augment a pointing gesture is difficult. For example, it may be difficult to verbally differentiate between or provide an accurate pointing gesture for similar objects in close proximity to one another. Additionally, if certain properties of an object cannot be described concisely, touching the object may be more cost-effective than attempting to describe the object.

Exhibiting – Exhibiting is a natural extension of touching where the object is grasped and lifted so that it can be observed by others (Clark, 2005). This gesture might be used when joint attention is obstructed due to referent location, making other gestures unusable. For example, objects concealed by other objects may prevent the listener from seeing the object, requiring an exhibiting gesture.

Grouping – Grouping offers a gesture similar to presenting in that the fingers are extended with the palm facing down. Instead of referring to a single block, however, grouping takes advantage of the larger area covered by the hand to reference those objects located underneath the hand. The speaker may use a circular hand motion—still in the grouping gesture—around the area they wish to indicate in cases where an area instead of an object is the referent. This gesture has also been implemented in interactive tabletop and wall touchscreen displays to highlight a group of objects (Vogel and Balakrishnan, 2004).

Sweeping – Similar to grouping, sweeping references one or more objects in a given area. A speaker utilizing sweeping will place their hand, with fingers extended, perpendicular to and above the surface to indicate a beginning boundary for referenced objects. The gesture then sweeps across additional referenced objects (Alibali et al., 1997).

6.1.2 Settings

While humans employ a variety of deictic gestures to direct attention to an object, each gesture has unique functional properties that might diminish its effectiveness in some settings. As robots start working alongside humans, it is expected that they will encounter similar settings as humans. In this section, I describe six settings that I believe can impact which gestures a robot should choose.

Distance from Referrer – The accuracy of a gesture may diminish when the distance between referrer and referent is larger, as listeners may make greater interpolations regarding where the speaker's hand is gesturing. In extremes, objects are located immediately in front of or substantially far away (e.g., the opposite end of a table) from the referrer. While some robots might be capable of reaching locations that would not be available to humans, there will always be situations where the robot will need to reference a distant referent.

Clustered Objects – Varied amounts of space exist between objects laid out on a table. This can vary from objects clustered together very closely to objects spread far apart. This setting also mimics the possibility of having one versus many objects, with one object effectively obtained by spreading objects far apart.

Noise – Many environments in which robots are expected to work, such as warehouses and assembly lines, can be noisy. Since deictic gestures are often accompanied by speech that can elaborate on the purpose of the gesture, noise might make some gestures more difficult to understand.

No Visibility – Often times, objects to which the robot wishes to draw attention may be in the referrer's line of sight but may not be visible to the listener. For

example, objects may be located in a container or behind a structure or object. In these cases, deictic gesturing may indicate to the listener that some object in the general area of the gesture is located outside their line of sight.

Ambiguity – During assembly tasks, pieces which initially look similar may differ in small ways, such as screws that have slightly different lengths and widths. These pieces may be difficult to differentiate verbally due to these subtle differences. Lack of adequate vocabulary may also hinder verbal differentiation and may also place significant cognitive burden on both the speaker and the listener.

Neutral – Those cases where there may not be any environmental factors affecting communication results in a neutral setting. Here, a diverse set of objects is nearby the referent with ample space between each object and in clear view of all involved parties.

These settings serve as a representative sample of the situations robots are expected to encounter, particularly in joint tasks with humans, making them appropriate contextualizations for better understanding how the affects of gestures change across settings.

6.2 Implementation

I chose to contextualize my implementation in an object referencing task, where the robot would refer to one or more wooden building blocks distributed on a workspace. The use of wooden building blocks in this task was inspired by Shah and Breazeal (2010). In the context of the wooden building block task, I created two workspaces of

blocks to accommodate my six settings and designed each of the gestures in every setting.

6.2.1 Gesture Design

I implemented my behaviors on the NAO robot. The NAO features six degrees of freedom in each arm: shoulder pitch, should roll, elbow yaw, elbow roll, wrist yaw, and finger pitch. The technical capabilities of the NAO enabled us to create accurate reproductions of each gesture in a variety of settings. I implemented the gestures on the NAO through *puppeteering*, a technique in which a designer manually guides the robot in executing a gesture while a motion capture program saves joint positions at each keyframe. These keyframes are later used to generate arm-motion trajectories. I puppeteered each gesture and manually edited the resulting motion profile as necessary in Choregraphe, a behavior authoring environment for the NAO. The gesture profiles were then saved on the robot to be executed by my experiment software.

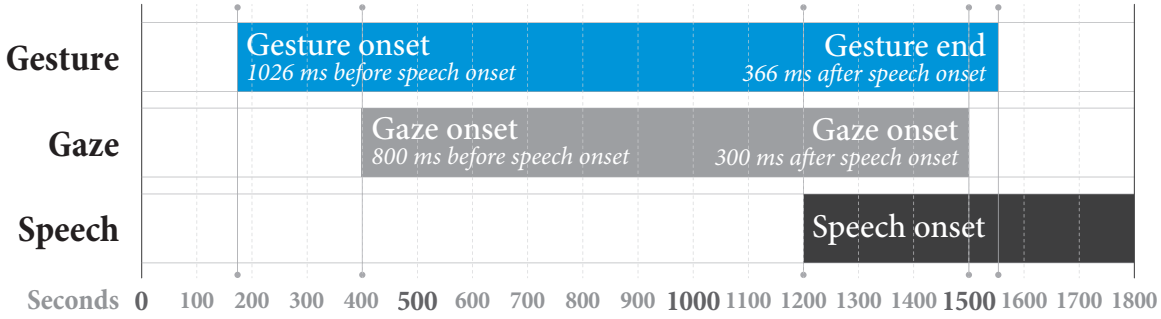

Figure 6.2: A model of the gesture-contingent gaze behavior implemented in my study. Start and end times are relative to the onset of speech.

As gaze is an integral part of a natural gesture, I implemented gesture-contingent gaze behavior as described in Huang and Mutlu (2013) for all of my gestures (see Figure 6.2). The gaze trajectory followed the robot's hand as it gestured to the block for all gestures.

6.2.2 Workspace Design

I designed a layout of wooden blocks for six settings that I divided onto two workspaces, which can be seen in Figure 6.3. Each workspace contained two sets of blocks, with one set on the left half of the workspace, and the second set on the right half. The first workspace displayed the ambiguity and no visibility settings, and the second workspace displayed the neutral, distant from referrer, clustered objects, and noise settings. The following are the descriptions of each setting from the robot's point of view:

- *Neutral*: An assortment of blocks arranged near the referrer. Blocks were spaced 1.5 to 2 in. (3.8 cm to 5.1 cm) from nearby blocks.
- *Distance from Referrer*: An assortment of blocks arranged far from the referent. Blocks were spaced 1.5 to 2 in. (3.8 cm to 5.1 cm) apart, and all blocks were at least 6.5 in. (16.5 cm) from the referrer.
- *Clustered Objects*: An assortment of blocks near the referrer. Blocks were spaced .5 in. (1.3 cm) from nearby blocks.
- *Noise*: Identical to the neutral setting, but with white noise of people talking loudly played from a nearby speaker.

– *No Visibility*: An assortment of blocks placed behind a 3.5 in. (8.9 cm) partition.

– *Ambiguity*: Blocks which were similar in color, length and shape were arranged near the referrer. Blocks were spaced 1.5 to 2 in. (3.8 cm to 5.1 cm) from nearby blocks.

6.3 Experimental Evaluation

To explore the effectiveness of gestures in different settings, I used the workspaces from my wooden blocks task to conduct a within-participants study of all gesture-setting combinations. For each condition, participants identified the blocks they believed the robot was referring to and rated the gesture on a number of items. The results indicate that setting has an impact on gesture.

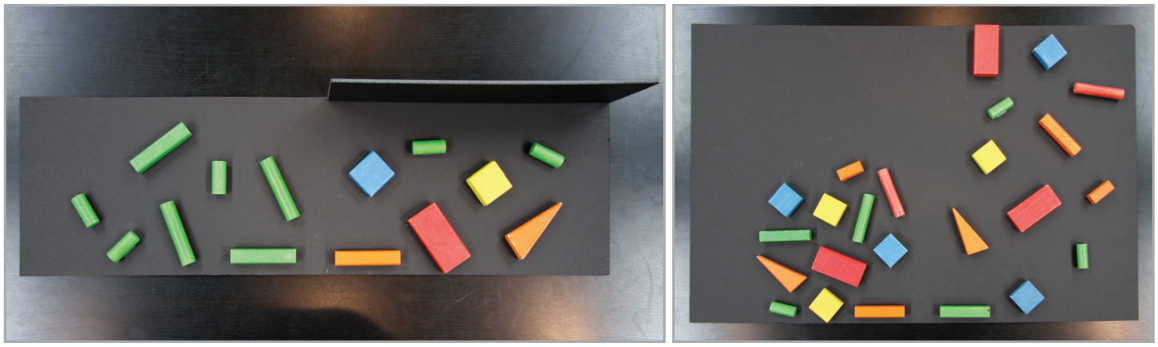

Figure 6.3: The two workspaces used to represent the six settings I explored. The left workspace displays the ambiguity (left) and the no visibility (right) settings. The right workspace displays the clustered objects setting (left), the distant from referrer setting (top right), and the neutral setting (bottom right), which was used for both the neutral and noise settings.

6.3.1 Experimental Design

To better understand the effects of gesture choice and setting on referential communication, I designed a within-participants study to explore every feasible combination of the gesture and setting factors described previously. In addition to the six gestures mentioned, I included two verbal-only baselines in my gesture factor. The first baseline involves the use of only *minimally articulated* verbal references that the robot uses in conjunction with gestures, such as "this block." This baseline follows results from Shah and Breazeal (2010), which showed that participants in a collaborative block building task often used generic referential statements in conjunction with gestures to bring attention to a block. The second baseline involves the use of *fully articulated*, descriptive speech to provide a complete description of the block to which the robot is referring, such as "the short green cylinder closest to you." Whereas the first baseline shows the outcome of eliminating gestures, this second baseline demonstrates the consequences of the robot engaging only in verbal communication. Since I designed gaze cues specifically for the gesture they accompanied, I eliminated gaze from my baselines.

In total, the combination of eight forms of communication (six gestures and two verbal baselines) and six settings resulted in 48 conditions. I eliminated two conditions—touching and exhibiting blocks at a distance—due to the physical impossibility of contact with these blocks, resulting in 46 conditions used in the study.

6.3.1.1 Task

Participants were asked to observe the robot as it referenced blocks situated on a table between the participant and robot. The participant observed 46 rounds of references made by the robot, where each round was one of the 46 conditions previously outlined. Rounds were broken down into two sets to allow for all of the settings to be displayed. The first set consisted of 30 rounds (neutral, distant from referrer, clustered objects, and noise settings) and the second of 16 rounds (no visibility and ambiguity settings). The order in which participants observed the two sets was balanced across participants, while the rounds within each set were randomly ordered. Additionally, to account for possible participant biases between the left and right arms, workspaces were flipped along the vertical axis for half of the participants. All possible presentations of the workspace were gender balanced. After the robot completed the action for a given round, the participant rated the robot's behavior on both objective and subjective scales on a one page questionnaire.

The experimental setup is shown in Figure 6.4. Participants were seated 2.5 feet (76.2 cm) away from the robot, with the workspace between them. The participant's questionnaires were placed between the participant and the workspace. A small speaker was placed next to the robot (out of view of the participant) to emit background noise of people talking during any conditions which involved noise.

6.3.2 Procedure

Following informed consent, participants were seated in the experiment room. The experimenter explained the task to the participant, started the robot, and left the

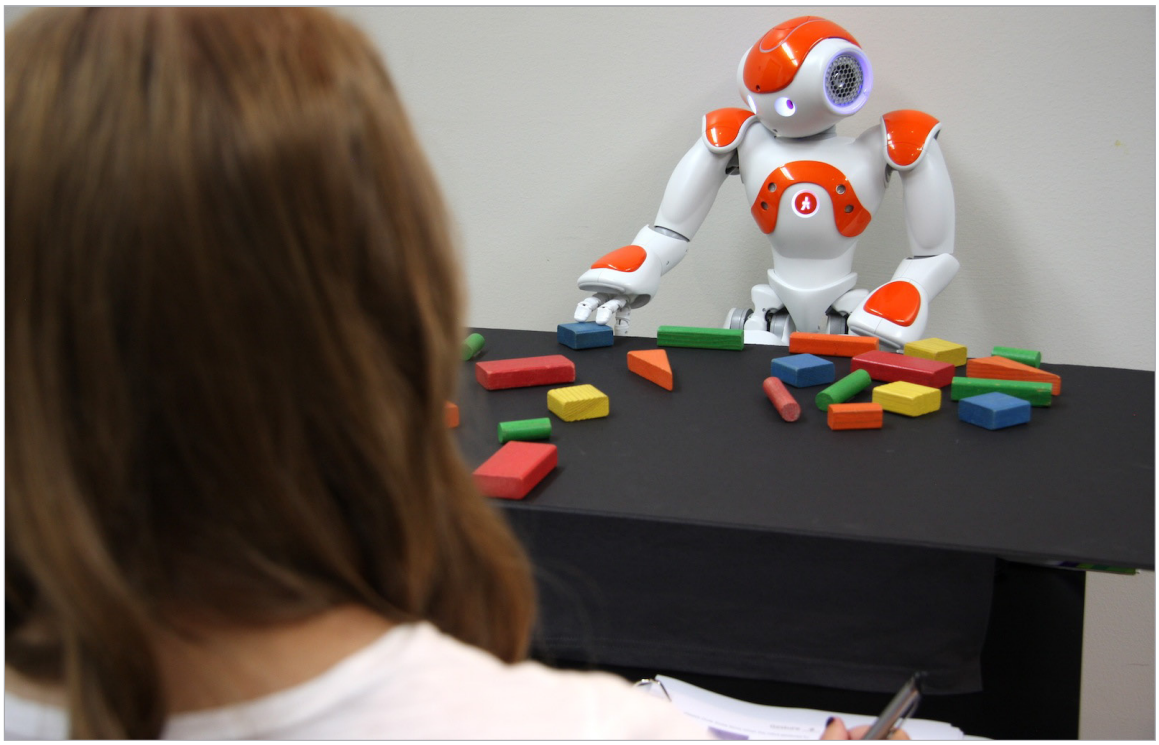

Figure 6.4: A participant evaluating the robot touching the blue block.

room. The robot initiated the interaction by giving an introduction, followed by starting the first round. During each round, the robot would choose one of the gesture and setting combinations to exhibit according to which set of settings was currently available. The robot would then perform the action associated with the particular combination. Upon completion of the robot's action, the participant would complete the questionnaire.

The top half of the questionnaire showed a picture of the workspace currently being used, where the participant would individually circle all blocks they believed the robot had referred to. On the bottom half of the questionnaire were six seven-point

rating items to measure the naturalness of the robot's gesture in that setting. When the participant was satisfied with their answers to the questionnaire, they would say "next" to advance the robot to the next round.

When the first workspace was completed, the experimenter would set up the second workspace and provide new questionnaires that reflected the new workspace layout. After the completion of both workspaces, participants were compensated $5 for their time. Participants took between 15 and 32 minutes to complete the task (M =23 minutes, 40 seconds, SD =3 minutes, 52.8 seconds).

I recruited 24 native English speakers (12 males, 12 females) with diverse majors and occupations and ages that ranged 18–46 (M = 22.7, SD = 5.92) from the University of Wisconsin–Madison.

6.3.3 Measures & Analysis

For each condition, participants completed a questionnaire in which they identified the blocks they believed the robot was referencing and answered six rating-scale questions on the robot's behavior. Participants identified blocks by circling referenced blocks on a picture of the workspace that was included on the questionnaire. As a measure of *accuracy*, I classified participant's identification of the blocks as either correct or incorrect based on whether the participant's answer exactly matched the blocks that the robot's gesture indicated, considering answers that were a superset or subset of the correct answer to be incorrect. My subjective measures assessed the perceived qualities of the gesture in the given setting. From the six questions asked, I constructed the following two scales (half of the items were reversed to prevent

response sets):

Perceived Effectiveness (Cronbach's $\alpha = .967$)

1. The robot used this gesture effectively.
2. The robot's gesture helped me to identify the object(s).
3. The robot's gesture was appropriate for the context.
4. The robot's gesture was easy to understand.

Naturalness (Cronbach's $\alpha = .790$)

1. The robot's gesture was humanlike.
2. The robot's gesture was fluid.

Data analysis involved a two-way analysis of variance (ANOVA), including gesture and setting as fixed effects. Tukey's honestly significant difference (HSD) test was used for pairwise comparisons.

6.4 Discussion

I discuss my most significant results below, first discussing gestures across all settings and then highlighting comparisons of gestures within each setting. Due to the large number of pairwise comparisons involved in my analysis, only Omnibus test results are reported in the paragraphs below, and pairwise comparisons are illustrated in Figures 6.5, 6.6, and 6.7.

6.4.0.1 Comparison of Gestures

A comparison of gestures across settings showed that gesture type had a significant effect on accuracy, $F(7, 1073) = 112.06$, $p < .001$ (Figure 6.5). The fully descriptive baseline was significantly less accurate than exhibiting and pointing, but significantly more accurate than sweeping and grouping. Exhibiting, touching, presenting, and pointing were all significantly more accurate than sweeping and grouping. Consistent with the results on accuracy, gesture type had a significant effect on the perceived effectiveness of the gesture, $F(7, 1073) = 134.37$, $p < .001$. Exhibiting and touching

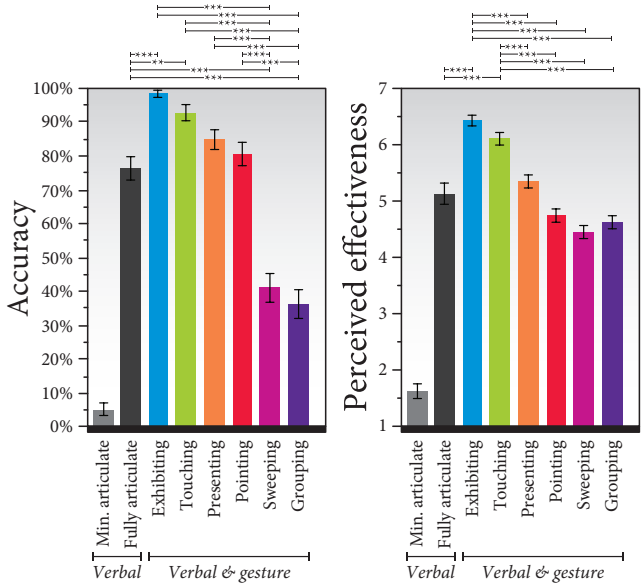

Figure 6.5: Results for both the accuracy of each gesture and the perceived effectiveness of each gesture across all settings. (***), (**) denotes $p < .001$, $p < .01$, respectively. Exhibiting and touching gestures were more accurate than the two baselines and the sweeping and grouping gestures and were perceived to be more effective than the two baselines and the other gestures.

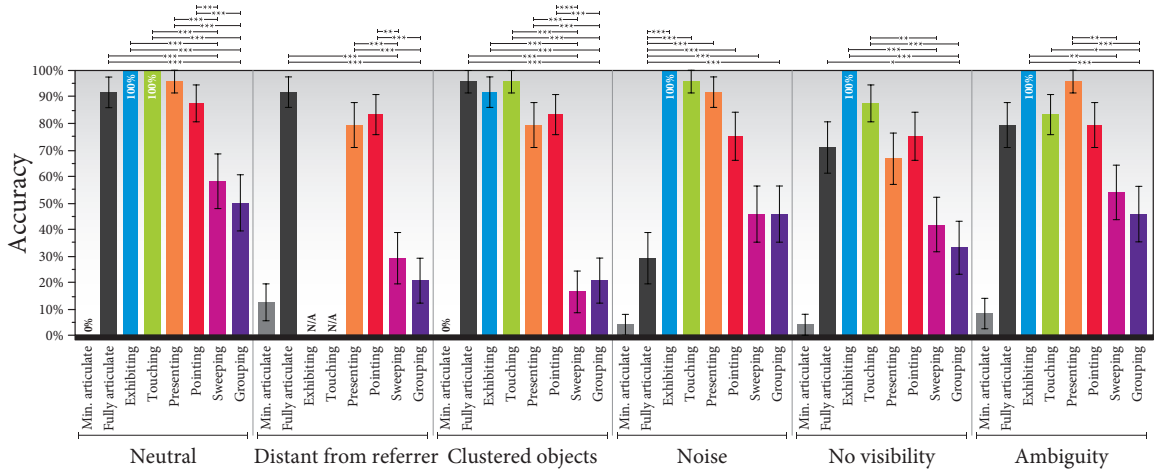

Figure 6.6: Results for the communicative accuracy of each gesture, displayed by setting. (***), (**), (*) denotes $p < .001$, $p < .01$, and $p < .05$, respectively. Exhibiting and touching were consistently more accurate than sweeping and grouping across the majority of settings.

were perceived as significantly more effective than the fully articulate baseline and the presenting, pointing, sweeping, and grouping gestures. All gestures were found to be fairly natural, with average ratings between 5.5 and 6.5 out of 7.

6.4.0.2 Comparison of Gestures by Setting

The following presents results for gestures within each setting. Pairwise comparisons for measures of accuracy and perceived effectiveness are illustrated in Figures 6.6 and 6.7, respectively.

Neutral – Gesture type had a significant effect on accuracy in the neutral setting, $F(7, 161) = 34.36$, $p < .001$. The fully articulate baseline, as Ill as the exhibiting, touching, presenting, pointing gestures, were all significantly more accurate in com-

municating the referent than sweeping and grouping were. Gesture type also had a significant affect on perceived effectiveness, $\mathsf{F}(7, 161) = 39.48$, $\mathsf{p} < .001$. The fully articulate baseline and the exhibiting, touching, and presenting gestures were perceived as significantly more effective than pointing, sweeping, and grouping.

Distant From Referrer – When referents were distant, gesture type had a significant effect on accuracy, $\mathsf{F}(5, 115) = 21.73$, $\mathsf{p} < .001$. The fully articulate baseline and the presenting and pointing gestures were all significantly more accurate than sweeping and grouping. Additionally, while the effectiveness of presenting and pointing fell by 16% and 5%, respectively, when compared to the neutral setting, sweeping and grouping observed greater losses of effectiveness at 27% and 30%, respectively. Gesture type also had a significant effect on perceived effectiveness in this setting, $\mathsf{F}(5, 115) = 19.83$,

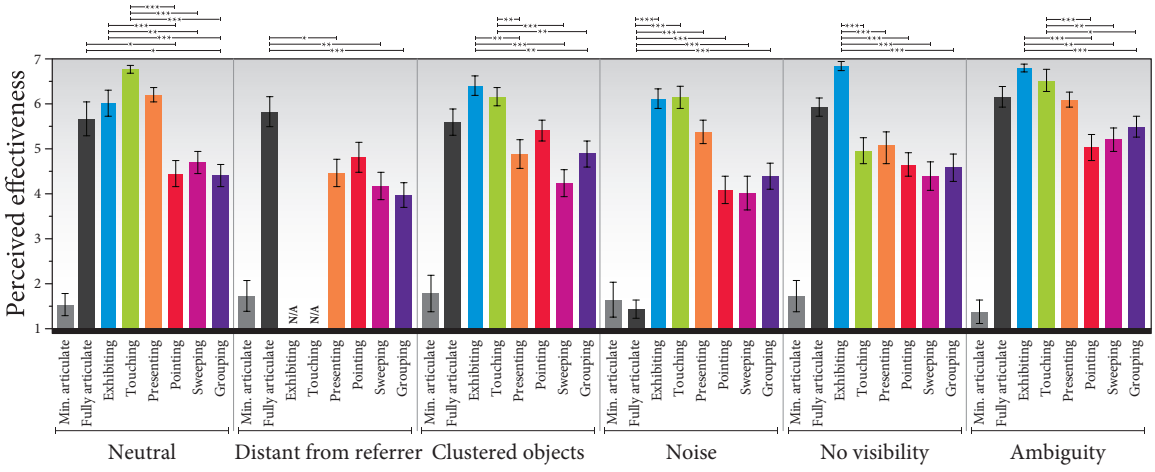

Figure 6.7: Results for the perceived effectiveness of each gesture, displayed by setting. (***), (**), (*) denote $\mathsf{p} < .001$, $\mathsf{p} < .01$, and $\mathsf{p} < .05$, respectively. Exhibiting and touching were consistently perceived to be more effective than presenting, pointing, sweeping, and grouping across the majority of the settings.

$p < .001$. The fully articulate baseline was perceived as significantly more effective than presenting, sweeping, and grouping.

Clustered Objects – Gesture type had a significant effect on accuracy, $F(7, 161) = 43.26$, $p < .001$, when objects were clustered. The fully articulate baseline and the presenting and pointing gestures were all significantly more accurate than sweeping and grouping. Exhibiting, touching, presenting, and pointing were all slightly less accurate than in the neutral setting, losing 5% to 15% accuracy. Sweeping and grouping saw larger drops compared to the neutral setting, losing 48% and 30% accuracy respectively. Perceived effectiveness was significantly affected by gesture type, $F(7, 161) = 25.86$, $p < .001$. Exhibiting and touching both were perceived as significantly more effective than presenting, sweeping, and grouping.

Noise – Gesture type also had a significant effect on accuracy in the noise setting, $F(7, 161) = 22.49$, $p < .001$. The fully articulate baseline was significantly more accurate than exhibiting, touching, presenting and pointing. Exhibiting, touching, and presenting were all significantly more accurate than pointing, sweeping, and grouping. Gesture type also had a significant effect on perceived effectiveness, $F(7, 161) = 38.54$, $p < .001$. The fully articulate baseline was perceived as significantly less effective than every gesture. Additionally, exhibiting and touching were perceived as significantly more effective than pointing, sweeping, and grouping.

No Visibility – Gesture type had a significant effect on accuracy in the no visibility setting, $F(7, 161) = 15.86$, $p < .001$. Exhibiting and touching were both significantly more accurate than sweeping and grouping. Additionally, exhibiting was the only level to not experience a drop in accuracy compared to the neutral setting. Gestures

type had a significant impact on perceived effectiveness, $\mathsf{F}(7, 161) = 40.18$, $\mathsf{p} < .001$. Exhibiting was perceived as significantly more effective than all other gestures. Additionally, the fully descriptive baseline was perceived as significantly more effective than pointing, sweeping, and grouping.

Ambiguity – Gesture type had a significant effect on accuracy under ambiguity, $\mathsf{F}(7, 161) = 15.52$, $\mathsf{p} < .001$. Exhibiting and presenting were significantly more accurate than sweeping and grouping. Gesture type also significantly affected perceived effectiveness, $\mathsf{F}(7, 161) = 59.40$, $\mathsf{p} < .001$. Exhibiting and touching were rated as significantly more effective than pointing, sweeping, and grouping.

My results offer implications for designing effective deictic gestures for robots. The paragraphs below summarize the most important results and discusses these implications.

6.4.1 Properties of Effective Gestures

When the gestures were compared without consideration of context against the fully articulated baseline, my six gestures organized into three groupings: referencing one object with physical contact, referencing one object without physical contact, and referencing multiple objects, with these groupings doing significantly better than, equivalent to, or significantly worse than the fully articulated baseline, respectively. Gestures that involved physical contact with the objects (exhibiting and touching) provided the most effective communication, rarely causing the participant to choose the incorrect block; the effectiveness of physical touch was confirmed in many of my settings as Ill. This finding confirms prior research on human deictics that report

that mothers choose to use physical touch to identify an object for their pre-verbal child due to the unambiguous nature of the gesture (Lempers, 1979). This behavior follows a failed attempt at pointing by the mother, leading her to try a more concrete gesture with a higher likelihood of success. These physical gestures alleviate the cognitive burden placed on those interpreting the gesture by eliminating uncertainty in the referent. My findings suggest that, in settings or tasks that require precise identification of objects, physical contact with the object provides the best chance of the listener correctly identifying the referent.

6.4.2 Setting and Gesture Accuracy

My study highlights instances in which the setting significantly affects the accuracy of gestures. Below, I discuss these results and their implications for designing robot behaviors.

Noise and Verbal Descriptions – The fully articulated baseline had comparable performance to exhibiting, touching, pointing, and presenting in all levels except noise. My findings showed that the fully articulated baseline performed much worse in the noise condition than in every other condition, while the accuracy of many of the gestures remained unchanged. This finding supports the use of gestures that come into contact with the object when fully articulated utterances are difficult to form.

Although the most prominent effect of the noise level was seen with the fully articulated baseline, even pointing and presenting were less effective than in the neutral setting, despite their perceived lack of reliance on utterances given the minimal information utterances included when coupled with gestures. While each

type of gesture maintained a similar motion profile across conditions, making it easier for participants to learn what the gesture was communicating across repeated viewings, it may be that the simple utterances that the robot used helped the participant to distinguish whether the robot was referring to one or many objects. For example, although sweeping and pointing gestures appear similar when the arm is extended, they follow different trajectories, pointing aiming toward a specific object and sweeping covering an area. In addition, the pointing gesture is accompanied by the phrase "this block," clearly indicating that only one block is being referenced, while the sweeping gesture is accompanied by the phrase "these blocks," suggesting more than one block is being referenced. With noise obscuring these phrases, participants may have doubted how many blocks they should select.

While my study did not look at the interaction between gesture and the complexity of the utterance, the combination of my findings on the fully articulate baseline and the pointing and presenting gestures in the noise setting suggest that the robot attempting to clarify non-physical gestures with verbal descriptions—a common human behavior (Fillmore, 1982)—would not improve the accuracy of the gesture.

Gestures for Obstructed Objects – In the no visibility level, where some blocks were obscured by a partition, the number of correctly identified objects was significantly lower for many gestures compared to the fully articulated baseline. Only the exhibiting gesture maintained its effectiveness, since the robot grasped the object and exhibited it above the partition for the participant to clearly see. In real-world applications, the robot can take the perspective of its user (Trafton et al., 2005) to determine whether a referent is obstructed and whether it is necessary to exhibit it for the

listener. In a real world setting, results for the other gestures in the obstructed level would be much lower, as participants in this study had the advantage of seeing an accurate representation of the layout of the workspace beyond the partition on their questionnaire. Although they could not see the gesture clearly, they could see enough of the arm motion to make educated guesses as to the robot's referent.

Diminishing Effectiveness of Multiple Object Gestures – The gestures that referred to multiple objects were consistently less effective than other gestures and the fully articulated baseline. This result is likely a product of the greater ambiguity inherent in these gestures. Such ambiguity occasionally led participants to include objects in the set that were not intended to be referenced. Because my objective measure only counted the answers that were a perfect match to the intended blocks as correct, participants' answers which were a superset or subset of the intended blocks were incorrect. The robot might correct the participant's understanding of which blocks should be included by engaging in repair, such as providing clarifications, which may be considered less costly than precisely identifying the correct blocks the first time. Expecting the listener to process and react to many objects they are expected to identify and manipulate would likely result in high cognitive load, leading to greater frustration with the robot (Morrison and Anglin, 2005). In the cases where identifying only the correct objects is imperative, the robot might use a gesture intended for one object multiple times.

In the *distant from referrer* and *clustered objects* settings, the low accuracy of gestures is compounded. I found that gestures referring to multiple objects were significantly less effective than gestures for a single object in both of these settings

and required more precision than other settings did. Prior work provides support for this observation; when objects are distant, humans use pointing gestures to indicate a spatial region rather than a specific object, instead relying on speech to convey the object (Bangerter, 2004). Likewise, when the robot finds itself working with objects that may be difficult for the listener to disambiguate, either due to distance from the robot or from other objects, the robot should rely on a combination of gesture accompanied with speech to help identify the referents.

Consistent Accuracy of Exhibiting and Touching – In five of the settings, exhibiting and touching maintained relatively consistent accuracy, almost always outperforming the remaining gestures and baselines. Only in the distant from referrer level were exhibiting and touching outperformed, and even then, this result was due to the physical impossibility of these gestures in that setting. These results seem to support the use of exhibiting and touching over pointing and presenting. However, while exhibiting and touching more consistently supported accurate identification of the referent, these gestures are more costly to execute, requiring the robot to physically lift and/or to relocate to be within physical touch of the object, which places limits on their use in real-world applications.

6.4.3 Limitations and Future Work

While I chose gestures and settings for my exploration based on a projection of what might best serve the design of robot behavior, there are other gestures and settings to explore. I did not explore how language, an integral part of deixis, influences the effectiveness of gestures in these settings. Preliminary work has explored some of the

issues regarding the influence of language on gestures (Lozano and Tversky, 2006), but a more comprehensive exploration of this area is needed.

The choice of the robot might also have an impact on how gestures are interpreted, as robot platforms vary in their ability to reproduce human gestures. For instance, while the NAO used in this study has articulated fingers that allow it to grasp and exhibit objects, these fingers are still substantially limited compared to human fingers. Other robots may not have working hands or individual fingers, which would affect how closely their gestures might resemble human gestures. Through iterative design, I sought to design the robot's gestures to mimic human gestures as closely as possible in terms of the intended communication. While I chose to implement the gestures used in this study through puppeteering, a Wizard-of-Oz technique, I hope to develop a robust gesture synthesis system that enables robots to autonomously generate accurate deictic gestures.

In the study, the robot's references included only one gesture. However, communicating complex ideas might require the use of a sequence of gestures. Although I expect my findings to generalize to independent evaluations of each gesture in a sequence, further examination is necessary to conclusively understand how gestures used in a sequence affect referential communication.

6.4.3.1 Gesture Adaptations for the NAO

While the NAO's six degrees of freedom enable us to accurately recreate many of the gestures, a few modifications were required due to the limitations of the platform, all of which involved the NAO's hands. The NAO has three fingers on each hand: two

fingers with an opposable thumb. The thumb's position is fixed such that it is always opposable, making it impossible to achieve a flat hand that is similar to what is seen in the presenting, touching, grouping and sweeping gestures. In these cases, I opened all three of the NAO's fingers to their fullest position to offer an intent similar to those suggested by human gestures.

Additionally, the NAO's fingers do not open and close independently of one another. Instead, extending a single finger requires that all fingers be extended. In the pointing gesture, which is often characterized by an extended index finger, I chose to extend all fingers. To differentiate pointing from presenting, which naturally has all fingers extended, I implemented presenting with the palm pointing entirely upward, while the palm was perpendicular to the table for pointing gestures.

6.5 Study Conclusions

Human collaborations involve the use of deictic gestures, allowing speakers to direct their collaborators' attention to objects in the environment while reducing cognitive load for themselves and their listeners. To function as competent collaborators, robots will need to use deictic gestures to effectively direct the attention of their users toward objects of joint interest, such as a collaborative robot using deictic gestures to draw a manufacturing worker's attention to a specific location in Section 1.1.2.2. Drawing on literature on human communication, I designed a set of deictic gestures for the NAO robot and specified a set of settings that provided a diverse set of conditions in which humans and robots might engage in deictic communication. I conducted

an exploratory human-robot interaction study that examined the communicative accuracy and perceived effectiveness of these gestures against two verbal baselines and how the setting of the communication affected these measures. The results suggest many design implications regarding the use of gestures in each setting, including what properties might make certain gestures more effective, the tradeoffs involved in referring to multiple objects, the effects of noise on verbal deictic references, and how gestures might function when objects are obscured. My findings suggest that future iterations of collaborative robots should adaptively choose deictic gestures to suit the needs of a particular context, enabling the robot to use effectively use deictic gestures across a variety of domains.

7 GENERAL DISCUSSION

In this dissertation, I have helped highlight some of the perceptions and expectations users have of their collaborative robots. Inspired by scenarios similar to my field work, I also studied three different collaborative behavioral cues, using hand-built models of human data implemented on a robotic platform. In this chapter, I discuss the high level points of this work, as well as highlight limitations that should be considered when extending or interpreting this research.

7.1 Summary and Synthesis

This dissertation contains four studies, each of which contributes new findings towards a better understanding of effective social behaviors for human-robot collaboration. Below, I summarize these studies, highlight high-level contributions resulting from these studies, and discuss how these contributions impact various domains and future work.

To review, the findings of each of my studies are outlined below:

- **Understanding the Reality of Collaborative Robots (Ch. 3):** This ethnographic examination of the implementation of collaborative robots into three manufacturing sites resulted in several insights. These insights included how the structure of a work cell changes when a collaborative robot is introduced, and the social ramifications of a robot co-worker. For example, operators who worked with the robot perceived the robot as a social entity, while other

stakeholders (i.e., management and maintenance staff) perceived the robot as a tool.

- **Speech Patterns (Ch. 4):** This study presented the design and implementation of *Interaction Blocks*, an authoring environment that enables users to prototype and test human-robot speech interactions through the use of design patterns. A user study highlighted the potential for the use of design patterns and the workflow that my authoring environment promotes to design, prototype, and evaluate interactions, enabling interaction designers to take advantage of patterns to synthesize complex interactions.

- **Instruction & Repair (Ch 5):** The use of two different strategies for providing instructions—grouping and summarization—demonstrated that, when the robot grouped instructions, participants completed the task faster but encountered more breakdowns. Additionally, summarization increased participant rapport with the robot.

- **Deictic Gestures (Ch 6):** In addition to providing a set of deictic gestures for robots, the results from this study suggest that there are many design implications regarding the use of gestures in different settings. These implications include what properties might make some gestures more effective than others, the tradeoffs to referring to multiple objects, and the use of deictic gestures when objects are obscured.

The results of this dissertation demonstrate the necessity of implementing human-like social cues on robots to achieve effective communication with human collaborative

partners, similar to the human-human scenario discussed in Section 2.1. Each study provides an example of how a particular social cue is used to communicate, such as deictic gestures or speech patterns. As a consequence of this communication, humans can then effectively coordinate their behavior with their robot counterparts. This finding is perhaps most obvious in the ethnographic study, where operators described how they used the robot's gaze to discern where the robot would next move, allowing the operators to work in a way that was complementary to the robot's actions. However, results from the laboratory studies can also be used to achieve similarly natural coordination. For example, the deictics study discussed the importance of deictic gestures in human-human communication, helping to achieve fluid coordination by augmenting or replacing speech, often performing as well as or better than verbal communication alone (Goldin-Meadow, 1999; McNeil, 1992). By understanding how to implement deictic gestures on robots and how gestures perform under particular conditions, robots will be similarly capable of using deictic gestures to achieve more natural and fluid coordination.

These findings also provide further support for the theories outlined in Section 2.1.1 when they are contextualized in human-robot collaboration. The theory of mental models posits that robots which display more humanlike behaviors will evoke a more natural and fluid response from their human partners. The findings of this dissertation support this theory, with human partners preferring robots that demonstrate humanlike behavior. For example, workers in the ethnographic study repeatedly referred to the robot's humanlike gaze patterns and humanlike morphology as allowing them to perceive the robot as humanlike. As a result, workers felt they could use the same

behaviors to work around the robot as they would a human co-worker. Theory of mind extends the idea of mental models, suggesting that humans will recognize intentional agents and coordinate their future actions to accommodate for the agent's actions. The ethnographic study specifically highlights how the addition of social cues to a collaborative robot can shape how humans perceive it, demonstrating how perceived intentionality can make nearby workers feel safer. Participants in the instruction and repair study recognized the intent of the robot to successfully lead them to completion of the task, enabling the participants to ask questions as needed of the robot to rectify breakdowns. Finally, the theory of embodied cognition suggests that robots should produce social cues that consider their own capabilities, morphology, and surrounding environment. The deictics study explored deictic gestures in a variety of settings (e.g., noise, objects distant from referrer), and found significant differences in how different gestures performed in different settings, suggesting the importance of embodied cognition in choosing how to employ social cues.

The studies presented in Chapters 4, 5 and 6 each utilized hand-built models from human-human data. This is counter to the trend in human-robot interaction research, where models are built with an alternative to the heuristics-based modeling presented in this dissertation. Instead, research often uses artificial intelligence techniques, such as Bayesian inference, to build models from human-human data (Broz et al., 2013; Hong et al., 2007; Huang and Mutlu, 2013; Lucignano et al., 2013), it was surprising to see how appropriate the hand-built models generated were to the tasks participants were presented with, and how these models were capable of revealing contextual differences in how behavioral cues should be used. This work suggests

that sometimes a simpler solution (i.e., hand-built models) can still present results that provide powerful suggestions for the future of collaborative robot design.

Another general finding, particularly from the studies presented in Chapters 4, 5 and 6 is the contextual nature of using collaborative behaviors. The results presented in this dissertation found that there is no single best strategic recommendation for each behavioral cue. Rather, this dissertation makes recommendations about when each strategy for a behavioral cue should be employed. While the idea of collaborative robots was initially accomplished through improved sensing, this dissertation acknowledges this improved sensing and requires that it be taken further to more fully realize the vision of collaborative robots working effectively with human partners. This finding also supports the need for more advanced sensing technology onboard robotic platforms, which is discussed further in Section 7.2.4.2.

While this work was initially contextualized in the manufacturing domain in Section 1.1, the findings presented in this dissertation are applicable to many additional collaborative scenarios. In general, these scenarios would involve a human and robot working together in a physical context toward some common goal, such as a robot instructor helping a human student in a lab, a robot assistant aiding a human in the home, or a small assistive robot supporting an astronaut in space. Each of these scenarios requires the robot to communicate effectively with their human partner, necessitating the use of various channels of communication, including speech, instructional strategies, and gestures. Additionally, some of these results might generalize to other situations outside of collaborative work where these same social cues might be needed. For example, the speech patterns developed in Chapter 4

were drawn from both conversational scenarios (e.g., interviews, storytelling) as well as collaborative scenarios, and should be applicable to some conversational settings. Deictic gestures can be used in a variety of conversational settings where it might be beneficial to draw attention to an item or space, such as a robotic medical assistant gesturing to specific instructions for a medication, potentially making these results informative outside of collaborative scenarios. Future work should explore how these behaviors translate to additional types of scenarios outside of collaborative scenarios.

The results presented here suggest key considerations for future work in human-robot collaboration. First, further research in the realm of human-robot collaboration should be cognizant of the role of the robot as a social agent, requiring similar social behaviors as humans to achieve collaboration that is effective on both objective and subjective measures. This insight encourages exploration of many facets of human collaborative behavior, including a broader range of social cues than explored here. As a corollary to this point, there will likely be instances where it is unclear which social cue to use, such as the nuanced results of the deictics work where different deictic gestures each had benefits and drawbacks. Future work on collaborative social cues should strive to provide insights into potential tradeoffs of different approaches. Additionally, a broader understanding of the tradeoffs in employing different social cues would present an opportunity to construct a framework for collaborative scenarios that could aid in understanding how results generalize to different settings, similar to frameworks built for other domains (Rae et al., 2015). For example, if we have results for a study performed with collaborative robots in the home, then those results might translate to other contexts that share similar traits (e.g., types of people the robot

interacts with, types of tasks performed), such as a retirement community, but not to contexts that differ on many traits, such as military contexts. This framework might also allow robot behavior designers to consider the tradeoffs of choosing different social behaviors to accomplish a goal, such as efficiency versus maintaining rapport.

In the remainder of this section, I discuss some of the limitations of this work, and opportunities for later work to build upon the research presented here. I focus on the research context, methodological challenges, generalizability, and technical challenges.

7.2 Limitations and Future Considerations

The studies in Chapters 4, 5, and 6 were inspired by scenarios such as the field study outlined in Chapter 3. However, at the beginning of this dissertation work, collaborative robots were not yet widely adopted, requiring me to choose behavioral cues to focus on and conduct initial studies based on hypothetical scenarios that collaborative robots might participate in. Section 7.2.1 will discuss this approach and the possible limitations of it.

In this work, I made the decision to approach my research of particular behavioral cues in a specific manner. In particular, I would first conduct some form of modeling by hand, based on data and prior human-human literature. Using the resulting model, I would examine that particular behavioral cue in a laboratory context by implementing the model on an autonomous robot platform to conduct human-robot interaction studies. The decision to conduct each step of the pipeline in this manner

has ramifications about how this work should be interpreted. Limitations of the methodology presented in this dissertation will be addressed in Section 7.2.2.

Each study presented in this dissertation was examined under only a single context. For example, each study only looked at a single research platform, examined only one type of experimental task, and only looked at each behavioral cue in a single laboratory context. These decisions raise questions about the generalizability of the results presented, and a discussion of how this work generalizes to other research platforms, experimental tasks, and other domains will be addressed in Section 7.2.3.

Finally, the work presented highlights many technical challenges, both that I personally encountered in conducting this research, but also technical challenges that will need to be overcome to make collaborative robots a reality. Technical challenges encountered in this work and an overview of how these challenges impact real-world collaboration will be discussed in Section 7.2.4.

7.2.1 Research Context

The studies in this work were inspired by hypothetical future scenarios that collaborative robots are envisioned to serve in, similar to the field study conducted the at the manufacturing plants presented in Chapter 3. While this work was inspired by these types of scenarios, collaborative robots were not yet widely adopted at the beginning of this work. As a result, the studies presented in Chapters 4, 5, and 6 were designed in anticipation of hypothetical future scenarios, rather than informed by the observation of collaborative robots in the real world, and the field study presented in Chapter 3 was conducted after the studies on specific behavioral cues. Although these

studies were not conducted in the ideal order (i.e., the ethnography first, followed by the studies on specific behavioral cues), results of the field study did confirm the need for work on the behaviors examined in Chapters 4, 5, and 6.

There are several technical challenges which must be overcome to realize the vision of collaborative robots, which will be addressed in Section 7.2.4. As the technology improves, collaborative robots will become capable of entering new domains, as well as working more effectively in domains they are already participating in. As these collaborative robots are equipped with new capabilities, future work should explore the new or modified challenges that must be resolved to more effectively integrate these robots as co-workers with human partners.

7.2.2 Methodological Challenges

In completing the work presented in this dissertation, I chose a specific set of methodologies to complete my work. In particular, I used hand-based modeling, examined each collaborative behavioral cue in the context of a laboratory setting, and tested those cues on autonomous robot platforms in a human-robot interaction experiment. Below, I address the limitations of the chosen methodological pipeline, particularly in light of alternate methodologies.

7.2.2.1 Modeling Approach

Human collaborative behaviors are composed of a wide range of variables that are highly interrelated. These variables might include external variables, such as the environmental context in which the collaboration takes place, and internal variables,

such as how multiple behavioral cues (e.g., gaze and gesture) are related. In this work, I chose to focus on modeling a single behavioral cue, such as the speech patterns presented in Chapter 4. Additional cues that are necessary to give the appearance of a humanlike robot, such as gaze, were then modeled uniformly across different implementations of that cue according to prior literature or observations from a human-human dataset. While this approach does not take into account additional variables, it does enable a systematic and focused understanding of each variable. However, how the findings from these studies will translate to a real-world system which must coordinate multiple behavioral cues simultaneously is unknown.

Another challenge in modeling human behavior is producing models with the correct level of detail. In this work, an iterative process was used to identify and capture the correct level of detail. This iteration occurred during the initial coding process of the human-human data. I would watch the videos multiple times, noting down interesting phenomena, in an attempt to capture the range of variables. These phenomena would be refined over the course of additional videos into codes that were then used to code all of the human-human data. Sometimes, after coding, further refinement would discard codes that were rare occurrences in the human-human dataset. While the resulting models for all three behavioral cues examined in this work accurately modeled human behavior and provided interesting insights into how a robot should behave during collaboration, the reliance on my own observations and modeling approach does leave room for additional variables that were not obvious to me. While the models presented in this work were sufficiently detailed to expose interesting findings about how human-robot collaboration should occur, future work

should focus on developing more nuanced modeling techniques to capture a greater range of variable in the model.

7.2.2.2 Autonomous Platforms

The studies outlined in Chapters 4 and 5 use autonomous robot platforms, with the robot acting based on human responses and actions. This autonomy was achieved through constraining the scenario the participants completed during the experimental task, limiting the possible scenarios and verbal responses the robot might encounter. The use of an autonomous robot system allows participants to engage in an interaction with the robot, creating a more authentic environment for the participant to evaluate.

While the autonomy of the system allows for more authentic interactions, it comes at the expense of realistic interactions. For example, in the study on giving instructions outlined in Chapter 5, participants were required to engage the robot verbally before the robot would offer any feedback on the participant's progress. This decision was made because the robot could not assess the participant's progress while actions were still being performed, as a human instructor might, due to sensing limitations of the system (e.g., the participant might be obstructing the sensor's ability to accurately detect the participant's progress). Similar to the technological challenges facing the widespread adoption of collaborative robots outlined in Section 7.2.1, several technological advances are needed to enable increasingly realistic autonomous studies. These issues will be addressed in Section 7.2.4.

7.2.3 Generalizability

In particular, this section will address the generalizability of the results regarding different research platforms, different tasks, and different research methodologies.

7.2.3.1 Research Platforms

While the robotic platforms explored in this work each possess a humanlike morphology, there are many variations between these platforms in how that morphology is realized. For example, the NAO and Baxter platforms each have a humanlike appearance. However, among other differences, the NAO has three articulated fingers that move together in a pinching motion, while Baxter's end effectors are two planks that move together in a pinching motion. Other platforms have similarly varying end effectors. This difference in end effectors has implications for how the deictic gestures presented in Chapter 6 will be implemented. For example, the NAO might be better able to present a pointing motion with its more humanlike fingers than Baxter. Some of the results from this dissertation, such as the work on speech patterns presented in Chapter 4 and the work on instruction and repair presented in Chapter 5, are platform agnostic. However, the field work presented in Chapter 3 focused on the Baxter platform. While Baxter has a humanlike morphology, other types of collaborative robots, such as collaborative robotic arms, only partially have a humanlike appearance. As a result, only some parts of this work that are less integrated with a humanlike morphology might be generalizable to these platforms, while other results may only be partially applicable. Additionally, the work on deictic gestures presented in Chapter 6 will require adapting the results for the particular robot platform they

are implemented on, taking into account the type of end effectors the platform has. Future work should consider a more thorough exploration of how these physical differences impact the ability of designers to implement collaborative behaviors and, in turn, how the effectiveness of these collaborative behaviors might change.

7.2.3.2 Experimental Tasks

A wide variety of tasks, all of which were modeled after a hypothetical collaborative scenario, were used across the three studies described in Chapters 4, 5, and 6. Each of these collaborative scenarios was meant to be representative of some, but not all, collaborative tasks robots might encounter. As a result, it is unknown how these results might translate to a new task, although some elements might translate well based on the similarity of core elements of the task structure (e.g., participant roles, task goal).

As part of integrating the results of this dissertation's studies into future experimental tasks, future work should explore a common design framework for collaborative robot tasks that might help discover caveats and clarifications for how the results of this dissertation's work should be applied.

7.2.3.3 Research Methodology

The two approaches to research methodology used in this work, ethnographic field studies and experimental laboratory studies, each offer different repercussions for the generalizability of the results produced. Below, I discuss how each methodology impacts generalizability, and where additional work might be needed to translate

results from this dissertation to new domains or applications.

Field Study Setting: The ethnographic field study presented in Chapter 3 allows for a holistic examination of the reality of implementing collaborative robots in the wild. This type of study enables researchers to focus on a specific setting, such as collaborative robots in manufacturing settings, in an attempt to gain insights about that particular setting. However, the focus on a single setting means that ethnographic studies are not meant to be generalizable. Rather, ethnographic studies serve as a way to gain deep, holistic insights into a particular setting, insights that might spur research questions in the studied setting, or even similar types of settings.

As collaborative robots become more prevalent in additional settings, as well as more technically sophisticated in how they are incorporated, future work should corroborate, refine, and add to the theories presented in the field study component of this dissertation. By seeking to better understand how humans are reacting to collaborative robots in a variety of settings, we can better understand the broader ramifications of collaborative robots across settings, as well as the challenges or nuances of implementing these robots into particular types of settings. For example, collaborative robots might be better received in situations where trust is less necessary, such as in the workplace, while collaborative robots in high trust situations, such as military or health settings, might be less welcome.

Laboratory Setting: The work presented in Chapters 4, 5, and 6 takes advantage of controlled laboratory settings to test the impact of various collaborative behaviors. This setting enables researchers to fix as many variables as possible in an attempt to identify significant differences between the conditions studied. However, this

approach presents threats to the ecological validity of the results, as experiments are not performed under real world conditions. These concerns about ecological validity must be weighed against the benefits from isolating and studying a single behavioral cue in a laboratory setting.

As a result of the above concerns, additional work is needed to translate these behaviors and validate the models and results in real world settings. Designers of robot behaviors must be capable of using the findings of the studies presented in this dissertation, as well as the contexts those studies were conducted under, to inform their design choices when building a new robotic system, rather than simply implementing the models presented.

7.2.4 Technical Challenges

While technological advancements now allow collaborative robots to work together with humans in a limited way, there are still many technical challenges facing the realization of collaborative robots. Below, I outline two categories of technical challenges: challenges I encountered in pursuit of my research, and broader challenges that prevent collaborative robots from fully working together with humans.

7.2.4.1 Technology for Autonomous Laboratory Studies

One challenge of building autonomous systems to test in laboratory settings is balancing the desire for a realistic, interactive system with the limitations of current technology. For example, speech generation and natural language processing are both crucial to to facilitating a fluid and realistic interaction. However, commercial

off-the-shelf products for natural language processing, such as the Microsoft Kinect or the natural language processing available on robotic platforms, still have limitations on the number of phrases they can recognize and their accuracy in distinguishing shorter phrases (e.g., "This one" "These ones"). Additionally, speech generation through a combination of a dialogue manager and text-to-speech technology is a similarly open-ended problem. These challenges place necessary limits on the open-ended nature of dialogue. To enable a fluid collaboration despite these limitations, experimental tasks are designed to curtail some of the uncertainty that can occur. For example, in the pipe task described in Chapter 5, the researcher verbally introduced each type of piece to participants prior to the experiment. This step encouraged participants to use the given name, rather than using their own descriptions of the pieces.

In addition to speech, the vision capabilities of autonomous robot platforms, achieved through the use of various sensors, are critical for processing changes in the environment, including the user's current state. In this work, vision was often achieved through the use of the stereo camera available on the Microsoft Kinect. While the Kinect was capable of sensing some desired properties of the environment (e.g., where the participant was during a specific point of time in the interaction, the placement of AR tags on the table), the current limitations of computer vision required designing the task to circumvent them. For example, in the pipe task from Chapter 5, the pipes used in the study had two small square boxes wrapped around them, with AR tags affixed to each face of the box. This ensured that the AR tags were readable by the vision system when layed on the table. Additionally, if the

participant wanted to ask a question about a piece they had chosen, such as saying "Is this the right pipe?" while holding the pipe, there was no guarantee that the participant would hold the pipe such that the vision system could recognize it. Thus, participants were instructed to lay pieces on the table prior to asking questions about them.

Future laboratory studies in human-robot collaboration would benefit from increased speech generational, natural language processing, and vision capabilities, either placed on board the robotic platform, or in the form of off-the-shelf products that can be used to augment and extend the robot's capabilities. These systems would allow researchers to ensure more natural interactions during experiments, without needing to try and curtail or control the participant's behaviors.

7.2.4.2 Technology for Real-World Collaborative Robots

While technological advances have enabled robots to collaborate with humans in specific domains, additional advancements are needed to fully realize this vision. One repercussion of the technical challenges outlined above is a limit to how interactive collaborative robots can be when placed in the real world. Although laboratory settings can circumvent some of the challenges of natural language processing through task design, real-world applications present a level of uncertainty that makes this solution untenable. Instead, current systems might either choose a select set of keywords or phrases that the robot will recognize for functional purposes, or might eliminate language completely in an effort to minimize user confusion about what the robot will and will not recognize. One challenge facing future designers of

collaborative robots is how to incorporate the recommendations for increased sociality presented in this dissertation given the expansive and uncertain nature of social dialogue. Currently, social dialogue systems tend to be situated in specific contexts to limit this variability (Bickmore and Cassell, 2005; Foster et al., 2012).

Additionally, challenges with vision systems will also need to be addressed to produce more effective collaborative robots. As shown in the field study, vision systems on real-world collaborative robots are not yet sufficient for working in a human environment that allows for slight variation. Instead, workers who incorporated these robots strove to remove sources of uncertainty, particularly pertaining to the vision system. Additionally, taking the input to a vision system and interpreting the environment correctly, such as recognizing gestures (Mitra and Acharya, 2007; Ren et al., 2013) or facial expression (Tian et al., 2011), can be difficult in real-time environments where data is often noisy. In addition to recognizing human state and intent, collaborative robots must also be capable of recognizing the objects they are working with. Current vision systems require significant training data for real-world object recognition. Future work on collaborative robots might require enabling those who use collaborative robots to train their robot to recognize objects the robot is expected to recognize and work with.

In light of the findings of this dissertation that behavioral cues should be applied with different strategies depending on context, it is crucial that sensing technologies improve. These improvements would allow robots to not only collaborate more fully and fluidly with their human partners, but would also enable these robots to more effectively utilize the models of collaborative behavioral cues presented in this work.

8 CONCLUSION

Technological advancements have helped realize the vision of collaborative robots, allowing robotic technologies to work with human partners both in the workplace and the home. The introduction of these robots has the potential to revolutionize the way people complete physical tasks, but little is known about what expectations users have of these robots, and how to design collaborative behaviors that allow robots to seamlessly and effectively work with their human partners.

In this dissertation, I focused on understanding the perceptions and expectations human users have for these collaborative robots, and then focusing on three behavioral cues—speech patterns, instruction and repair, and deictic gestures—designed models for these cues that could be implemented and tested on robotic platforms. Specifically, this research has four contributions to the field of human-robot collaboration:

- **Understanding of Real-World Collaborative Behaviors**: Through studying an environment where collaborative robots have already been introduced, a grounded understanding of the needs, desires, and expectations users have for collaborative robots.

- **Computational Models of Human Collaborative Behaviors**: Models of how specific social cues are used, drawn from the literature and human-human observations, that can be implemented on a robot.

- **Contextualization of Behaviors and an Understanding of their Role in Facilitating Interaction**: An understanding of how the use of social cues

might differ depending on contextual factors, and how this variation supports a range of collaborative outcomes (e.g., time spent on task, rapport).

– **Tools to Facilitate Developing and Testing Human-Robot Collaborations**: The synthesis of tools to enable a wide range of users to implement social behaviors on collaborative robots.

The success of the hand-built models in this work suggests that this approach, as compared to artificial intelligence techniques, has the potential to also produce valid models that can be of use in classifying human behavior and in designing collaborative robots. The results from this work also highlight the need for contextual awareness when deciding how to employ a behavioral cue, suggesting that robotic technologies could benefit from more sophisticated sensing technologies to achieve this awareness. While future work will undoubtedly improve upon the results presented here, this dissertation provides a foundation to build upon in understanding the implications of collaborative robots working with humans, and in designing more refined collaborative behaviors to enable robots to be more effective task partners.

REFERENCES

ABB Group. 2013. Dual-arm concept robot. www.abb.com. Accessed: 2014-09-23.

Aggarwal, Praveen, and Rajiv Vaidyanathan. 2003. The perceived effectiveness of virtual shopping agents for search vs. experience goods. *Advances in Consumer Research* 30:347–348.

Aleven, Vincent A, and Kenneth R Koedinger. 2002. An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science* 26(2):147–179.

Alibali, Martha W, Lucia M Flevares, and Susan Goldin-Meadow. 1997. Assessing knowledge conveyed in gesture: Do teachers have the upper hand? *Journal of Educational Psychology* 89(1):183.

Andrist, Sean, Bilge Mutlu, and Michael Gleicher. 2013a. Conversational gaze aversion for virtual agents. In *Intelligent Virtual Agents*, 249–262. Springer.

Andrist, Sean, Erin Spannan, and Bilge Mutlu. 2013b. Rhetorical robots: making robots more effective speakers using linguistic cues of expertise. In *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction*, 341–348. IEEE Press.

Argyle, Michael. 2013. *Bodily communication*. Routledge.

Astington, Janet W. 1993. *The child's discovery of the mind*, vol. 31. Harvard University Press.

Bangerter, Adrian. 2004. Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science* 15(6):415–419.

Bangor, Aaron, Philip T Kortum, and James T Miller. 2008. An empirical evaluation of the system usability scale. *International Journal of Human–Computer Interaction* 24(6):574–594.

Baron-Cohen, Simon. 1997. *Mindblindness: An essay on autism and theory of mind*. MIT press.

Benbasat, Izak, and Weiquan Wang. 2005. Trust in and adoption of online recommendation agents. *Journal of the Association for Information Systems* 6(3): 4.

Berelson, Bernard. 1952. *Content analysis in communication research*. Free press.

Bicho, Estela, Wolfram Erlhagen, Luis Louro, and Eliana Costa e Silva. 2011. Neuro-cognitive mechanisms of decision making in joint action: A human–robot interaction study. *Human Movement Science* 30(5):846–868.

Bickmore, Timothy, and Justine Cassell. 2005. Social dialogue with embodied conversational agents. In *Advances in Natural Multimodal Dialogue Systems*, 23–54. Springer.

Blakemore, Sarah-Jayne, and Jean Decety. 2001. From the perception of action to the understanding of intention. *Nature Reviews Neuroscience* 2(8):561–567.

Blaylock, Nate, James Allen, and George Ferguson. 2003. Managing communicative intentions with collaborative problem solving. In *Current and New Directions in Discourse and Dialogue*, 63–84. Springer.

Bolinger, Dwight. 1983. Intonation and gesture. *American Speech* 156–174.

Boucher, Jean-David, Ugo Pattacini, Amelie Lelong, Gerard Bailly, Frederic Elisei, Sascha Fagel, Peter Ford Dominey, and Jocelyne Ventre-Dominey. 2012. I reach faster when i see you look: gaze effects in human–human and human–robot face-to-face cooperation. *Frontiers in Neurorobotics* 6.

Brandstätter, Veronika, Angelika Lengfelder, and Peter M Gollwitzer. 2001. Implementation intentions and efficient action initiation. *Journal of Personality and Social Psychology* 81(5):946.

Bratman, Michael. 1987. *Intention, plans, and practical reason*. Harvard University Press.

Breazeal, Cynthia. 2003. Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies* 59(1):119–155.

Breazeal, Cynthia, Cory D Kidd, Andrea Lockerd Thomaz, Guy Hoffman, and Matt Berlin. 2005. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005)*, 708–713. IEEE.

Breazeal, Cynthia, and Brian Scassellati. 1999. A context-dependent attention system for a social robot. *rn* 255:3.

Brennan, Susan E, Xin Chen, Christopher A Dickinson, Mark B Neider, and Gregory J Zelinsky. 2008. Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition* 106(3):1465–1477.

Brogårdh, Torgny. 2007. Present and future robot control development: An industrial perspective. *Annual Reviews in Control* 31(1):69–79.

Brooks, Andrew G, and Cynthia Breazeal. 2006. Working with robots and objects: Revisiting deictic reference for achieving spatial common ground. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-Robot Interaction*, 297–304. ACM.

Broz, Frank, Illah Nourbakhsh, and Reid Simmons. 2013. Planning for human–robot interaction in socially situated tasks. *International Journal of Social Robotics* 5(2): 193–214.

Bryan, Ronnie, Pietro Perona, and Ralph Adolphs. 2012. Perspective distortion from interpersonal distance is an implicit visual cue for social judgments of faces.

Carroll, John M, Nancy S Anderson, and Judith Reitman Olson. 1987. *Mental models in human-computer interaction: Research issues about what the user of software knows*. 12, National Academies.

Caselli, Maria Cristina. 1990. Communicative gestures and first words. In *From Gesture to Language in Hearing and Deaf Children*, 56–67. Springer.

Clark, Eve V. 2009. *First language acquisition*. Cambridge University Press.

Clark, Herbert H. 1992. *Arenas of language use*. University of Chicago Press.

———. 1996. *Using language*, vol. 1996. Cambridge University Press.

———. 2005. Coordinating with each other in a material world. *Discourse Studies* 7(4-5):507–525.

Craik, Kenneth James Williams. 1967. *The nature of explanation*. CUP Archive.

Croft, William. 2000. *Explaining language change: an evolutionary approach*. Pearson Education.

d'Andrade, Roy. 1987. *A folk model of the mind*. Cambridge University Press.

Datta, Chandan, Chandimal Jayawardena, I Han Kuo, and Bruce A MacDonald. 2012. Robostudio: A visual programming environment for rapid authoring and customization of complex services on a personal service robot. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2352–2357. IEEE.

Doshi, Anup, and Mohan M Trivedi. 2009. On the roles of eye gaze and head dynamics in predicting driver's intent to change lanes. *Intelligent Transportation Systems, IEEE Transactions on* 10(3):453–462.

Dragan, Anca, and Siddhartha Srinivasa. 2014. Familiarization to robot motion. In *Proceedings of the 2014 acm/ieee international conference on human-robot interaction*, 366–373. ACM.

Ekman, Paul, and Wallace V Friesen. 1967. Head and body cues in the judgment of emotion: A reformulation. *Perceptual and Motor Skills* 24(3):711–724.

Emerson, Robert M, Rachel I Fretz, and Linda L Shaw. 2011. *Writing ethnographic fieldnotes*. University of Chicago Press.

Feinfield, Kristin A, Patti P Lee, Eleanor R Flavell, Frances L Green, and John H Flavell. 1999. Young children's understanding of intention. *Cognitive Development* 14(3):463–486.

Fillmore, Charles J. 1982. Towards a descriptive framework for spatial deixis. *Speech, Place and Action: Studies in Deixis and Related Topics* 31–59.

Fogg, Brian J. 1998. Persuasive computers: perspectives and research directions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 225–232. ACM Press/Addison-Wesley Publishing Co.

Fong, Terrence, Charles Thorpe, and Charles Baur. 2003. Collaboration, dialogue, human-robot interaction. In *Robotics Research*, 255–266. Springer.

Forlizzi, Jodi. 2007. How robotic products become social products: an ethnographic study of cleaning in the home. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, 129–136. ACM.

Forlizzi, Jodi, and Carl DiSalvo. 2006. Service robots in the domestic environment: a study of the Roomba vacuum in the home. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, 258–265. ACM.

Foster, Mary Ellen, Andre Gaschler, Manuel Giuliani, Amy Isard, Maria Pateraki, and Ronald Petrick. 2012. Two people walk into a bar: Dynamic multi-party social interaction with a robot agent. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction*, 3–10. ACM.

Foster, Mary Ellen, Manuel Giuliani, Amy Isard, Colin Matheson, Jon Oberlander, and Alois Knoll. 2009. Evaluating description and reference strategies in a cooperative human-robot dialogue system. In *Ijcai*, 1818–1823.

Franklin, DF, Roger E Kahn, Michael J Swain, and R James Firby. 1996. Happy patrons make better tippers: Creating a robot waiter using perseus and the animate agent architecture. In *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition*, 253–258. IEEE.

Fussell, Susan R, Sara Kiesler, Leslie D Setlock, and Victoria Yew. 2008. How people anthropomorphize robots. In *Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction*, 145–152. ACM.

Gallese, Vittorio, and Alvin Goldman. 1998. Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences* 2(12):493–501.

Glas, D, Satoru Satake, Takayuki Kanda, and Norihiro Hagita. 2012. An interaction design framework for social robots. In *Robotics: Science and Systems*, vol. 7, 89.

Glaser, Barney G, Anselm L Strauss, and Elizabeth Strutzel. 1968. The discovery of grounded theory: strategies for qualitative research. *Nursing Research* 17(4):364.

Goetz, Jennifer, Sara Kiesler, and Aaron Powers. 2003. Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *The 12th IEEE International Workshop on Robot and Human Interactive Communication*, 55–60. IEEE.

Goffman, Erving. 1983. The interaction order: American sociological association, 1982 presidential address. *American Sociological Review* 1–17.

Goldin-Meadow, Susan. 1999. The role of gesture in communication and thinking. *Trends in Cognitive Sciences* 3(11):419–429.

Gonsior, Barbara, Dirk Wollherr, and Martin Buss. 2010. Towards a dialog strategy for handling miscommunication in human-robot dialog. In *Ro-man*, 264–269. IEEE.

Gray, Jesse, Cynthia Breazeal, Matt Berlin, Andrew Brooks, and Jeff Lieberman. 2005. Action parsing and goal inference using self as simulator. In *Ieee international workshop on robot and human interactive communication*, 202–209. IEEE.

Griffin, Zenzi M. 2001. Gaze durations during speech reflect word selection and phonological encoding. *Cognition* 82(1):B1–B14.

Grosz, Barbara J, and Sarit Kraus. 1996. Collaborative plans for complex group action. *Artificial Intelligence* 86(2):269–357.

Grosz, Barbara J, and Candace L Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics* 12(3):175–204.

Grudin, Jonathan. 1988. Why cscw applications fail: problems in the design and evaluation of organizational interfaces. In *Proceedings of the 1988 ACM Conference on Computer-Supported Cooperative Work*, 85–93. ACM.

———. 1994. Groupware and social dynamics: Eight challenges for developers. *Communications of the ACM* 37(1):92–105.

Haddadin, Sami, A Albu-Schaffer, Alessandro De Luca, and Gerd Hirzinger. 2008. Collision detection and reaction: A contribution to safe physical human-robot interaction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3356–3363. IEEE.

Harrison, Simon. 2011. The creation and implementation of a gesture code for factory communication. In *Proceedings of Gestures and Speech Interaction*.

Hato, Yasuhiko, Satoru Satake, Takayuki Kanda, Michita Imai, and Norihiro Hagita. 2010. Pointing to space: modeling of deictic interaction referring to regions. In *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction*, 301–308. IEEE Press.

Hegel, Frank, Manja Lohse, and Britta Wrede. 2009. Effects of visual appearance on the attribution of applications in social robotics. In *The 18th ieee international symposium on robot and human interactive communication*, 64–71. IEEE.

Helweg-Larsen, Marie, Stephanie J Cunningham, Amanda Carrico, and Alison M Pergram. 2004. To nod or not to nod: An observational study of nonverbal com-

munication and status in female and male college students. *Psychology of Women Quarterly* 28(4):358–361.

Hirst, Graeme, Susan McRoy, Peter Heeman, Philip Edmonds, and Diane Horton. 1994. Repairing conversational misunderstandings and non-understandings. *Speech Communication* 15(3):213–229.

Hoey, Jesse, Pascal Poupart, Craig Boutilier, and Alex Mihailidis. 2005. Pomdp models for assistive technology. In *Proc. AAAI Fall Symposium on Caring Machines: AI in Eldercare*.

Hong, Jin-Hyuk, Youn-Suk Song, and Sung-Bae Cho. 2007. Mixed-initiative human–robot interaction using hierarchical Bayesian networks. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* 37(6):1158–1164.

Hsiao, Kai-yuh, Nikolaos Mavridis, and Deb Roy. 2003. Coupling perception and simulation: Steps towards conversational robotics. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 1, 928–933. IEEE.

Huang, Chien-Ming, and Bilge Mutlu. 2012. Robot behavior toolkit: generating effective social behaviors for robots. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, 25–32. ACM.

———. 2013. Modeling and evaluating narrative gestures for humanlike robots. In *Robotics: Science and Systems*.

Humphrey, Curtis M, Christopher Henk, George Sewell, Brian W Williams, and Julie A Adams. 2007. Assessing the scalability of a multiple robot interface. In

ACM/IEEE International Conference on Human-Robot Interaction (HRI), 239–246. IEEE.

Imai, Michita, Tetsuo Ono, and Hiroshi Ishiguro. 2003. Physical relation and expression: Joint attention for human-robot interaction. *IEEE Transaction on Industrial Electronics* 50(4):636–643.

Jancovic, MerryAnn, Shannon Devoe, and Morton Wiener. 1975. Age-related changes in hand and arm movements as nonverbal communication: Some conceptualizations and an empirical exploration. *Child Development* 922–928.

Johnson-Laird, Philip N. 1983. *Mental models: Towards a cognitive science of language, inference, and consciousness*. 6, Harvard University Press.

Jokinen, Kristiina. 2009. Gaze and gesture activity in communication. In *Universal Access in Human-Computer Interaction: Intelligent and Ubiquitous Interaction Environments*, 537–546. Springer.

Kahn, Peter H, Nathan G Freier, Takayuki Kanda, Hiroshi Ishiguro, Jolina H Ruckert, Rachel L Severson, and Shaun K Kane. 2008. Design patterns for sociality in human-robot interaction. In *Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction*, 97–104. ACM.

Kanda, Takayuki, Rumi Sato, Naoki Saiwaki, and Hiroshi Ishiguro. 2007. A two-month field trial in an elementary school for long-term human–robot interaction. *IEEE Transactions on Robotics* 23(5):962–971.

Kendon, Adam. 2004. *Gesture: Visible action as utterance*. Cambridge University Press.

Kita, Sotaro. 2003. Pointing: A foundational building block of human communication. *Pointing: Where Language, Culture, and Cognition Meet* 1–8.

Kobayashi, Yoshinori, Takashi Shibata, Yosuke Hoshi, Yoshinori Kuno, Mai Okada, and Keiichi Yamazaki. 2010. "i will ask you" choosing answerers by observing gaze responses using integrated sensors for museum guide robots. In *RO-MAN*, 652–657. IEEE.

Kobsa, Alfred, Jürgen Allgayer, Carola Reddig, Norbert Reithinger, Dagmar Schmauks, Karin Harbusch, and Wolfgang Wahlster. 1986. Combining deictic gestures and natural language for referent identification. In *Proceedings of the 11th Coference on Computational Linguistics*, 356–361. Association for Computational Linguistics.

Kock, S., J. Bredahl, P. J. Eriksson, M. Myhr, and K. Behnisch. 2013. Taming the robot: better safety without higher fences. www.abb.com. Doc no. 9AKK105152A2830. Accessed: 2014-09-23.

Kock, S, T Vittor, B Matthias, H Jerregård, M Källman, I Lundberg, R Mellander, and M Hedelind. 2011. A robot concept for scalable, flexible assembly automation. In *Proceedings of IEEE International Symposium on Assembly and Manufacturing*, 25–27.

Kooijmans, Tijn, Takayuki Kanda, Christoph Bartneck, Hiroshi Ishiguro, and Norihiro Hagita. 2006. Interaction debugging: an integral approach to analyze human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, 64–71. ACM.

Kopp, Stefan, Lars Gesellensetter, Nicole C Krämer, and Ipke Wachsmuth. 2005. A conversational agent as museum guide–design and evaluation of a real-world application. In *Intelligent Virtual Agents*, 329–343. Springer.

Koulouri, Theodora, and Stanislao Lauria. 2009. Exploring miscommunication and collaborative behaviour in human-robot interaction. In *Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 111–119. Association for Computational Linguistics.

Kuno, Yoshinori, Kazuhisa Sadazuka, Michie Kawashima, Keiichi Yamazaki, Akiko Yamazaki, and Hideaki Kuzuoka. 2007. Museum guide robot based on sociological interaction analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1191–1194. ACM.

Lakoff, George, and Mark Johnson. 1980. *Metaphors we live by*. University of Chicago Press.

Landis, J Richard, and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 159–174.

Langellier, Kristin M. 1989. Personal narratives: Perspectives on theory and research. *Text and Performance Quarterly* 9(4):243–276.

Lee, Min Kyung, Sara Kiesler, Jodi Forlizzi, and Paul Rybski. 2012. Ripple effects of an embedded social agent: a field study of a social robot in the workplace. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 695–704. ACM.

Lee, Min Kyung, Sara Kiesler, Jodi Forlizzi, Siddhartha Srinivasa, and Paul Rybski. 2010. Gracefully mitigating breakdowns in robotic services. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction*, 203–210. IEEE.

Lee, Sau-lai, Ivy Yee-man Lau, Sara Kiesler, and Chi-Yue Chiu. 2005. Human mental models of humanoid robots. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, 2767–2772. IEEE.

Lempers, Jacques D. 1979. Young children's production and comprehension of nonverbal deictic behaviors. *The Journal of Genetic Psychology* 135(1):93–102.

Leslie, Alan M. 1987. Pretense and representation: The origins of "theory of mind.". *Psychological review* 94(4):412.

Liu, Phoebe, Dylan F Glas, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2013. It's not polite to point: generating socially-appropriate deictic behaviors towards people. In *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction*, 267–274. IEEE Press.

Ljungblad, Sara, Jirina Kotrbova, Mattias Jacobsson, Henriette Cramer, and Karol Niechwiadowicz. 2012. Hospital robot at work: something alien or an intelligent

colleague? In *Proceedings of the ACM 2012 conference on Computer-Supported Cooperative Work*, 177–186. ACM.

Lohse, Manja, Frederic Siepmann, and Sven Wachsmuth. 2014. A modeling framework for user-driven iterative design of autonomous systems. *International Journal of Social Robotics* 6(1):121–139.

Lozano, Sandra C, and Barbara Tversky. 2006. Communicative gestures facilitate problem solving for both communicators and recipients. *Journal of Memory and Language* 55(1):47–63.

Lucignano, Lorenzo, Francesco Cutugno, Silvia Rossi, and Alberto Finzi. 2013. A dialogue system for multimodal human-robot interaction. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*, 197–204. ACM.

Malle, Bertram F, and Joshua Knobe. 1997. The folk concept of intentionality. *Journal of Experimental Social Psychology* 33(2):101–121.

Mantovani, Giuseppe. 1996. Social context in hci: A new framework for mental models, cooperation, and communication. *Cognitive Science* 20(2):237–269.

Matarazzo, Joseph D, and Arthur N Wiens. 1972. *The interview: Research on its anatomy and structure*. Transaction Publishers.

Matthias, Bjoern, Soenke Kock, Henrik Jerregard, Mats Kallman, Ivan Lundberg, and Roger Mellander. 2011. Safety of collaborative industrial robots: Certification possibilities for a collaborative assembly robot concept. In *IEEE International Symposium on Assembly and Manufacturing*, 1–6. IEEE.

McNeil, David. 1992. Hand and mind.

Meltzoff, Andrew N. 1995. Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Developmental psychology* 31(5):838.

Meyer, Antje S, Astrid M Sleiderink, and Willem JM Levelt. 1998. Viewing and naming objects: Eye movements during noun phrase production. *Cognition* 66(2): B25–B33.

Miklósi, Ádam, and Krisztina Soproni. 2006. A comparative analysis of animals' understanding of the human pointing gesture. *Animal Cognition* 9(2):81–93.

Mirnig, Nicole, and Manfred Tscheligi. 2015. Comprehension, coherence, and consistency: Essentials of robot feedback. In *Robots that Talk and Listen*, ed. Judith A Markowitz. De Gruyter.

Mitra, Sushmita, and Tinku Acharya. 2007. Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews* 37(3):311–324.

Moon, Youngme. 2000. Intimate exchanges: Using computers to elicit self-disclosure from consumers. *Journal of Consumer Research* 26(4):323–339.

Morrison, Gary R, and Gary J Anglin. 2005. Research on cognitive load theory: Application to e-learning. *Educational Technology Research and Development* 53(3): 94–104.

Murphy, Catherine M. 1978. Pointing in the context of a shared activity. *Child Development* 371–380.

Mutlu, Bilge, Sean Andrist, and Allison Sauppé. 2015. Enabling human-robot dialogue. In *Robots that Talk and Listen*, ed. Judith A Markowitz. De Gruyter.

Mutlu, Bilge, and Jodi Forlizzi. 2008. Robots in organizations: the role of workflow, social, and environmental factors in human-robot interaction. In *3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 287–294. IEEE.

Mutlu, Bilge, Jodi Forlizzi, and Jessica Hodgins. 2006. A storytelling robot: Modeling and evaluation of human-like gaze behavior. In *IEEE-RAS International Conference on Humanoid Robots*, 518–523. IEEE.

Mutlu, Bilge, Toshiyuki Shiwa, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2009. Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction*, 61–68. ACM.

Nass, Clifford, and Kwan Min Lee. 2000. Does computer-generated speech manifest personality? an experimental test of similarity-attraction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 329–336. ACM.

Nass, Clifford, and Youngme Moon. 2000. Machines and mindlessness: Social responses to computers. *Journal of Social Issues* 56(1):81–103.

Nicolescu, Monica N, and Maja J Mataric. 2003. Linking perception and action in a control architecture for human-robot domains. In *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, 10–pp. IEEE.

O'Daniel, M, and AH Rosenstein. 2008. Chapter 33. professional communication and team collaboration. patient safety and quality: An evidence-based handbook for nurses. *Patient safety and quality: An evidence-based handbook for nurses. Agency for Healthcare Research and Quality, Rockville, MD.*

Okuno, Yusuke, Takayuki Kanda, Michita Imai, Hiroshi Ishiguro, and Norihiro Hagita. 2009. Providing route directions: design of robot's utterance, gesture, and timing. In *4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 53–60. IEEE.

O'Leary, Maria J, and Cynthia Gallois. 1985. The last ten turns: Behavior and sequencing in friends' and strangers' conversational findings. *Journal of Nonverbal Behavior* 9(1):8–27.

Olson, Gary M, Thomas W Malone, and John B Smith. 2001. *Coordination theory and collaboration technology.* Psychology Press.

Peltason, Julia. 2010. Modeling human-robot-interaction based on generic interaction patterns. *AAAI Fall Symposium: Dialog with Robots.*

Perlin, Ken. 2002. Improving noise. In *ACM Transactions on Graphics (TOG)*, vol. 21, 681–682. ACM.

Quigley, Morgan, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y Ng. 2009. Ros: an open-source robot operating system. In *ICRA Workshop on Open Source Software*, vol. 3, 5.

Rae, Irene, Gina Venolia, John C Tang, and David Molnar. 2015. A framework for understanding and designing telepresence. In *Proceedings of the 18th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 1552–1566. ACM.

Reigeluth, Charles M, M David Merrill, Brent G Wilson, and Reginald T Spiller. 1980. The elaboration theory of instruction: A model for sequencing and synthesizing instruction. *Instructional Science* 9(3):195–219.

Ren, Zhou, Junsong Yuan, Jingjing Meng, and Zhengyou Zhang. 2013. Robust part-based hand gesture recognition using kinect sensor. *IEEE Transactions on Multimedia* 15(5):1110–1120.

Rethink Robotics. 2012. Baxter safety and compliance overview. www.rethinkrobotics.com. Accessed: 2014-09-23.

Rosch, Eleanor, Evan Thompson, and Francisco J Varela. 1992. *The embodied mind: Cognitive science and human experience*. MIT Press.

Rutter, Derek R, and Kevin Durkin. 1987. Turn-taking in mother–infant interaction: An examination of vocalizations and gaze. *Developmental Psychology* 23(1):54.

Sacks, Harvey, Emanuel A Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *language* 696–735.

Salem, Maha, Stefan Kopp, Ipke Wachsmuth, Katharina Rohlfing, and Frank Joublin. 2012. Generation and evaluation of communicative robot gesture. *International Journal of Social Robotics* 4(2):201–217.

Sauppé, Allison, and Bilge Mutlu. 2014a. Design patterns for exploring and prototyping human-robot interactions. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, 1439–1448. CHI '14, New York, NY, USA: ACM.

———. 2014b. Effective task training strategies for instructional robots. In *Proceedings of the 10th Annual Robotics: Science and Systems Conference*.

———. 2014c. Robot deictics: How gesture and context shape referential communication. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, 342–349. HRI '14, New York, NY, USA: ACM.

———. 2015a. Effective task training strategies for human and robot instructors. *Autonomous Robots* 1–17.

———. 2015b. The social impact of a robot co-worker in industrial settings. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 3613–3622. CHI '15, New York, NY, USA: ACM.

Scassellati, Brian. 1999. Imitation and mechanisms of joint attention: A developmental structure for building social skills on a humanoid robot. In *Computation for metaphors, analogy, and agents*, 176–195. Springer.

———. 2002. Theory of mind for a humanoid robot. *Autonomous Robots* 12(1): 13–24.

Schauerte, Boris, and Gernot A Fink. 2010. Focusing computational visual attention in multi-modal human-robot interaction. In *International conference on multimodal interfaces and the workshop on machine learning for multimodal interaction*, 6. ACM.

Scheeff, Mark, John Pinto, Kris Rahardja, Scott Snibbe, and Robert Tow. 2002. Experiences with sparky, a social robot. In *Socially intelligent agents*, 173–180. Springer.

Schegloff, Emanuel A. 1972. Sequencing in conversational openings. *Directions in sociolinguistics* 346–380.

Schegloff, Emanuel Abraham. 1967. The first five seconds: The order of conversational opening. Ph.D. thesis, University of California.

Scherer, Klaus R, Judy Koivumaki, and Robert Rosenthal. 1972. Minimal cues in the vocal communication of affect: Judging emotions from content-masked speech. *Journal of Psycholinguistic Research* 1(3):269–285.

Sebanz, Natalie, and Guenther Knoblich. 2009. Prediction in joint action: What, when, and where. *Topics in Cognitive Science* 1(2):353–367.

Seedhouse, Paul. 1999. The relationship between context and the organisation of repair in the L2 classroom. *IRAL. International review of applied linguistics in language teaching* 37(1):59–80.

Shah, Julie, and Cynthia Breazeal. 2010. An empirical analysis of team coordination behaviors and action planning with application to human–robot teaming. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 52(2):234–245.

Siino, Rosanne, and Pamela J Hinds. 2005. Robots, gender & sensemaking: Sex segregation's impact on workers making sense of a mobile autonomous robot. In *Ieee international conference on robotics and automation*, vol. 3, 2773. IEEE; 1999.

Skehan, Peter. 1996. A framework for the implementation of task-based instruction. *Applied linguistics* 17(1):38–62.

St Clair, A, Ross Mead, and Maja J Mataric. 2011. Investigating the effects of visual saliency on deictic gesture production by a humanoid robot. In *Ro-man, 2011 ieee*, 210–216. IEEE.

Staggers, Nancy, and Anthony F. Norcio. 1993. Mental models: concepts for human-computer interaction research. *International Journal of Man-machine studies* 38(4): 587–605.

Staudte, Maria, and Matthew W Crocker. 2009. Visual attention in spoken human-robot interaction. In *Proceedings of the 4th acm/ieee international conference on human robot interaction*, 77–84. ACM.

Steinfeld, Aaron, Terrence Fong, David Kaber, Michael Lewis, Jean Scholtz, Alan Schultz, and Michael Goodrich. 2006. Common metrics for human-robot interaction. In *Proceedings of the 1st acm sigchi/sigart conference on human-robot interaction*, 33–40. ACM.

Stivers, Tanya. 2008. Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation. *Research on language and social interaction* 41(1): 31–57.

Sugiyama, Osamu, Takayuki Kanda, Michita Imai, Hiroshi Ishiguro, and Norihiro Hagita. 2007. Natural deictic communication with humanoid robots. In *Intelligent robots and systems, 2007. iros 2007. ieee/rsj international conference on*, 1441–1448. IEEE.

Sung, Ja-Young, Lan Guo, Rebecca E Grinter, and Henrik I Christensen. 2007. *ï¿½my roomba is ramboï¿½: Intimate home appliances*. Springer.

Sweller, John. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive science* 12(2):257–285.

Szafir, Daniel, and Bilge Mutlu. 2013. Artful: adaptive review technology for flipped learning. In *Proceedings of the sigchi conference on human factors in computing systems*, 1001–1010. ACM.

Tanaka, Fumihide, and Takeshi Kimura. 2009. The use of robots in early education: A scenario based on ethical consideration. In *Ro-man*, 558–560.

Tanaka, Fumihide, and Javier R Movellan. 2006. Behavior analysis of children's touch on a small humanoid robot: Long-term observation at a daily classroom over three months. In *Robot and human interactive communication, 2006. roman 2006. the 15th ieee international symposium on*, 753–756. IEEE.

Tannen, Deborah. 1989. *Talking voices: Repetition, dialogue, and imagery in conversational discourse*, vol. 26. Cambridge University Press.

Taylor, Steven J, and Robert Bogdan. 1998. *Introduction to qualitative research methods: A guidebook and resource* . John Wiley & Sons Inc.

Tian, Yingli, Takeo Kanade, and Jeffrey F Cohn. 2011. Facial expression recognition. In *Handbook of face recognition*, 487–519. Springer.

Tomasello, Michael. 1995. Joint attention as social cognition. *Joint attention: Its origins and role in development* 103–130.

Tonietti, Giovanni, Riccardo Schiavi, and Antonio Bicchi. 2005. Design and control of a variable stiffness actuator for safe and fast physical human/robot interaction. In *Robotics and automation, 2005. icra 2005. proceedings of the 2005 ieee international conference on*, 526–531. IEEE.

Torrey, Cristen, Aaron Powers, Susan R Fussell, and Sara Kiesler. 2007. Exploring adaptive dialogue based on a robot's awareness of human gaze and task progress. In *Proceedings of the acm/ieee international conference on human-robot interaction*, 247–254. ACM.

Torrey, Cristen, Aaron Powers, Matthew Marge, Susan R Fussell, and Sara Kiesler. 2006. Effects of adaptive robot dialogue on information exchange and social relations. In *Proceedings of the 1st acm sigchi/sigart conference on human-robot interaction*, 126–133. ACM.

Trafton, J Gregory, Nicholas L Cassimatis, Magdalena D Bugajska, Derek P Brock, Farilee E Mintz, and Alan C Schultz. 2005. Enabling effective human-robot interaction using perspective-taking in robots. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on* 35(4):460–470.

Vertesi, Janet. 2008. Seeing like a rover: embodied experience on the mars exploration rover mission. In *CHI'08 Extended Abstracts on Human Factors in Computing Systems*, 2523–2532. ACM.

Vogel, Daniel, and Ravin Balakrishnan. 2004. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*, 137–146. ACM.

Wainer, Joshua, David J Feil-Seifer, Dylan A Shell, and Maja J Mataric. 2007. Embodiment and human-robot interaction: A task-based perspective. In *The 16th IEEE International Symposium on Robot and Human interactive Communication*, 872–877. IEEE.

Wilson, John R, and Andrew Rutherford. 1989. Mental models: Theory and application in human factors. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 31(6):617–634.

Wilson, Margaret. 2002. Six views of embodied cognition. *Psychonomic bulletin & review* 9(4):625–636.

You, Zhen-Jia, Chi-Yuh Shen, Chih-Wei Chang, Baw-Jhiune Liu, and Gwo-Dong Chen. 2006. A robot as a teaching assistant in an English class. In *Sixth International Conference on Advanced Learning Technologies*, 87–91. IEEE.

Zahn, Christopher J. 1984. A reexamination of conversational repair. *Communications Monographs* 51(1):56–66.

Zheng, Kuanhao, Dylan F Glas, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2013. Supervisory control of multiple social robots for navigation. In *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction*, 17–24. IEEE Press.

Ziefle, Martina, and Susanne Bay. 2004. Mental models of a cellular phone menu: Comparing older and younger novice users. In *Mobile Human-Computer Interaction*, 25–37. Springer.